FYST17 Lecture 9
Statistical methods in Particle Physics

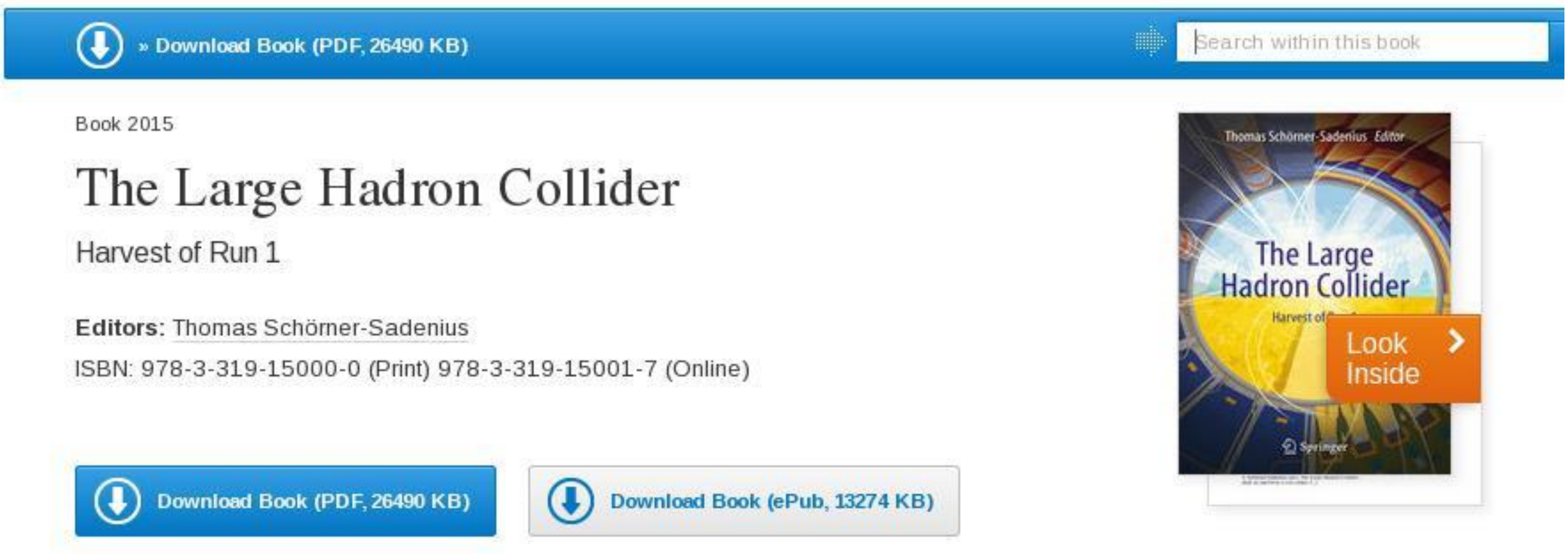
Thanks to J. Morris, S. Menzemer

Outline

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

Link:

http://link.springer.com/book/10.1007/978-3-319-15001-7

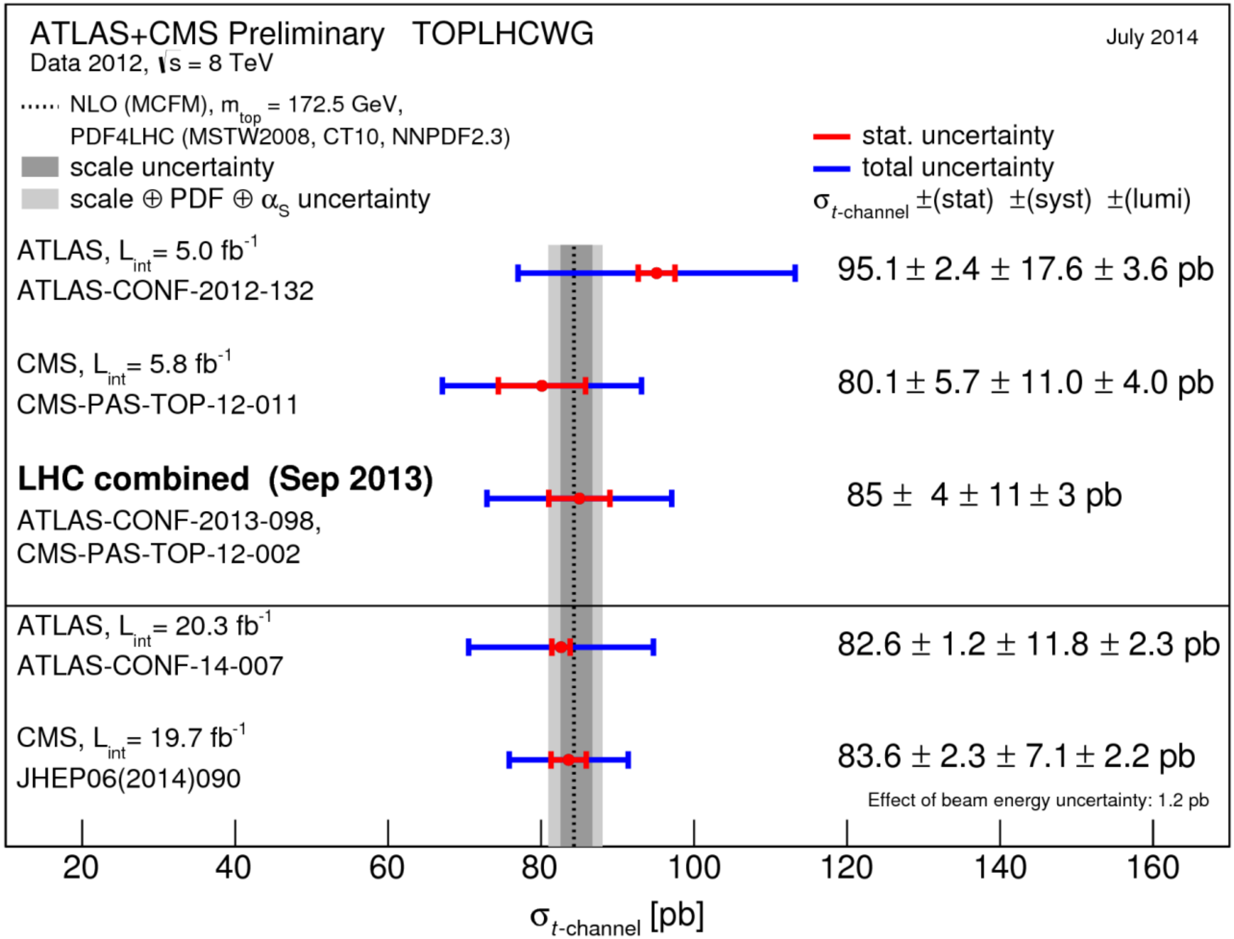


What is a systematic uncertainty?

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

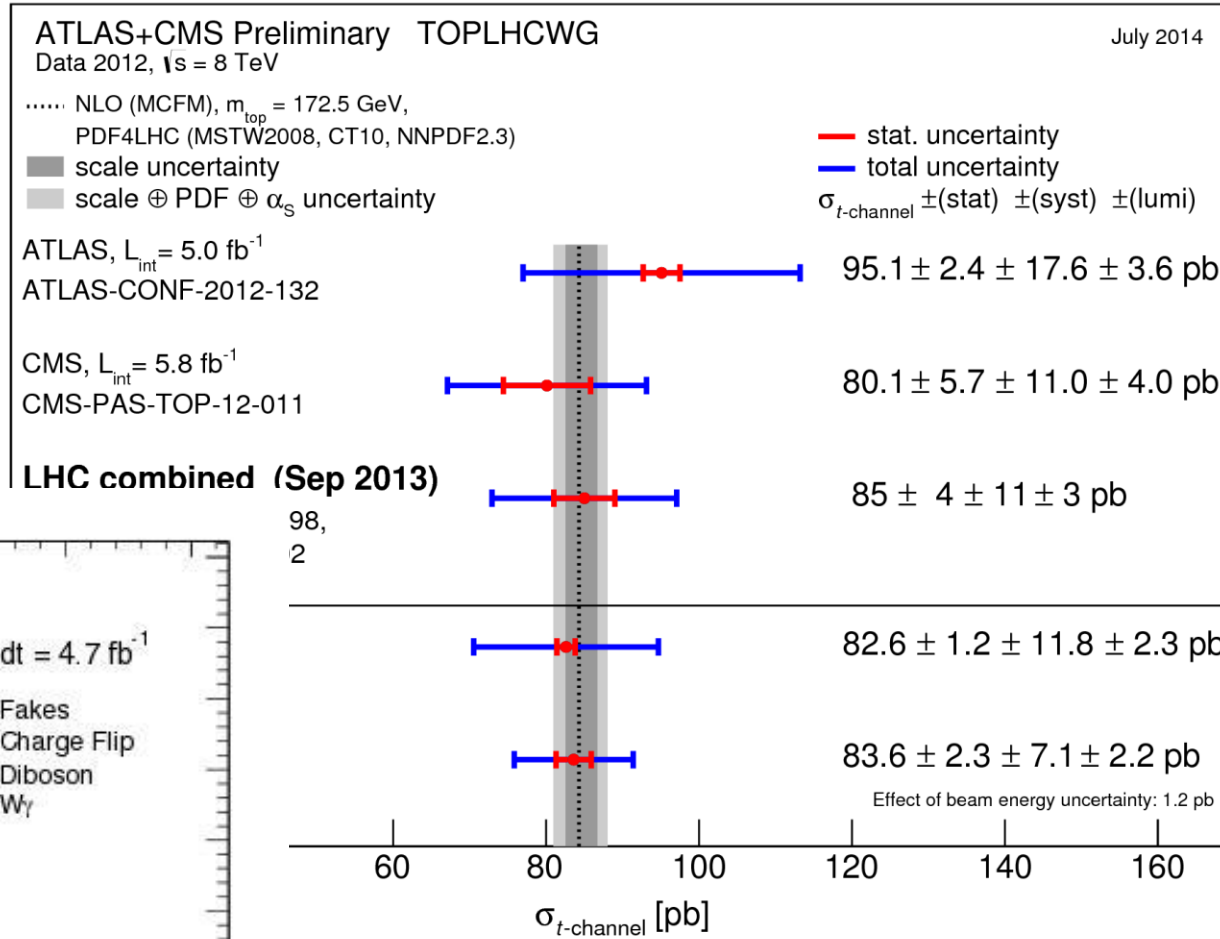
- Badly known backgrounds
- Badly known detector resolutions
- Wrong calibrations
- Badly known acceptances or efficiencies
- Preferred outcomes
- External factors, such as theory uncertainties on cross sections etc
- Other biases ….

Examples

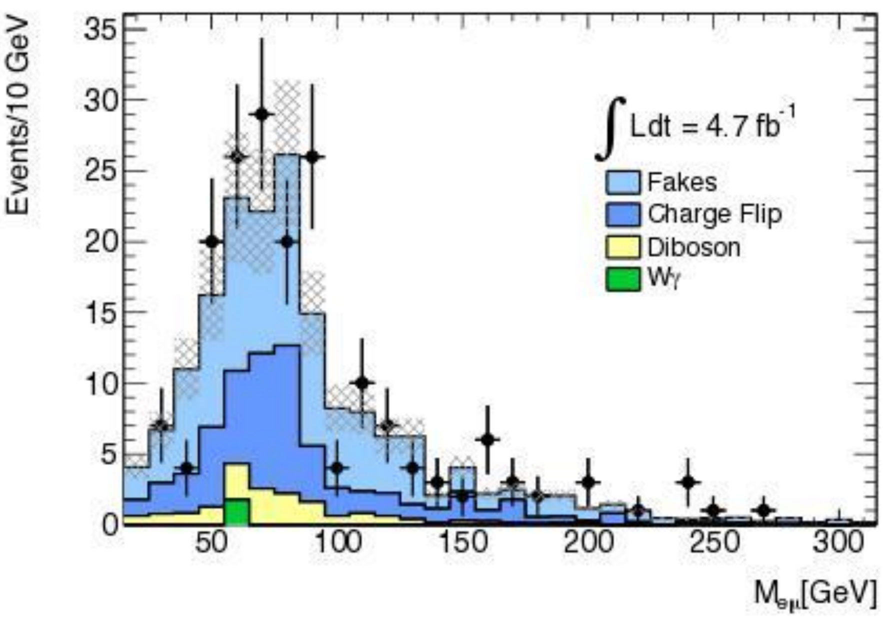


ATLAS+CMS Preliminary TOPLHCWG
July 2014
Data 2012, √s = 8 TeV
NLO (MCFM), m_top = 172.5 GeV,
PDF4LHC (MSTW2008, CT10, NNPDF2.3)
scale uncertainty
scale ⊕ PDF ⊕ α_S uncertainty
stat. uncertainty
total uncertainty
σ_t-channel ±(stat) ±(syst) ±(lumi)
ATLAS, L_int = 5.0 fb^-1
ATLAS-CONF-2012-132
95.1 ± 2.4 ± 17.6 ± 3.6 pb
CMS, L_int = 5.8 fb^-1
CMS-PAS-TOP-12-011
80.1 ± 5.7 ± 11.0 ± 4.0 pb
LHC combined (Sep 2013)
ATLAS-CONF-2013-098,
CMS-PAS-TOP-12-002
85 ± 4 ± 11 ± 3 pb
ATLAS, L_int = 20.3 fb^-1
ATLAS-CONF-14-007
82.6 ± 1.2 ± 11.8 ± 2.3 pb
CMS, L_int = 19.7 fb^-1
JHEP06(2014)090
83.6 ± 2.3 ± 7.1 ± 2.2 pb
Effect of beam energy uncertainty: 1.2 pb
20
40
60
80
100
120
140
160
σ_t-channel [pb]

Examples



ATLAS+CMS Preliminary TOPLHCWG
July 2014
Data 2012, √s = 8 TeV
NLO (MCFM), m_top = 172.5 GeV,
PDF4LHC (MSTW2008, CT10, NNPDF2.3)
scale uncertainty
scale ⊕ PDF ⊕ α_S uncertainty
stat. uncertainty
total uncertainty
σ_t-channel ±(stat) ±(syst) ±(lumi)
ATLAS, L_int = 5.0 fb⁻¹
ATLAS-CONF-2012-132
95.1 ± 2.4 ± 17.6 ± 3.6 pb
CMS, L_int = 5.8 fb⁻¹
CMS-PAS-TOP-12-011
80.1 ± 5.7 ± 11.0 ± 4.0 pb
LHC combined (Sep 2013)
98,
2
85 ± 4 ± 11 ± 3 pb
82.6 ± 1.2 ± 11.8 ± 2.3 pb
83.6 ± 2.3 ± 7.1 ± 2.2 pb
Effect of beam energy uncertainty: 1.2 pb
60
80
100
120
140
160
σ_t-channel [pb]



Events/10 GeV
35
30
25
20
15
10
5
0
∫ Ldt = 4.7 fb⁻¹
Fakes
Charge Flip
Diboson
Wγ
50
100
150
200
250
300
M_eμ[GeV]

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

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

More data will not help!

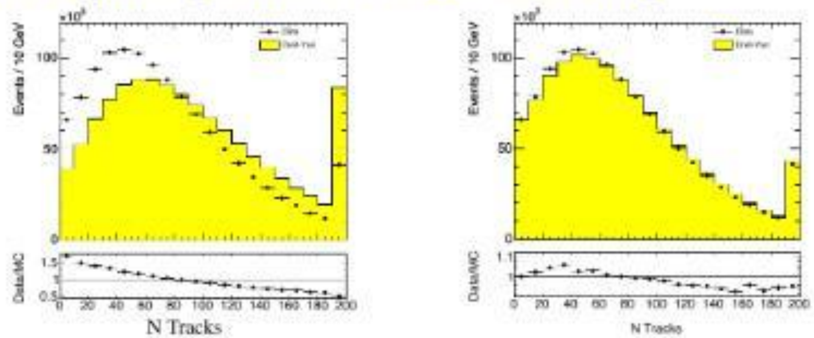
How are systematics estimated?!

- 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 to simulation

Correcting MC

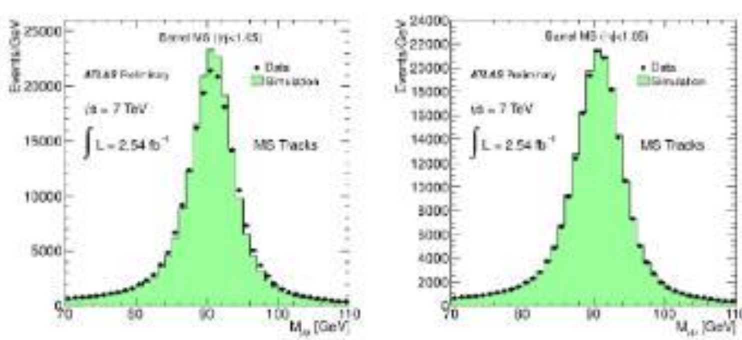
Use event weighting and 4-vector smearing



(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

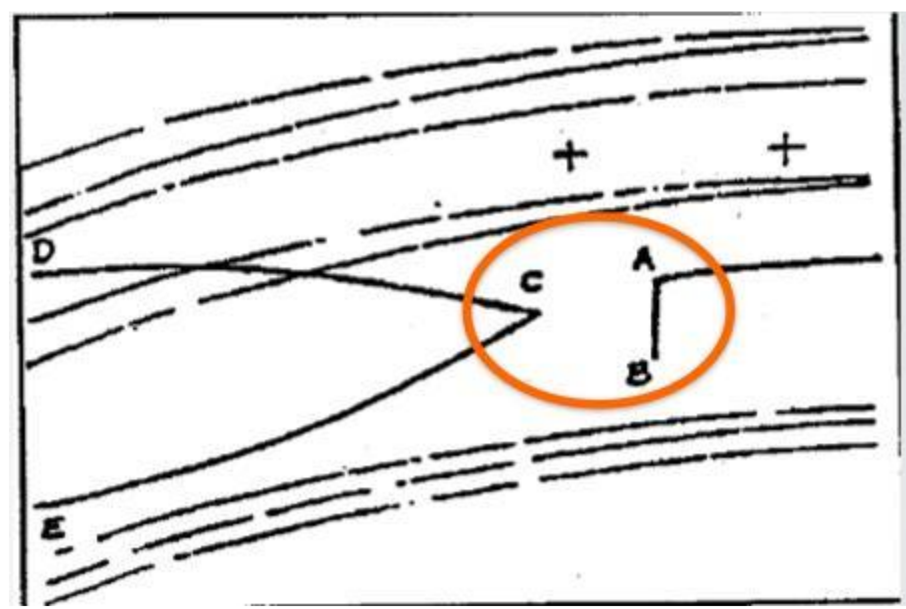


How to look for a particle

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

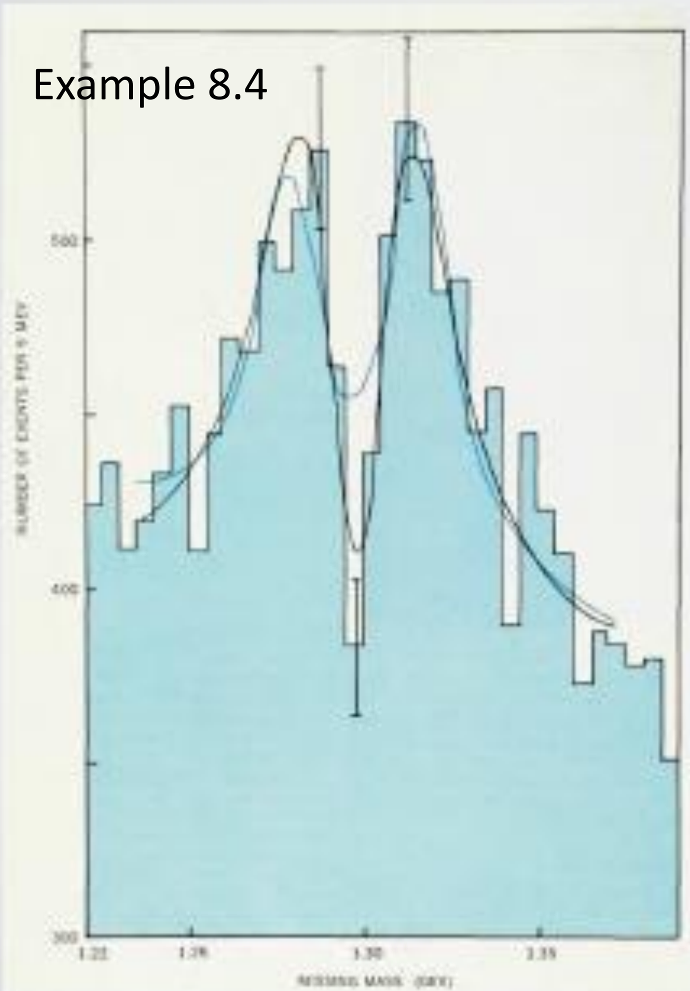
   (for instance, $K^0 \rightarrow \pi^+\pi^-$, displaced vertex)

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



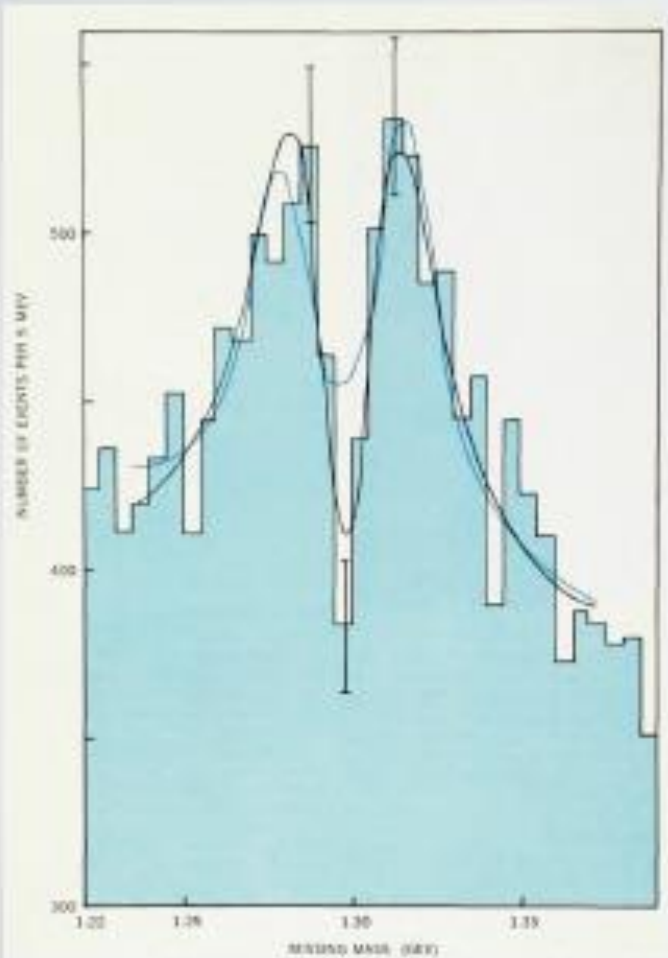
A CAUTIONARY TALE: ONE PEAK OR TWO?

Example 8.4



- CERN experiment in late 1960s observed A_2 mesons
- Particle appeared to be a doublet
- Statistical significance of split is very high
- There is really only one particle!!

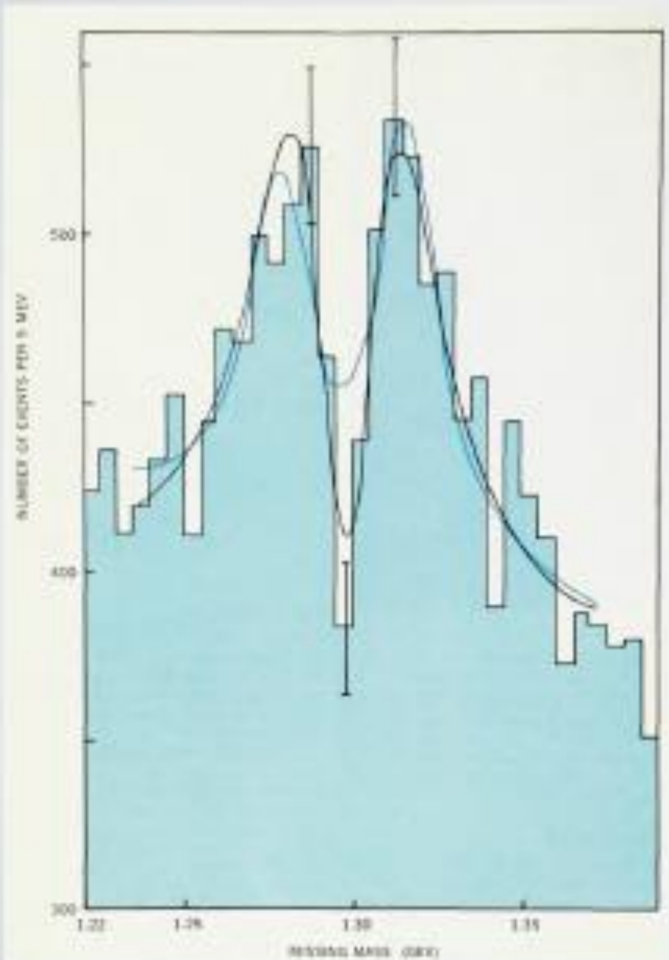
A CAUTIONARY TALE: HOW DID THIS HAPPEN?



- In an early run, a dip showed up. It was a statistical fluctuation, but people noticed it and suspected it might be real.

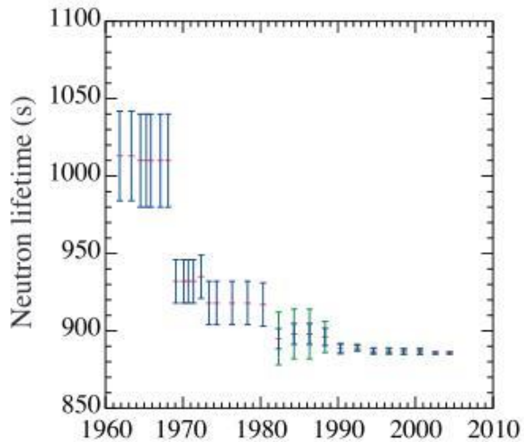
- Subsequent runs were looked at as they came in. If no dip showed up, the run was investigated for problems. There's usually a minor problem somewhere in a complicated experiment, so most of these runs were cut from the sample.

A CAUTIONARY TALE: HOW DID THIS HAPPEN?

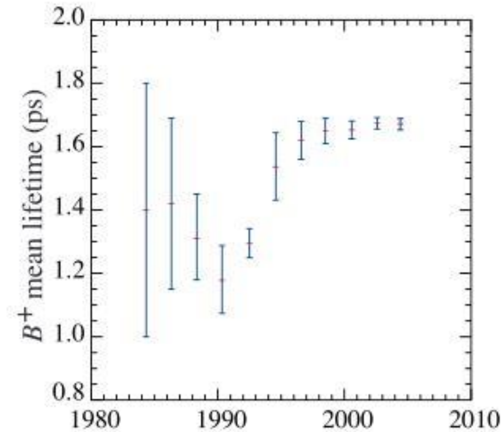


- When a dip appeared, they didn't look as carefully for a problem.
- So an insignificant fluctuation was boosted into a completely wrong "discovery."
- Lesson: Don't let result influence which data sets you use/want.

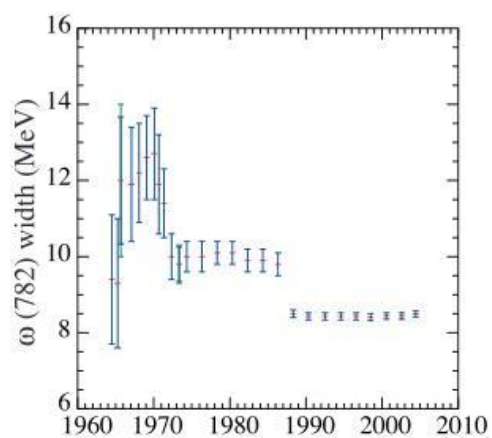
PDG history plots!



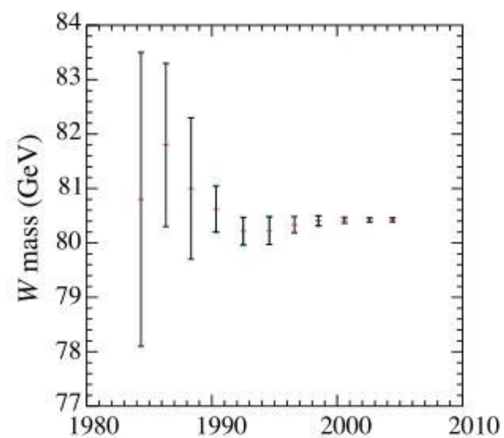
(a) Neutron lifetime vs. Publication Date



(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*

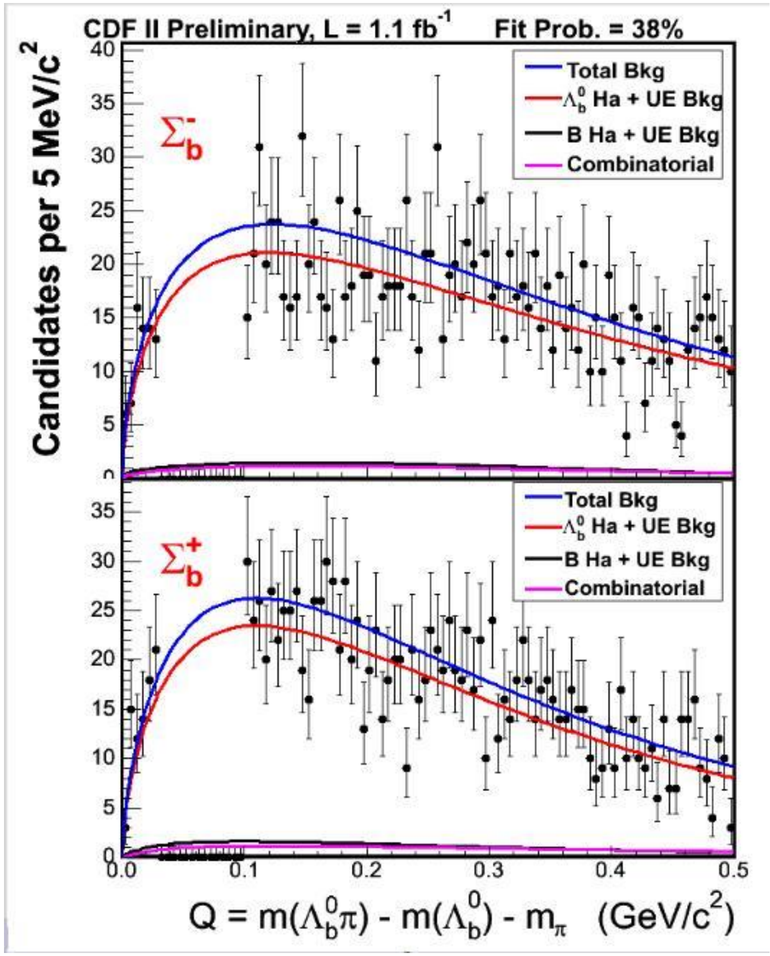
”Blind” analysis

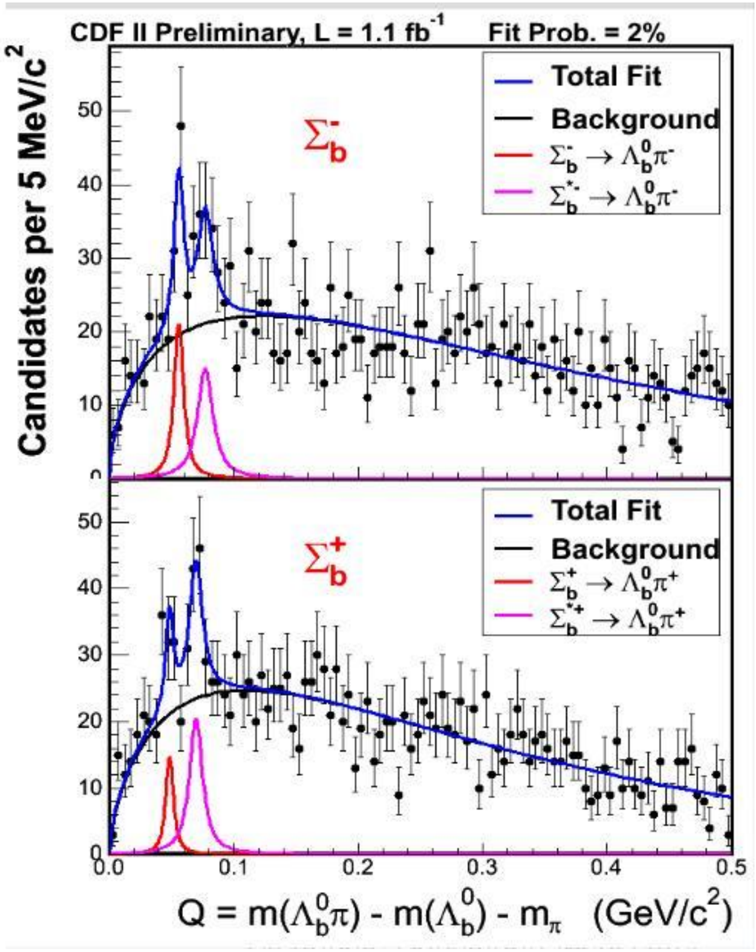
Simply put, avoid looking at a potential signal (in data) as long as possible, to minimize biases

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: Σ_b observation CDF



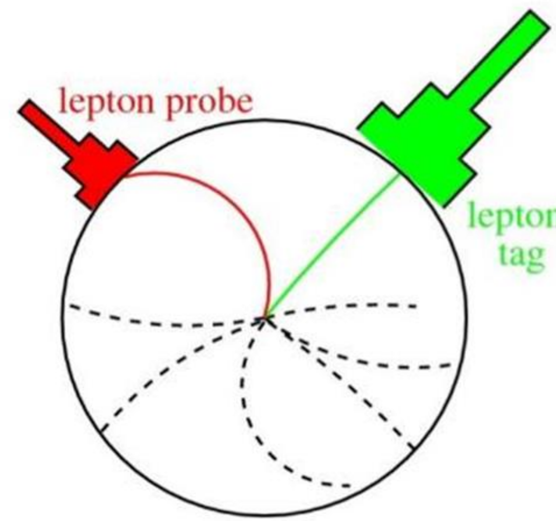


Estimations directly from data

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 $Z \to ee$, $W \to \mu\nu$
  - What to use decay products from well known particles
  - Know that $Z \to ee$ 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

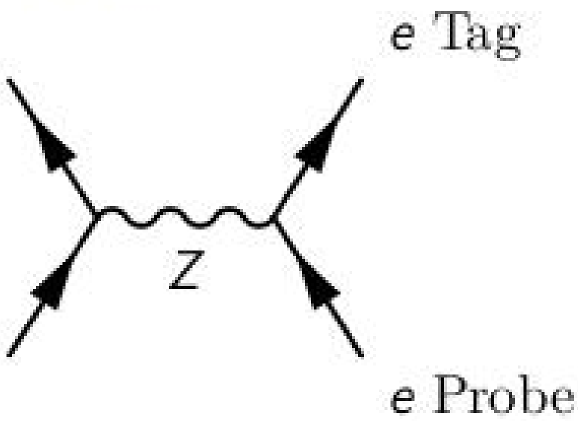
Tag and probe

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 : $Z \to ee$ example



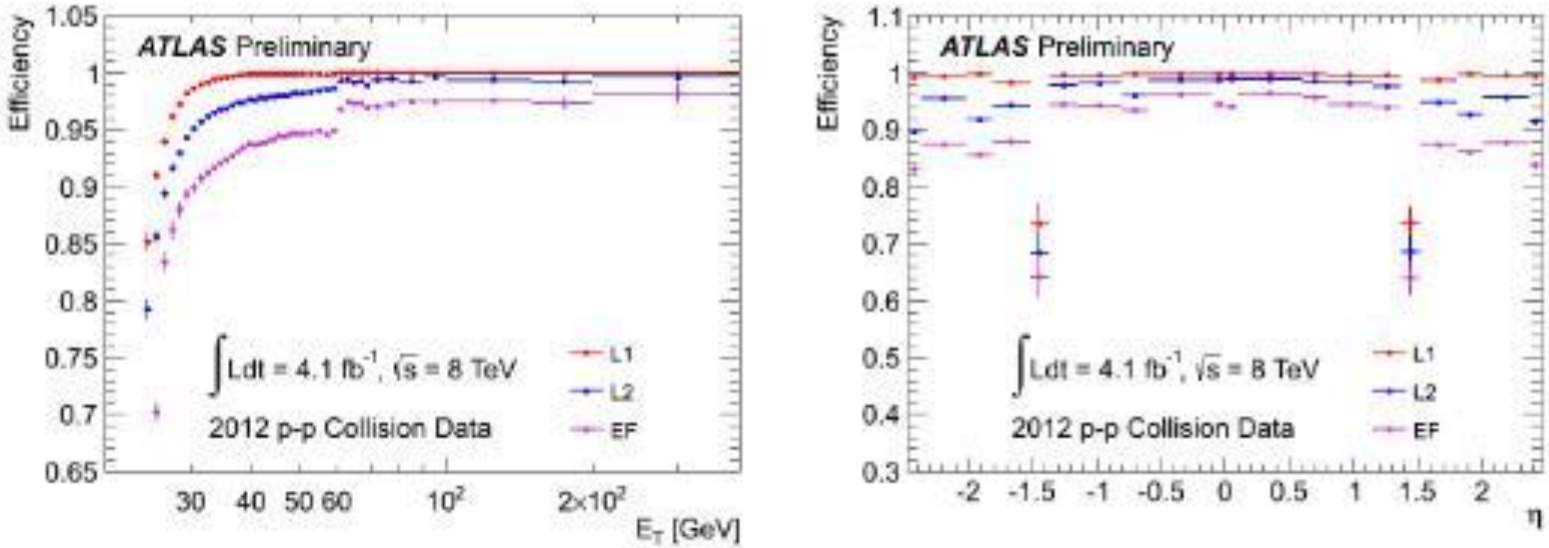
- Tag electron:
  - Matched to Trigger
  - Tight reconstruction
  - Satisfies all conditions for being an excellent electron
- Probe electron:
  - Minimal selection

Use probe to determine efficiencies

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

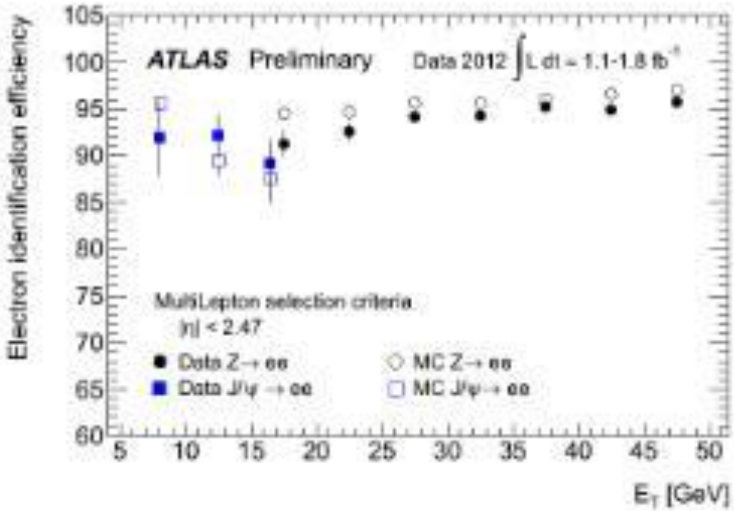
$$\text{For Example :: Trigger Efficiency} = \frac{\text{Probe Fires Trigger}}{\text{All probes}}$$

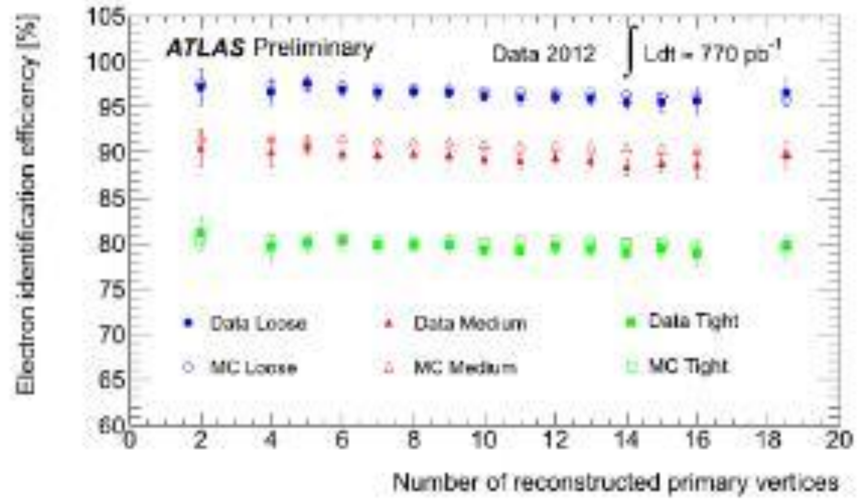
Tag & Probe : Electron trigger efficiency



- $Z \to ee$ 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





- $J/\Psi \to ee$ and $Z \to ee$ 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 $\sigma(pp \rightarrow \text{jets}) >> \sigma(\text{EWK} \rightarrow \text{leptons})$
- 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

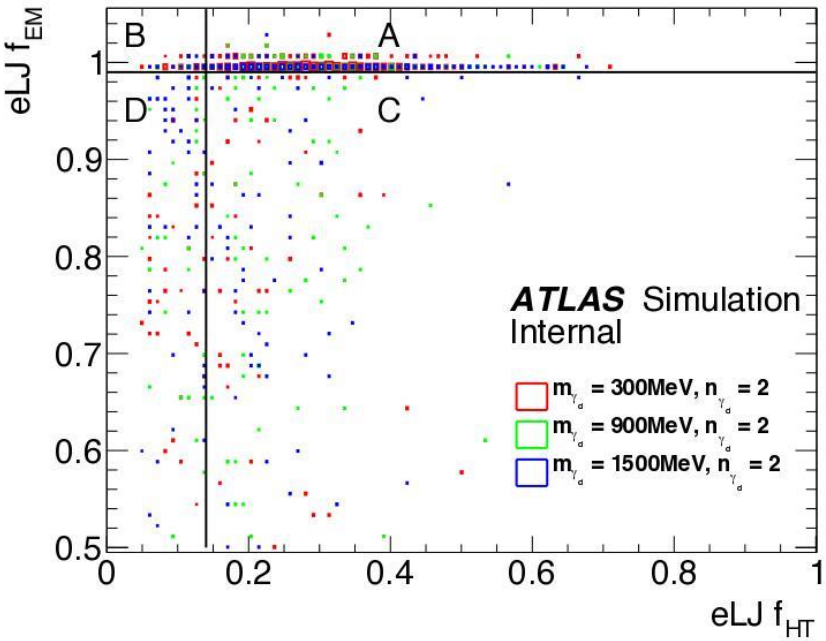
Region A = Signal concentrated region

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

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

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

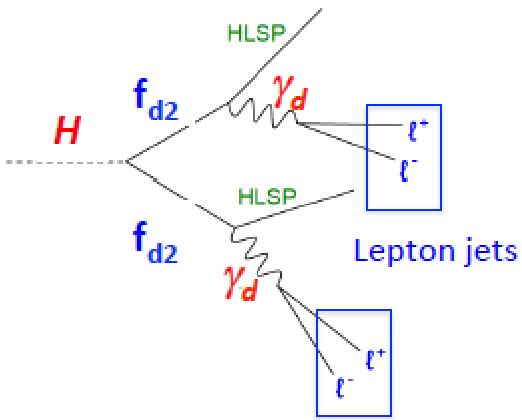
In realistic cases (with signal also in B, C;,D) use a likelihood to estimate relative rates in the 4 regions.



Example from the lepton-jets search (arXiv 1511.05542)

Recently published search for dark photons, dark fermions. Model to explain PAMELA positron excess

Signal ”polution” exists, thus a likelihood fit used. Performance:



QCD 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

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: $N^L = N_R^L + N_F^L$; $N^T = \varepsilon_R N_R^L + \varepsilon_F N_F^L$

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:

$w = \frac{\varepsilon_F \varepsilon_R}{\varepsilon_R - \varepsilon_F}$ 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: $w_{TT} = r_1 f_2\, w_{RF} + f_1\, r_2\, w_{FR} + f_1 f_2 w_{FF}$

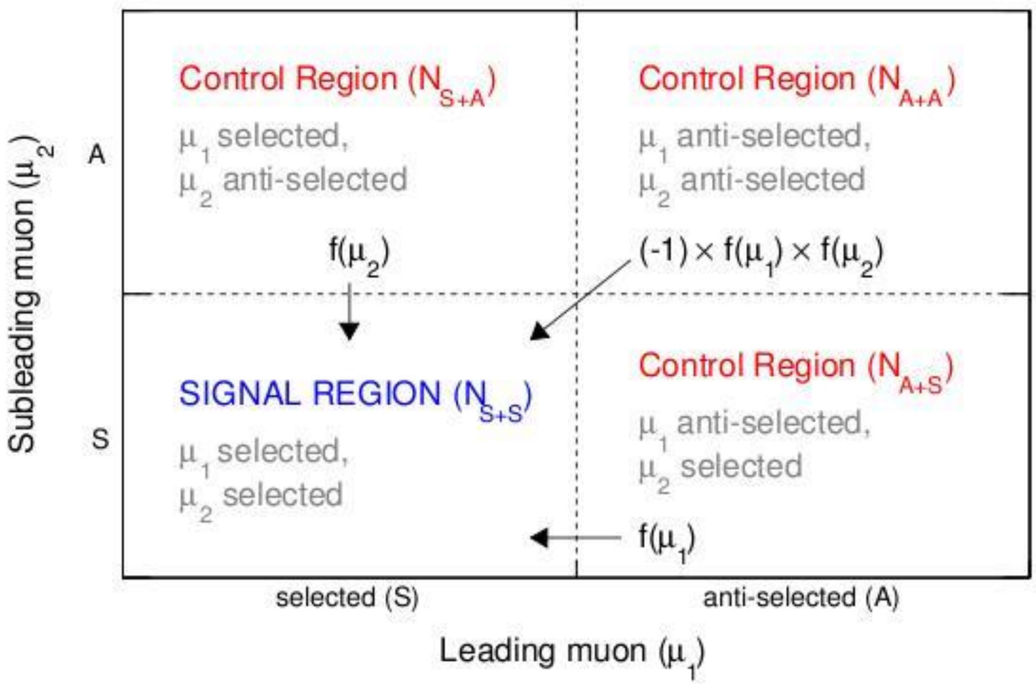
Fake factors

Define data control region inverting some selection criteria, then extrapolate this into signal region: $f \equiv \frac{N_{selected}}{N_{Anti-selected}}$ where

f= function(p_T , η)

Example with two muons:

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



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

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

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 <u>more than one method </u>– 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