Search and Discovery Statistics in HEP
LECTURE 3

Eilam Gross, Weizmann Institute of Science

This presentation would have not been possible without the tremendous help of the following people throughout many years

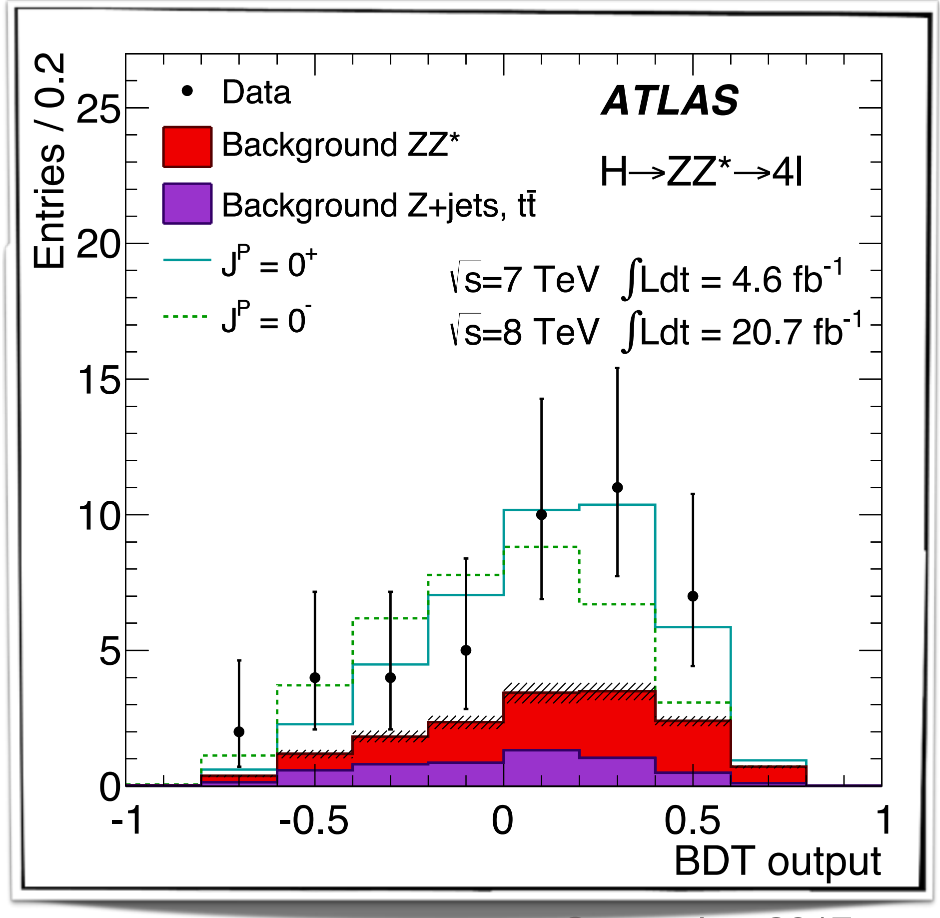
Louis Lyons, Alex Read, Glen Cowan ,Kyle Cranmer
Ofer Vitells & Bob Cousins

Application of CIs and q^{NP} test statistic

$$q^{NP} = -2\ln\frac{L(H_0)}{L(H_1)} = \sum_{bins} -2\ln\frac{L_i(0^+)}{L_i(0^-)}$$

$$L_i(0^+) = Pois(n_i; n_i^{0^+}) = \frac{\left(n_i^{0^+}\right)^{n_i} e^{-n_i^{0^+}}}{n_i!}$$

Can you tell O^+ from 0^-?



Test Spin 0 parity

$$H_0 = 0^+$$

$$H_1 = 0^-$$

$$q^{NP} = -2\ln\frac{L(H_0)}{L(H_1)}$$

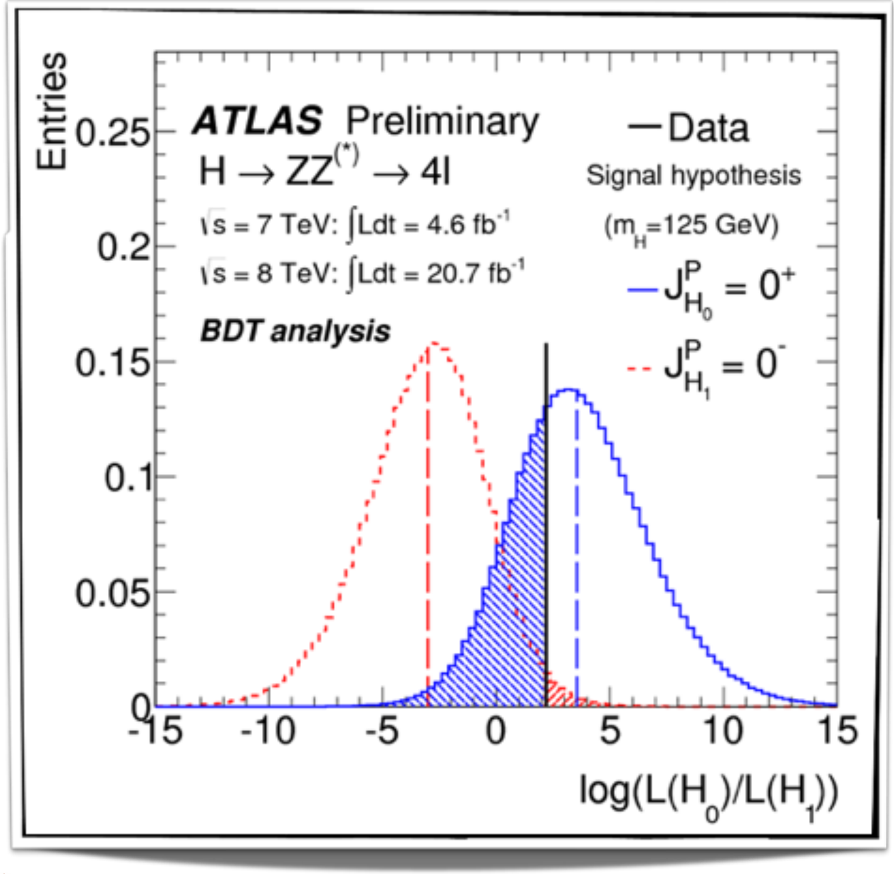
$p_{H_1}(\exp | H_0) = 0.37\%$,

$p_{H_1}(obs) = 1.5\%$

$p_{H_0}(obs) = 31\%$

$p_{H_1}^{CL_s}(obs) = 2.2\%$

$$p_{H_1}^{CL_s} = \frac{p_{H_1}}{1-p_{H_0}} = \frac{1.5\%}{1-0.31} = 2.2\%$$



Which means
$J^{p}=0^-$ is excluded at the
97.8% CL in favour of $J^{p}=0^+$

$H_1 like$ $H_0 like$

DISCOVERY

Case Study:

Higgs Discovery

Basic Definition: Signal Strength

- We normally relate the signal strength to its expected Standard Model value, i.e.

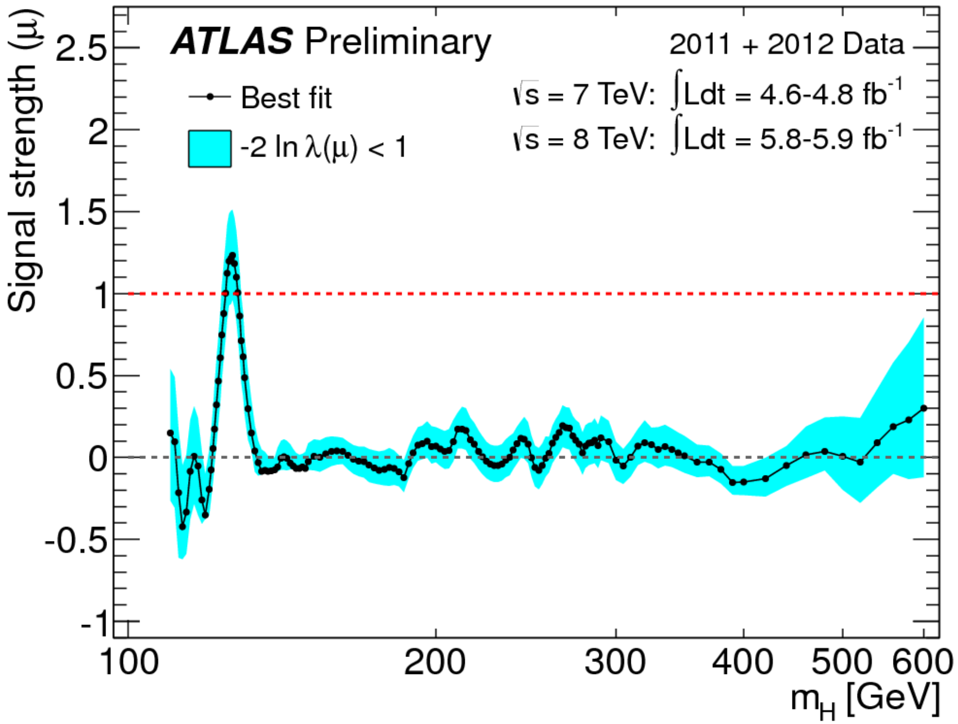
$$\mu(m_H) = \frac{\sigma(m_H)}{\sigma_{SM}(m_H)}$$

$$\hat{\mu}(m_H) = \text{MLE of } \mu$$

Introducing the Heartbeat

$$\mu(m_H) = \frac{\sigma(m_H)}{\sigma_{SM}(m_H)}$$

$$\hat{\mu}(m_H) = \text{MLE of } \mu$$



Reminder: The test statistic

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0\,, \\ 0 & \hat{\mu} < 0\,, \end{cases}$$

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

- Downward fluctuations of the background do not serve as an evidence against the background
- Upward fluctuations of the signal do not serve as an evidence against the signal

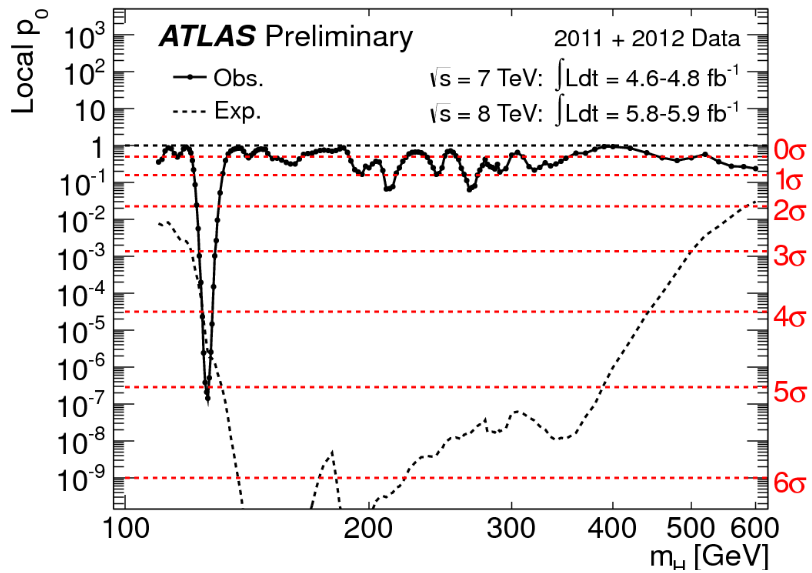
p0 and the expected p0

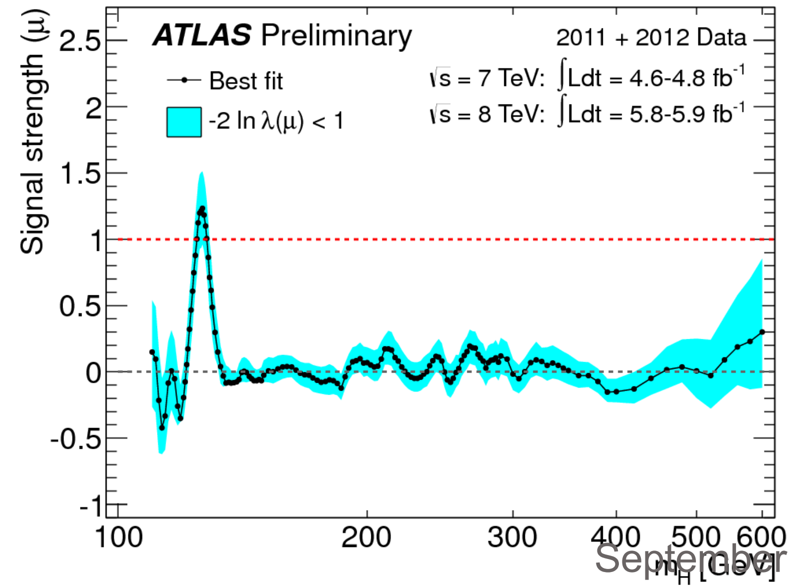
$$p_0 = \int_{q_{0,\text{obs}}}^{\infty} f(q_0|0)\,dq_0$$

p_0 is the probability to observe a less BG like result (more signal like) than the observed one

Small p0 leads to an observation

A tiny p0 leads to a discovery





$$p = \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = 1 - \Phi(Z)$$

$$Z = \Phi^{-1}(1-p)$$

Distribution of q0 (discovery)

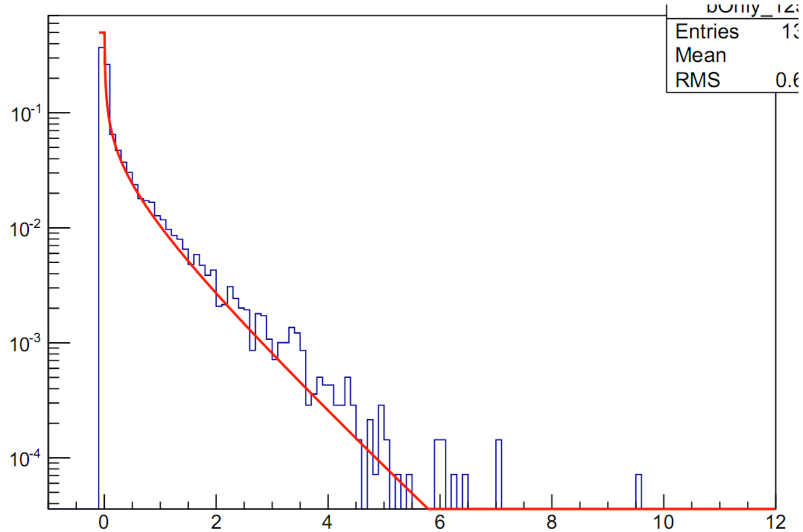
- We find

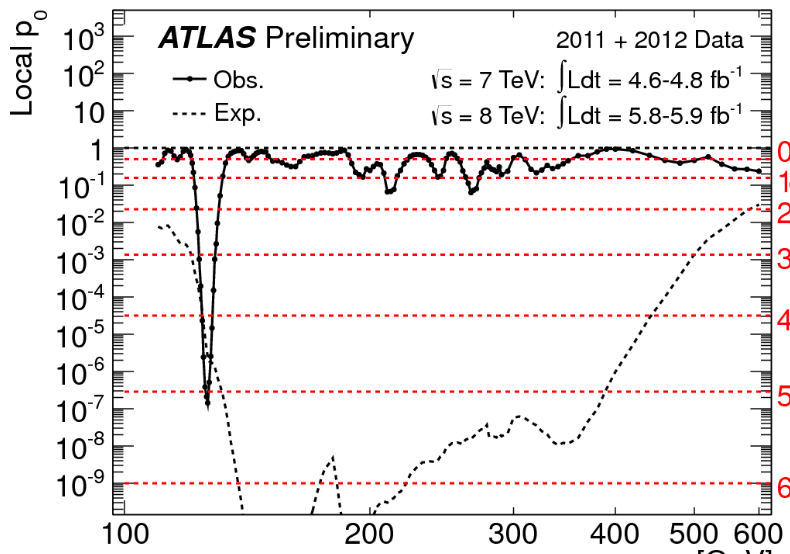
$$p_0 = \int_{q_{0,\mathrm{obs}}}^{\infty} f(q_0|0)\, dq_0 \,.$$

$$f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2} \,.$$

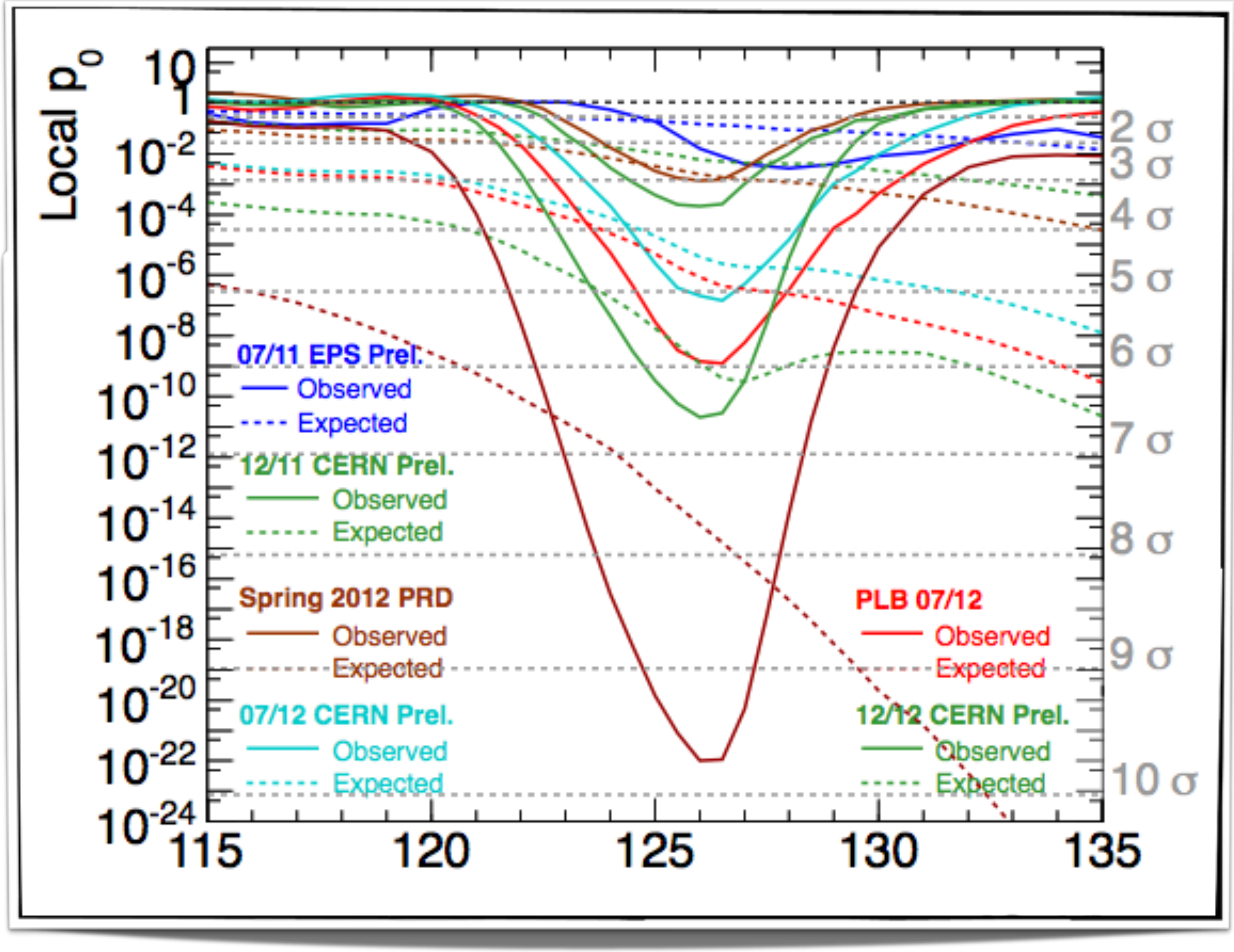
$$Z_0 = \Phi^{-1}(1 - p_0) = \sqrt{q_0} \,.$$

- q_0 distribute as half a delta function at zero and half a chi squared. $q_{0,obs} = q_{0,obs}(m_H) \rightarrow p_0 = p_0(m_H)$





p0



Local p_0
10
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
10^{-20}
10^{-22}
10^{-24}
115
120
125
130
135
2 σ
3 σ
4 σ
5 σ
6 σ
7 σ
8 σ
9 σ
10 σ
07/11 EPS Prel.
Observed
Expected
12/11 CERN Prel.
Observed
Expected
Spring 2012 PRD
Observed
Expected
07/12 CERN Prel.
Observed
Expected
PLB 07/12
Observed
Expected
12/12 CERN Prel.
Observed
Expected

The New $s/\sqrt{b}$

The new $s/\sqrt{b}$

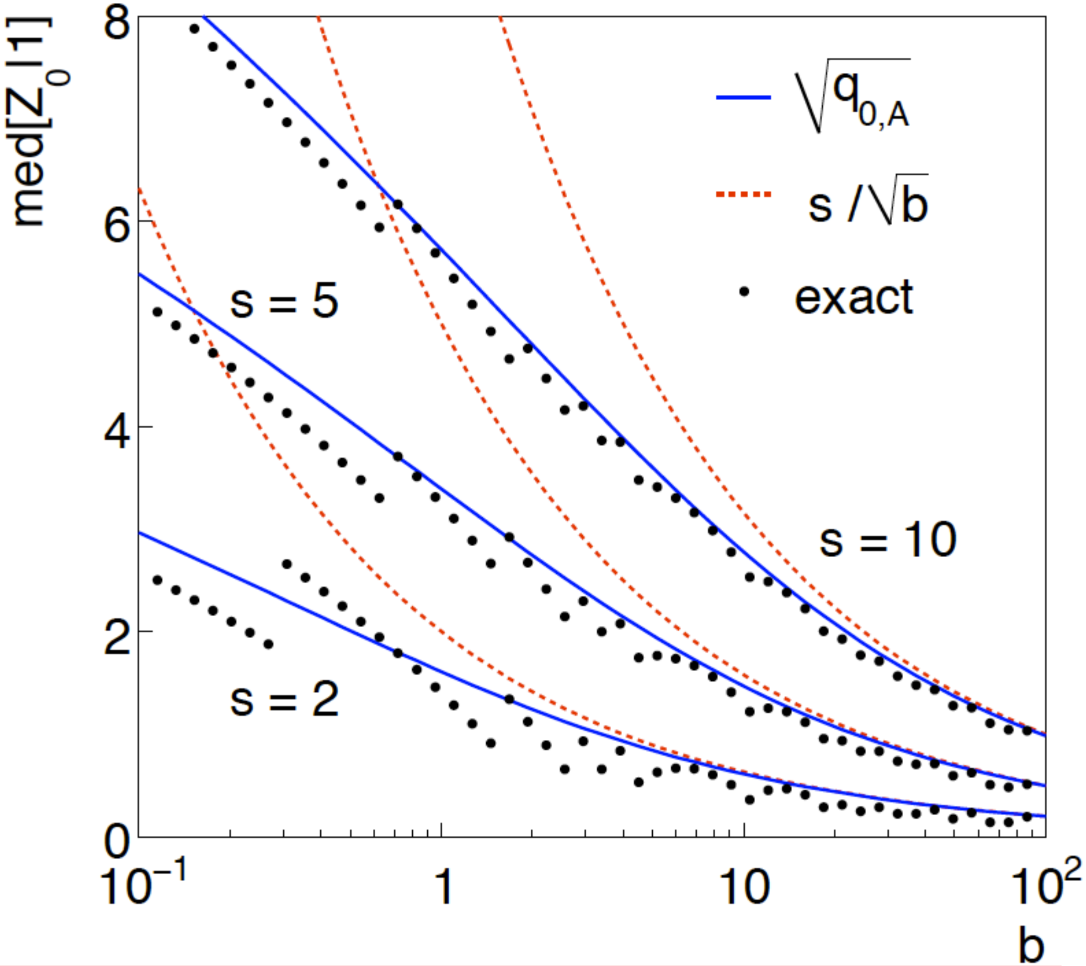
$$Z_A = \sqrt{q_{0,A}}$$

$$\mathrm{med}[Z_0|1] = \sqrt{q_{0,\mathrm{A}}} = \sqrt{2\left((s+b)\ln(1+s/b) - s\right)}$$

$$Z_A = \sqrt{q_{0,A}} \xrightarrow{s/b \ll 1} \frac{s}{\sqrt{b}} + O(s/b)$$

The New $s/\sqrt{b}$

$s/\sqrt{b}$?



The new $s/\sqrt{b}$

$$\mathrm{med}[Z_0|1] = \sqrt{q_{0,\mathrm{A}}} = \sqrt{2\left((s+b)\ln(1+s/b) - s\right)}$$

Taking Background Systematics into Account

- The intuitive explanation of $s/\sqrt{b}$ is that it compares the signal, s, to the standard deviation of n assuming no signal, $\sqrt{b}$.
- Now suppose the value of b is uncertain, characterized by a standard deviation σ_b.
- A reasonable guess is to replace $\sqrt{b}$ by the quadratic sum of $\sqrt{b}$ and σ_b, i.e.,

$$b \pm \Delta \cdot b \Rightarrow \sigma_b = \sqrt{\left(\sqrt{b}\right)^2 + (\Delta \cdot b)^2} = \sqrt{b + \Delta^2 b^2}$$

$$s/\sqrt{b} \Rightarrow s/\sqrt{b(1+b\Delta^2)} \xrightarrow{L\to\infty} \frac{s/b}{\Delta}$$

$$\frac{s/b}{\Delta} \geq 5 \rightarrow s/b \geq 0.5 \text{ for } \Delta \sim 10\%$$

If s/b<0.5 we will never be able to make a discovery

But even that formula can be improved using the Asimov formalism

Significance with systematics

- We find (G. Cowan)

$$Z_{\mathrm{A}} = \left[2 \left((s+b) \ln \left[\frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right) \right]^{1/2}$$

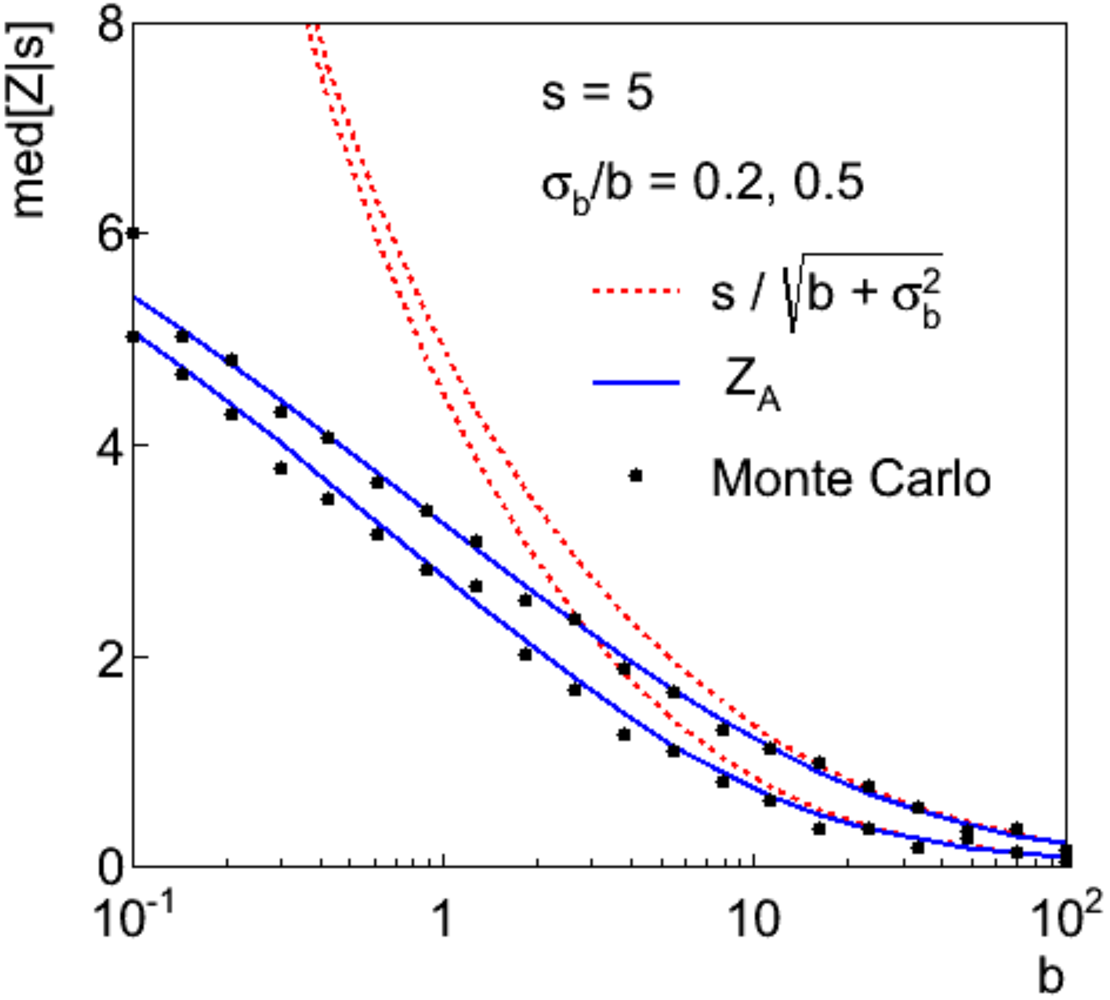
Expanding the Asimov formula in powers of s/b and $\sigma_b{}^2/b$ gives

$$Z_{\mathrm{A}} = \frac{s}{\sqrt{b+\sigma_b^2}} \left(1 + \mathcal{O}(s/b) + \mathcal{O}(\sigma_b^2/b) \right)$$

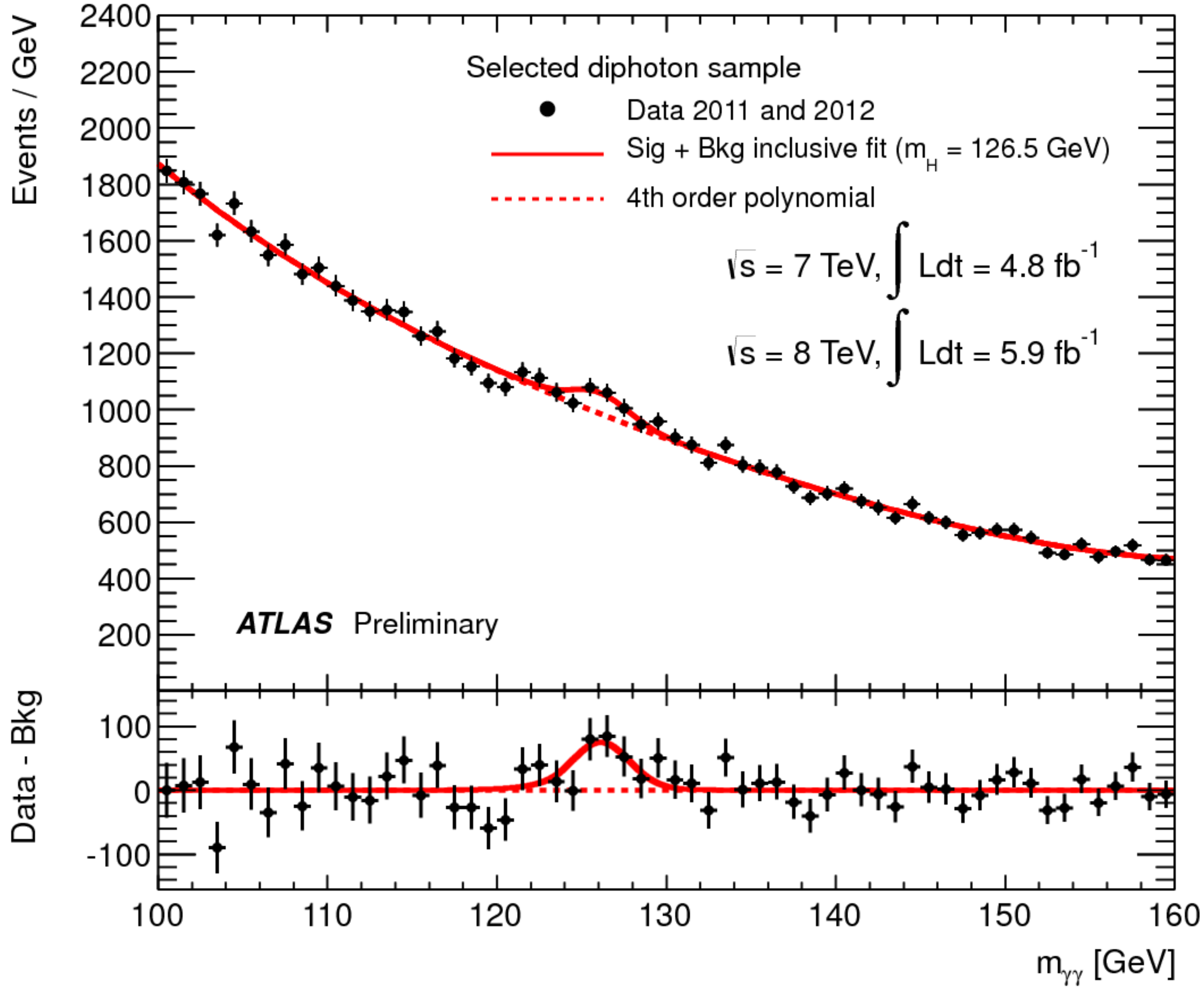
- So the "intuitive" formula can be justified as a limiting case

of the significance from the profile likelihood ratio test evaluated with the Asimov data set.

Significance with systematics

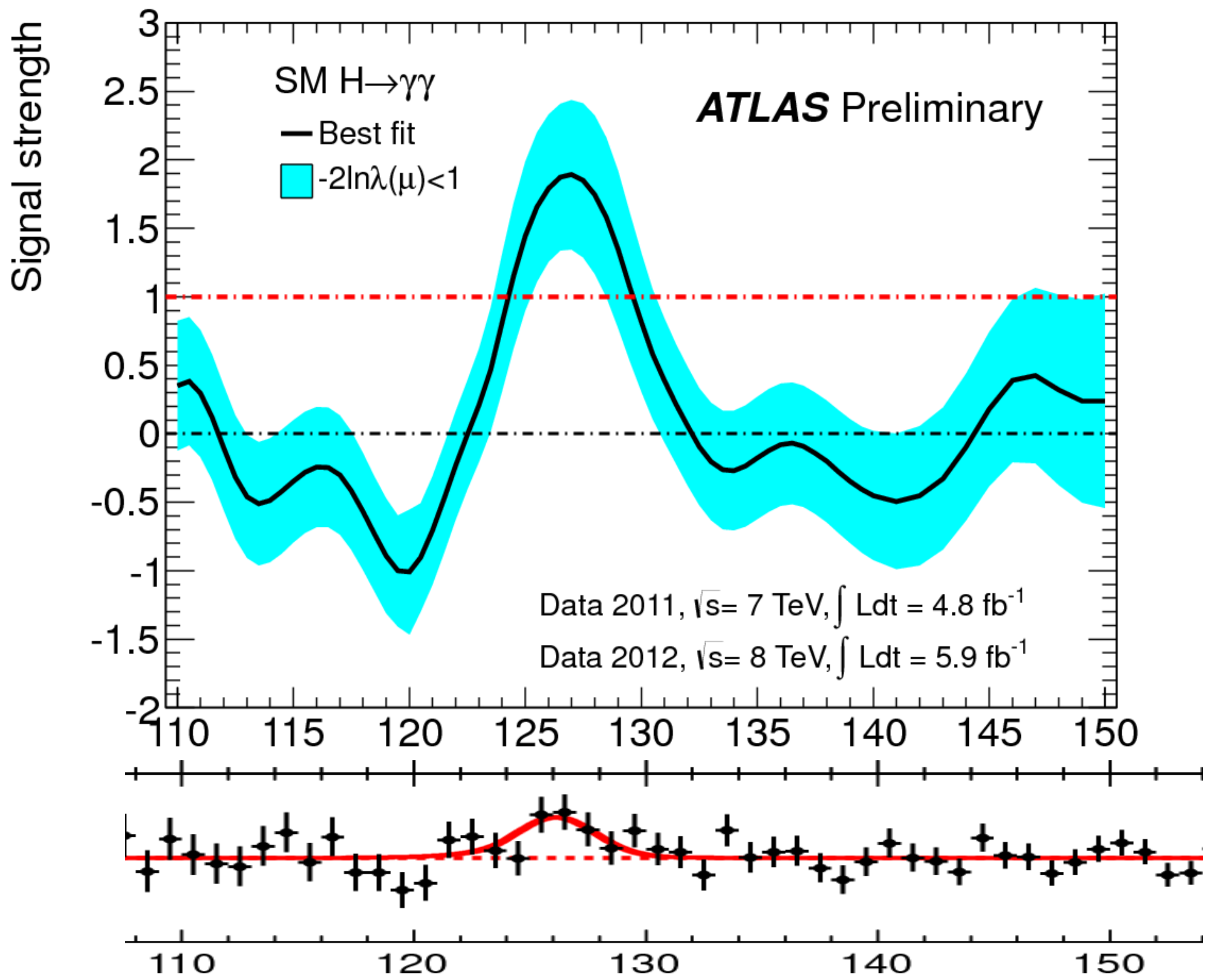


Example: H→γγ

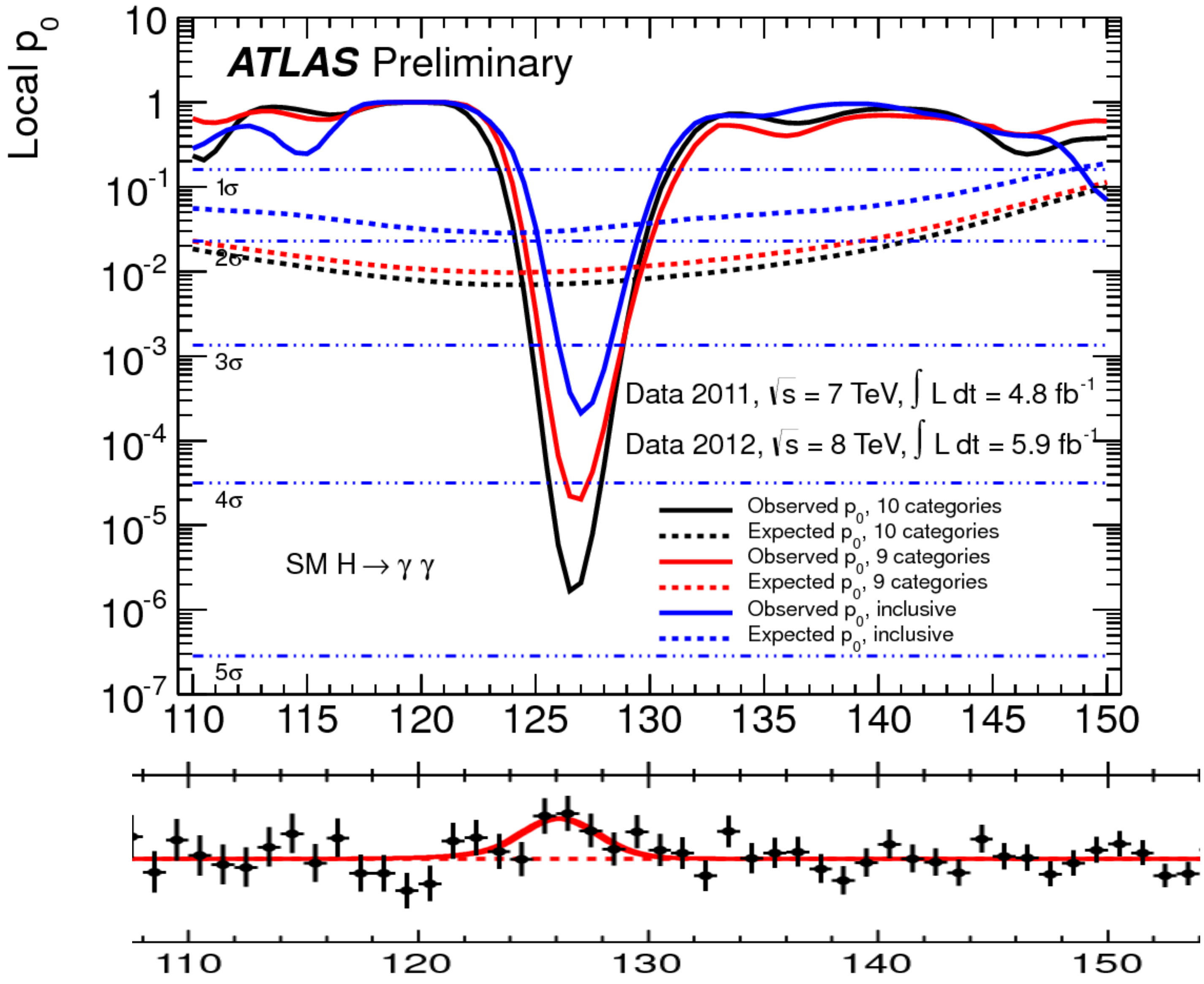


Events / GeV
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
Selected diphoton sample
Data 2011 and 2012
Sig + Bkg inclusive fit (m_H = 126.5 GeV)
4th order polynomial
$\sqrt{s}$ = 7 TeV, $\int$ Ldt = 4.8 fb^{-1}
$\sqrt{s}$ = 8 TeV, $\int$ Ldt = 5.9 fb^{-1}
ATLAS Preliminary
Data - Bkg
100
0
-100
100
110
120
130
140
150
160
$m_{\gamma\gamma}$ [GeV]

H→γγ


Signal strength
SM H→γγ
Best fit
-2lnλ(μ)<1
ATLAS Preliminary
Data 2011, √s= 7 TeV, ∫ Ldt = 4.8 fb-1
Data 2012, √s= 8 TeV, ∫ Ldt = 5.9 fb-1
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
110
115
120
125
130
135
140
145
150
110
120
130
140
150

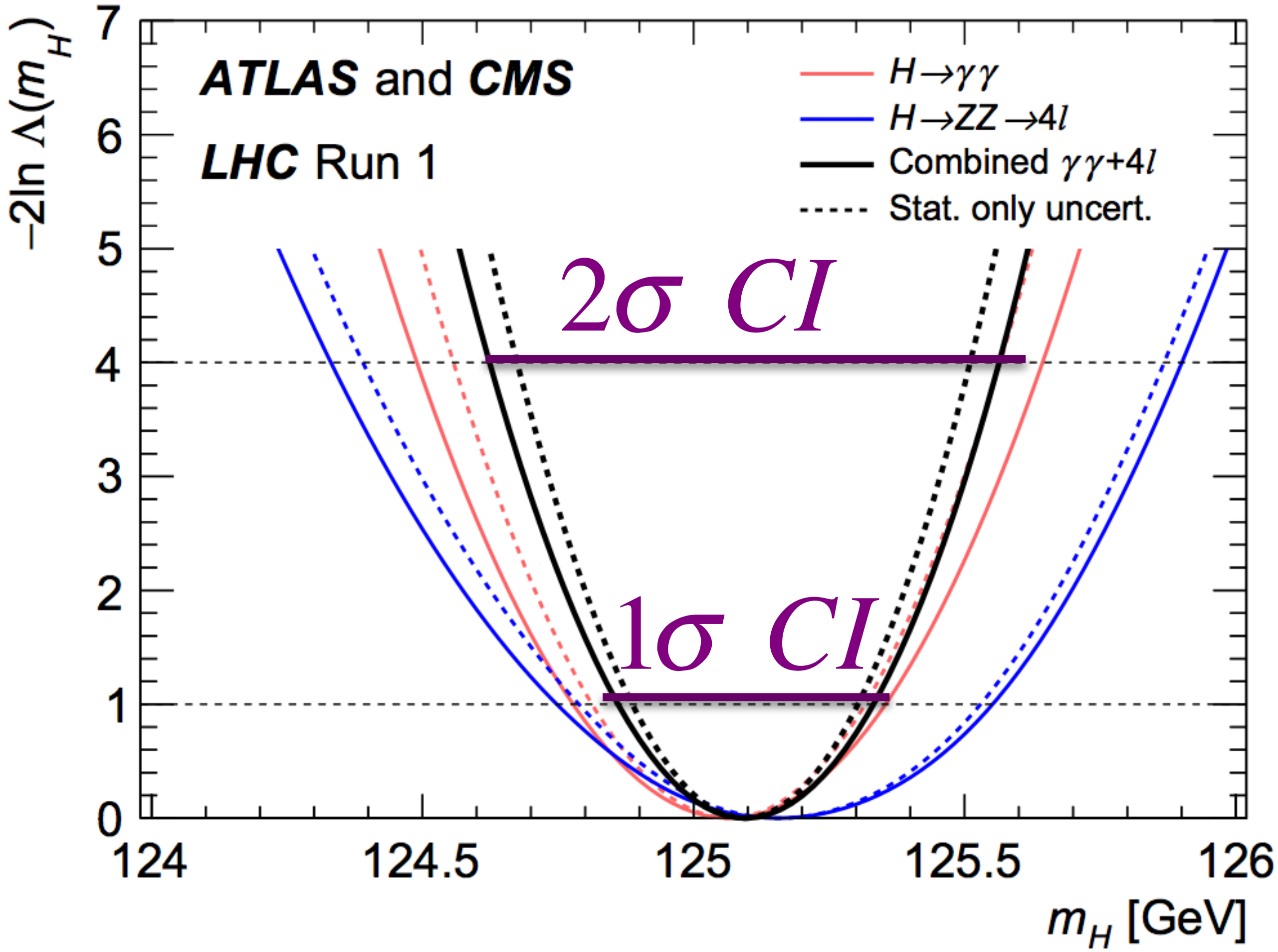
H→γγ



Local p_0
ATLAS Preliminary
Data 2011, $\sqrt{s}$ = 7 TeV, $\int$ L dt = 4.8 fb^{-1}
Data 2012, $\sqrt{s}$ = 8 TeV, $\int$ L dt = 5.9 fb^{-1}
SM H → γ γ
Observed p_0, 10 categories
Expected p_0, 10 categories
Observed p_0, 9 categories
Expected p_0, 9 categories
Observed p_0, inclusive
Expected p_0, inclusive
1σ
2σ
3σ
4σ
5σ
10
1
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
110 115 120 125 130 135 140 145 150
110 120 130 140 150

Obtaining the Syst Error

$$\sigma_{syst} = \sqrt{\sigma_{tot}^2 - \sigma_{stat}^2}$$



Measurements

Case studies: ATLAS and CMS mass and coupling combinations

PL in obtaining the mass

$$\Lambda(\boldsymbol{\alpha}) = \frac{L(\boldsymbol{\alpha}, \hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\alpha}))}{L(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}})} \qquad t_\alpha = -2 ln \Lambda(\alpha)$$

$$\Lambda(m_H) = \frac{L(m_H \,,\, \hat{\hat{\mu}}^{\gamma\gamma}_{ggF+t\bar{t}H(+b\bar{b}H)}(m_H) \,,\, \hat{\hat{\mu}}^{\gamma\gamma}_{VBF+VH}(m_H) \,,\, \hat{\hat{\mu}}^{ZZ}(m_H) \,,\, \hat{\hat{\boldsymbol{\theta}}}(m_H))}{L(\hat{m}_H, \hat{\mu}^{\gamma\gamma}_{ggF+t\bar{t}H(+b\bar{b}H)} \,,\, \hat{\mu}^{\gamma\gamma}_{VBF+VH}, \hat{\mu}^{ZZ}, \hat{\boldsymbol{\theta}})}$$

Scan the test statistic $\mathrm{t}_\alpha = t(\alpha)$

find $\hat{\alpha}$

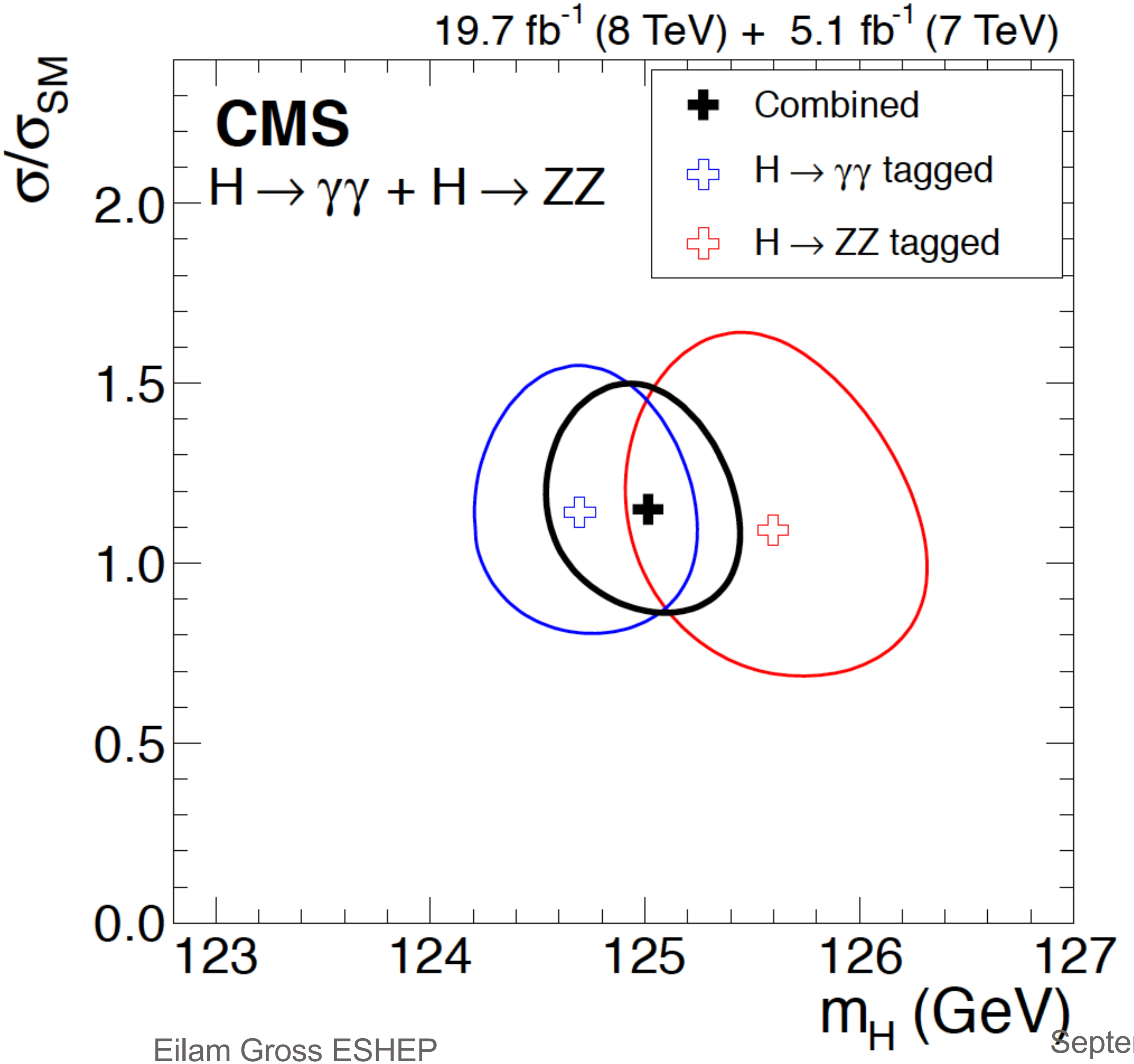
$t(\hat{\alpha} \pm N\sigma_{\hat{\alpha}}) = N^2$

A case of 2 poi

- In order to address the values of the signal strength and mass of a potential signal that are simultaneously consistent with the data, the following profile likelihood ratio is used:

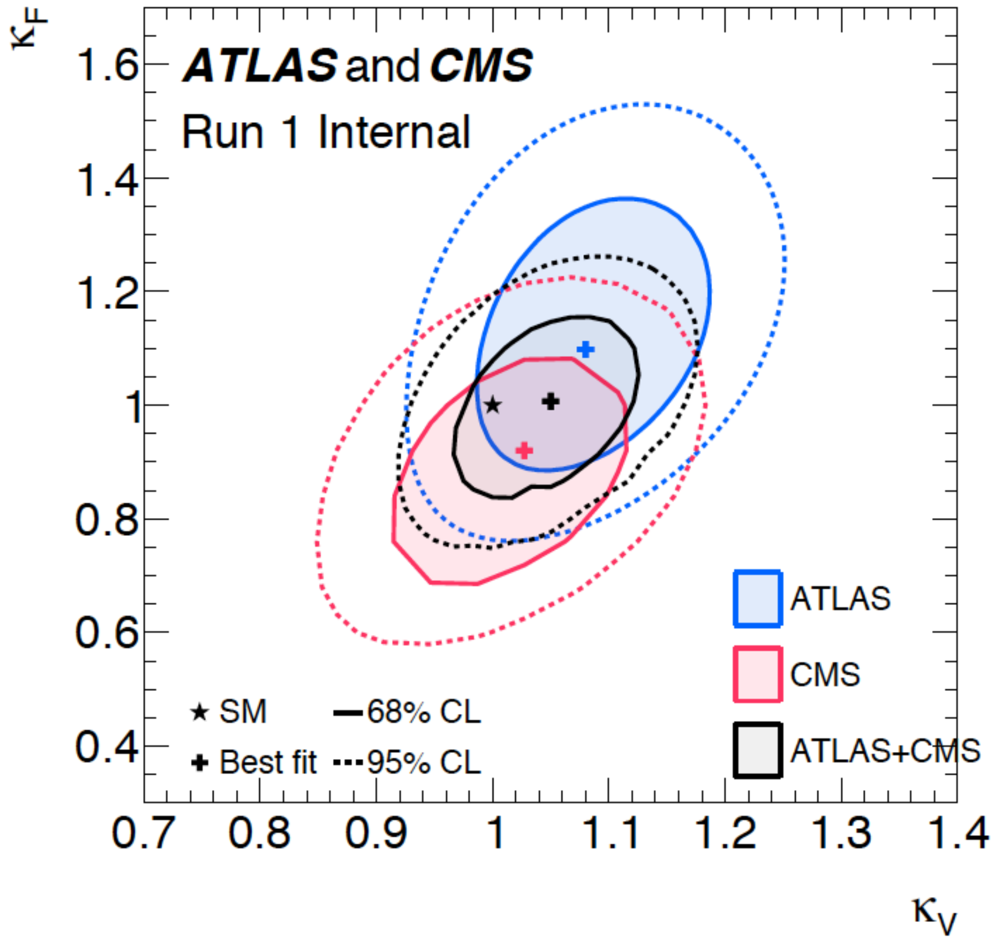
$$\lambda(\mu, m_H) = \frac{L(\mu, m_H, \hat{\hat{\theta}}(\mu, m_H))}{L(\hat{\mu}, \hat{m}_H, \hat{\theta})}$$

- In the presence of a signal, this test statistic will produce closed contours about the best fit point $(\hat{\mu}, \hat{m}_H)$;
- The 2D LR behaves asymptotically as a Chis squared with 2 DOF (Wilks' theorem) so the derivation of 68% and 95% CL cintours is easy, but care must be taken; The projection of 2D CI are not 1D CI!


19.7 fb-1 (8 TeV) + 5.1 fb-1 (7 TeV)
CMS
H → γγ + H → ZZ
Combined
H → γγ tagged
H → ZZ tagged
σ/σSM
2.0
1.5
1.0
0.5
0.0
123
124
125
126
127
mH (GeV)

PL in obtaining the Couplings

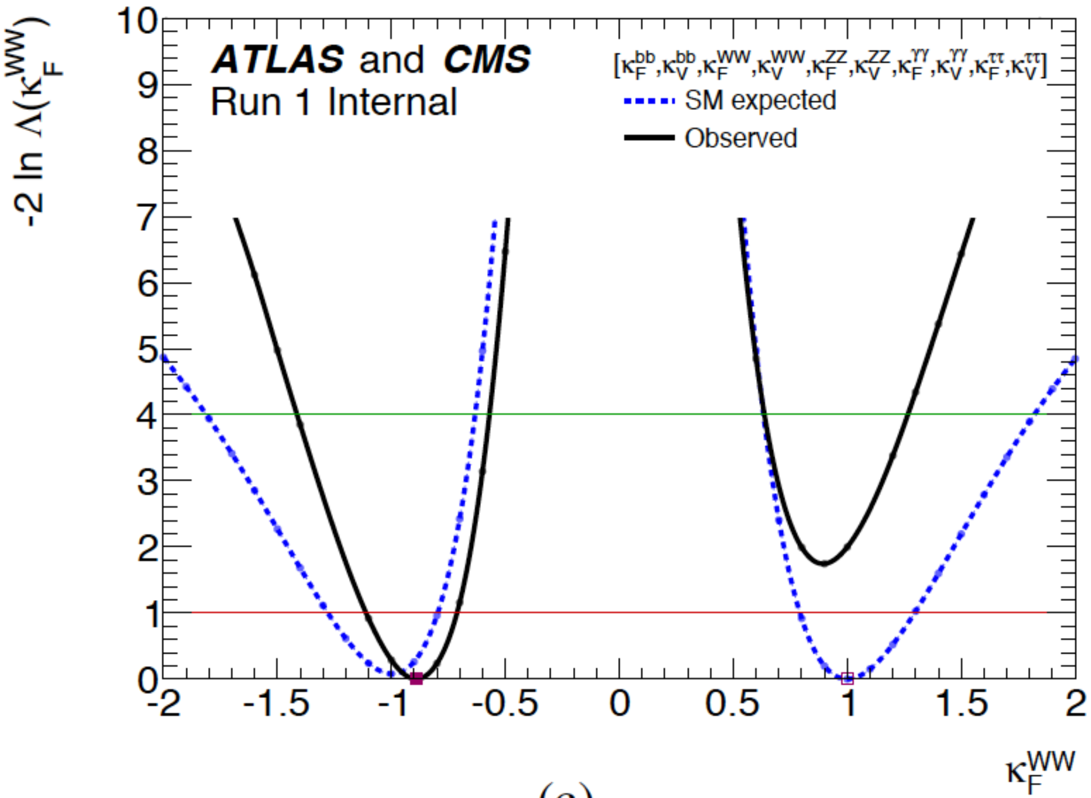
$$\Lambda(\kappa_F,\kappa_V) = \frac{L(\kappa_F,\kappa_V,\hat{\hat{\vec{\theta}}}(\kappa_F,\kappa_V))}{L(\hat{\kappa}_F,\hat{\kappa}_V,\hat{\vec{\theta}})} .$$



68% CI is a tricky issue

Is the WW a better measurement than the combination?

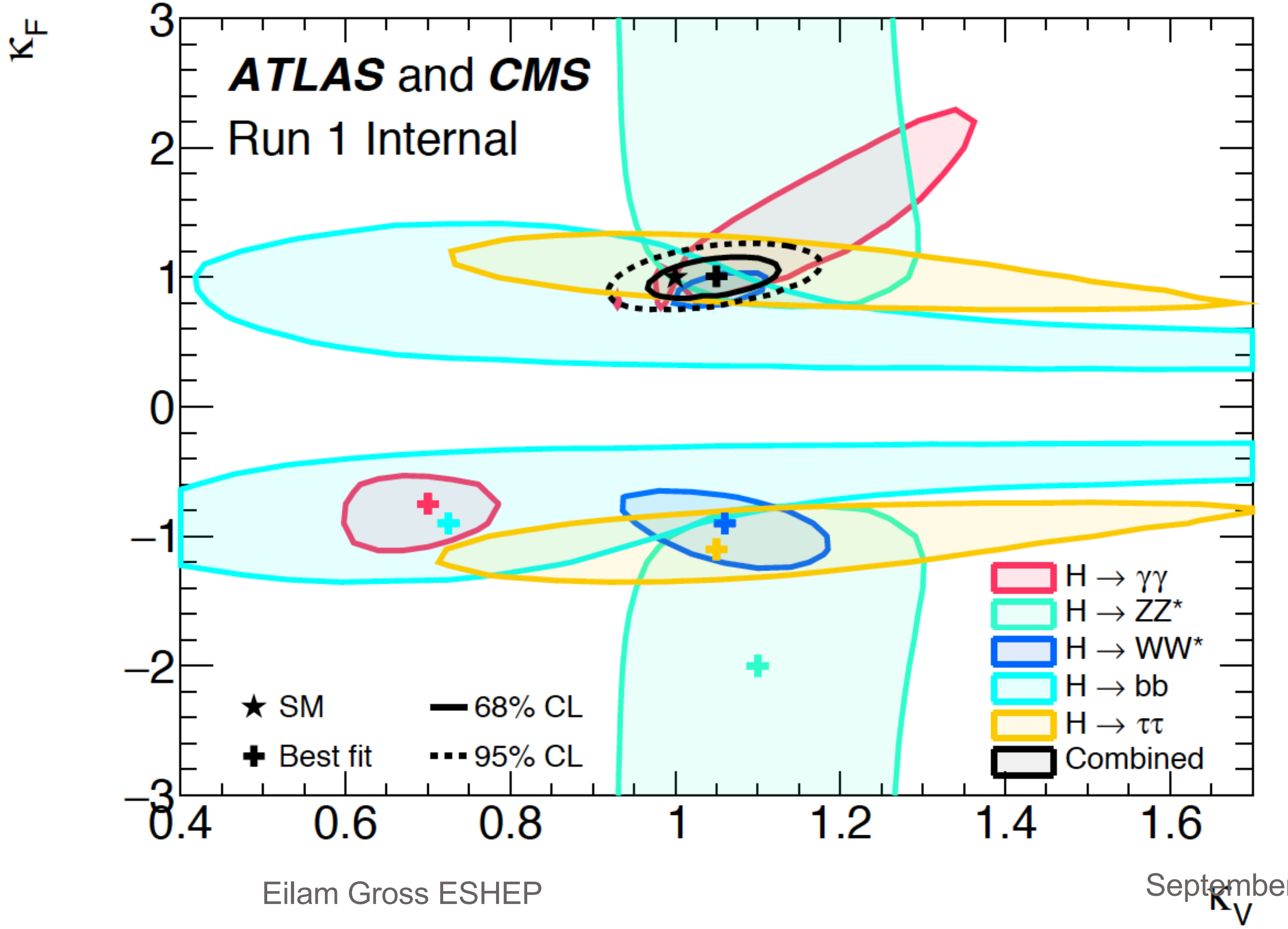
1D CI
Is not
2D CI



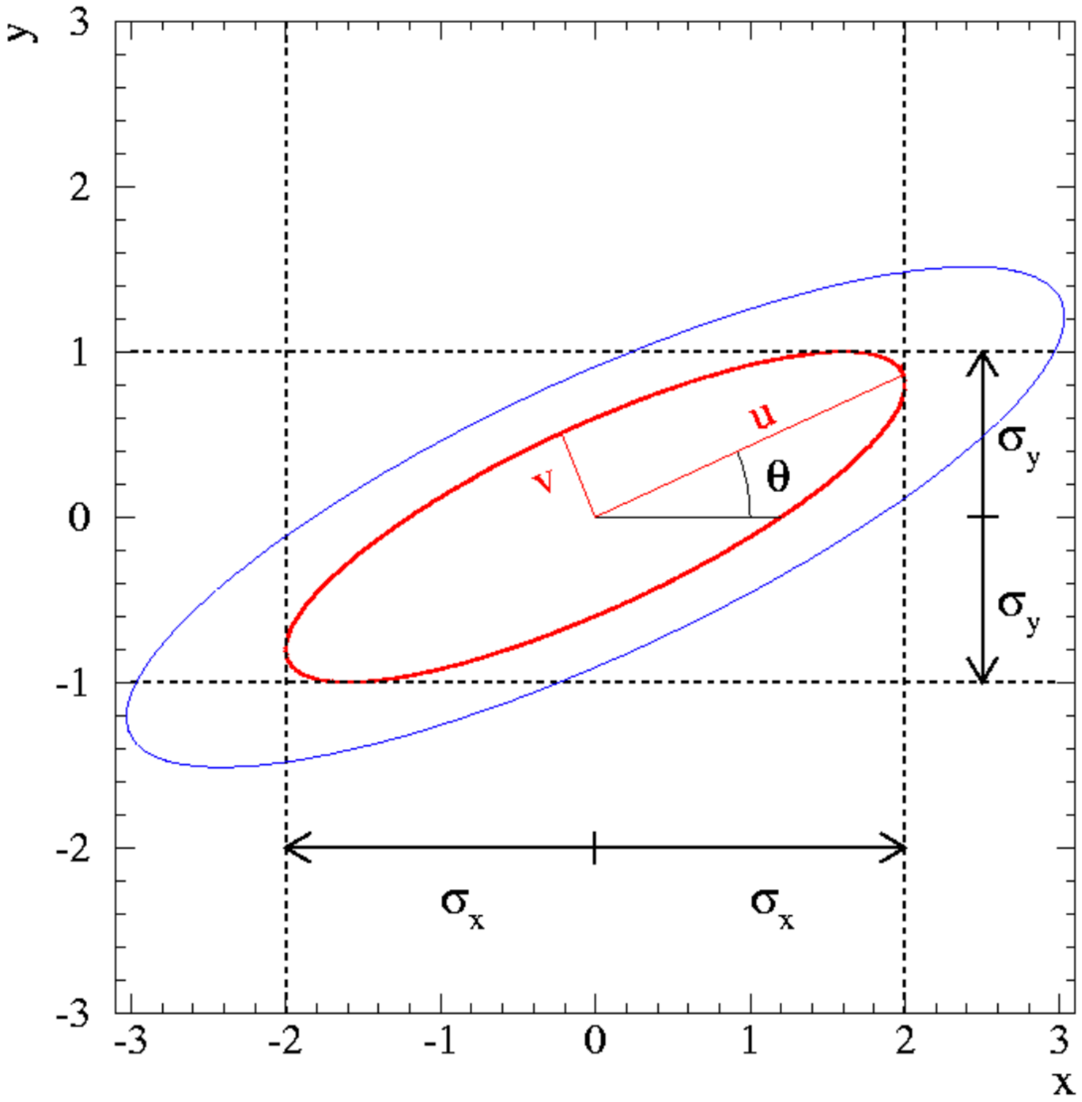
(c)

68% CI is a tricky issue

Is the WW a better measurement than the combination?



1D vs 2D Confidence Interval



$$\Delta\chi^2 = 1$$

$$\Delta\chi^2 = 2.3 \quad (68\%\ CL)$$

Multidimensional PL
A Tutorial

A toy case with 1-3 poi

3 cases studied

$1 poi : \mu$ while ϵ, A, b *profiled*

$2 poi : \mu, \epsilon$ *profile* A and b

3poi: μ, ϵ, A profile b

$$n = \mu \epsilon A s + b$$

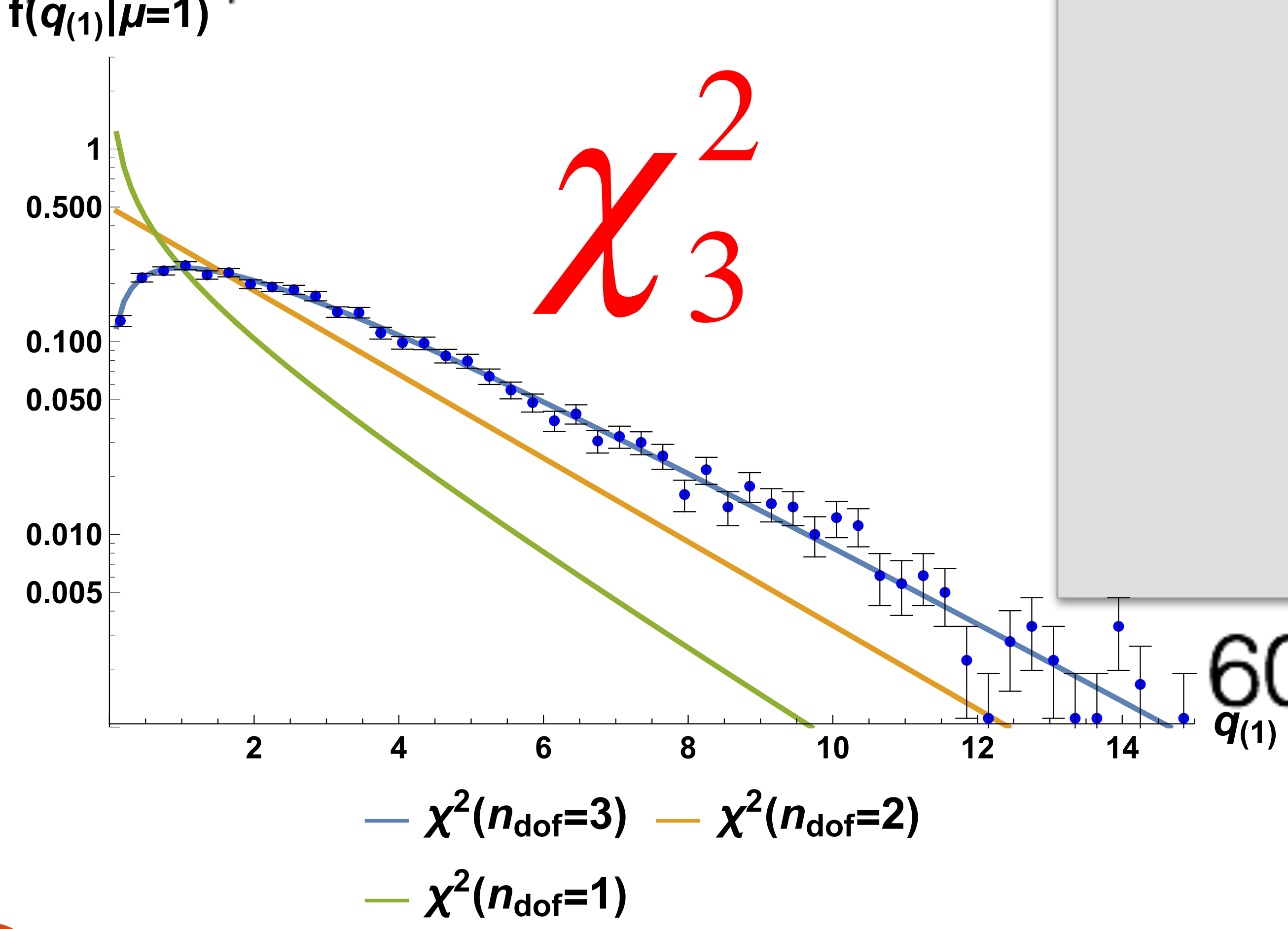
$$L = L(\mu, \epsilon, A, b)$$

$$L(\mu, \varepsilon, A) = \frac{(\mu \varepsilon A s + b)^n}{n!} e^{-(\mu \varepsilon A s + b)} \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-(\varepsilon_{meas} - \varepsilon)^2 / 2\sigma_\varepsilon^2} \frac{1}{\sigma_b \sqrt{2\pi}} e^{-(b_{meas} - b)^2 / 2\sigma_b^2} \frac{1}{\sigma_A \sqrt{2\pi}} e^{-(A_{meas} - A)^2 / 2\sigma_A^2}$$

A toy case with 3 poi

$$L(\mu,\varepsilon,A)=\frac{(\mu\varepsilon As+b)^n}{n!}e^{-(\mu\varepsilon As+b)}\frac{1}{\sigma_\varepsilon\sqrt{2\pi}}e^{-(\varepsilon_{meas}-\varepsilon)^2/2\sigma_\varepsilon^2}\frac{1}{\sigma_b\sqrt{2\pi}}e^{-(b_{meas}-b)^2/2\sigma_b^2}\frac{1}{\sigma_A\sqrt{2\pi}}e^{-(A_{meas}-A)^2/2\sigma_A^2}$$

three parameters of interest (profiling only b)
non-profiled parameters set to their real value



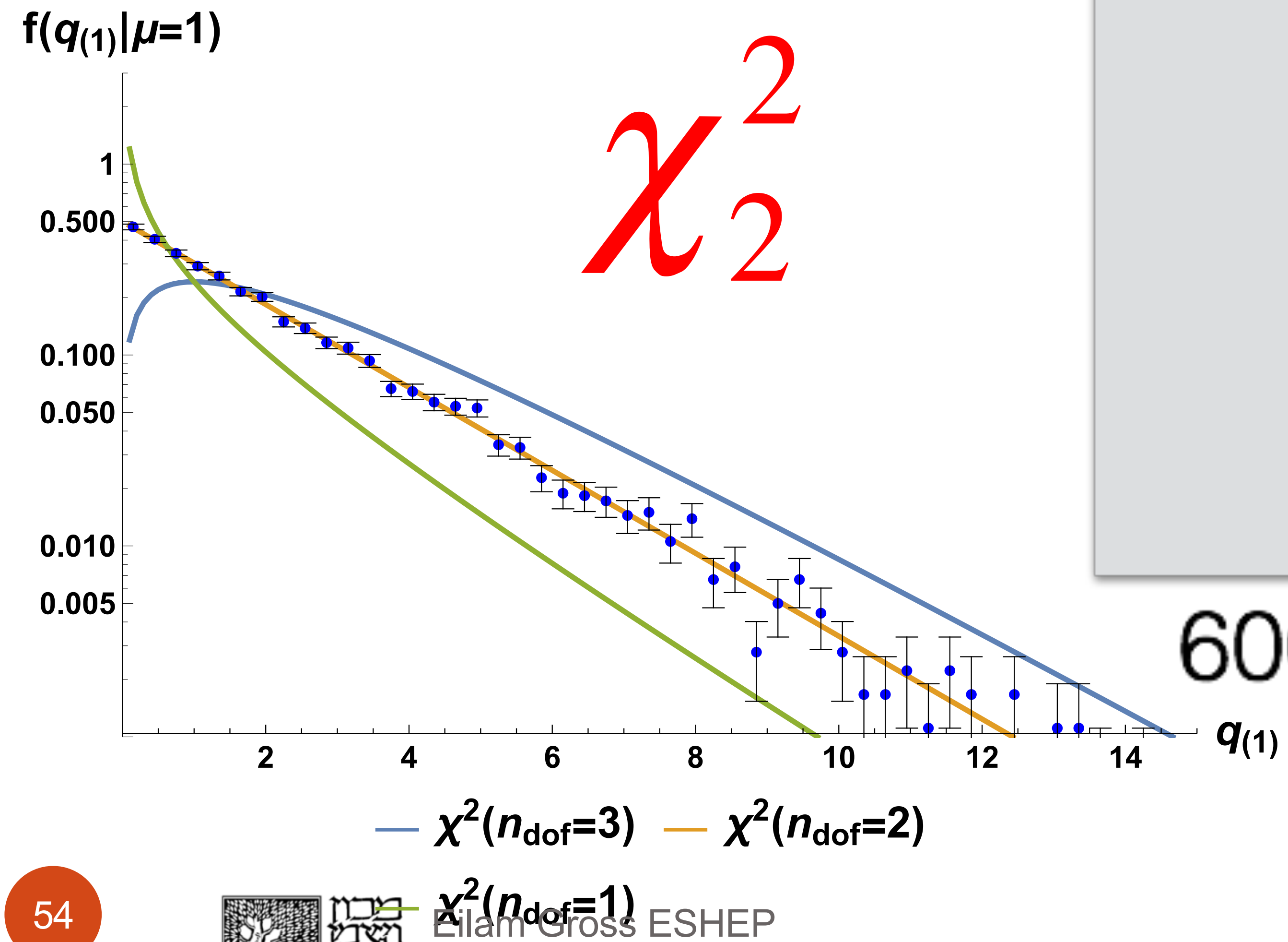
background = 100
signal = 90
$\varepsilon = 0.5$
A=0.7
σ_ε **= 0.05**
σ_b = 10
σ_A **= 0.2**

6000 events

A toy case with 2 poi

$$L(\mu,\varepsilon,A)=\frac{(\mu\varepsilon As+b)^n}{n!}e^{-(\mu\varepsilon As+b)}\frac{1}{\sigma_\varepsilon\sqrt{2\pi}}e^{-(\varepsilon_{meas}-\varepsilon)^2/2\sigma_\varepsilon^2}\frac{1}{\sigma_b\sqrt{2\pi}}e^{-(b_{meas}-b)^2/2\sigma_b^2}\frac{1}{\sigma_A\sqrt{2\pi}}e^{-(A_{meas}-A)^2/2\sigma_A^2}$$

two parameters of interest (profiling A and b)
non-profiled parameters set to their real value



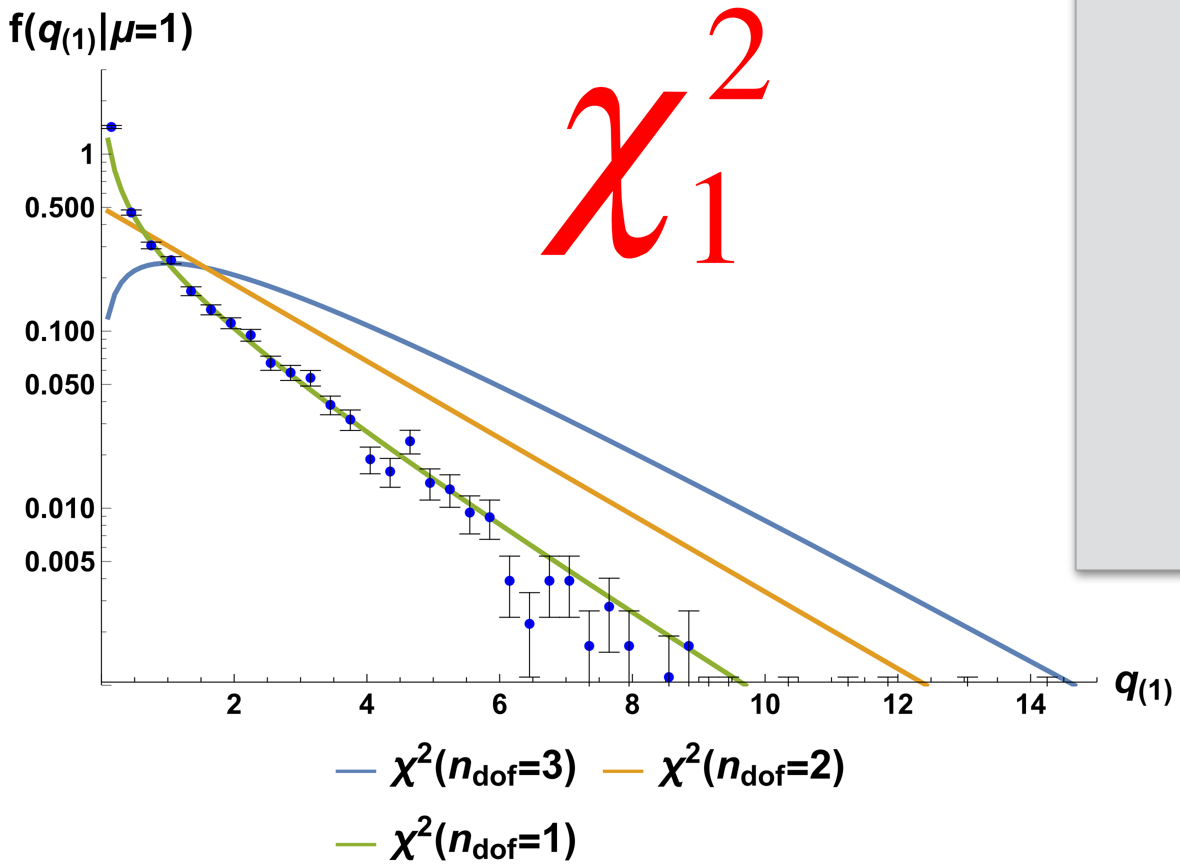
background = 100
signal = 90
ε = 0.5
A=0.7
= 0.05
= 10
= 0.2

6000 events

A toy case with 1 poi

$$L(\mu,\varepsilon,A)=\frac{(\mu\varepsilon As+b)^n}{n!}e^{-(\mu\varepsilon As+b)}\frac{1}{\sigma_\varepsilon\sqrt{2\pi}}e^{-(\varepsilon_{meas}-\varepsilon)^2/2\sigma_\varepsilon^2}\frac{1}{\sigma_b\sqrt{2\pi}}e^{-(b_{meas}-b)^2/2\sigma_b^2}\frac{1}{\sigma_A\sqrt{2\pi}}e^{-(A_{meas}-A)^2/2\sigma_A^2}$$

one parameter of interest (profiling ε A and b)
non-profiled parameters set to their real value



background = 100
signal = 90
ε = 0.5
A=0.7
σ_ε **= 0.05**
σ_b = 10
σ_A **= 0.2**

6000 events

Significance

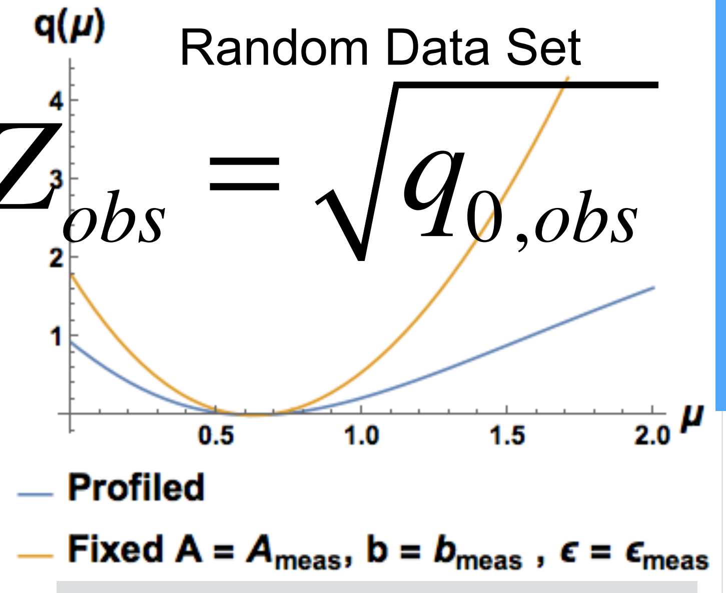
random data set

$b_{meas} = 106.84$
$\varepsilon_{meas} = 0.523$
$A_{meas} = 0.477$
$\mu_{meas} = 0.629$
$n_{meas} = 121$

For the fixed data set
The Nuisance Parameters
Are fixed to their nominal values.
The likelihood are more parabolic,
yet, never symmetric
The asymptotic hold!

Random Data Set

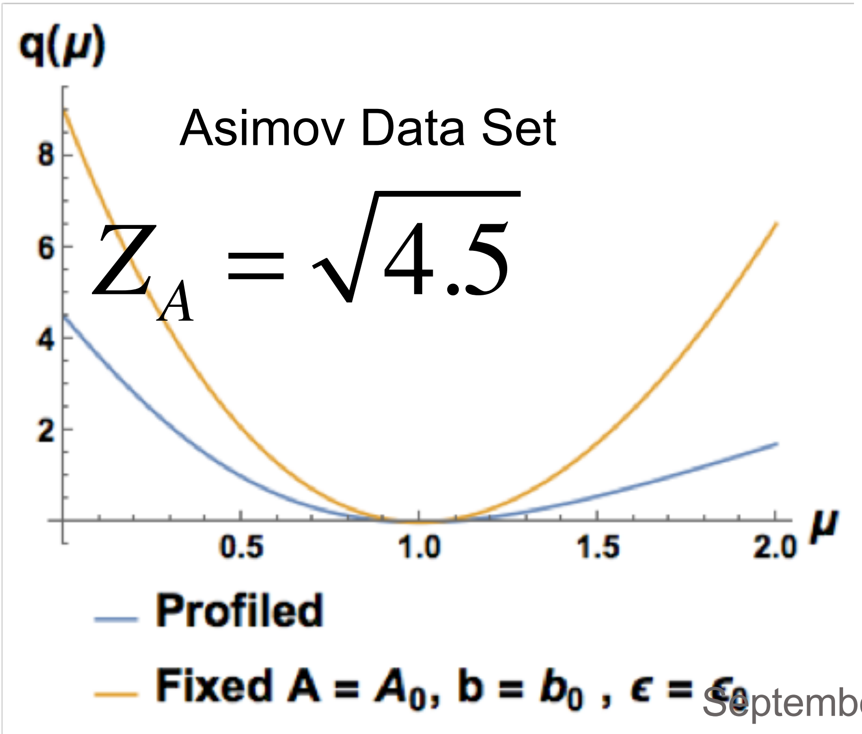
$$Z_{obs} = \sqrt{q_{0,obs}}$$



background = 100
signal = 90
$\varepsilon = 0.5$
A=0.7
σ_ε **= 0.05**
$\sigma_b = 10$
σ_A **= 0.2**

Asimov Data Set

$$Z_A = \sqrt{4.5}$$



Pulls and Ranking of NPs

The pull of θ_i is given by $\dfrac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0}$

without constraint $\sigma\left(\dfrac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0}\right) = 1 \quad \left\langle \dfrac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0} \right\rangle = 0$

It's a good habit to look at the pulls of the NPs and make sure that Nothing irregular is seen

In particular one would like to guarantee that the fits do not over constrain a NP in a non sensible way

Asimov

s=90

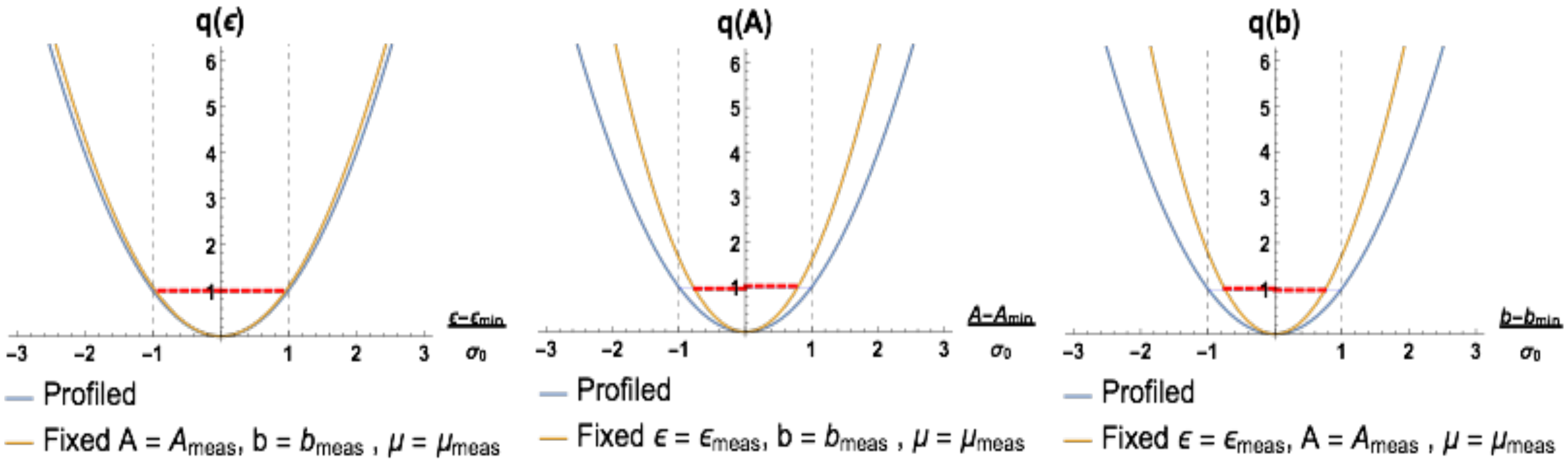
b_{meas} = 100
ε_{meas} = 0.5
A_{meas} =0.7
μ_{meas} = 1
n_{meas}=μsεA+b=131.5

reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90

σ_0

σ_ε = 0.05
σ_b = 10
σ_A = 0.2

To get the pulls:
–scan $q(\epsilon)$
–Find $\hat{\epsilon}$
–Find σ_ϵ^+ and σ_ϵ^- i.e. the poitive and negative error bar substituting $q(\epsilon)=1$



With the Asimov data sets we find perfect pulls for the profiled scans
But not for the fix scans!

Random Data Set

n_{meas} = 132

b_{meas} = 103.208

ϵ_{meas} = 0.465459

A_{meas} = 0.487107

μ_{meas} = 1.41099

reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90

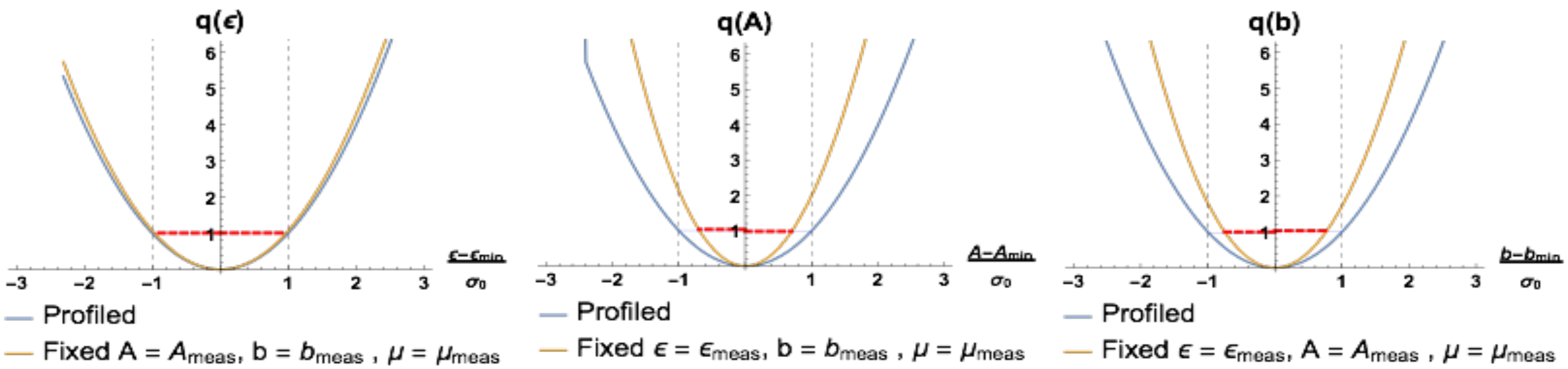
σ_0

σ_ε = 0.05

σ_b = 10

σ_A = 0.2

To get the pulls:
–scan q(ϵ)
–Find $\hat{\epsilon}$
–Find σ_ϵ^+ and σ_ϵ^- i.e. the poitive and negative error bar substituting q(ϵ)=1



With the random data sets we find perfect pulls for the profiled scans
But not for the fix scans!

Back to Asimov: Find the Impact of a NP

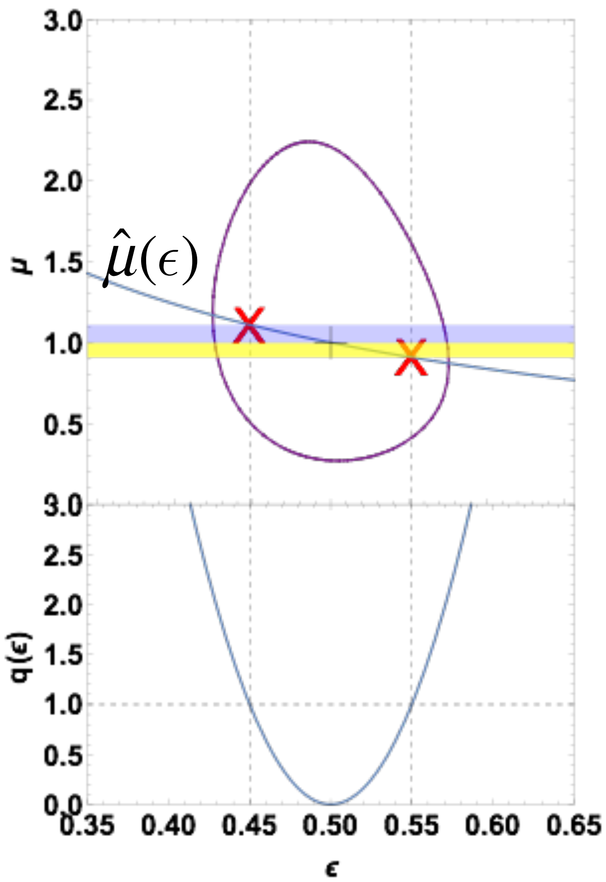
b_{meas} = 100
ε_{meas} = 0.5
A_{meas} =0.7
μ_{meas} = 1
$n_{meas}=\mu s\varepsilon A+b=131.5$

reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90

σ_0
σ_ε = 0.05
σ_b = 10
σ_A = 0.2

To get the impact of a Nuisance Parameter in order to rank them:

Say we want the impact of ϵ
–Scan q(ϵ), profiling all other NPs
–Find $\hat{\epsilon}$
–(note that $\hat{\mu}_{\hat{\epsilon}} = \hat{\mu}$)
–Find $\hat{\mu}_{\hat{\epsilon}\pm\sigma_\epsilon^\pm} = \hat{\hat{\mu}}_{\hat{\epsilon}\pm\sigma_\epsilon^\pm}$
–The impact is given by $\Delta\mu^\pm = \hat{\hat{\mu}}_{\hat{\epsilon}\pm\sigma_\epsilon^\pm} - \hat{\mu}$



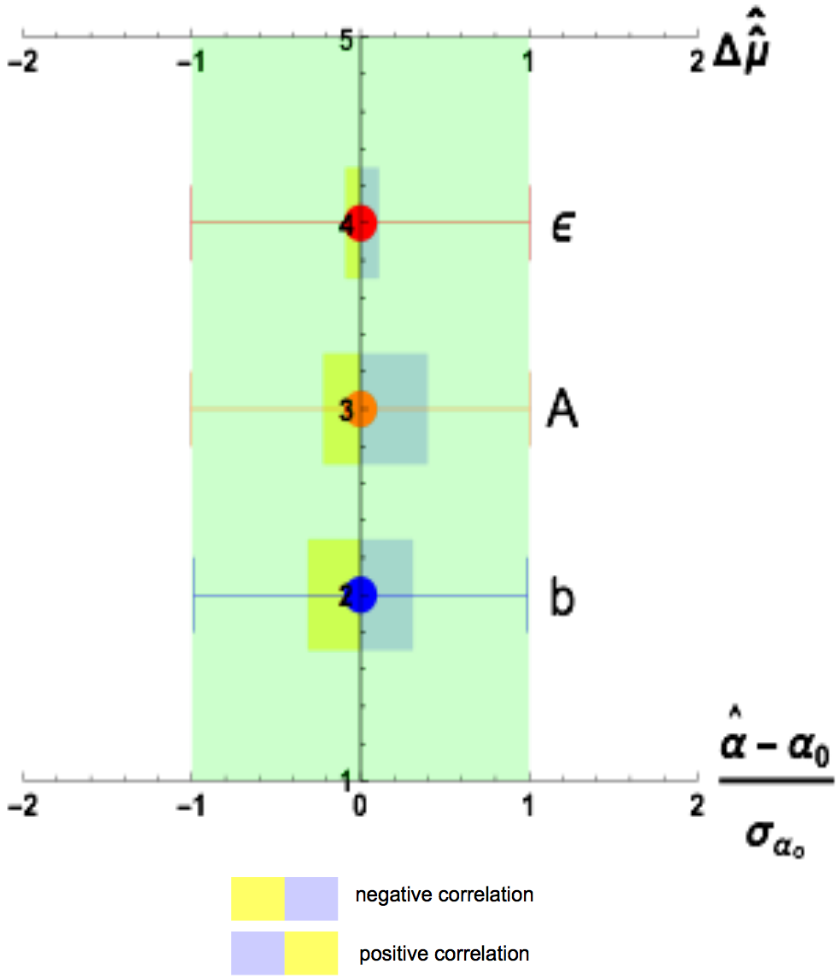
Asimov: SUMMARY of Pulls and Impact

b_{meas} = 100
ε_{meas} = 0.5
A_{meas} =0.7
μ_{meas} = 1
$n_{meas}=\mu s \varepsilon A+b=131.5$

reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90

σ_0

$\sigma_\varepsilon = 0.05$
$\sigma_b = 10$
$\sigma_A = 0.2$



Random Data Set: SUMMARY of Pulls and Impact

n_{meas} = 132

b_{meas} = 103.208

ϵ_{meas} = 0.465459

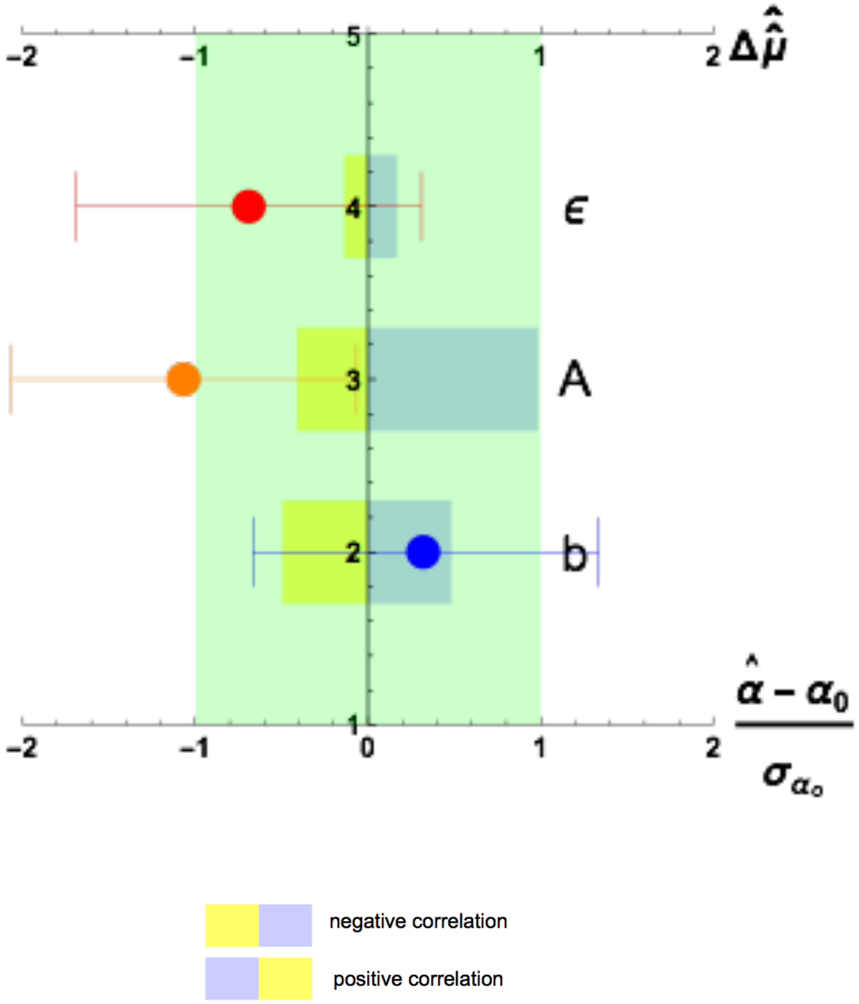
A_{meas} = 0.487107

μ_{meas} = 1.41099

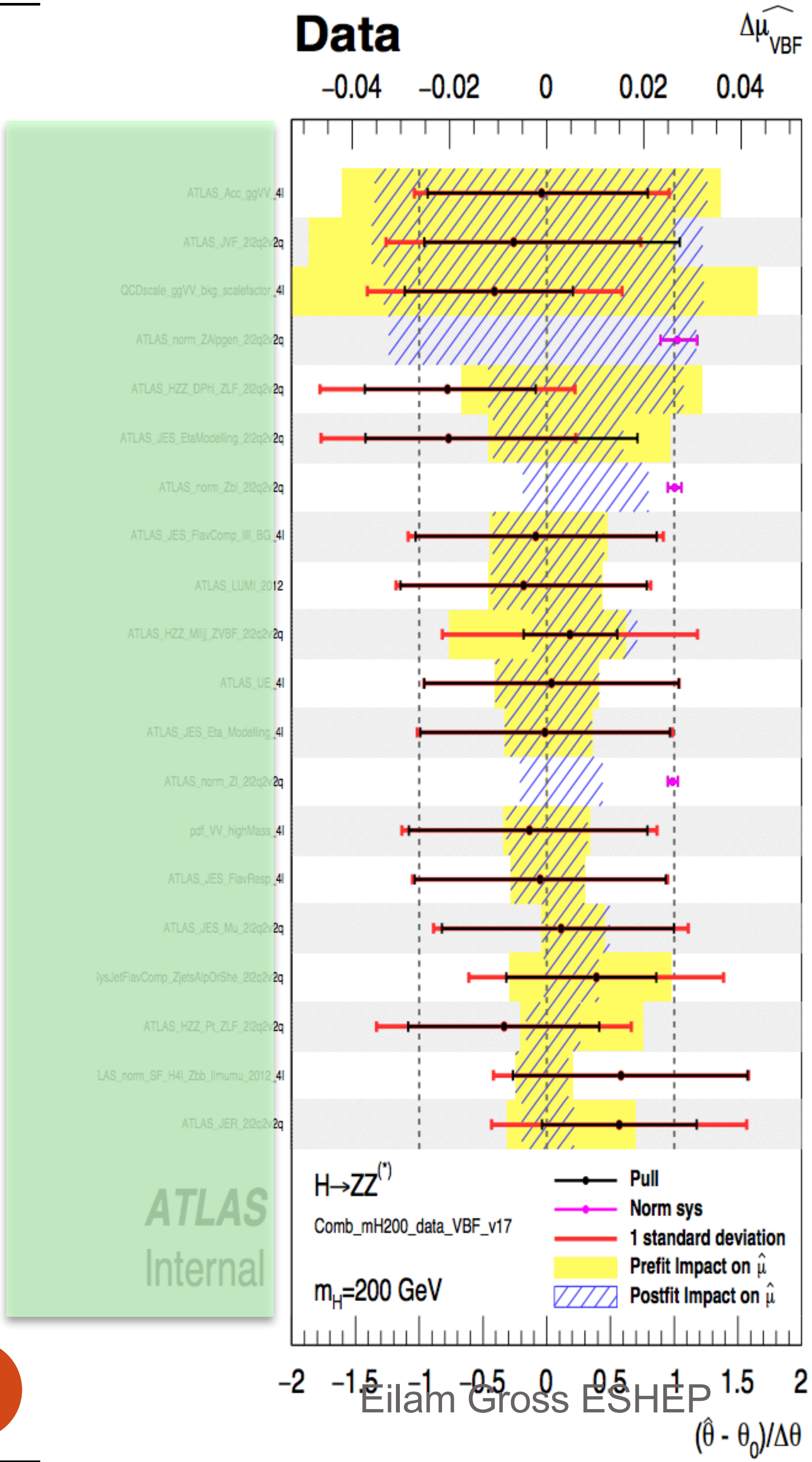
reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90

σ_0

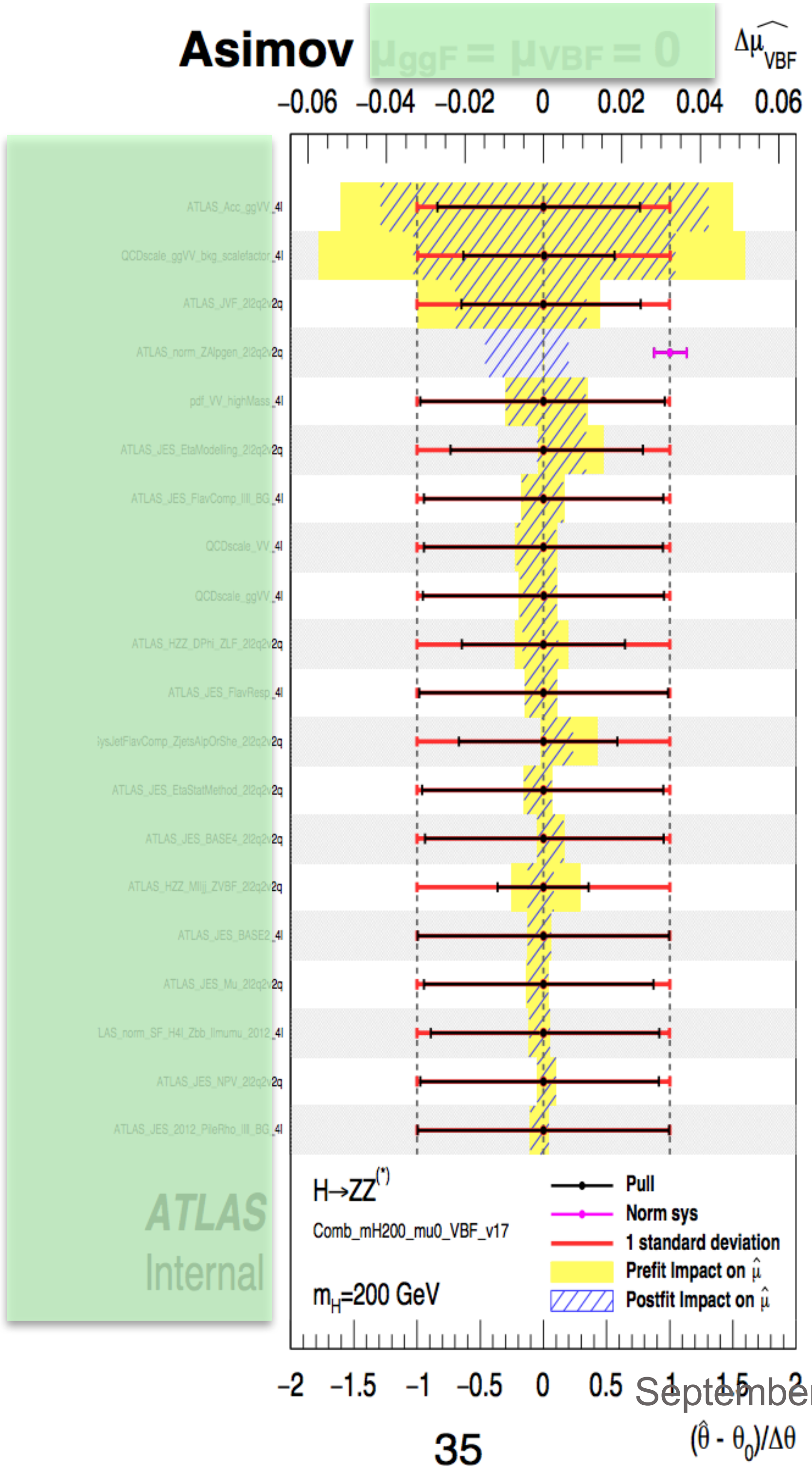
σ_ε = 0.05
σ_b = 10
σ_A = 0.2



Real Examples



Data
$\widehat{\Delta\mu}_{VBF}$
ATLAS
Internal
$H\rightarrow ZZ^{(*)}$
Comb_mH200_data_VBF_v17
m_H=200 GeV
Pull
Norm sys
1 standard deviation
Prefit Impact on $\hat{\mu}$
Postfit Impact on $\hat{\mu}$
$(\hat{\theta} - \theta_0)/\Delta\theta$



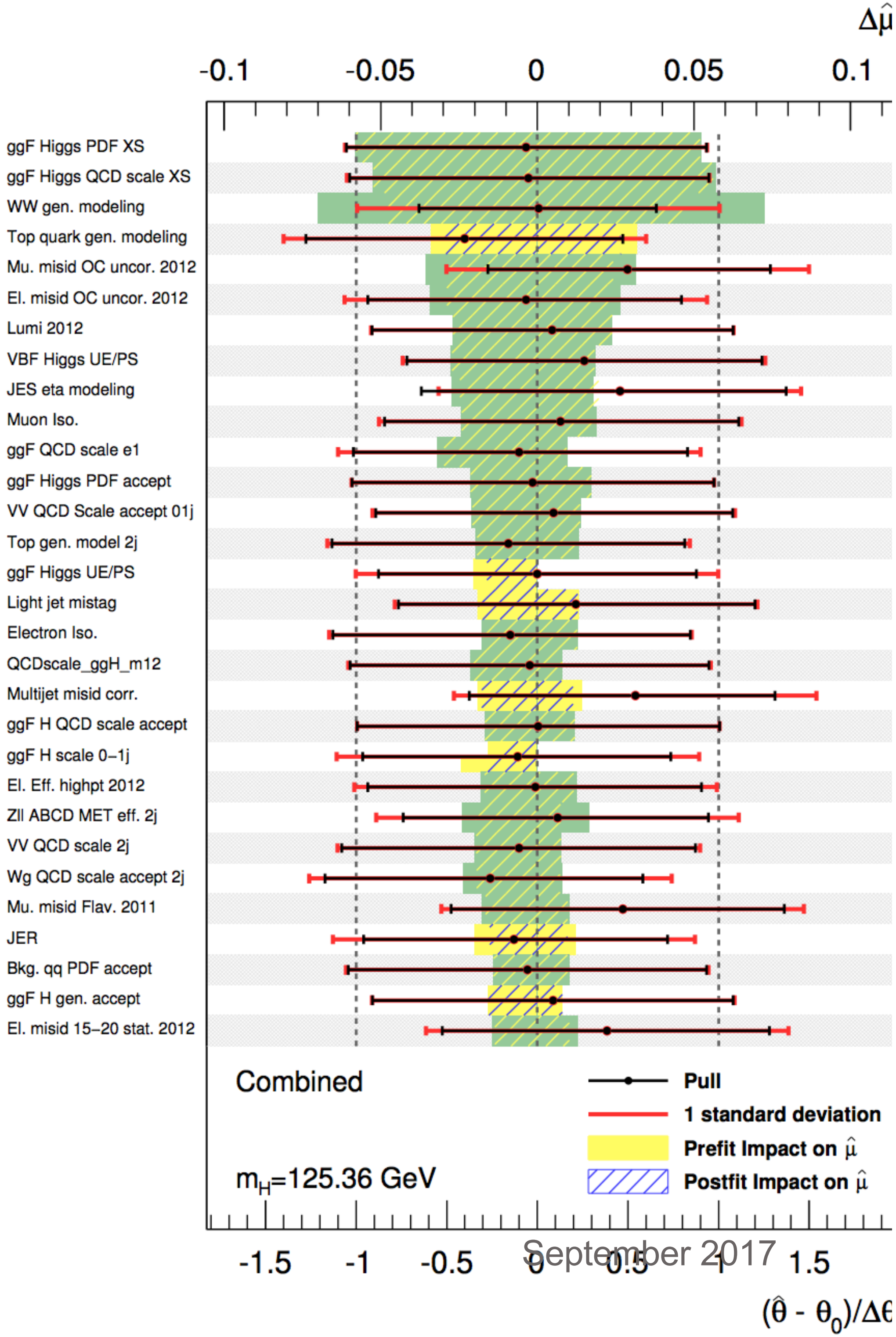
Asimov
$\widehat{\Delta\mu}_{VBF}$
ATLAS
Internal
$H\rightarrow ZZ^{(*)}$
Comb_mH200_mu0_VBF_v17
m_H=200 GeV
Pull
Norm sys
1 standard deviation
Prefit Impact on $\hat{\mu}$
Postfit Impact on $\hat{\mu}$
$(\hat{\theta} - \theta_0)/\Delta\theta$

Pulls and Ranking (

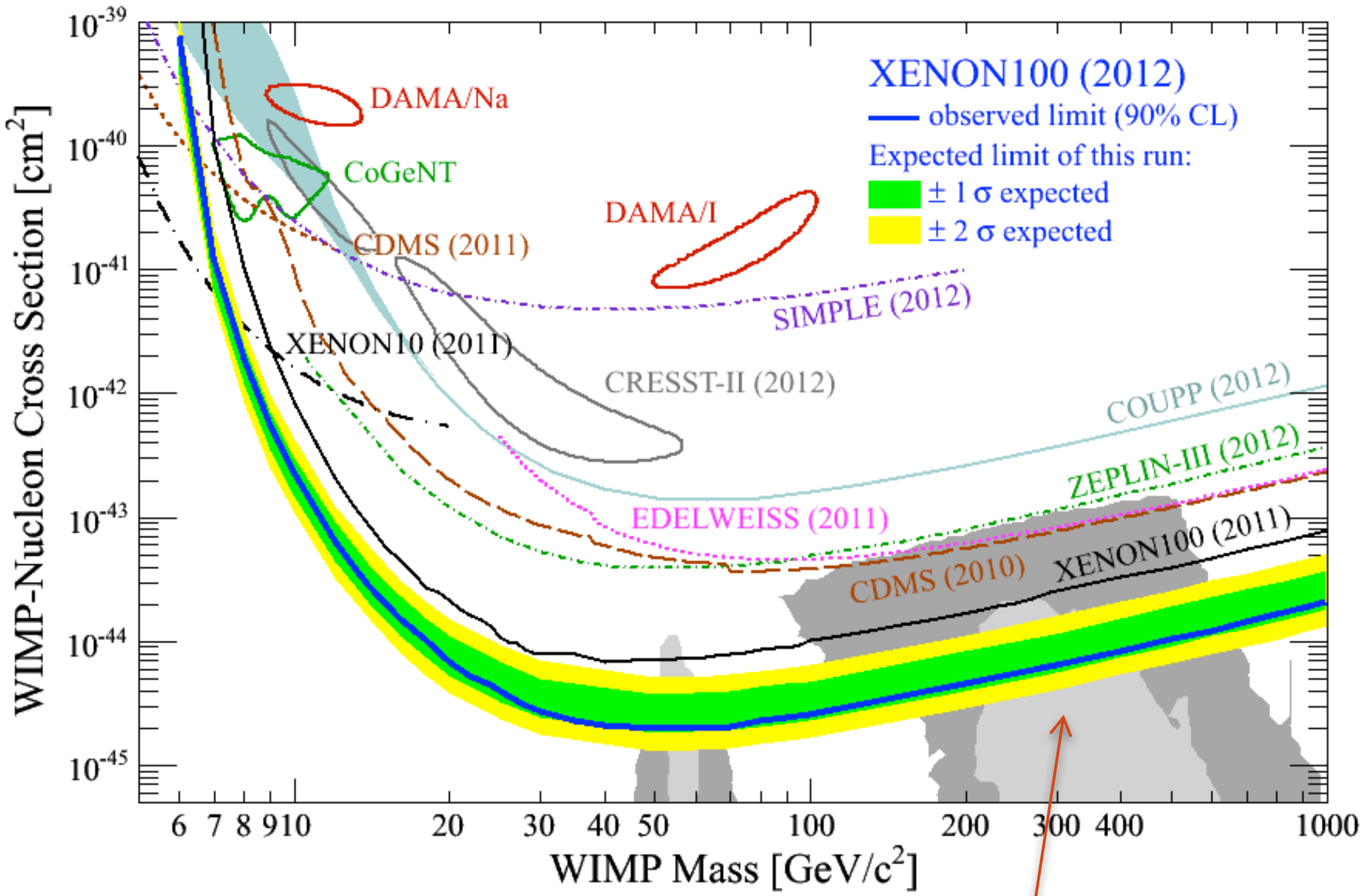
Ranking θ_i by its effect in the NP

$$\Delta\mu^{\pm} = \hat{\hat{\mu}}_{\hat{\epsilon}\pm\sigma_{\epsilon}^{\pm}} - \hat{\mu}$$

By ranking we can tell which NPs are the important ones and which can be pruned



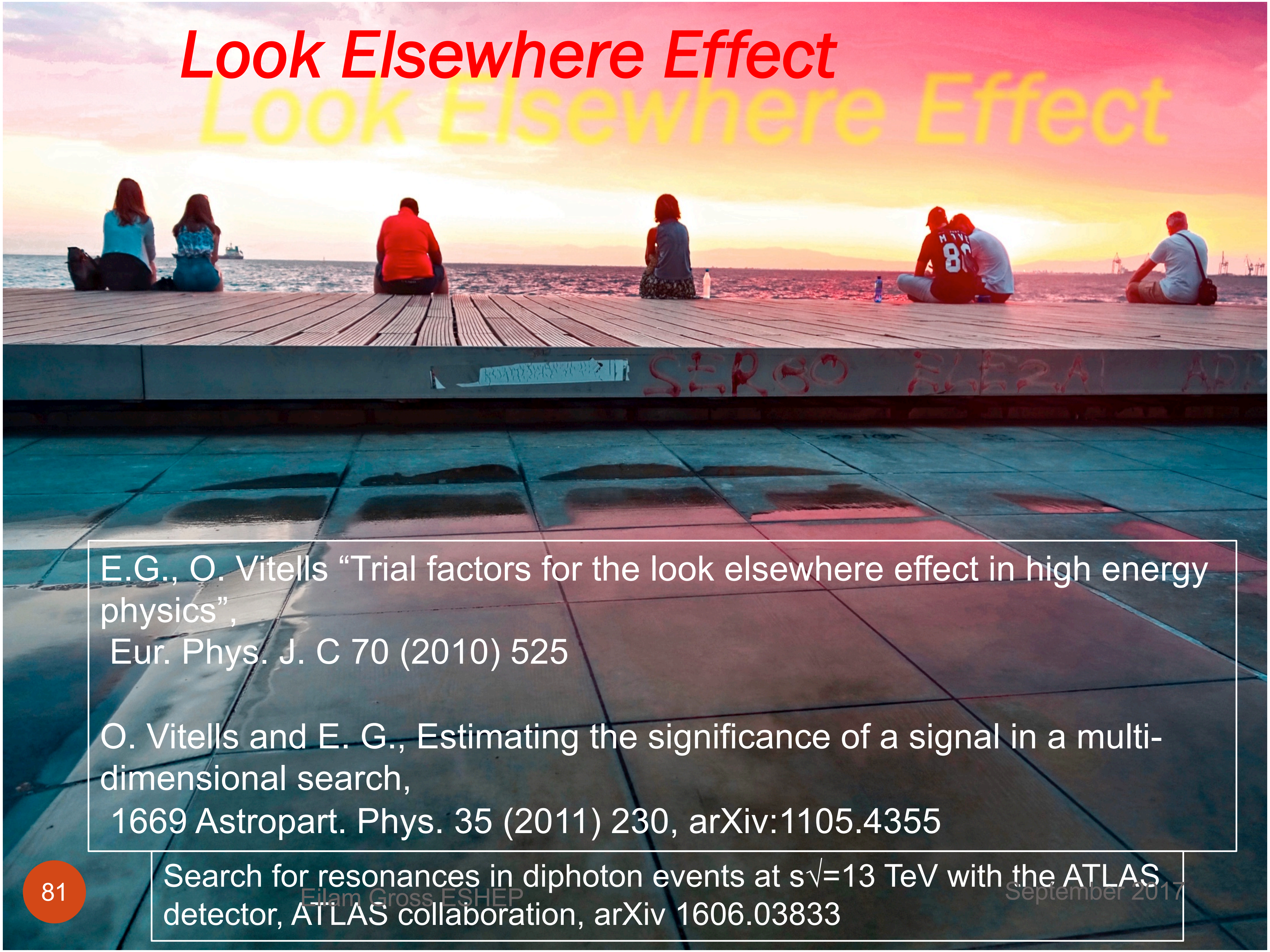
Implications in Astro-Particle Physics



The lack of events in spite of an expected background allows us to set a better limit than the expected

Look Elsewhere Effect

Look Elsewhere Effect



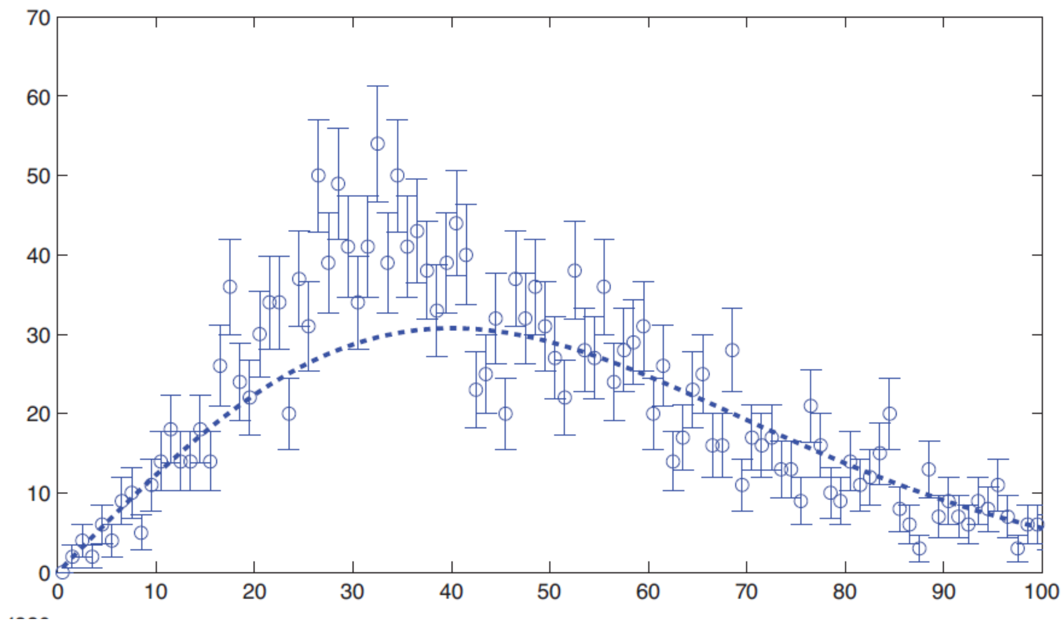
E.G., O. Vitells "Trial factors for the look elsewhere effect in high energy physics",
Eur. Phys. J. C 70 (2010) 525

O. Vitells and E. G., Estimating the significance of a signal in a multi-dimensional search,
1669 Astropart. Phys. 35 (2011) 230, arXiv:1105.4355

Search for resonances in diphoton events at s√=13 TeV with the ATLAS detector, ATLAS collaboration, arXiv 1606.03833

Look Elsewhere Effect

Is there a signal here?



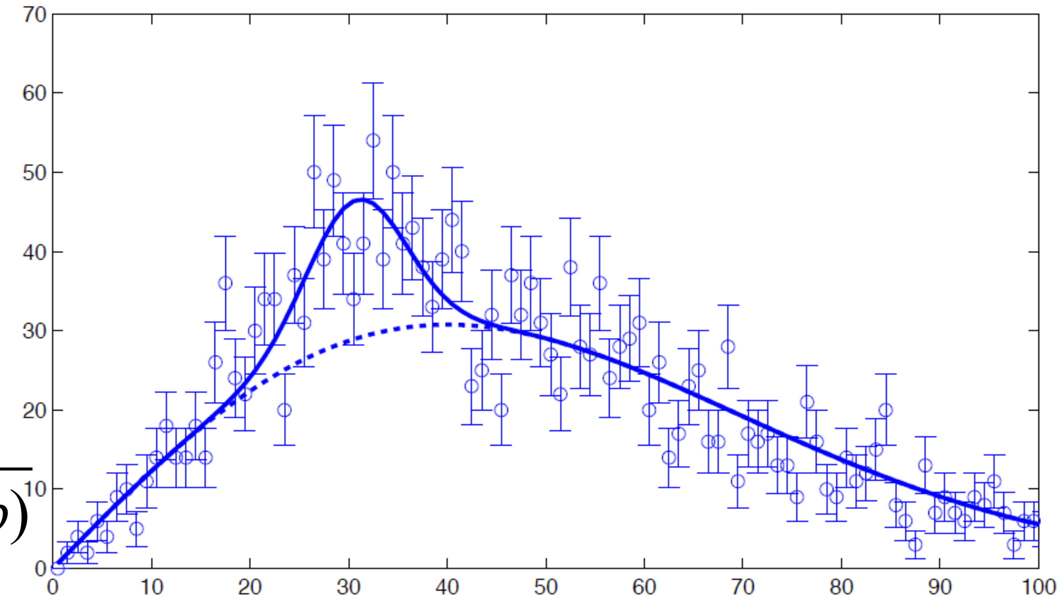
Look Elsewhere Effect

Looks like @ m=30

What is its significance?

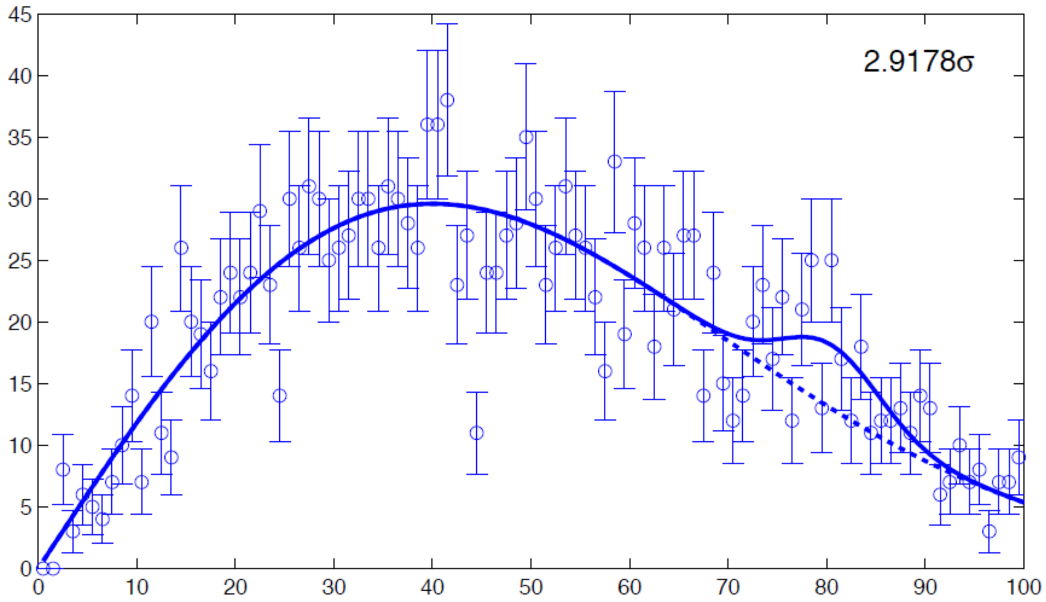
What is your test statistic?

$$q_{fix,obs} = -2\ln\frac{L(b)}{L(\hat{\mu}s(m=30)+b)}$$



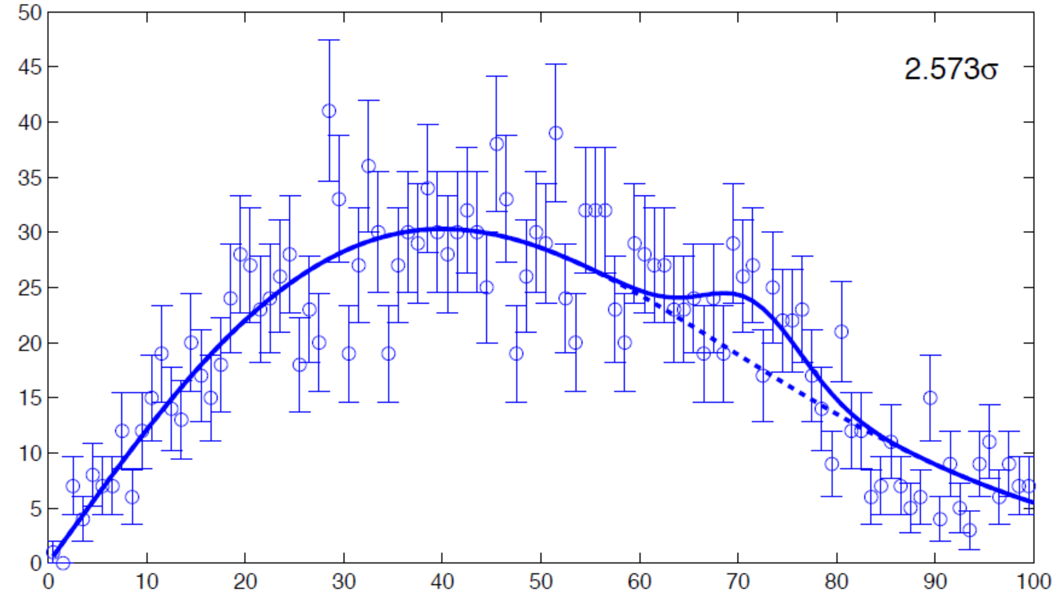
Look Elsewhere Effect

Would you ignore this signal, had you seen it?



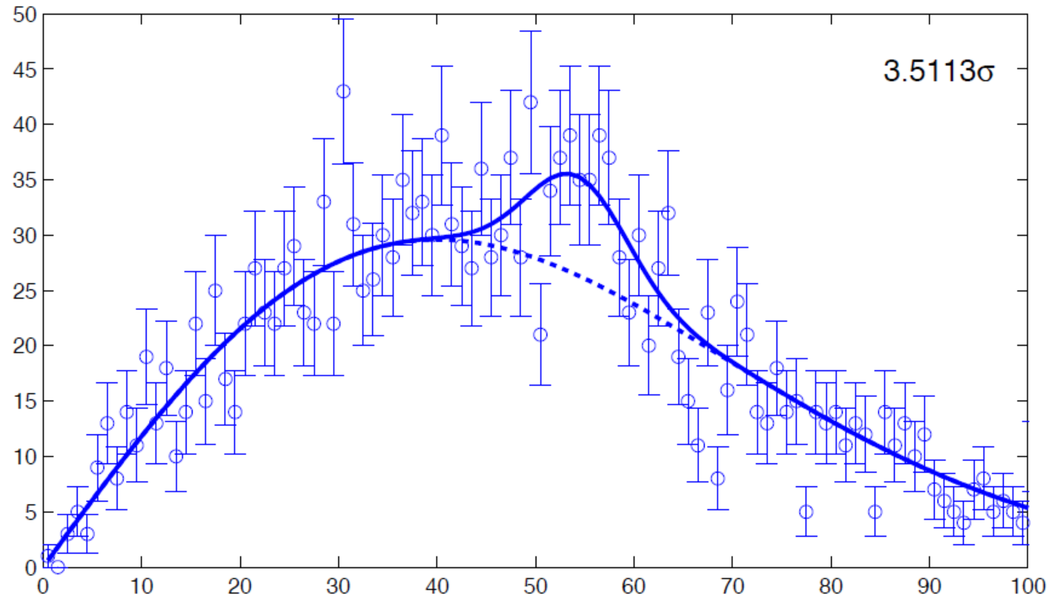
Look Elsewhere Effect

Or this?



Look Elsewhere Effect

Or this?

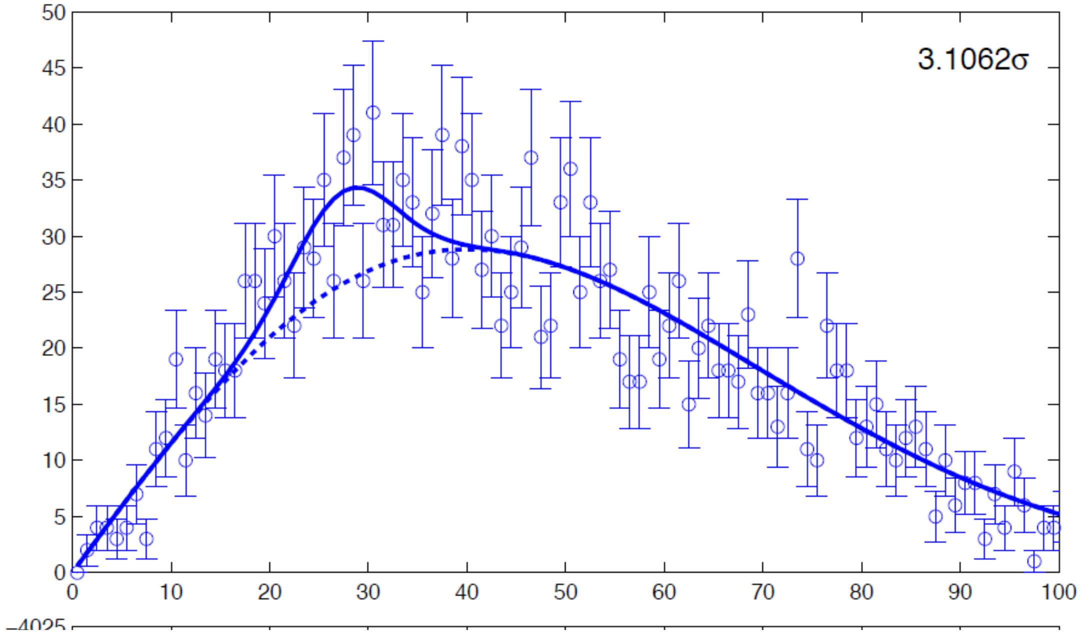


Look Elsewhere Effect

Or this?

Obviously NOT!

ALL THESE
“SIGNALS” ARE
BG
FLUCTUATIONS



Look Elsewhere Effect

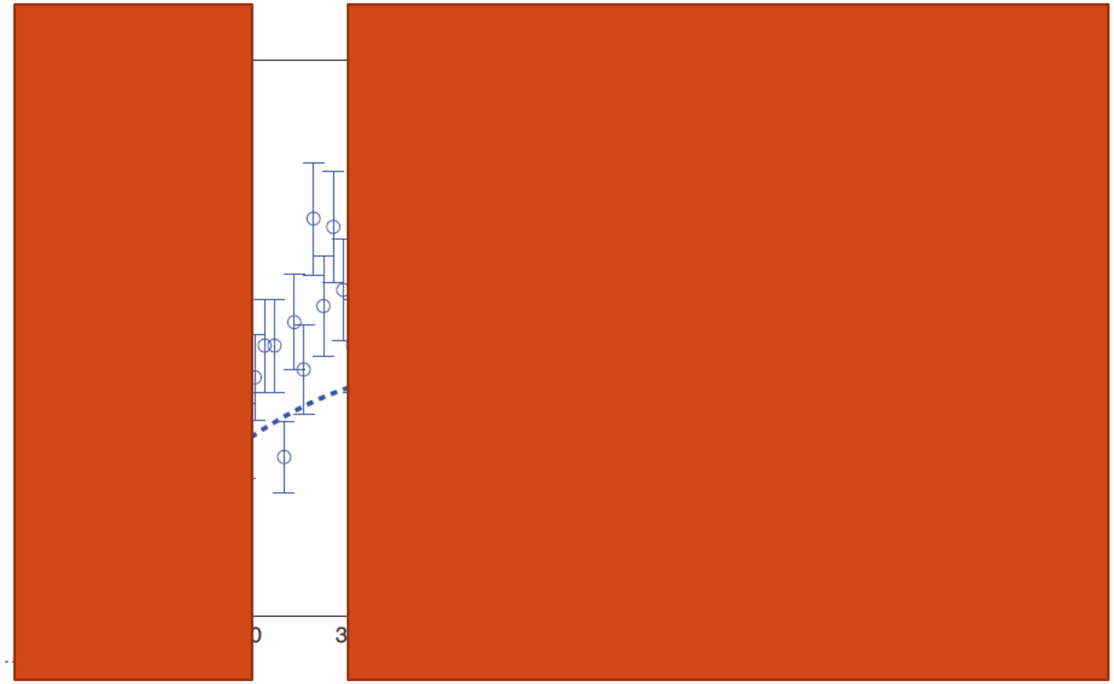
Having no idea where the signal might be there are two options

OPTION I:

scan the mass range in pre-defined steps and test any disturbing fluctuations

(do not confuse me with the facts)

Perform a fixed mass analysis at each point



$$q_{fix,obs}(\hat{\mu}) = -2\ln\frac{L(b)}{L(\hat{\mu}s(m)+b)}$$

Look Elsewhere Effect

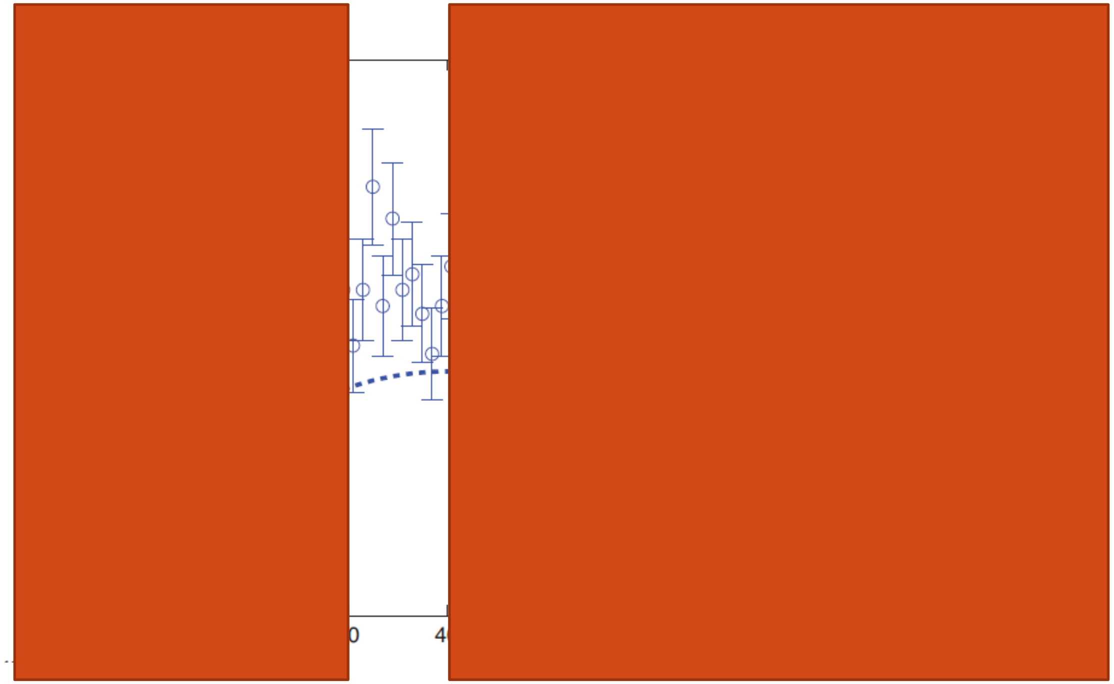
Having no idea where the signal might be there are two options

OPTION I:

scan the mass range in pre-defined steps and test any disturbing fluctuations

(do not confuse me with the facts)

Perform a fixed mass analysis at each point

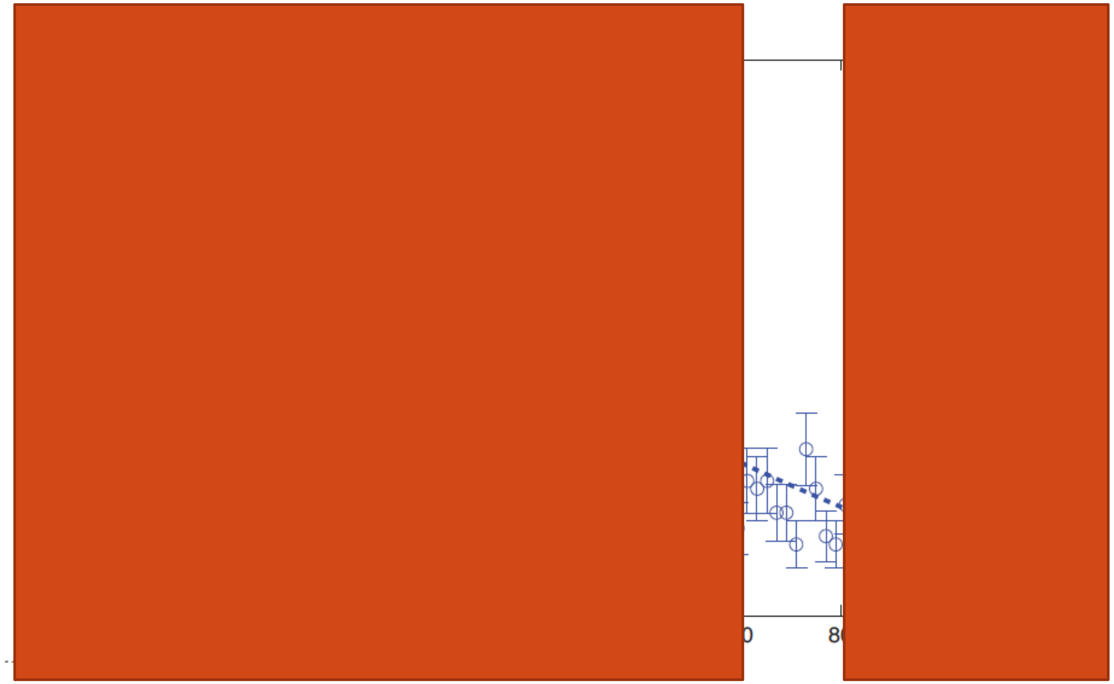


$$q_{fix,obs}(\hat{\mu}) = -2\ln\frac{L(b)}{L(\hat{\mu}s(m)+b)}$$

Look Elsewhere Effect

The scan resolution must be less than the signal mass resolution

Assuming the signal can be only at one place, pick the one with the smallest p-value (maximum significance)



$$q = \max_m \left\{ q_{fix,obs}(\hat{\mu}) \right\} = \max_m \left\{ -2\ln \frac{L(b)}{L(\hat{\mu}s(m)+b)} \right\}$$

$$= \min_m \left\{ p - value \right\}$$

Look Elsewhere Effect

Test statistic

$$q_{fix,obs} = -2\ln\frac{L(b)}{L(\hat{\mu}s(m=30)+b)}$$

What is the p-value?

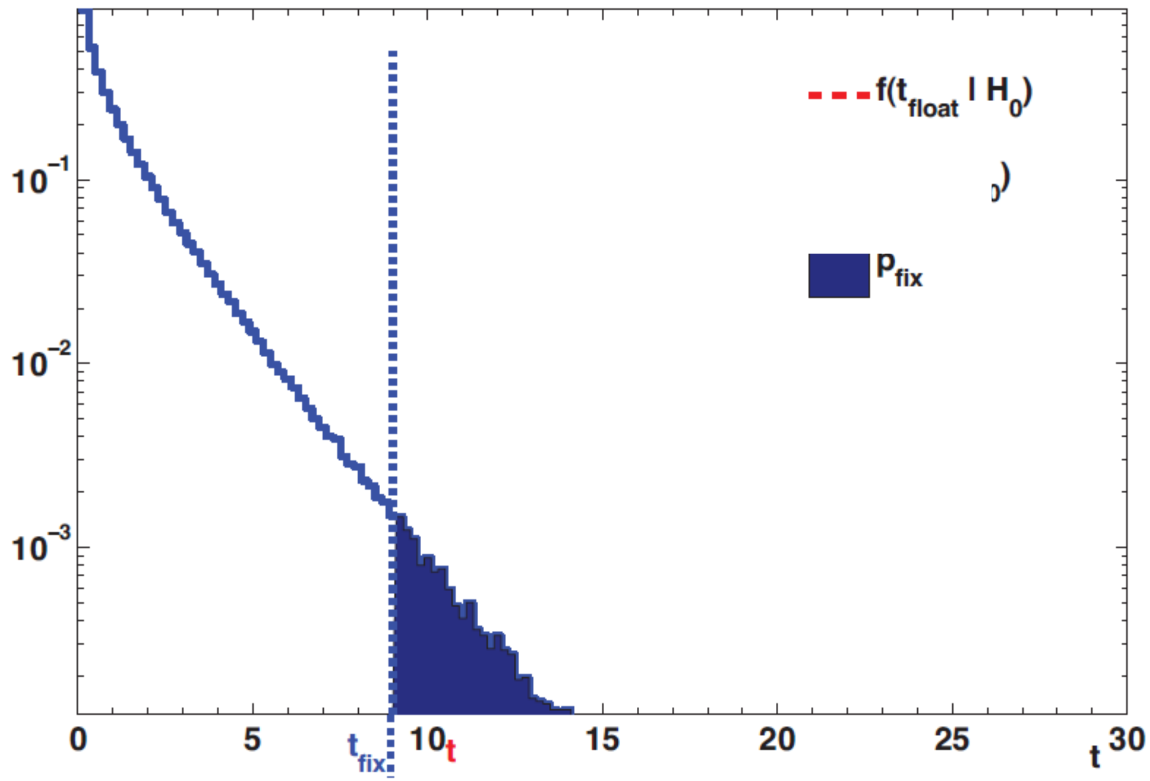
generate the PDF

$$f(q_{fix}\,|\,H_0)$$

and find the **p-value**

Wilks theorem:

$$f(q_{fix}\,|\,H_0) \sim \chi_1^2$$



$$p_{fix} = \int_{q_{fix,obs}}^{\infty} f(q_{fix}\,|\,H_0)\,dq_{fix}$$

Look Elsewhere Effect: Floating Mass

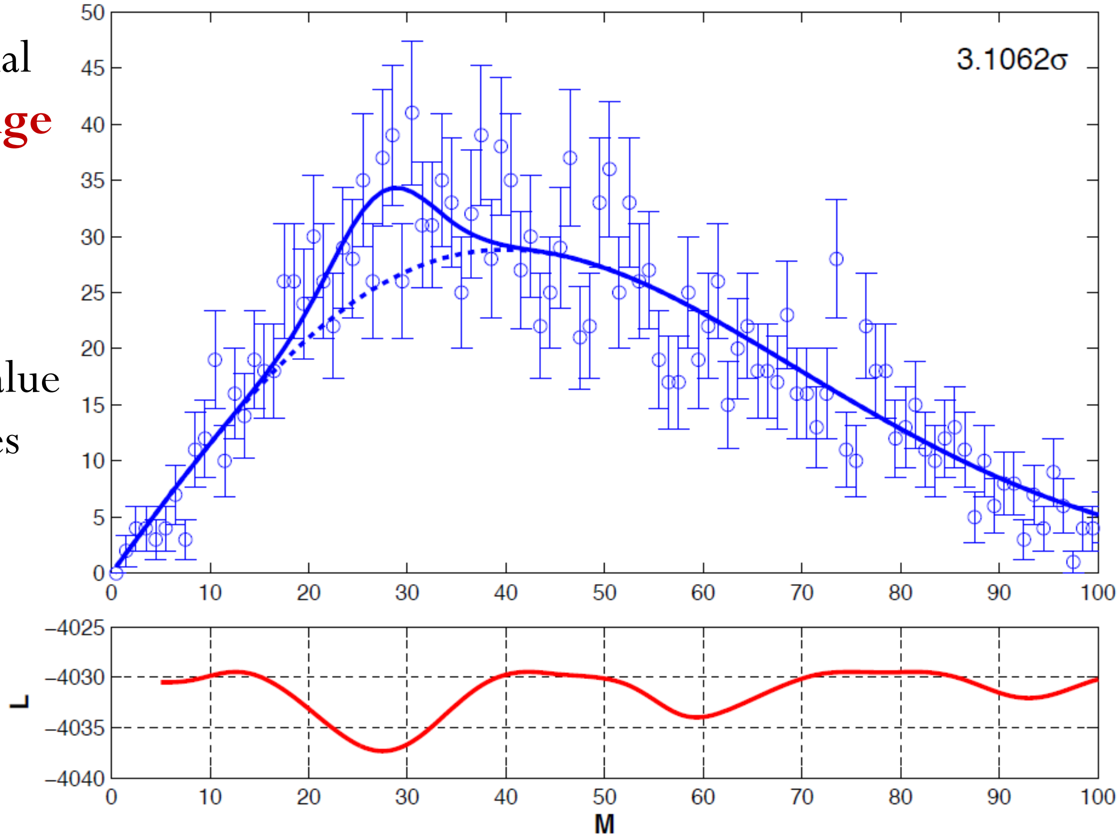
Option II:

Leave the mass floating

Having no idea where the signal might be you would allow the signal to be anywhere in the **search range** and use a modified test statistic

For the same observation, the p-value increases because more possibilities are opened

$$q_{float,obs}(\hat{\mu},\hat{m}) = -2\ln\frac{L(b)}{L(\hat{\mu}s(\hat{m})+b)}$$



Look Elsewhere Effect

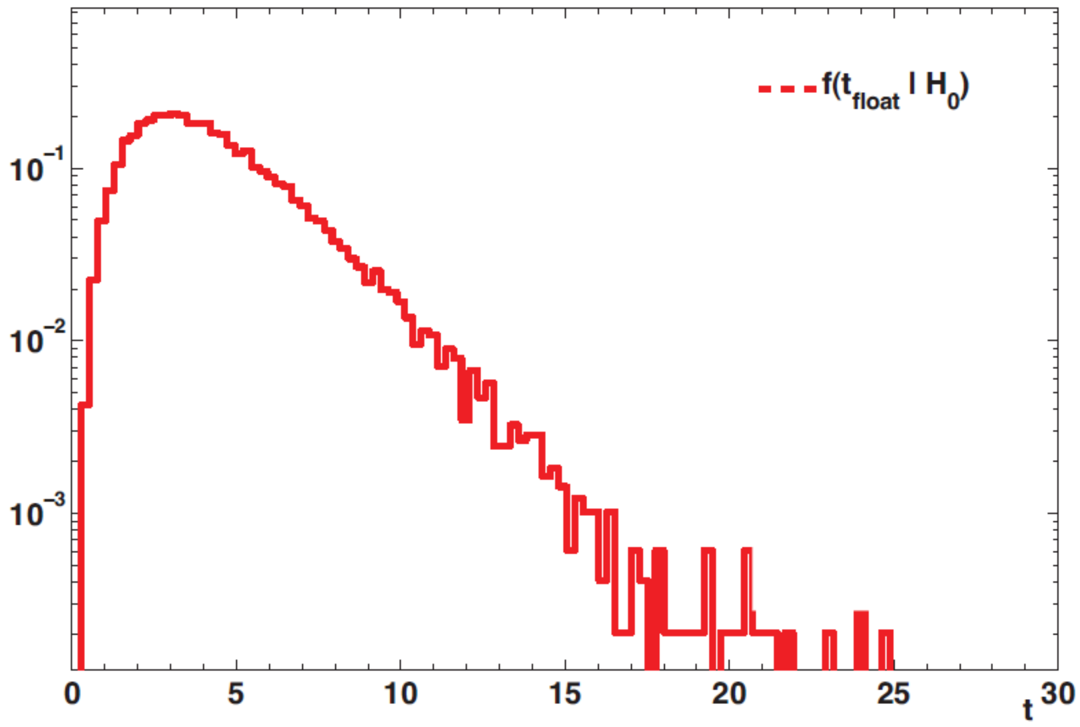
the test statistis

$$q_{float,obs}(\hat{\mu},\hat{m}) = -2\ln\frac{L(b)}{L(\hat{\mu}s(\hat{m})+b)}$$

The null hypothesis PDF

$$f(q_{float} \mid H_0)$$

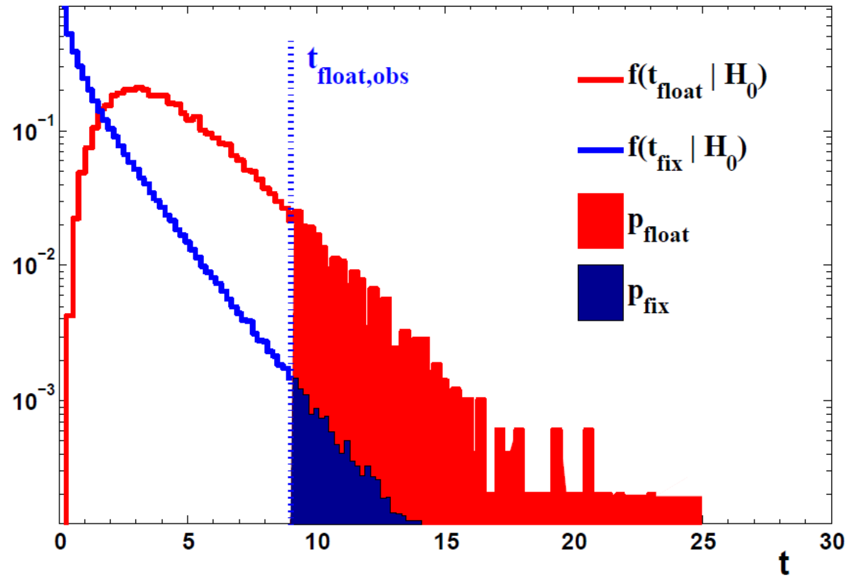
does not follow a chi-squared with 2dof because there are multiple minima depending on the size of the search range and resolution



trial#

Assume a maximal local fluctua
mass $\hat{m} = 30$

We can calculate the following



$$q_{fix,obs} = q_{float,obs} = -2\ln\frac{L(b)}{L(\hat{\mu}s(\hat{m} = m = 30) + b)}$$

$$p_{fix} = \int_{q_{obs}} f(q_{fix} \mid H_0)\, dq_{fix} < p_{float} = \int_{q_{obs}} f(q_{float} \mid H_0)\, dt_{float}$$

$$\boxed{trial\# = \frac{\int_{q_{obs}} f(q_{float} \mid H_0)\, dt_{float}}{\int_{q_{obs}} f(q_{fix} \mid H_0)\, dt_{fix}} = \frac{p_{float}}{p_{fix}} > 1}$$

Can we analytically calculate the trial# (or p_float)?

Define the Problem

- Let $n = \mu s(m,\Gamma) + b$
- m,Γ are nuisance parameters undefined under the null hypothesis $\mu = 0$
- What is the pdf of

$$\hat{q}_0 \equiv q_0(\hat{m},\hat{\Gamma}) = -2\log\frac{\mathcal{L}(\mu=0)}{\mathcal{L}(\hat{\mu},\hat{m},\hat{\Gamma})} = \max_{m,\Gamma}[q_0(m,\Gamma)]$$

  under the null hypothesis
- To generalize the problem , let θ be the nuisance parameter, undefined under the null hypothesis, and let us try to find out the pdf of $\boxed{\hat{q}_0 \equiv q_0(\hat{\theta}) = -2\log\frac{\mathcal{L}(\mu=0)}{\mathcal{L}(\hat{\mu},\hat{\theta})} = \max_{\theta}[q_0(\theta)]}$

  for which we want to calculate $\boxed{\text{p-value}=P(\max_{\theta}[q_0(\theta)] \geq u),\ u = Z^2}$

The profile-likelihood test statistic

(with a nuisance parameter that is not defined under the Null hypothesis)

- Consider the teststatistic:

$$q_0(\theta) = -2\log\frac{\mathcal{L}(\mu=0)}{\mathcal{L}(\hat{\mu},\theta)} \qquad \begin{array}{l} H_0 : \mu = 0 \\ H_1 : \mu > 0 \end{array}$$

μ="signal strength"

- For some <u>fixed</u> θ, $q_0(\theta)$ has (asymptotically) a chi^2 distribution with one degree of freedom by Wilks' theorem.

- $q_0(\theta)$ is a *<u>chi^2 random field</u>* over the space of θ (a random variable indexed by a continuous parameter(s)). we are interested in

$$\hat{q}_0 \equiv q_0(\hat{\theta}) = -2\log\frac{\mathcal{L}(\mu=0)}{\mathcal{L}(\hat{\mu},\hat{\theta})} = \max_{\theta}[q_0(\theta)]$$

$\hat{\theta}$ is the **global** maximum point

- For which we want to know what is the p-value

$$\text{p-value=P}(\max_{\theta}[q_0(\theta)] \geq u),\ \ u = Z^2$$

The profile-likelihood test statistic

(with a nuisance parameter that is not defined under the Null hypothesis)

- Usually we only look for 'positive' signals
  (downward fluctuations of the BG are not considered as evidence against the BG)

$$q_0(\theta) = \begin{cases} -2\log \dfrac{\mathcal{L}(\mu=0)}{\mathcal{L}(\hat{\mu},\theta)} & \hat{\mu} > 0 \\ 0 & \hat{\mu} \le 0 \end{cases}$$

$q_0(\theta)$ is 'half chi^2'

[H. Chernoff, Ann. Math. Stat. 25, 573578 (1954)]

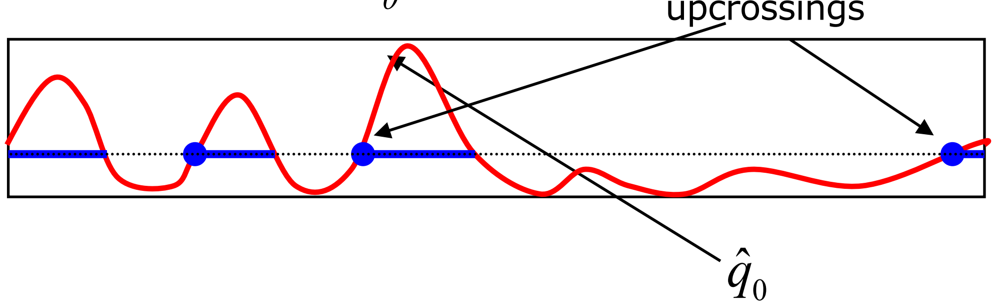
The p-value just get divided by 1/2

- Or equivalently consider $\hat{\mu}$ as a gaussian field

(since $q_0(\theta) = \left(\dfrac{\hat{\mu}(\theta)}{\sigma}\right)^2$)

Random fields (1D)

- In 1 dimension: points where the field values become larger then *u* are called *upcrossings.*
- $\max_{\theta}[q_0(\theta)]$



- The probability that the global maximum is above the level u is called *exceedance probability*. (p-value of $\hat{q}_0$)

$$P(\max_{\theta}[q_0(\theta)] \geq u)$$

Random fields

- Fortunately, quite a lot of statistical literature on the properties of random fields

[3] R.J. Adler and A.M. Hasofer, *Level Crossings for Random Fields*,Ann. Probab. **4**, Number 1 (1976), 1-12.

[4] R.J. Adler, *The Geometry of Random Fields*, New York (1981), Wiley, ISBN: 0471278440.

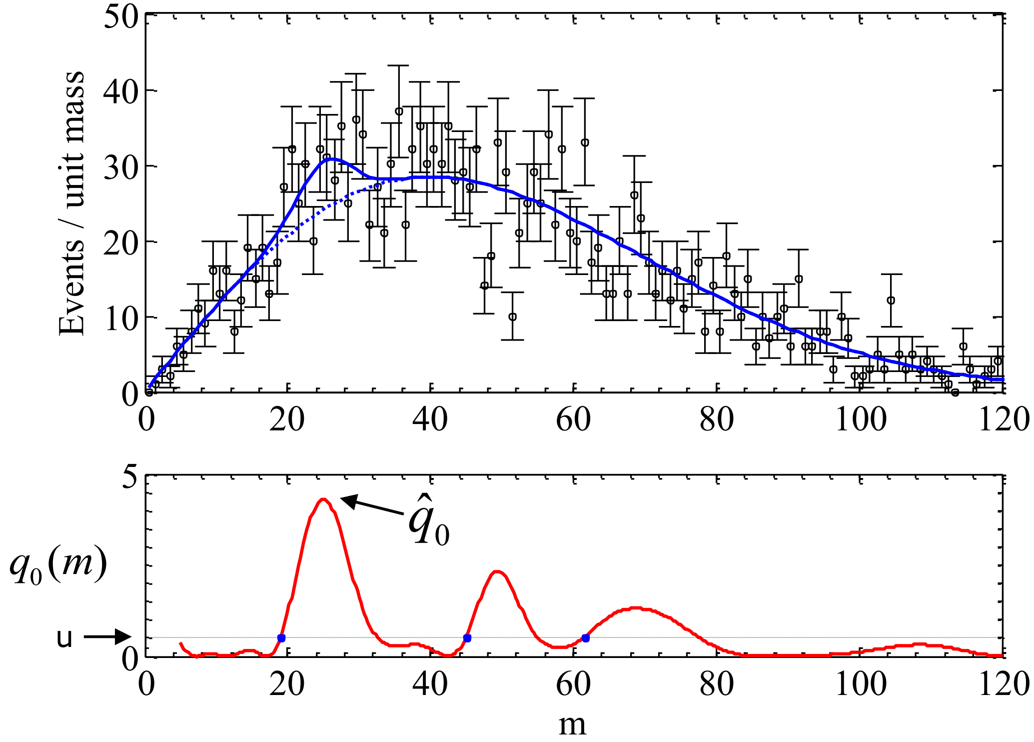
[5] K.J. Worsley, S. Marrett, P. Neelin, A.C. Vandal, K.J. Friston and A.C. Evans, *A Unified Statistical Approach for Determining Significant Signals in Location and Scale Space Images of Cerebral Activation*, Human Brain Mapping **4** (1996) 58-73.

[6] R.J. Adler and J.E. Taylor, *Random Fields and Geometry* , Springer Monographs in Mathematics (2007). ISBN: 978-0-387-48112-8.

[9] J. Taylor, A. Takemura and R.J. Adler, *Validity of the expected Euler characteristic heuristic*, Ann. Probab. 33 (2005) 1362-1396.

Applications in Cosmology, Brain mapping, Oceanography …

The 1-dimensional case



For a chi[2] random field, the expected number of *upcrossings* of a level *u* is given by: [Davies,1987]

$$E[N_u] = \mathcal{N}_1 e^{-u/2}$$

Note the inequality:
$E[N_u] \geq \mathrm{P}(N_u > 0)$

To have the global maximum above a level *u:*

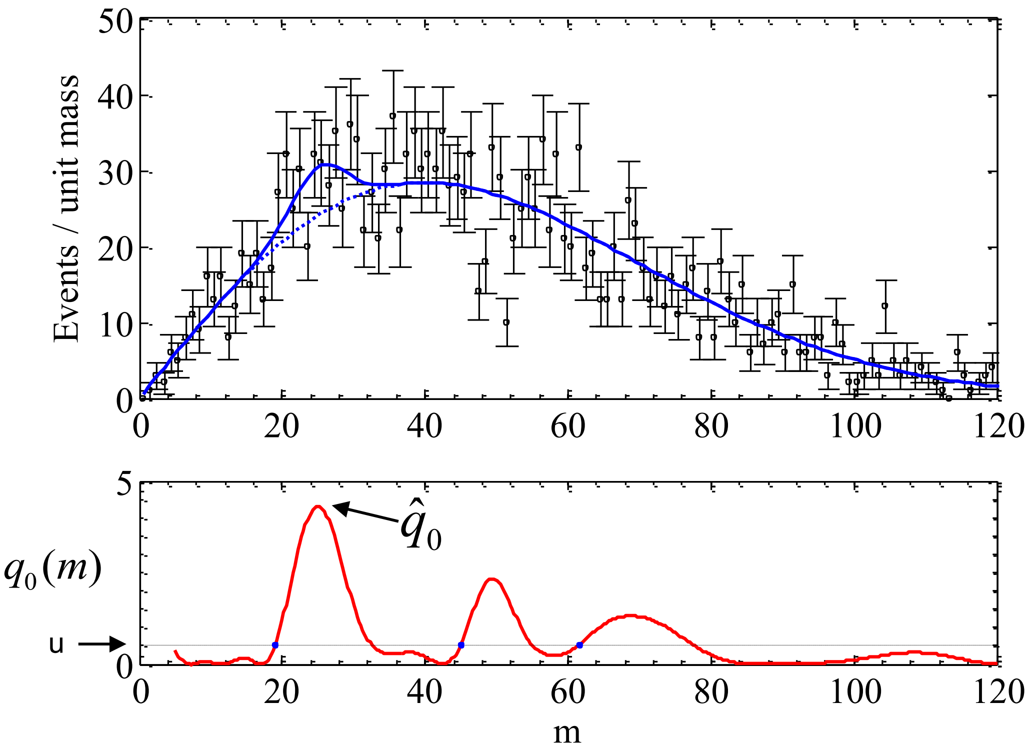
- Either have at least one upcrossing (*N_u>0*) ***or*** have $q_0>u$ at the origin *($q_0(0)>u$)* **:**

$$P(\hat{q}_0 > u) \leq P(N_u > 0) + P(q_0(0) > u)$$
$$\leq E[N_u] + P(q_0(0) > u)$$

[R.B. Davies, Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika **74**, 33–43 (1987)]

Becomes an equality for large *u*

The 1-dimensional case



$$E[N_u] = \mathcal{N}_1 e^{-u/2}$$

The only unknown is $\mathcal{N}_1$ which can be estimated from the average number of upcrossings at some low reference level

$$E[N_u] = N_1 e^{-u/2}$$

$$E[N_{u_0}] = N_1 e^{-u_0/2}$$

$$N_1 = E[N_{u_0}] e^{u_0/2}$$

$$E[N_u] = E[N_{u_0}] e^{(u_0-u)/2}$$

$$\mathrm{P}(q_0 > u) \le E[N_u] + \mathrm{P}(q_0(0) > u)$$

$$= \mathcal{N}_1 e^{-u/2} + \frac{1}{2} P(\chi_1^2 > u) = E[N_{u_0}] e^{(u_0-u)/2} + \frac{1}{2} P(\chi_1^2 > u)$$

$$p_{global} = E[N_{u_0}] e^{(u_0-u)/2} + p_{local}$$

Example: The 750 GeV Resonance

Spin 0 2015

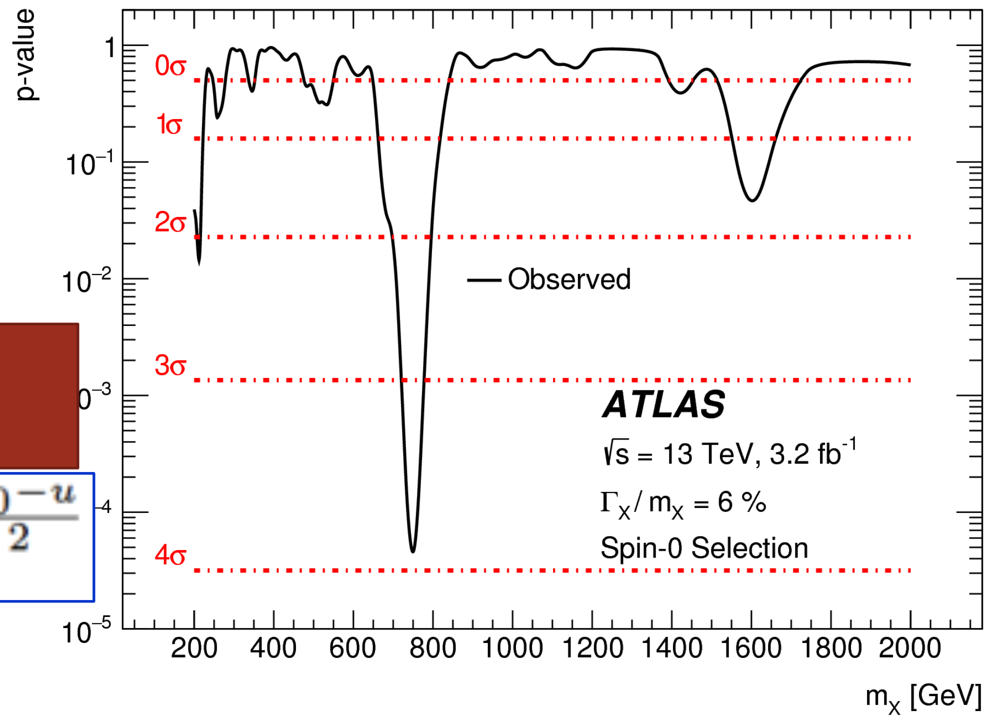
Largest significance

$m_x \sim 750\text{GeV}, \Gamma_x \sim 45\text{GeV}(6$

Local Z = 3.9σ

Any peak with Z>3.8σ
with m=500-2000 will draw our attention

$$P_{global}(u) \approx p_{local}(u) + E(n_{u_0})e^{\frac{u_0 - u}{2}}$$



$$p_{local} = 5 \cdot 10^{-5}$$

$$u_0 = 0$$

$$n_{u_0} = 7 \pm 2.6$$

$$u = Z^2 = 3.9^2 = 15.2$$

$$p_{global} = 5 \cdot 10^{-5} + (7 \pm 2.6)e^{-15.2/2} = (2.2 - 4.8)10^{-3}$$

$$Z_{global} \sim 2.7 \pm 0.1\sigma$$

The LEE is even stronger when you consider another dimension (the width range (0-10%m) should also be taken into account)

A real life example

$$\mathrm{P}(q_0 > u) \le E[N_u] + \mathrm{P}(q_0(0) > u)$$

$$E[N_u] = \mathcal{N}_1 e^{-u/2}$$

$$\mathcal{N}_1 \cong \left\langle N_{u_0} \right\rangle e^{u_0/2}$$

$$\mathrm{P}(q_0 > u) = \mathcal{N}_1 e^{-u/2} + \frac{1}{2} P(\chi_1^2 > u)$$

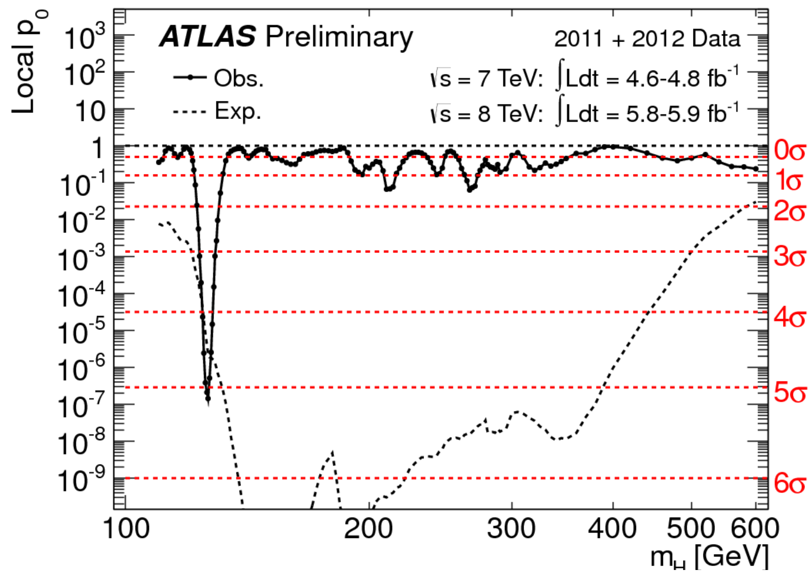
$$p_{global} = \mathcal{N}_1 e^{-u/2} + p_{local}$$

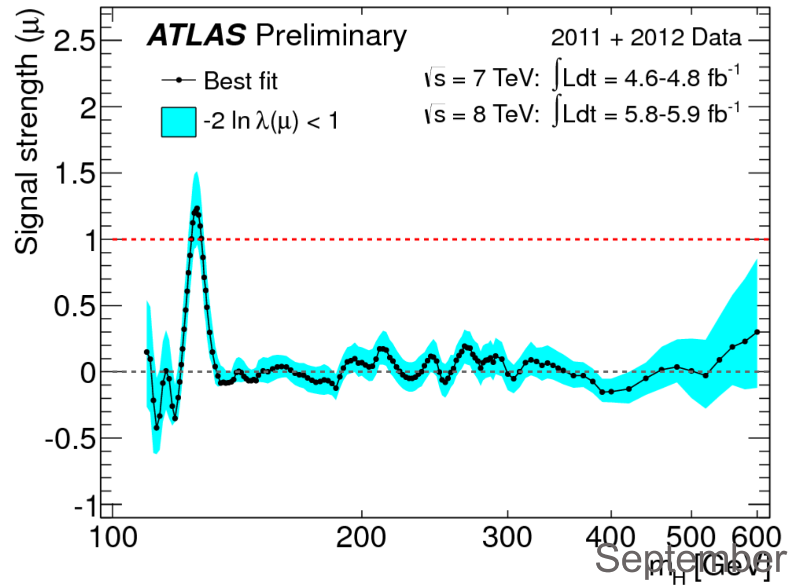
$$p_{global} = \left\langle N_{u_0} \right\rangle e^{\frac{u_0 - u}{2}} + p_{local}$$

$$N_{u_0=0} = 9 \pm 3$$

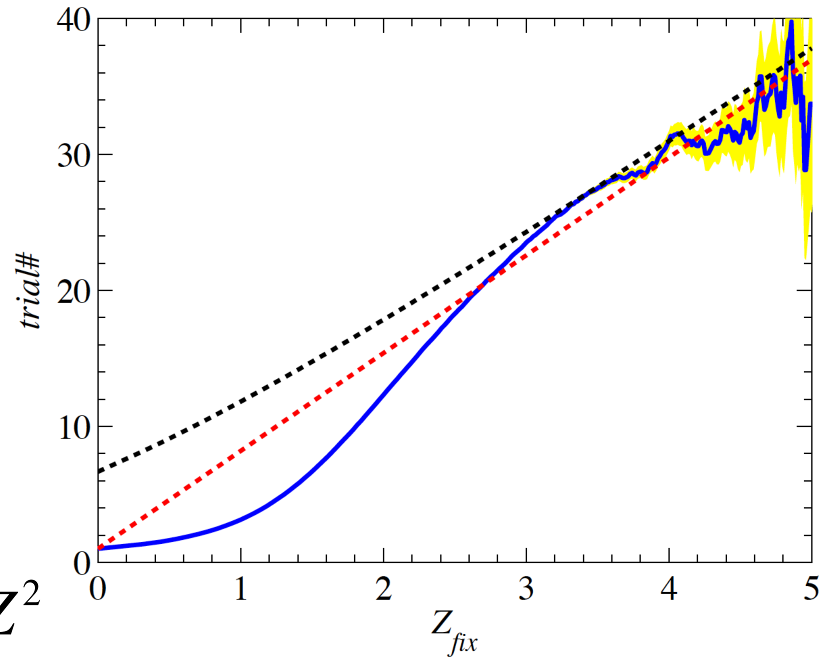
$$p_{global} = 9 \cdot \mathrm{e}^{-25/2} + O(10^{-7}) = 3.3 \cdot 10^{-5}$$

$5\sigma \to 4\sigma$ trial#~100





Trial Factor



$$u = Z^2$$

The Trial factor is given by

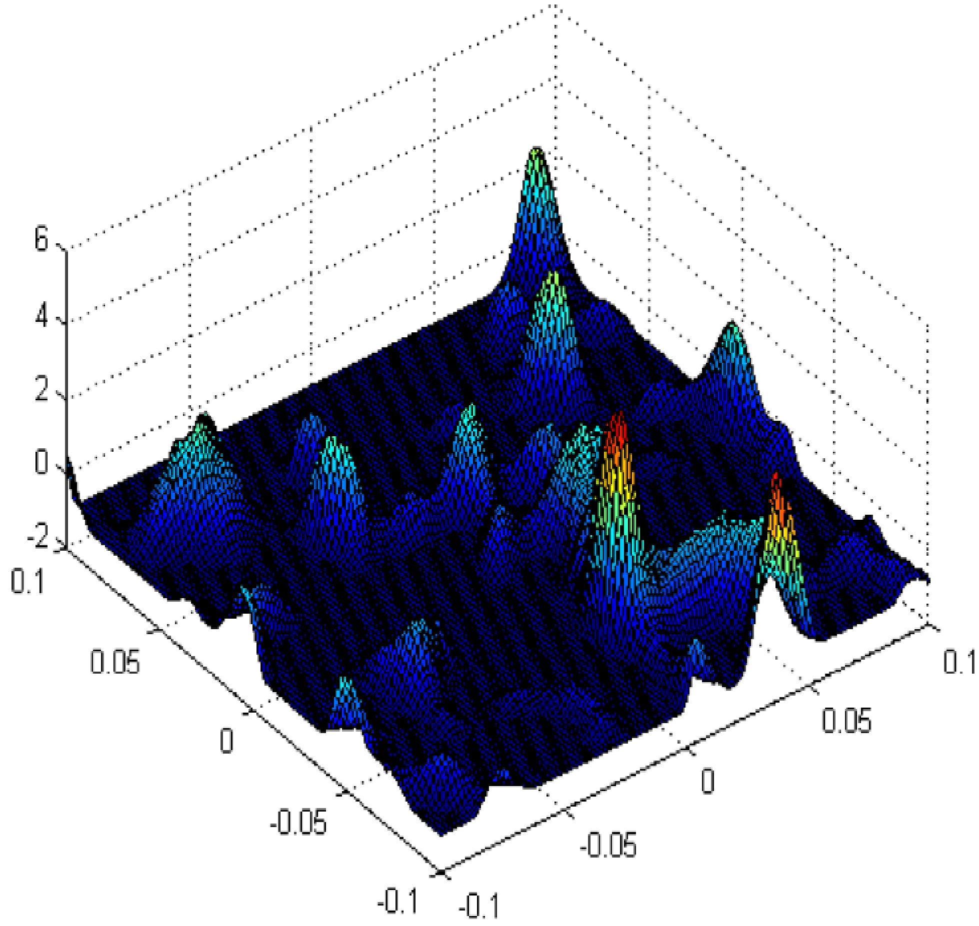
$$trial\,\# = \frac{p_{global}}{p_{local}} \leq \frac{\mathcal{N}_1 e^{-u/2} + \frac{1}{2}P(\chi_1^2 > u)}{\frac{1}{2}P(\chi_1^2 > u)}$$

$\mathcal{N}_1$ Is the number of independent search regions

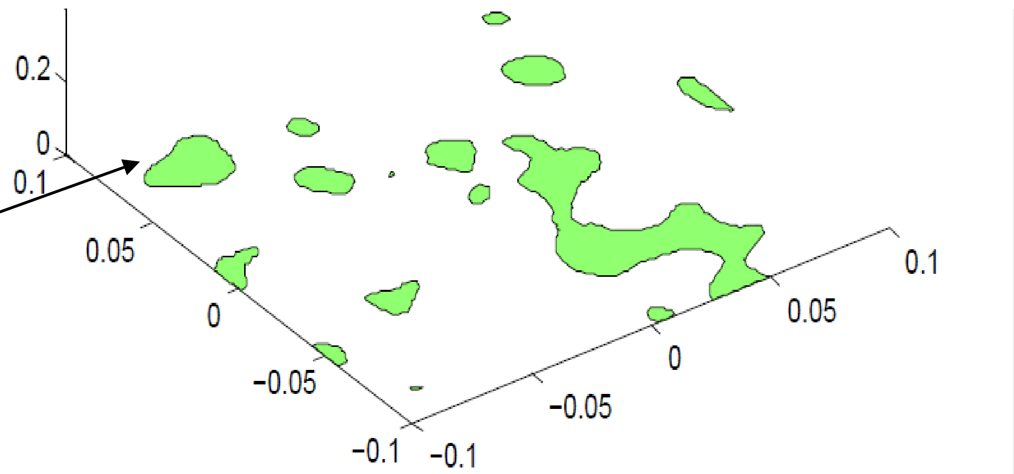
$$trial\,\#(u >> 1) = 1 + \sqrt{\frac{\pi}{2}}\mathcal{N}_1 Z_{fix} \approx \sqrt{\frac{\pi}{2}}\mathcal{N}_1 Z_{fix}$$

Random fields (>1 D)

- The set of points where the field has values larger then some number *u* is called the *excursion set* A_u above the level *u*.

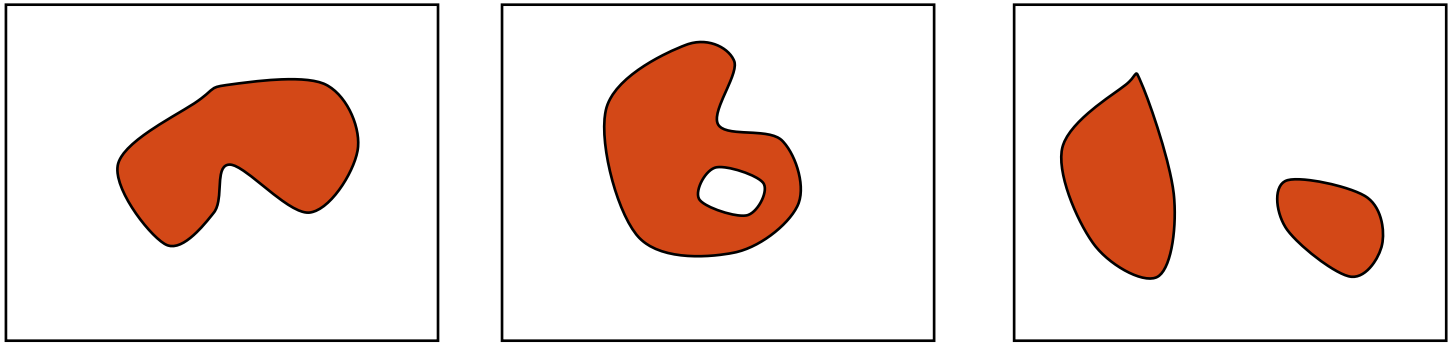


Excursion set

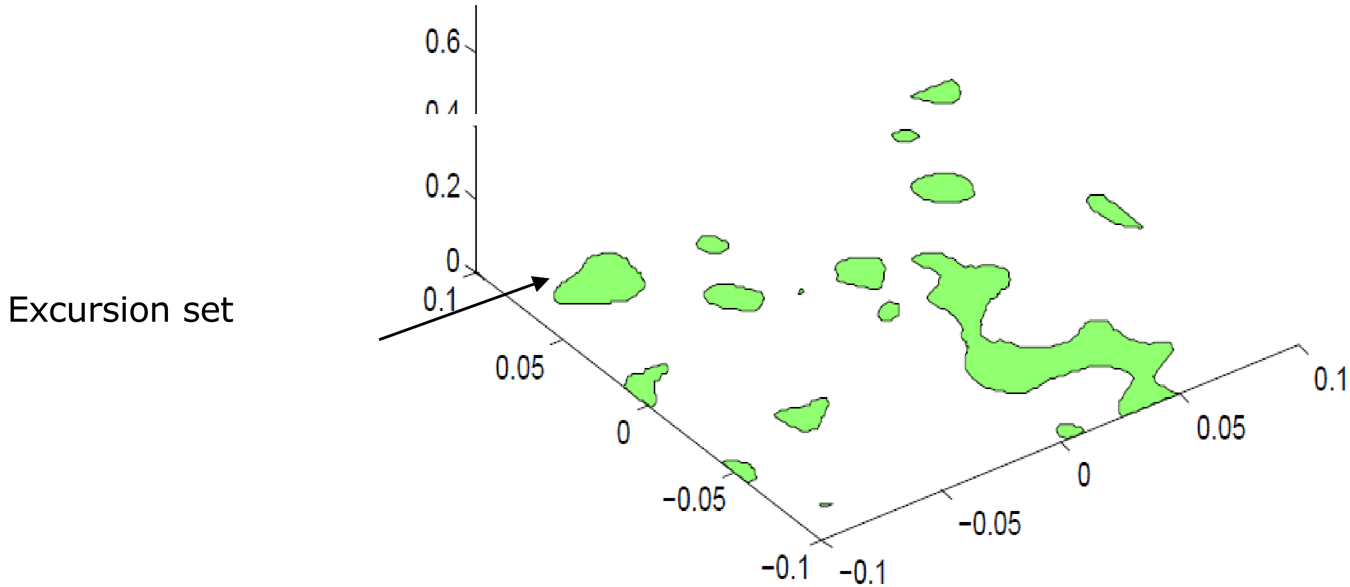


Euler characteristic

- Number of disconnected components minus number of `holes'



$\varphi=1$ $\varphi=0$ $\varphi=2$



2-d example: search for neutrino sources (IceCube)

For a chi² field in 2 dimensions:

$$E[\varphi(A_u)] = \frac{1}{2}\mathrm{P}(\chi^2 > u) + (\mathcal{N}_1 + \mathcal{N}_2\sqrt{u})e^{-u/2}$$

Estimate E[φ] at two levels, e.g. 0 and 1, and solve for $\mathcal{N}_1$ and $\mathcal{N}_2$

From 20 bkg. Simulations:

$$\langle\varphi_0\rangle = 33.5 \pm 2$$

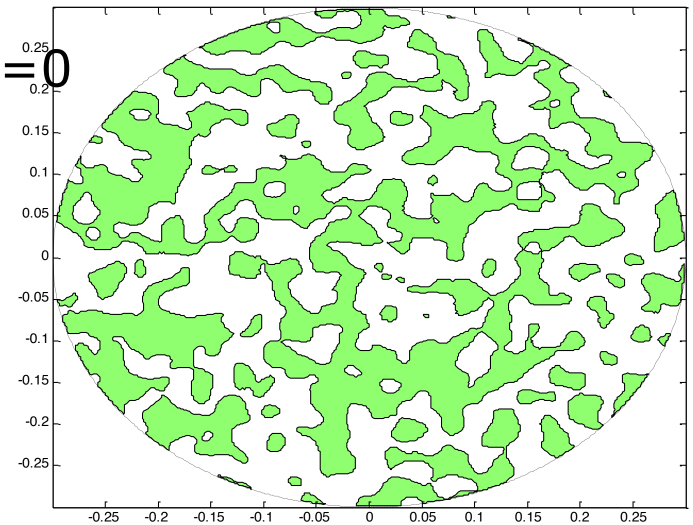
$$\langle\varphi_1\rangle = 94.6 \pm 1.3$$

$$\Downarrow$$

$$\mathcal{N}_1 = 33 \pm 2$$

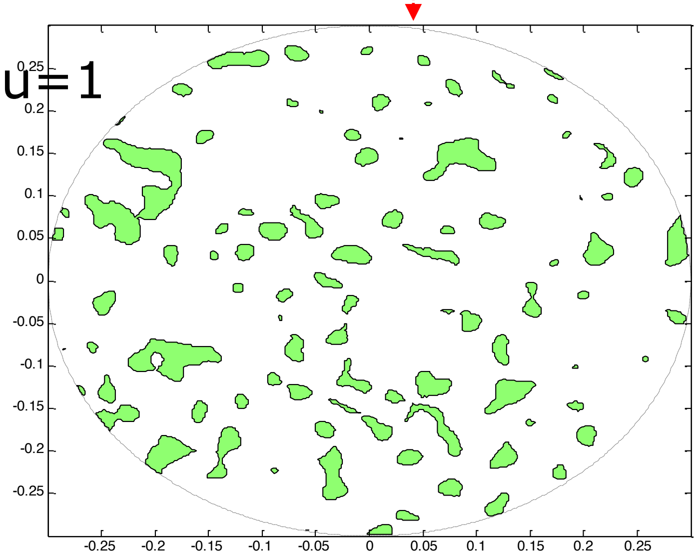
$$\mathcal{N}_2 = 123 \pm 3$$

u=0



φ=35

u=1



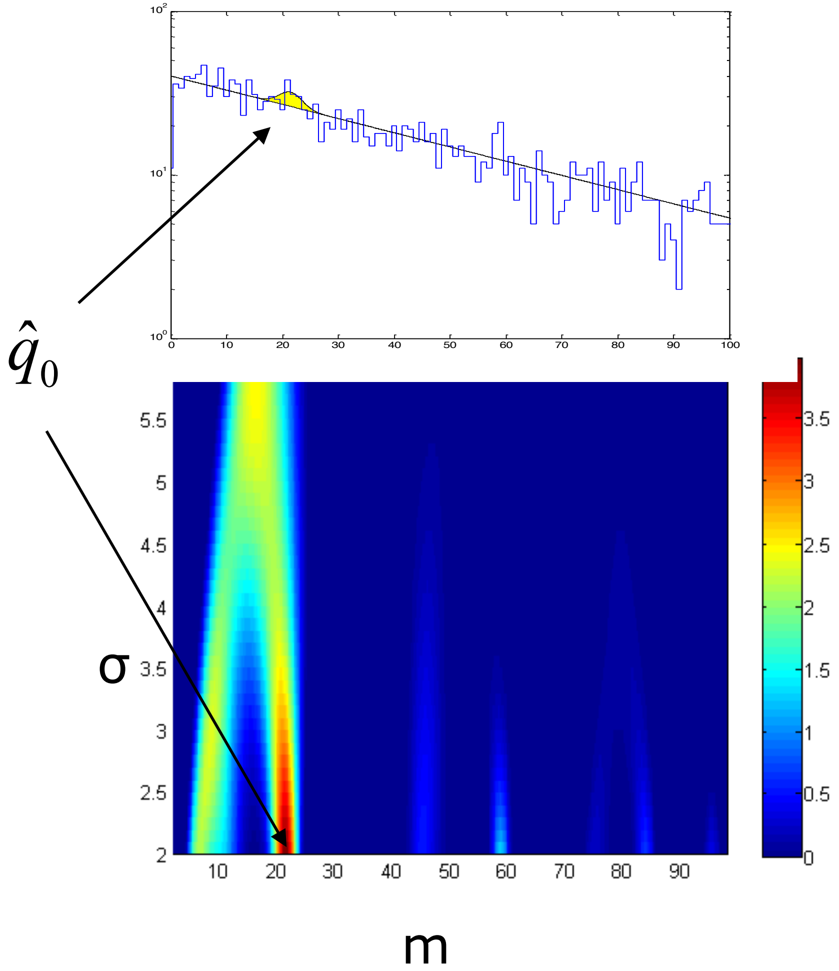
φ=95

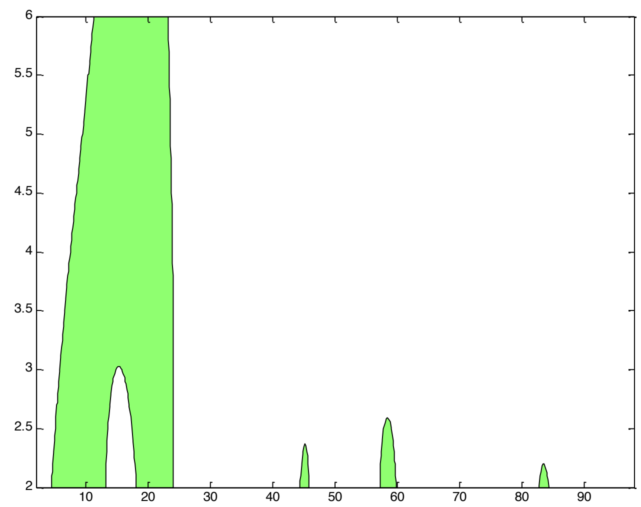
2-D exapmle #2: resonance search with unknown width

- Gaussian signal on exponential backgrou
- Toy model : 0<m<100 ,   2<σ<6
- Unbinned likelihood:

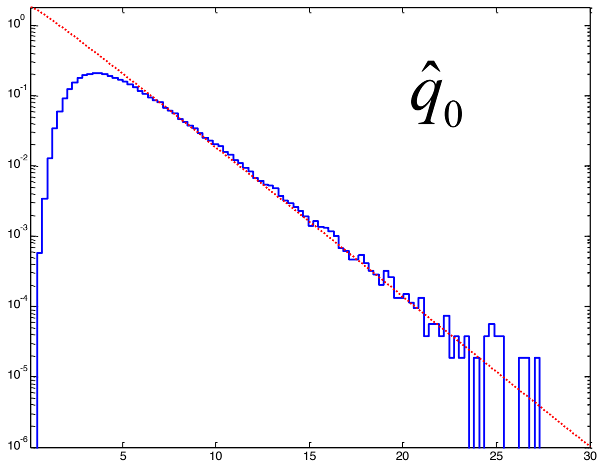
$$\mathcal{L} = \prod_i \frac{N_s f_s(x_i) + N_b f_s(x_i)}{N_s + N_b} \times Poiss(N \mid N_s + N_b)$$

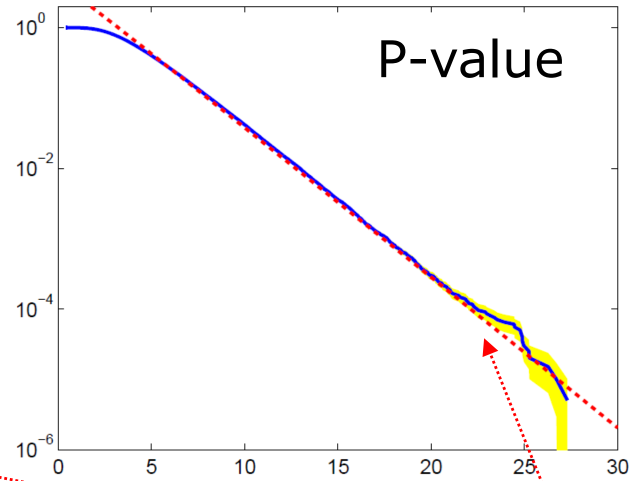
$$f_s(x;m,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad f_b(x) = ce^{-cx}$$





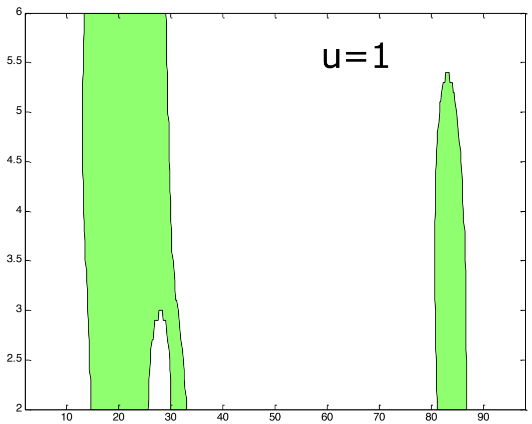
2-D exapmle #2: resonance search with unknown width



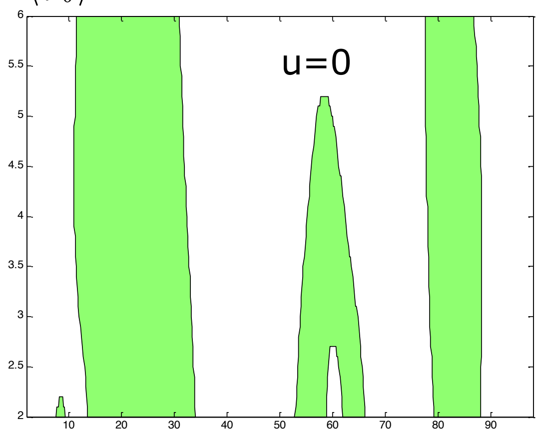


Excellent approximation above the ~2σ level

$\langle \varphi_1 \rangle = 3 \pm 0.16$



$\langle \varphi_0 \rangle = 4.5 \pm 0.2$



$$E[\varphi(A_u)] = \frac{1}{2}\mathrm{P}(\chi^2 > u) + (\mathcal{N}_1 + \mathcal{N}_2\sqrt{u})e^{-u/2}$$

$$\mathcal{N}_1 = 4 \pm 0.2$$

$$\mathcal{N}_2 = 0.7 \pm 0.3$$

The 750 GeV saga

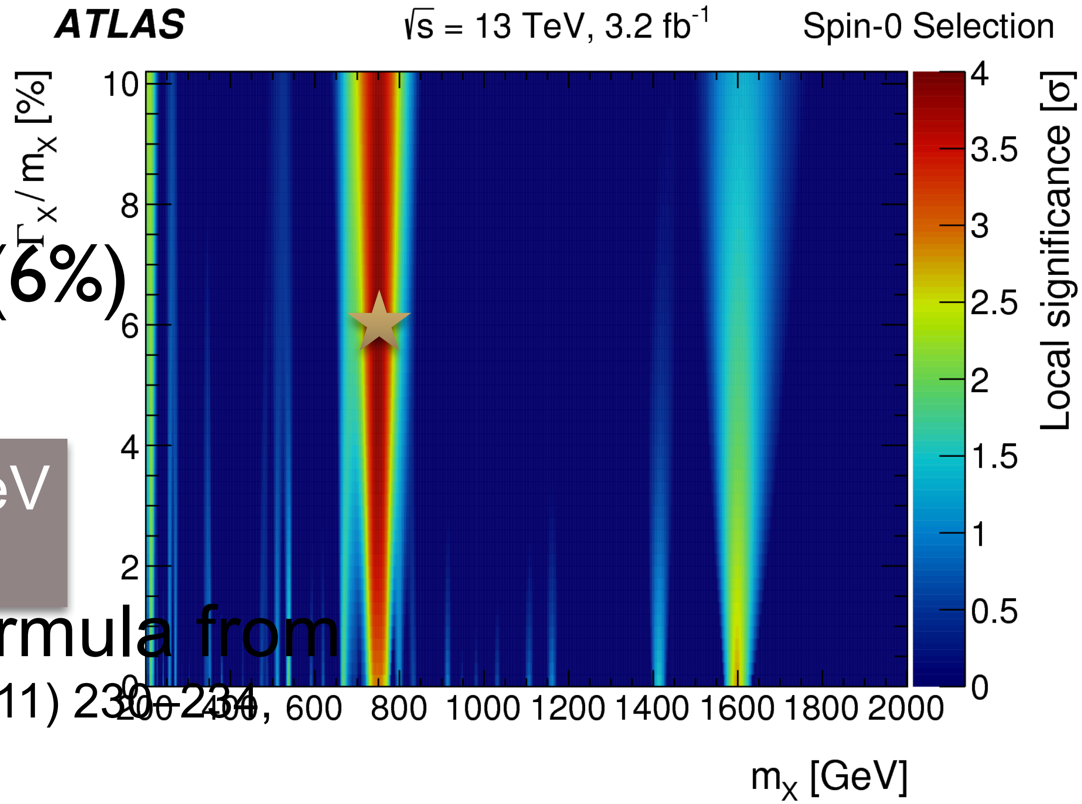
2015

2D Scan
Largest significance

$m_x \sim 750GeV, \Gamma_x \sim 45GeV(6\%)$

Local Z = **3.9σ**

m=200-2000 GeV
Γx/mx=0-10%



Use toys or asymptotic formula from
O. Vitells et. al. Astropart. Phys. 35 (2011) 230-234,
arXiv:1105.4355

$$Z_{local} = 3.9\sigma$$

$$Z_{global} = 2.1\sigma$$

2.1σ is not something to write home about

Addendum

CCGV Useful Formulae – The Bands

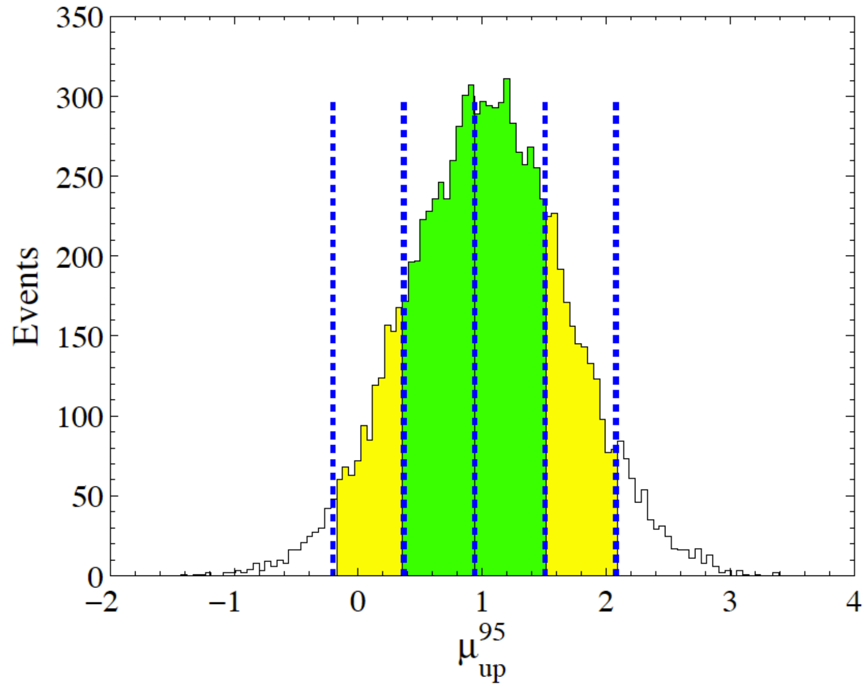
$$\mu_{up}^{med} = \hat{\mu} + \sigma_{\mu_{up}^{med}} \phi^{-1}(1-\alpha)$$

$$\alpha = 0.05 \rightarrow \phi^{-1}(1-\alpha) = \phi^{-1}(0.95) = 1.64$$

$$\mu_{up+N\sigma} = \hat{\mu} + \sigma_{up+N\sigma}\left(\phi^{-1}(1-\alpha) + N\right)$$

$$\sigma_{up+N\sigma}^2 = \frac{\mu_{up+N\sigma}^2}{q_{up+N\sigma,A}}$$

Distribution of the upper limit with background only experiments

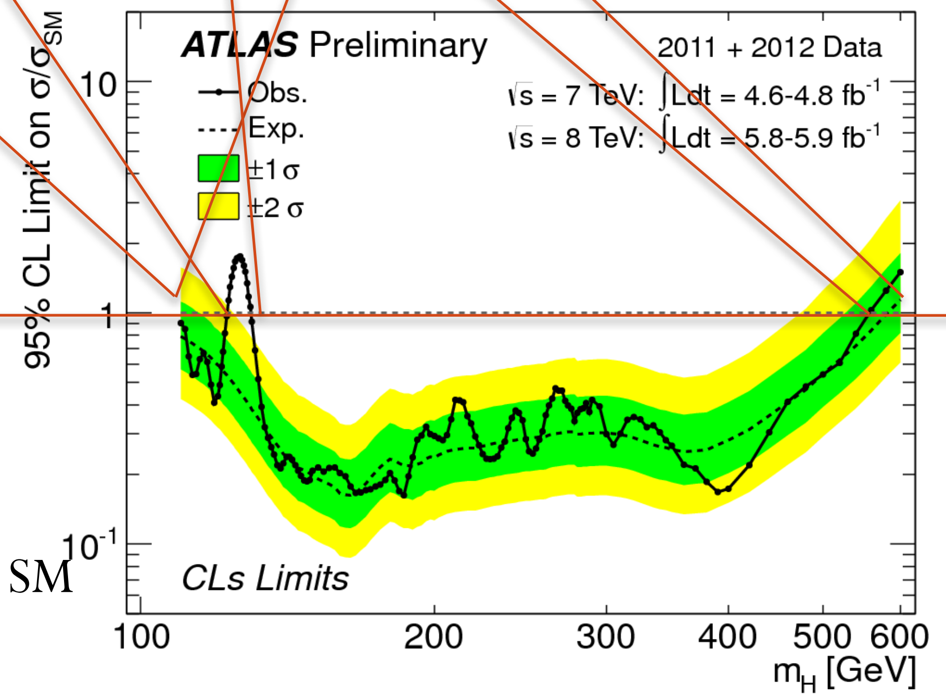


The Asimov data set is n=b
-> median upper limit

Understanding the Brazil Plot

The expected 95% CL exclusion region covers the m_H range from 110 GeV to 582 GeV. The observed 95% CL exclusion regions are from 110 GeV to 122.6 GeV and 129.7 GeV to 558 GeV. The addition of

- $\mu_{up}=\sigma(m_H)/\sigma_{SM}(m_H)<1 \rightarrow \sigma(m_H)<\sigma_{SM}(m_H) \rightarrow$ SM m_H excluded



- The line $\mu up=1$ corresponds to CLs=5% (p'_s=5%)
- The smaller $\mu up<1$ is, the exclusion of a SM Higgs is deeper $\rightarrow$ p's<5%, p's=CLs $\rightarrow$ CL=1-p's>95%