

FYST17 Lecture 9

More statistical methods in Particle Physics

Thanks to J. Morris, S. Menzemer

Suggested reading: Statistics book chap 8

Outline

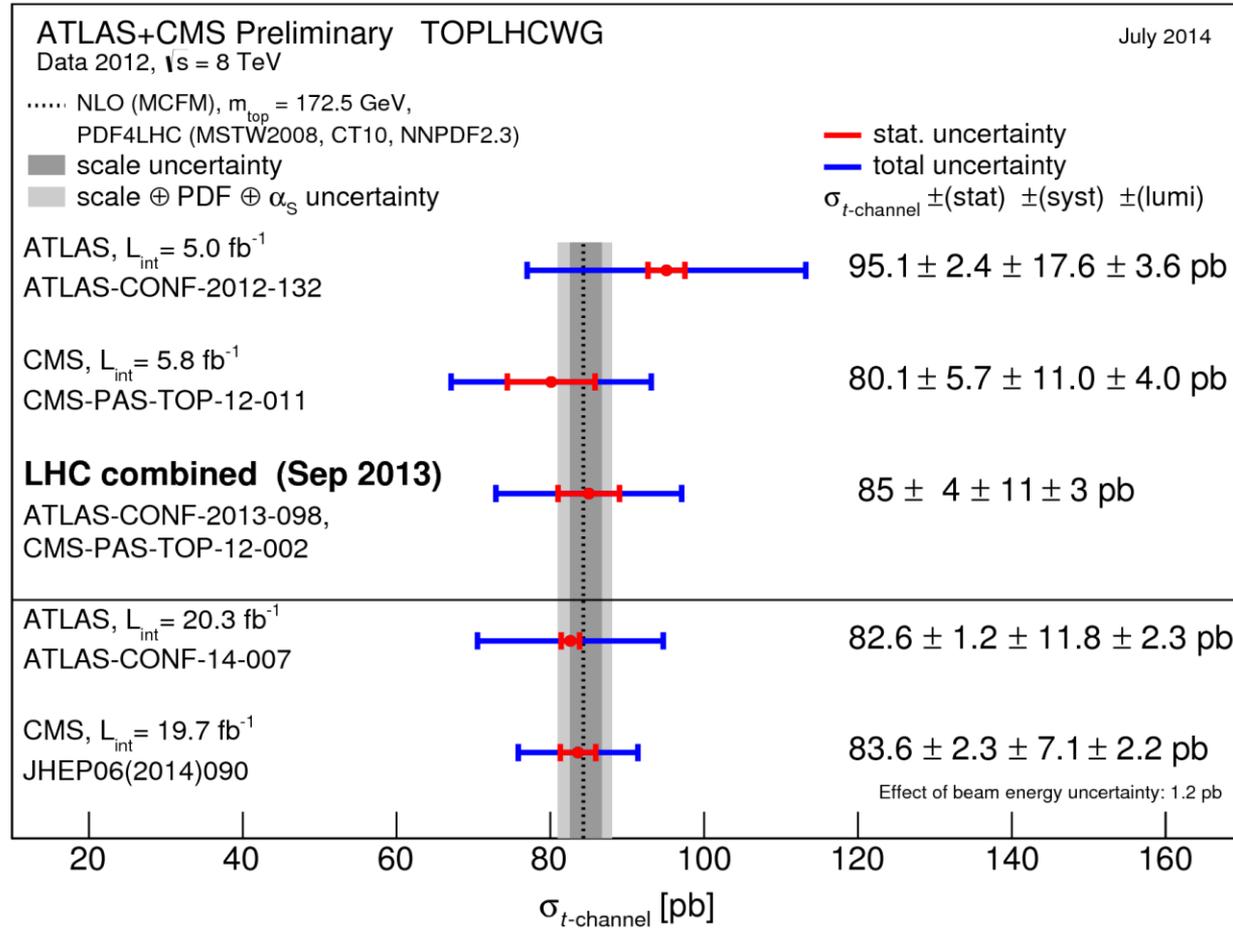
- Systematic uncertainties
 - Definition, examples
- The a_2 mass splitting measurement
- "Blind" analysis
- Estimating efficiencies
- Estimating backgrounds

What is a systematic uncertainty?

All uncertainties that are not directly due to the statistics of the data. For instance:

- Badly known backgrounds
- Badly known detector resolutions
- Wrong calibrations
- Badly known acceptances or efficiencies
- Preferred outcomes
- External factors, such as theory uncertainties on cross sections etc
- Other biases

Examples



Examples

ATLAS+CMS Preliminary TOPLHCWG

July 2014

Data 2012, $\sqrt{s} = 8$ TeV

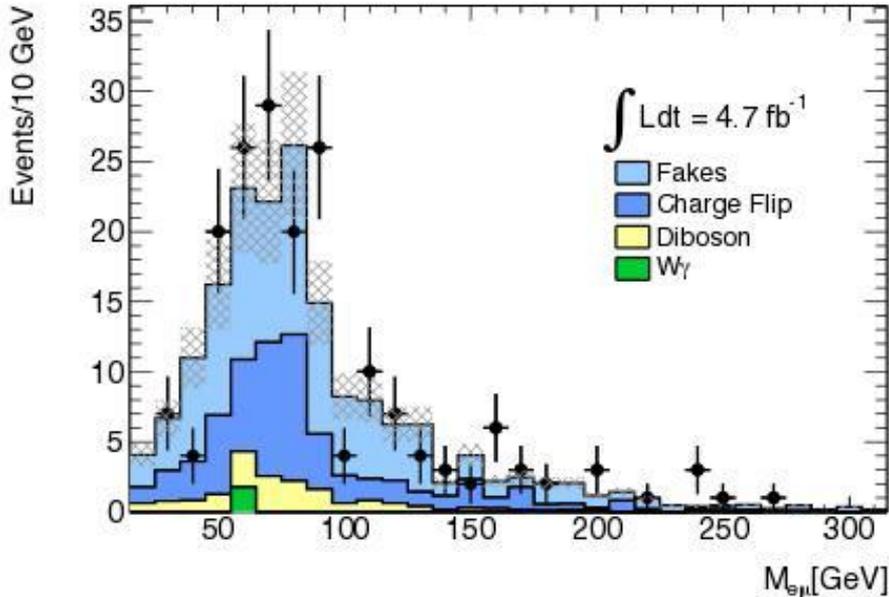
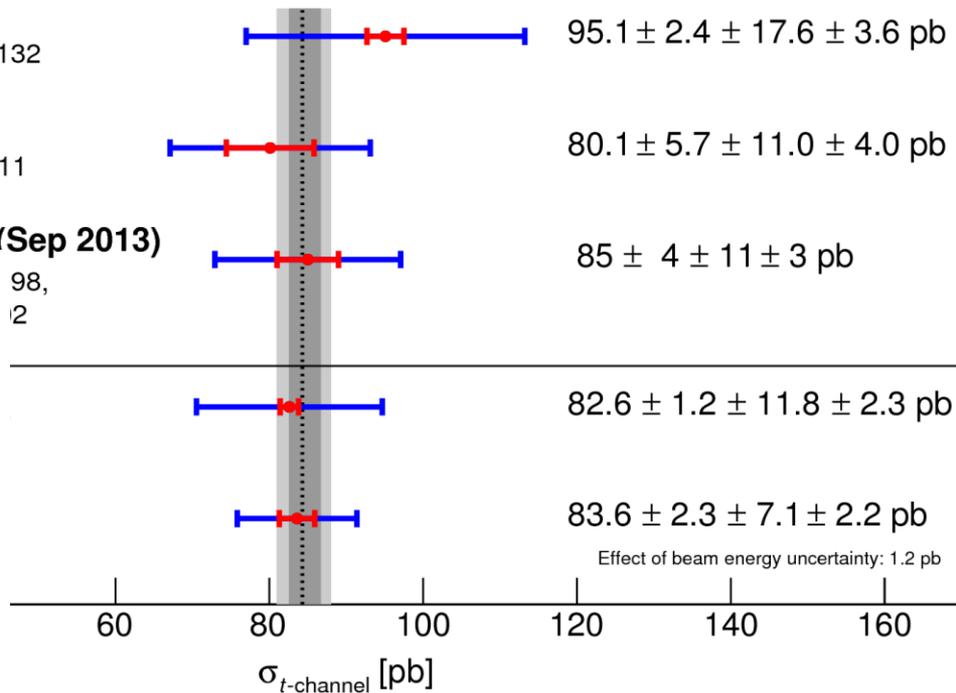
..... NLO (MCFM), $m_{top} = 172.5$ GeV,
 PDF4LHC (MSTW2008, CT10, NNPDF2.3)
 ■ scale uncertainty
 ■ scale \oplus PDF \oplus α_s uncertainty

— stat. uncertainty
 — total uncertainty
 $\sigma_{t\text{-channel}} \pm(\text{stat}) \pm(\text{syst}) \pm(\text{lumi})$

ATLAS, $L_{int} = 5.0 \text{ fb}^{-1}$
 ATLAS-CONF-2012-132

CMS, $L_{int} = 5.8 \text{ fb}^{-1}$
 CMS-PAS-TOP-12-011

LHC combined (Sep 2013)



Statistical (random) vs systematic uncertainties

Statistical	Systematic
No preferred direction	Bias on the measurement. Only one direction
Changes with each data point. Taking more data reduces the mean	Stays the same for each measurement. More data won't help you
Gaussian model is usually good (some exceptions counting experiments)	Gaussian model is usually terrible but we use it anyway

Example:

Statistical Systematic

$$M_{\text{top}} = 173.34 \pm 0.36 \pm 0.67 \text{ GeV}$$

More data will not help!

Evaluating uncertainties

- There are, broadly speaking, 4 main categories of uncertainty
- Statistical
 - Determined entirely by how many signal events you have
- Luminosity
 - We only know it to $\approx \cancel{4\%} \mathbf{2\% !}$
- Experimental
 - Background estimation
 - Uncertainties from re-weighting MC
 - Uncertainties from smearing MC
- Theoretical modelling uncertainties
 - Generator, PDF, parton shower, ISR/FSR

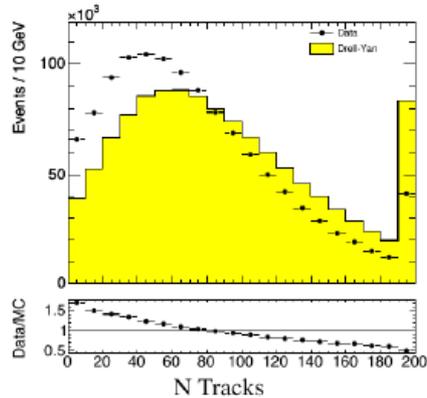
These are usually called
the systematic
uncertainties

How are systematics estimated?!

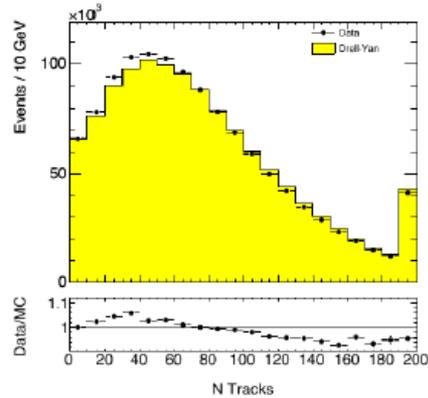
- No standard recipe! Some examples:
- If amount of material important, check simulation with different amount of material
- If efficiencies important, try varying nominal values with $\pm 1\sigma$ and see the effect
 - This is standard, test effect of changes in analysis procedure (for instance different fit window)
- Compare data and simulation in general to see differences
- Divide data up in periods with different conditions and compare
- ...
- A bit of an art, actually

Example: systematic difference data & MC

Use event weighting and 4-vector smearing

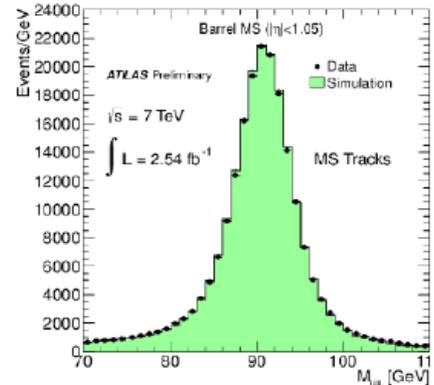
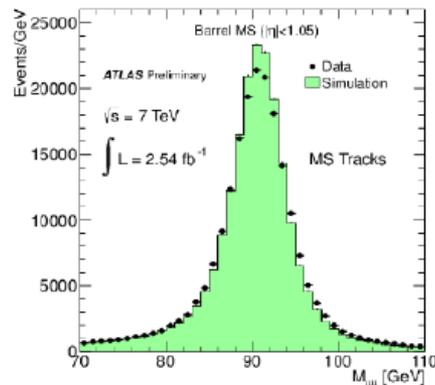


(a) $\langle\mu\rangle$ rescale factor= 1 (no rescale)



(b) $\langle\mu\rangle$ rescale factor= 0.83

- Re-weight MC events to achieve same efficiency in MC as in data
- Smear MC 4-vectors to achieve same resolution in MC as in data



Example: PDF uncertainties

The PDF4LHC prescription

PDF = **P**article **D**ensity **F**unctions of the proton

- Test 3 different PDF sets
 - CTEQ (Nominal), MSTW and NNPDF
- Each PDF set gives an event weight, based on EventNumber
- The first weight is the PDF nominal value, followed by n different tests varying a PDF eigenvector $\pm 1\sigma$
- Uncertainties are calculated differently for each PDF set
- Total uncertainty is $\frac{1}{2}$ the envelope

	PDF Set		
	CTEQ	MSTW	NNPDF
N Weights	53	41	101
N PDF eigenvectors varied $\pm 1\sigma$	26	20	50
Uncertainty method	Symmetric Hessian	Asymmetric Hessian	Standard Deviation

- If:
 - X_0 = Nominal PDF cross section (first weight, not our $\sigma_{t\bar{t}}$)
 - X_i^\pm = PDF eigenvector test i , varying the eigenvector by $\pm 1\sigma$
 - \bar{X} = Mean of all PDF eigenvectors (Used for NNPDF)
 - N = All PDF eigenvectors. This does not include the first value

• Then :

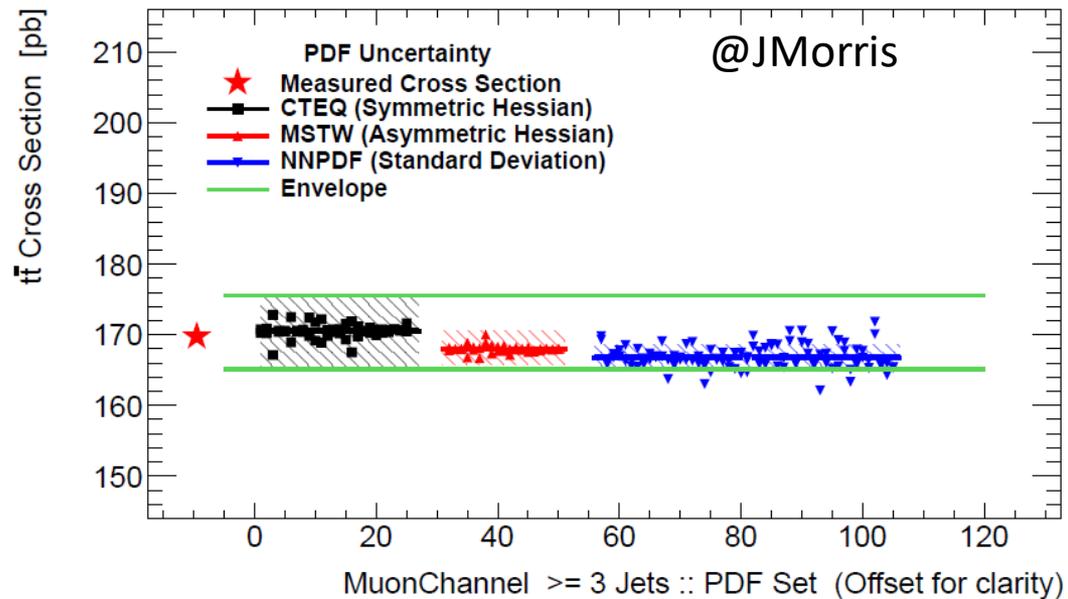
- CT10 $\alpha = \frac{1}{2} \sqrt{\sum_{i=1}^{N/2} (X_i^+ - X_i^-)^2}$

- MSTW $\alpha^\uparrow = \sqrt{\sum_{i=1}^N (X_i - X_0)^2}$: if $(X_i - X_0) > 0$

- MSTW $\alpha^\downarrow = \sqrt{\sum_{i=1}^N (X_i - X_0)^2}$: if $(X_i - X_0) < 0$

- NNPDF $\alpha = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\bar{X} - X_i)^2}$

Resulting in this →
variation



Example: systematics table for a $t\bar{t}$ mass measurement

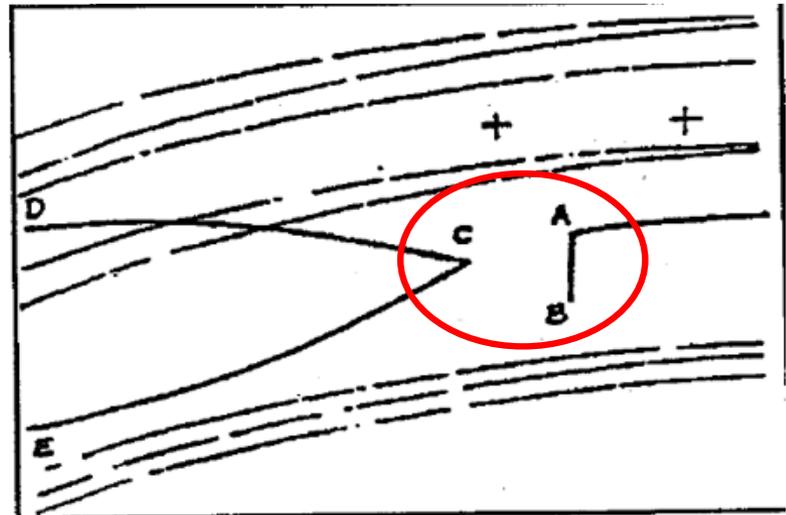
Source	Relative cross section uncertainty [%]		
	e+jets	μ +jets	Combined
Statistical Uncertainty	± 1.5	± 1.3	± 1.0
<i>Object selection</i>			
Lepton energy resolution	+0.4 /-0.3	+0.2 /-0.1	+0.2 /-0.1
Lepton reco, ID, trigger	+2.4 /-2.5	+1.5 /-1.5	+1.7 /-1.8
Jet energy scale	+3.8 /-4.3	+3.2 /-3.6	+3.5 /-3.8
Jet energy resolution	± 0.2	± 0.5	± 0.2
Jet reconstruction efficiency	± 0.06	± 0.06	± 0.06
Jet vertex fraction	+1.2 /-1.4	+1.2 /-1.4	+1.2 /-1.4
E_T^{miss} uncertainty	± 0.06	± 0.08	± 0.07
SMT muon reco, ID	± 1.3	± 1.3	± 1.3
SMT muon χ_{match}^2 efficiency	± 0.6	± 0.6	± 0.6
<i>Background estimates</i>			
Multijet normalisation	± 5.2	± 3.9	± 4.4
W+jet normalisation	± 5.2	± 5.7	± 5.5
Other bkg normalisation	± 0.2	± 0.2	± 0.1
Other bkg systematics	+1.6 /-1.5	+2.5 /-2.0	+2.2 /-1.8
<i>Signal simulation</i>			
$b \rightarrow \mu X$ Branching ratio	+2.9 /-3.0	+2.9 /-3.1	+2.9 /-3.1
ISR/FSR	± 2.4	± 0.9	± 1.5
PDF	± 3.2	± 3.0	± 3.1
NLO generator	± 3.2	± 3.2	± 3.2
Parton shower	± 2.2	± 2.2	± 2.2
Total systematics	± 11.2	± 10.2	± 10.5
Integrated luminosity	± 3.8	± 3.8	± 3.8

How to look for a particle

- 1) Look in high-energy collisions for events with multiple output particles that could be decay products.

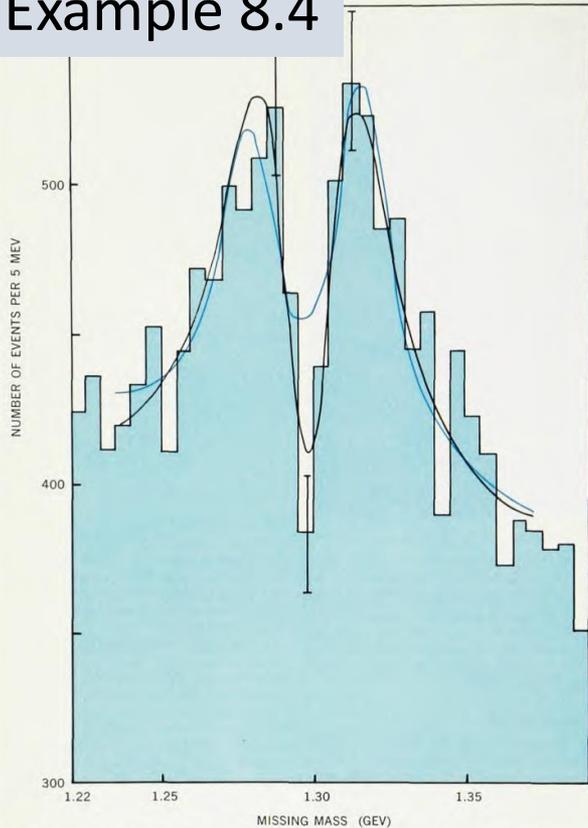
(for instance, $K^0 \rightarrow \pi^+\pi^-$, displaced vertex)

- 2) Reconstruct invariant mass from assumed decay products
- 3) Make a histogram of the masses
- 4) Look for a peak indicating a state of well-defined mass



A Cautionary Tale: One or two peaks?

Example 8.4



Fits to the two-peak structure of data from the CERN missing-mass and boson spectrometer group for the A_2 , 1965–68. The black curve is the fit for two coherent Breit-Wigner amplitudes or a dipole (these two can not be distinguished); the colored curve is the fit for two incoherent Breit-Wigner curves. The incoherent case is ruled out by a confidence level of 0.2%, while the coherent and dipole fits have equal confidence levels of 40%. Figure 4

for both the A_2^L and the A_2^H gives a confidence level greater than 10%, while the assignment $1^-, 1^+$ (s-wave), 1^+ (d), 2^- (p), and 2^- (f) all give a confidence level much less than 0.01%. The ratio of resonance to $\rho\pi$ and 3π background, which must be known for

the reaction $\pi^+ p \rightarrow pA_2^+$. In none of their three final states ($\rho^0\pi^+$, $\gamma\pi^+$ and $K^+K_1^0$) do they see a splitting, though they claim to have an experimental resolution of about ± 5 MeV. They base their conclusion of no splitting on the sum of $K^+K_1^0$ and $\gamma\pi^+$ events and in-

MeV using their background and resonance assumptions, one obtains an acceptable confidence level for the dipole. One also obtains an acceptable dipole fit over the whole mass spectrum if one assumes a second-order background. Furthermore, one has to note that the extremely crucial background behavior at both ends of the spectrum is based on 2–6 events per 10-MeV bin. The same procedures increase the confidence level for a dipole in the $\rho^0\pi^+$ events by a considerable amount. Aside from statistics and background considerations, one must bear in mind the very general fact that it is much easier *not* to see a splitting than to see it, because of a variety of resolution-killing effects that are normally hard to track down, both in counter and bubble-chamber experiments.

Exciting new results on the neutral A_2 were reported, at the Kiev International High Energy Conference in September, by T. Massam of the group at CERN headed by A. Zichichi. In the first reported observation of the splitting in A_2^0 , the CERN counter group measured the recoil neutron in the charge-exchange reaction



at a beam momentum of 3.2 GeV/c. They saw a marked dip at the center of the A_2^0 . Confidence levels for a single peak, incoherent double peak and dipole were 1%, 23% and 67% respectively.

Dependence of splitting

To arrive at some conclusions concerning the A_2 splitting we will look for variables the effect may depend on. The dependence or independence might give a clue to the nature of the A_2 . We will discuss the possible dependence of the A_2 splitting on four quantities: bombarding energy, final state, production reaction and momentum transfer.

The effect of symmetric splitting has been observed significantly at threshold and at 6 and 7 GeV/c by the CERN counter group. At the present time, no data shed doubt on the invariance of the A_2 splitting with bombarding energy.

The most significant data show the effect in missing mass with negative charge, or $\rho\pi$. There is considerable indication for the splitting in the $K\bar{K}$ final state (CERN counter and bubble-chamber groups) and some indications for $\gamma\pi$ (the counter group). We therefore conclude that the A_2 splitting does not

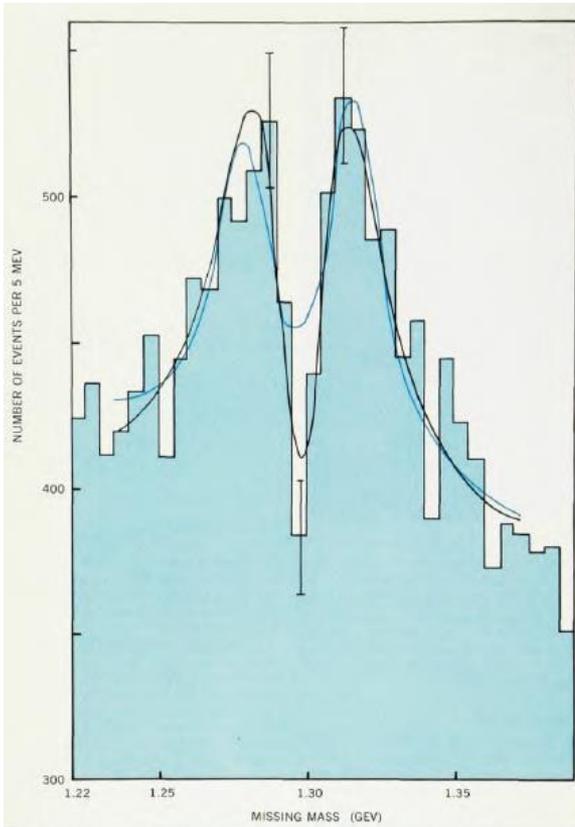
Late 1960's:

CERN experiment observes A_2 mesons. Appeared to be a doublet – with two mass peaks!

Statistical significance of split very high!

But there really is only 1 particle here – what went wrong?

What went wrong?



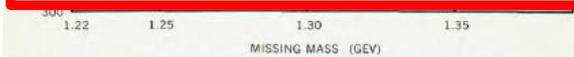
- The dip was noticed already in an early run. Likely a statistical fluctuation but experimenters suspected it was real
- Therefore, in the subsequent runs, this was looked into.
 - **If the run did not show the dip, the run was looked into further.** There is always something to point your finger at if you are looking for problems (especially at a complicated experiment) and thus many of the runs without dip were declared faulty and *removed* from the dataset.
 - **Runs with some downward fluctuation were less carefully investigated, and usually not removed** ⇒ the insignificant fluctuation got a boost
- Voila, they suddenly had a “fake” peak!

What went wrong?

- The dip was noticed already in an early run.

Morale: ***Never remove data.*** If you suspect a problem, fix it, and start over.

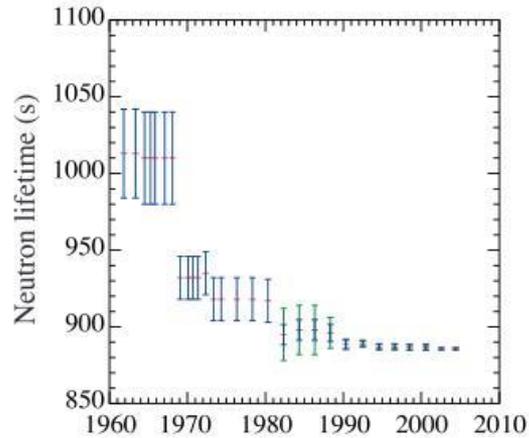
In a complicated experiment this is not always possible; some runs *are* actually bad etc. and should be discarded. But make that decision before embarking on analysis. Do not let the results influence the data you use!



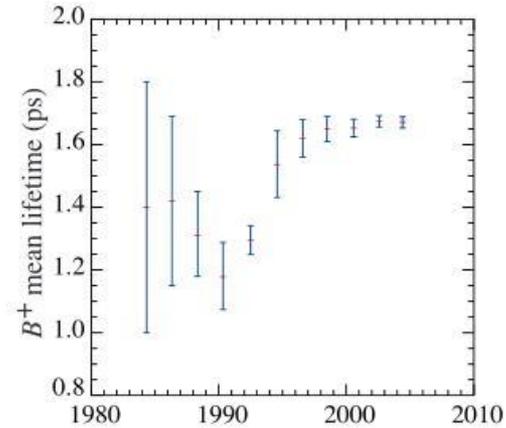
removed → the insignificant fluctuation got a boost

- Voila, they suddenly had a “fake” peak!

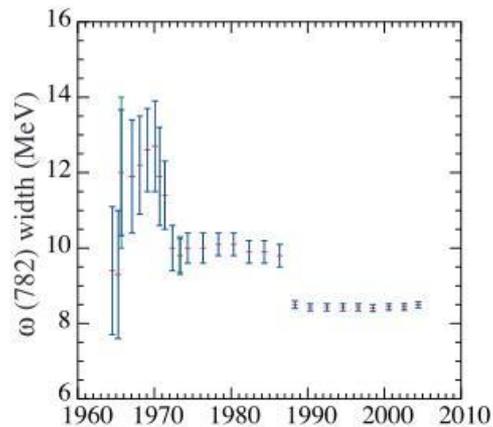
PDG history plots!



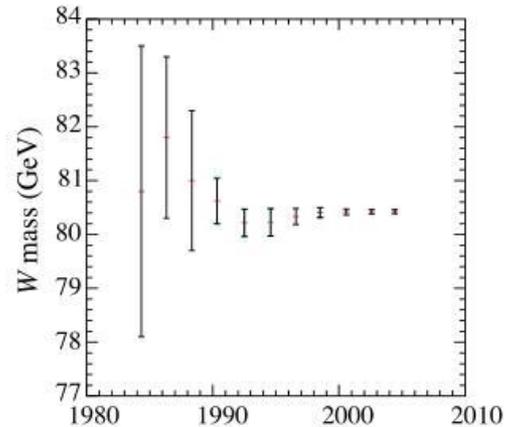
(a) Neutron lifetime vs. Publication Date



(b) B^+ lifetime vs. Publication Date



(c) ω width vs. Publication Date



(d) W mass vs. Publication Date

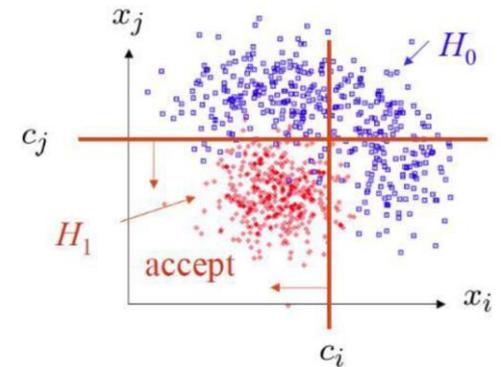
Reducing systematics?

- Easier once you have first estimate
 - *which sources are important and which negligible*
 - *More knowledge means more precise estimates*
- Take advantage of measurements where certain systematics cancel out
 - *Measure ratios and differences*
- Design analysis in more unbiased way
 - *“blind” analysis*

”Blind” analysis

Simply put, avoid looking at a potential signal (in data) as long as possible, to minimize biases

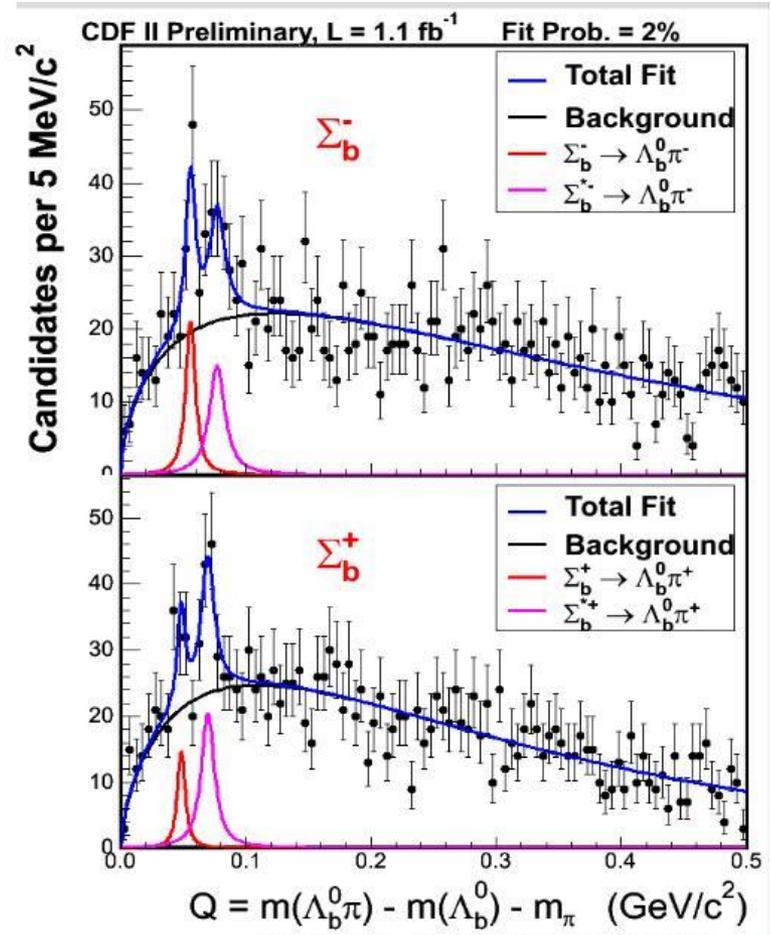
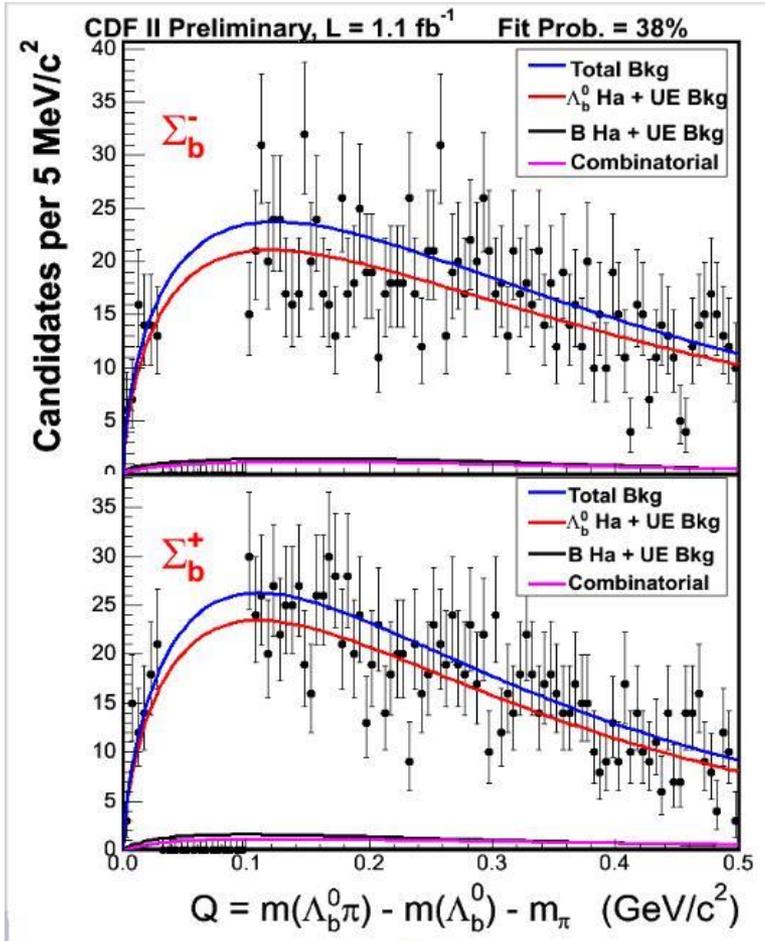
Most analyses are performed this way



Blind analysis doesn't mean:

- you never look at the data
- you can't correct a mistake if you find one
- the analysis is necessarily correct—it's merely blind!
- conversely, a non-blind analysis doesn't necessarily give the wrong answer, but it does leave open the risk of bias.

Example: Λ_b observation CDF

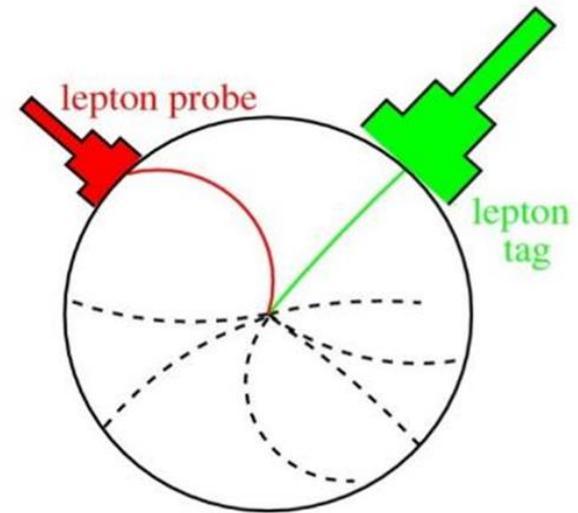


Estimations directly from data

To reduce systematics from data/ simulation differences, some estimates (or additional weights applied to MC) are taken directly from data ("data-driven")

Two examples much in use:

- Efficiencies
- Multi-jet background



Tag and probe

A very common methodology for *ATLAS*

- Tag & Probe is a method to study analysis objects (jets, μ , etc)
- Study trigger, reconstruction efficiencies
- Data and MC often disagree in many places
 - Determine many different weights, often binned in p_T and η
- Use Standard Model candles like $Z \rightarrow ee$, $W \rightarrow \mu\nu$
 - What to use decay products from well known particles
 - Know that $Z \rightarrow ee$ has 2 electrons
 - Know the invariant mass of a Z very well
 - “Tag” one electron and study the other, “Probe” electron
- Use data-driven methods to determine the detector response
 - No need for MC in determining trigger, reco efficiencies
 - MC generally only used to determine weights
- In *ATLAS* weights are often called **Scale Factors**

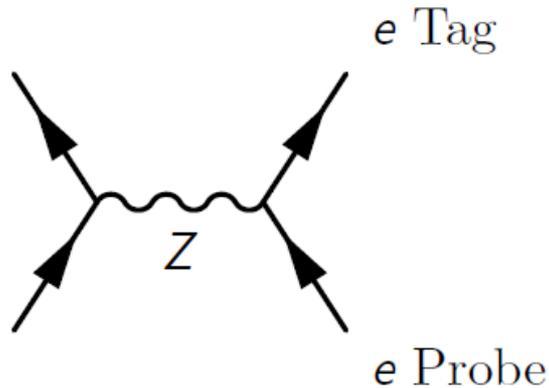
Tag and probe

A very common methodology for *ATLAS*

Of course this only works when the quantities under study are not correlated between the two electrons!!!

- Know the invariant mass of a Z very well
- “Tag” one electron and study the other, “Probe” electron
- Use data-driven methods to determine the detector response
 - No need for MC in determining trigger, reco efficiencies
 - MC generally only used to determine weights
- In *ATLAS* weights are often called **Scale Factors**

Tag & Probe : $Z \rightarrow ee$ example

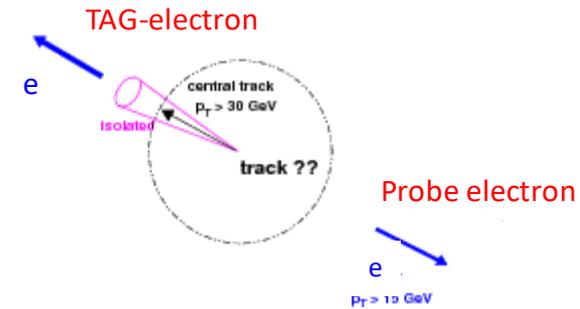


Invariant mass should be Z mass!

Use probe to determine efficiencies

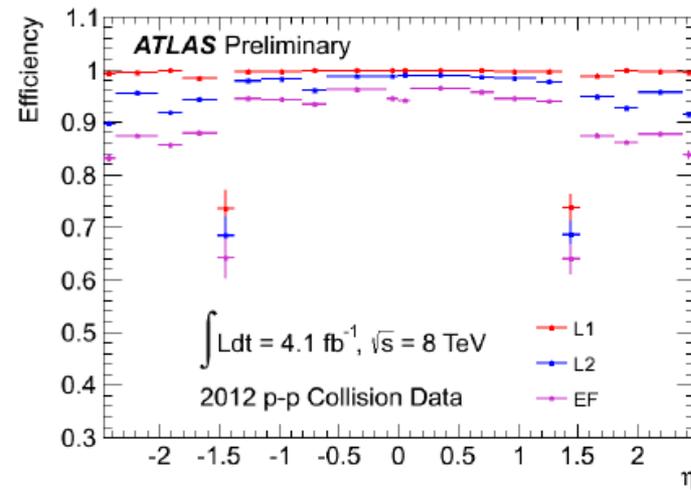
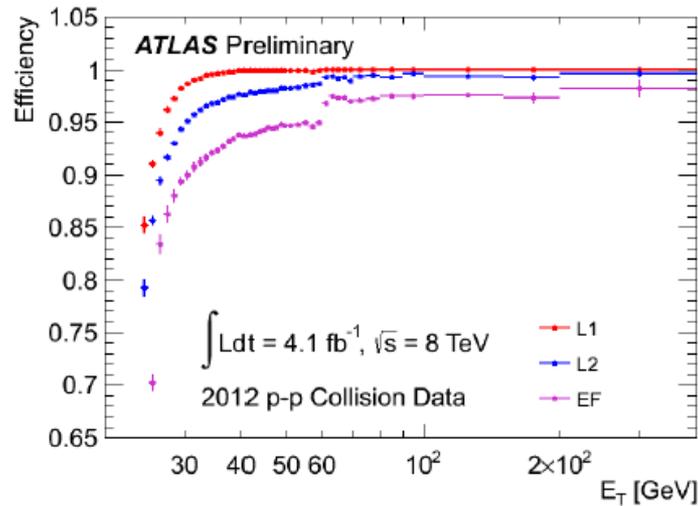
$$\text{Some Efficiency} = \frac{\text{Probe Passes Selection}}{\text{All probes}}$$

$$\text{For Example} :: \text{Trigger Efficiency} = \frac{\text{Probe Fires Trigger}}{\text{All probes}}$$



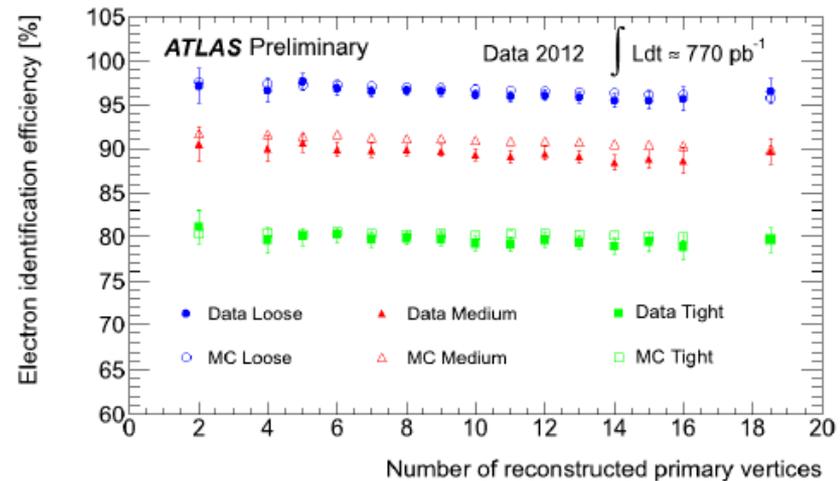
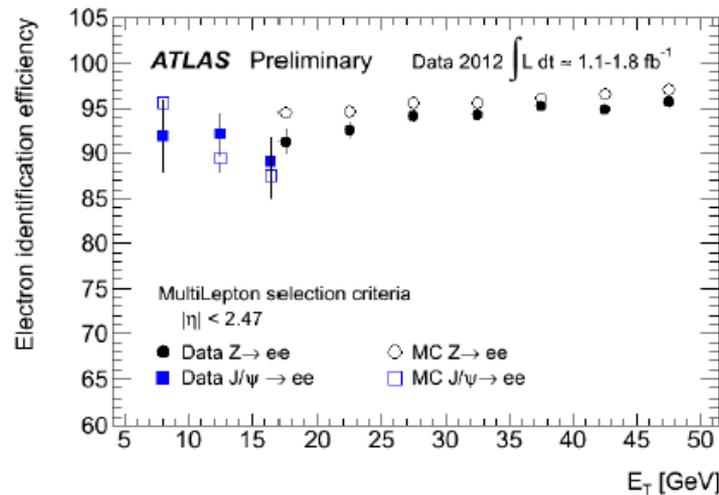
- Tag electron:
 - Matched to Trigger
 - Tight reconstruction
 - Satisfies all conditions for being an excellent electron
- Probe electron:
 - Minimal selection

Tag & Probe : Electron trigger efficiency



- $Z \rightarrow ee$ event selection applied
- Tag electron used to select events
 - No requirement on probe to pass trigger
- Probe electron asked “Did you fire the trigger?”
- Trigger efficiency determined for L1, L2 and EF
- Shown as a function of electron E_T and η

Tag & Probe : Electron Identification efficiency



- $J/\psi \rightarrow ee$ and $Z \rightarrow ee$ event selection applied
- Tag electron used to select events
 - No requirement on probe to pass identification cuts
- Probe electron asked “Did you pass ID selection?”
- Shown as a function of electron E_T and number of vertices
- Note how MC does not match data \Rightarrow weights are required

Data-driven background estimation

In some cases unrealistic to simulate the background

- For instance multijet production faking leptons
- Low probability but $\sigma(\text{pp} \rightarrow \text{jets}) \gg \sigma(\text{pp} \rightarrow \text{EWK} \rightarrow \text{leptons})$
- Would need HUGE MC samples and understand all details in detector + hadronization with precision

Thus, for these often use data-driven methods instead.

Some standard methods:

- "ABCD" methods
- Matrix method
- Fake factor method

ABCD method

Two uncorrelated variables for each channel divided up into 4 regions in that parameter space

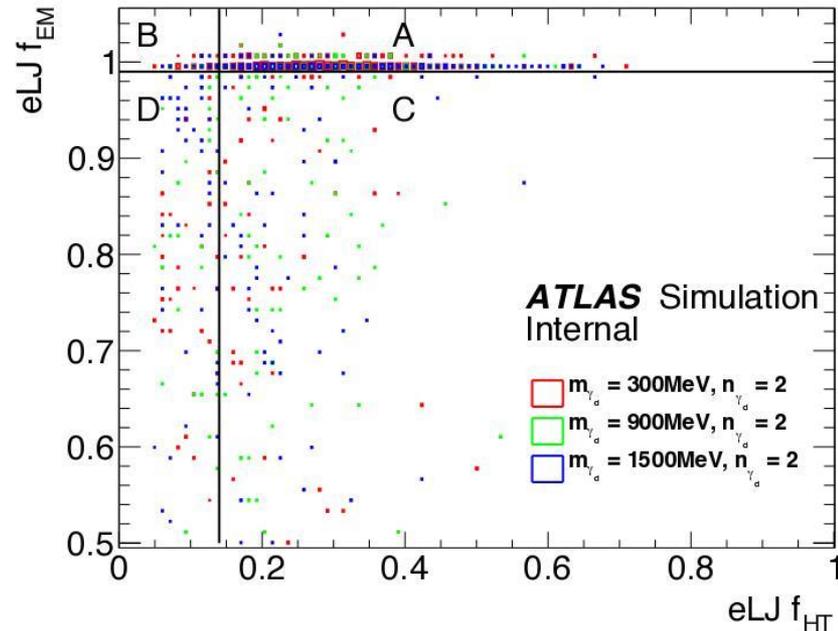
Region A = Signal concentrated region

Regions B, C, D = background concentrated regions (control regions)

Amount of QCD bkg due to hadronic jets in A can be estimated as:

$$N_A = \frac{N_B \times N_C}{N_D}$$

In realistic cases (with signal also in B, C, D) use a likelihood to estimate relative rates in the 4 regions.



Example from the lepton-jets search (JHEP 03 (2016) 026)

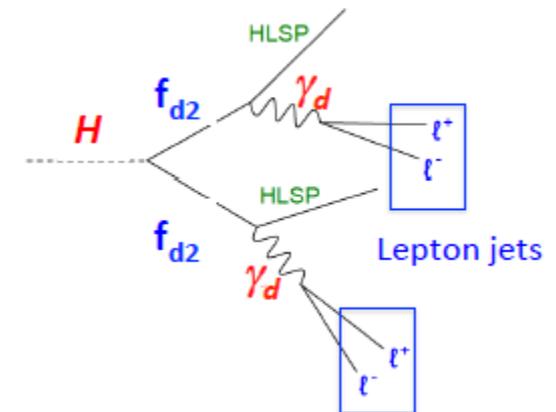
Recently (sort of) published search for dark photons, dark fermions. Model to explain PAMELA positron excess

Signal "pollution" exists, thus a

likelihood fit used. Performance:

QCD \mathcal{L} predicted

Channel	ABCD likelihood method	Total background in region A	Observed events from the data
eLJ-eLJ	2.9 ± 0.9	4.4 ± 1.2	6
muLJ-muLJ	2.9 ± 0.6	4.4 ± 1.1	4
eLJ-muLJ	6.7 ± 1.4	7.1 ± 1.4	2
eLJ-emuLJ	7.8 ± 2.0	7.8 ± 2.0	5
muLJ-emuLJ	20.2 ± 4.5	20.3 ± 4.5	14
emuLJ-emuLJ	1.3 ± 0.8	1.9 ± 0.9	0



The Matrix method

Built from two rates:

The real rate: probability that a real lepton identified as a loose lepton gets identified as a tight lepton

The fake rate: probability that a real jet identified as a loose leptons is identified as tight lepton

Single lepton selection: the # of loose and tight leptons can be written as: $N^L = N_R^L + N_F^L$; $N^T = \varepsilon_R^L N_R + \varepsilon_F^L N_F$

Where ε 's are the fraction of events that pass from loose to tight

These are measured in control data samples, depends on kinematics and jet type

In the end results in weights given to each event:

$$W = \frac{\varepsilon_F \varepsilon_R}{\varepsilon_R - \varepsilon_F} \text{ if it fails loose cuts and } \frac{\varepsilon_F}{\varepsilon_R - \varepsilon_F} (\varepsilon_R - 1)$$

otherwise

The Matrix method

The matrix when selecting events with two leptons::

$$\begin{pmatrix} N_{TT} \\ N_{TL} \\ N_{LT} \\ N_{LL} \end{pmatrix} = M(r_1, r_2, f_1, f_2) \begin{pmatrix} N_{RR} \\ N_{RF} \\ N_{FR} \\ N_{FF} \end{pmatrix}$$

$$M = \begin{pmatrix} r_1 r_2 & r_1 f_2 & f_1 r_2 & f_1 f_2 \\ r_1(1-r_2) & r_1(1-f_2) & f_1(1-r_2) & f_1(1-f_2) \\ (1-r_1)r_2 & (1-r_1)f_2 & (1-f_1)r_2 & (1-f_1)f_2 \\ (1-r_1)(1-r_2) & (1-r_1)(1-f_2) & (1-f_1)(1-r_2) & (1-f_1)(1-f_2) \end{pmatrix}$$

- We can measure all N_{LL} , N_{TL} , N_{LT} , and N_{TT} and then invert the matrix to get N_{RF} , N_{FR} , and N_{FF} , which is the contribution of fakes to the di-electron selection.
- To use the method, we need to measure **the real** and **the fake rate!**

Estimated background given by event weight: $w_{TT} = r_1 f_2 w_{RF} + f_1 r_2 w_{FR} + f_1 f_2 w_{FF}$

Fake factors

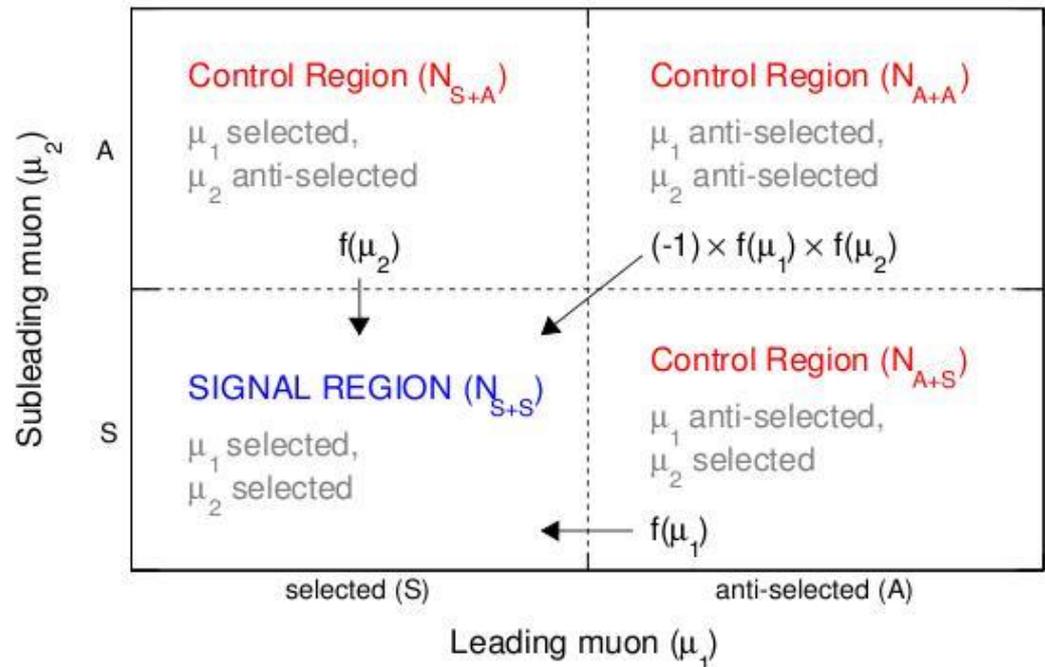
Define data control region inverting some selection criteria, then

extrapolate this into signal region: $f \equiv \frac{N_{selected}}{N_{Anti-selected}}$

where $f = \text{function}(p_T, \eta)$

Example with two muons:

$$N_{\text{multijet}} = \sum_{i=1}^{N(A+S)} f(\mu) + \sum_{i=1}^{N(S+A)} f(\mu) + \sum_{i=1}^{N(A+A)} f(\mu)$$



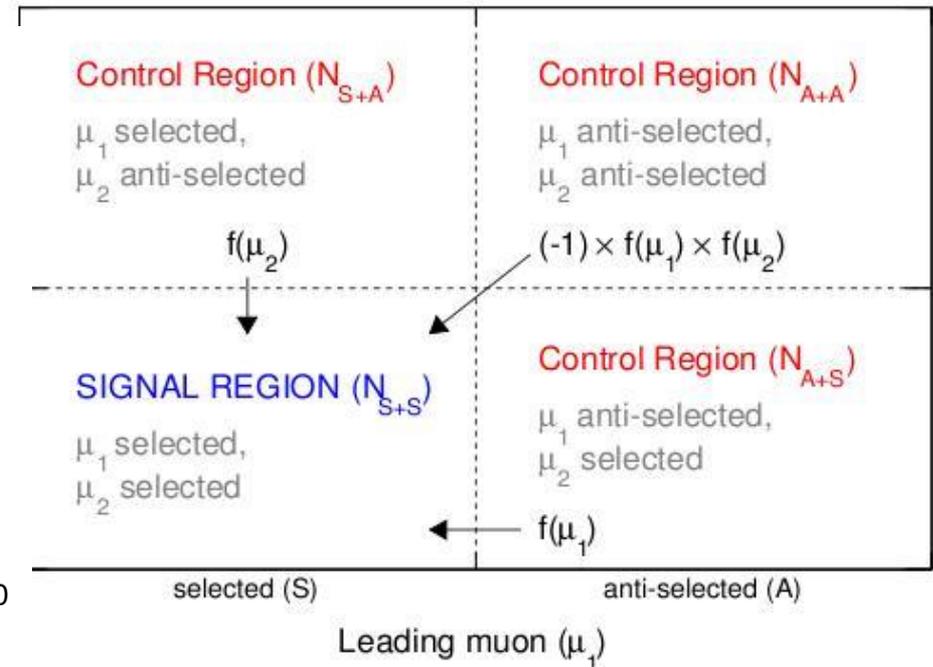
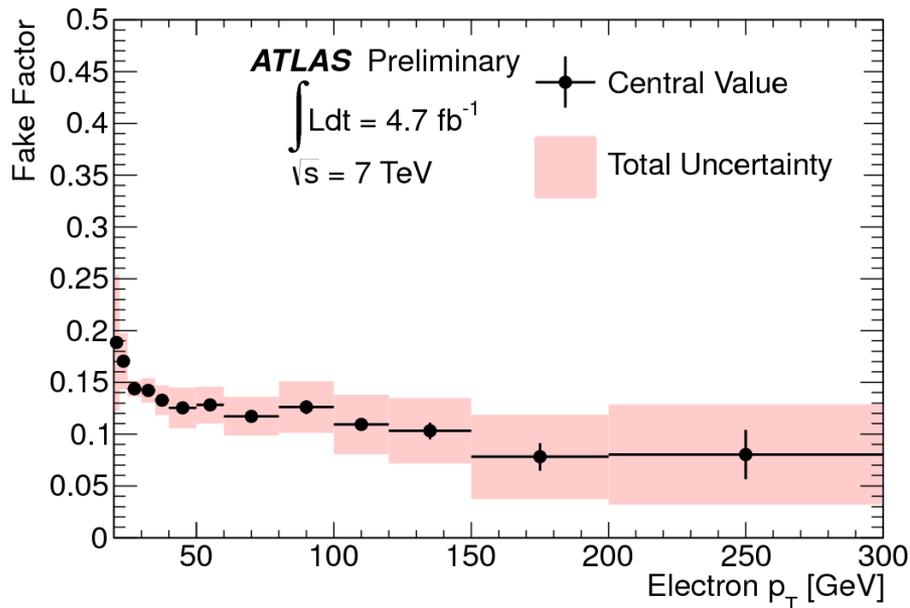
Needs independent sample for measuring f , as well as corrections for other backgrounds

Fake factors

Define data control region inverting some selection criteria, then

extrapolate this into signal region:
$$f \equiv \frac{N_{selected}}{N_{Anti-selected}}$$

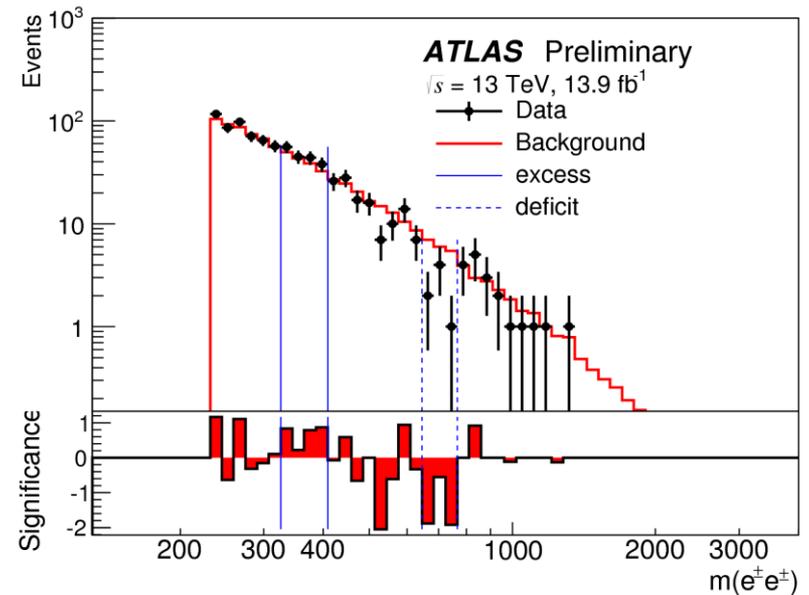
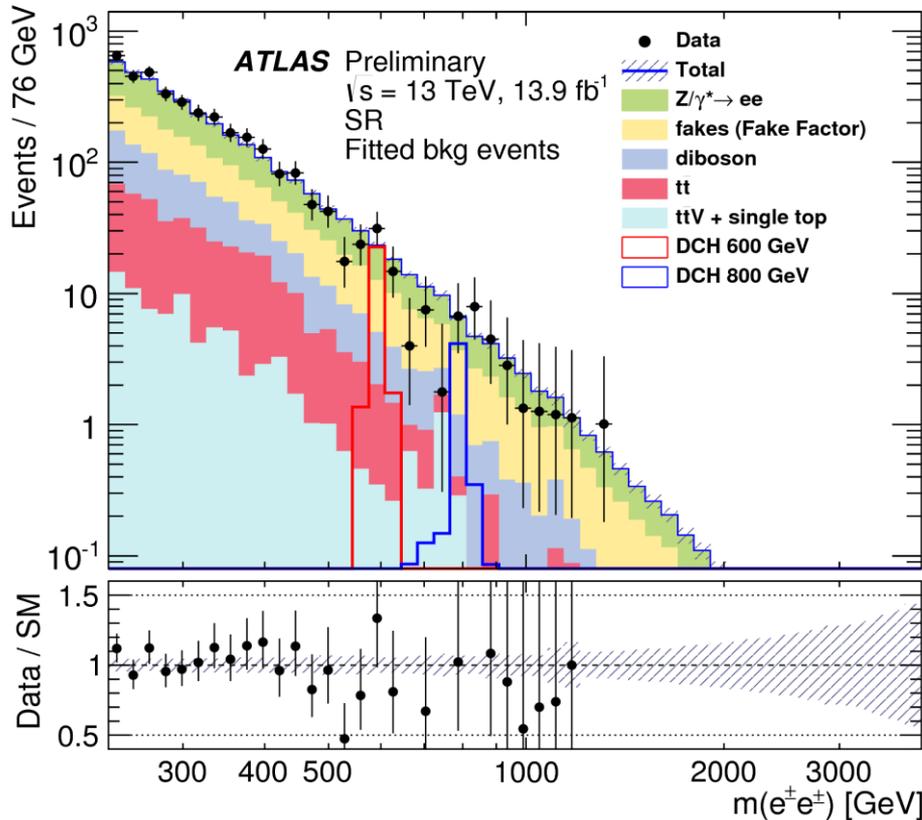
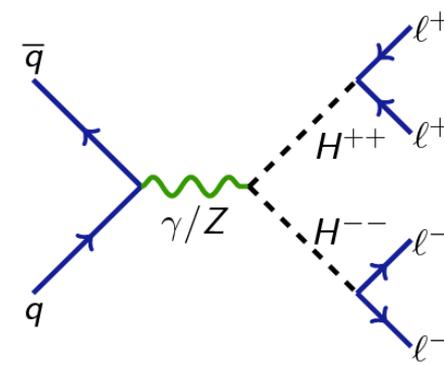
where $f = \text{function}(n_{+}, n_{-})$



Needs independent sample for measuring f , as well as corrections for other backgrounds

Examples of results

(ATLAS-CONF-2016-051)



Excess: 1.5σ , p-value 0.9
 Deficit: 1.3σ , p-value 0.09

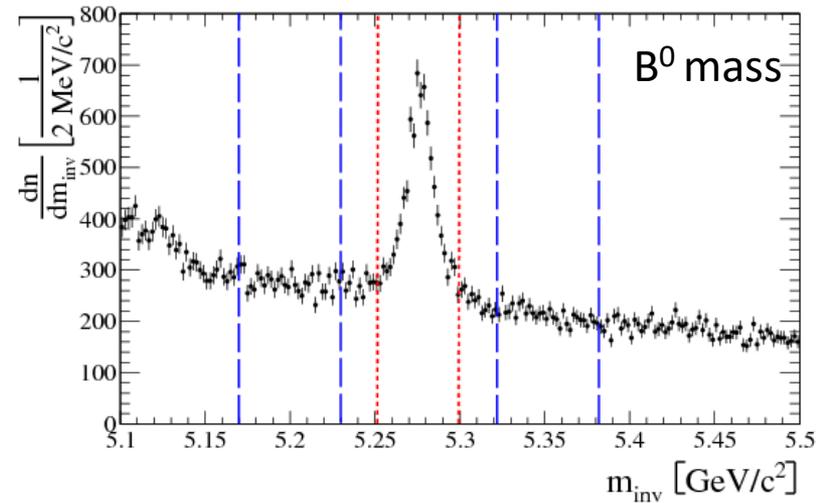
This search sets a limit of doubly-charged higgs (DCH) mass between 420 GeV and 530 GeV (depending on the couplings)

Pros and cons

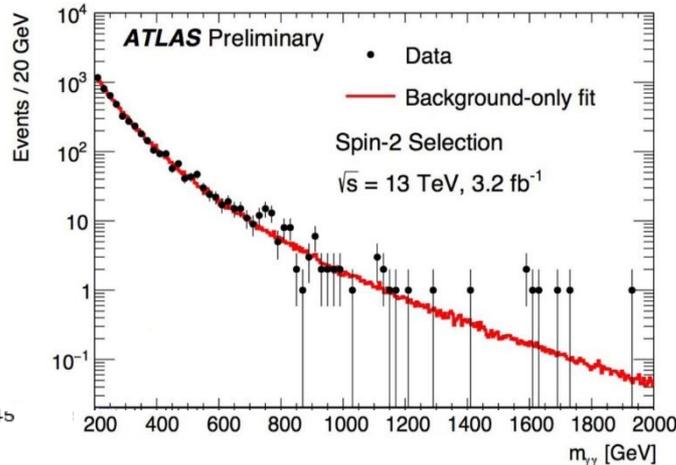
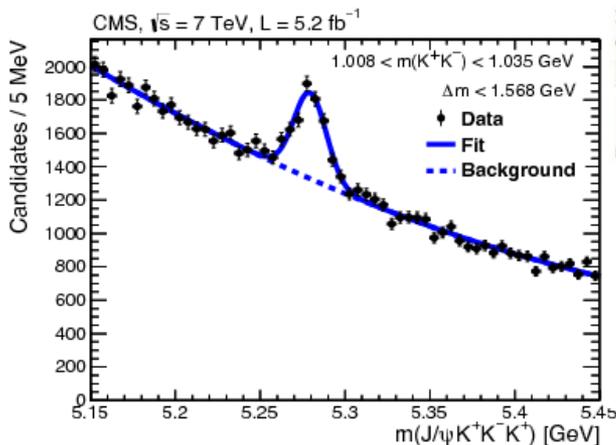
- ABCD method
 - Simple, if applicable
 - Hard to find the best, uncorrelated variables, and to test validity of method in advance
- Matrix method:
 - Precise, in theory
 - In reality, lots of efficiencies to be measured – i.e. potentially correlated or large uncertainties
 - Overlaps between different types of backgrounds hard to distinguish
- Fake factors
 - "simplified" matrix method
 - Some precision lost
 - How to define appropriate control regions

Alternatives in special cases

The *sidebands* can be used to estimate the background under a peak



A smooth, high statistics background can be fitted:



2015 diphoton bump

This you can try yourself: in ROOT library find macro rf_fit_for_peak.cc

Gaussian peak on pol background: the J/ψ mass peak

$$\text{Signal } G(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

μ is fixed at J/ψ mass 3.15 GeV
 $\sigma \cong 0.05$ GeV, allowed to float

$$\text{Background } P(x) = ax + bx^2$$

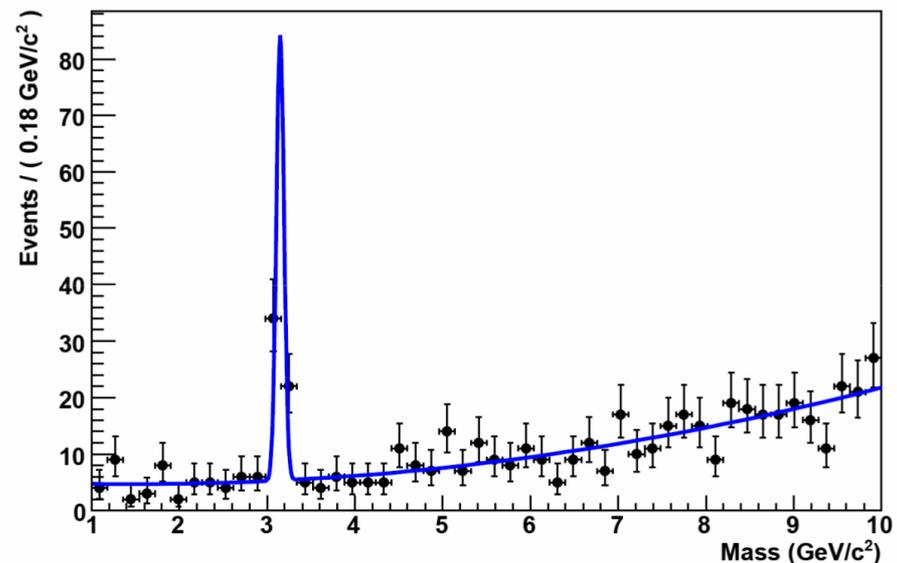
i.e. the total pdf is $N_{signal} G(x, \mu, \sigma) + N_{background} P(x)$

Do the fit! :

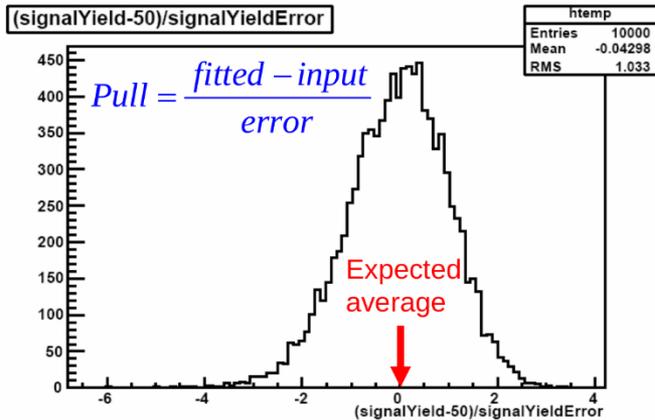
$$\sigma = 0.043 \pm 0.005$$

$$a = -0.14^{+0.13}_{-0.08}$$

$$b = 0.045 \pm 0.008$$



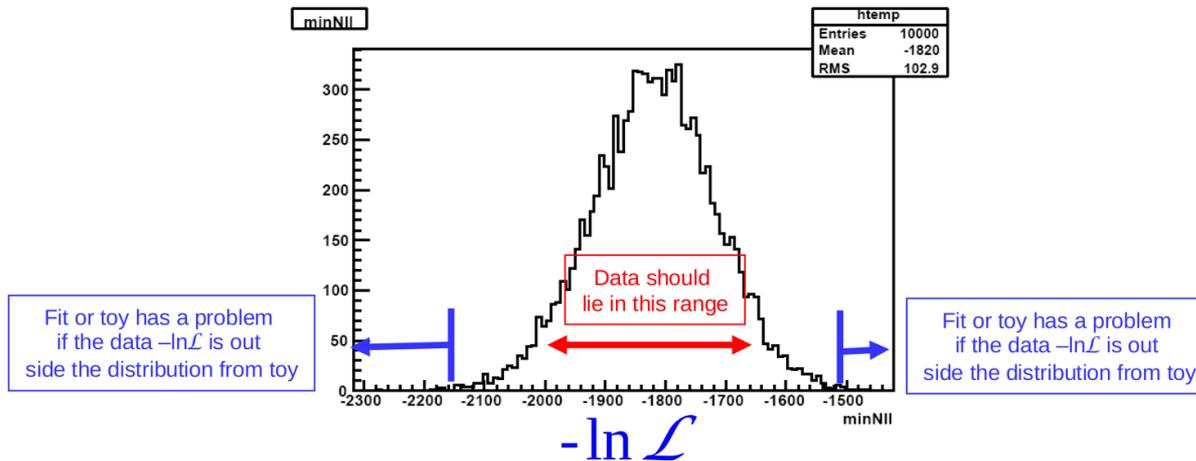
Checks: Run MC simulations ("Toy MC") to validate!



← Checking if error on signal yield reasonable

Log likelihood test: is $-\ln \mathcal{L}_{DATA}$ compatible with $-\ln \mathcal{L}_{MC}$?

$$\mathcal{L} = e^{-(N_{\text{signal}} + N_{\text{background}})} \prod_{i=1}^{n \text{ events}} N_{\text{signal}} \text{PDF}_{\text{signal}} + N_{\text{background}} \text{PDF}_{\text{background}}$$



Background estimation cont.

- *Optimal strategy depends on the specific analysis!*
 - Simulation or data-driven, or a combination?
 - Which data-driven method
- More methods than shown here (for instance template method often used) and variations over the “standard” methods
- In some cases we use more than one method – very useful to get a real estimate of systematic uncertainties in either methods
 - (but of course time consuming)

Summary

- Systematic uncertainties important – *can be your dominant source of uncertainty!*
- Hard to estimate – no recipe
 - Nevertheless we do have some go-to procedures
 - Self critical attitude (paranoia?!) can help uncover hidden systematics
- To decrease potential biases, most analyses are performed as “blind” analyses
- Statistical methods come in different disguises
 - Efficiencies and background estimates are sources of systematic uncertainties

Links to ROOT framework

<https://root.cern.ch> and

https://root.cern.ch/notebooks/HowTos/HowTo_ROOT-Notebooks.html

Python flavour

In order to use ROOT in a Python notebook, we first need to import the ROOT module. During the import, all notebook related functionalities are activated.

In [1]: import ROOT

Welcome to ROOTaaS 6.05/01

Now we are ready to use [PyROOT](#). For example, we create a histogram.

In [2]:

```
h = ROOT.TH1F("gauss","Example histogram",100,-4,4)
```

```
h.FillRandom("gaus")
```

Next we create a [canvas](#), the entity which holds graphics primitives in ROOT.

In [3]:

```
c = ROOT.TCanvas("myCanvasName","The Canvas Title",800,600)
```

```
h.Draw()
```

For the histogram to be displayed in the notebook, we need to draw the canvas.

In [4]:

```
c.Draw()
```

