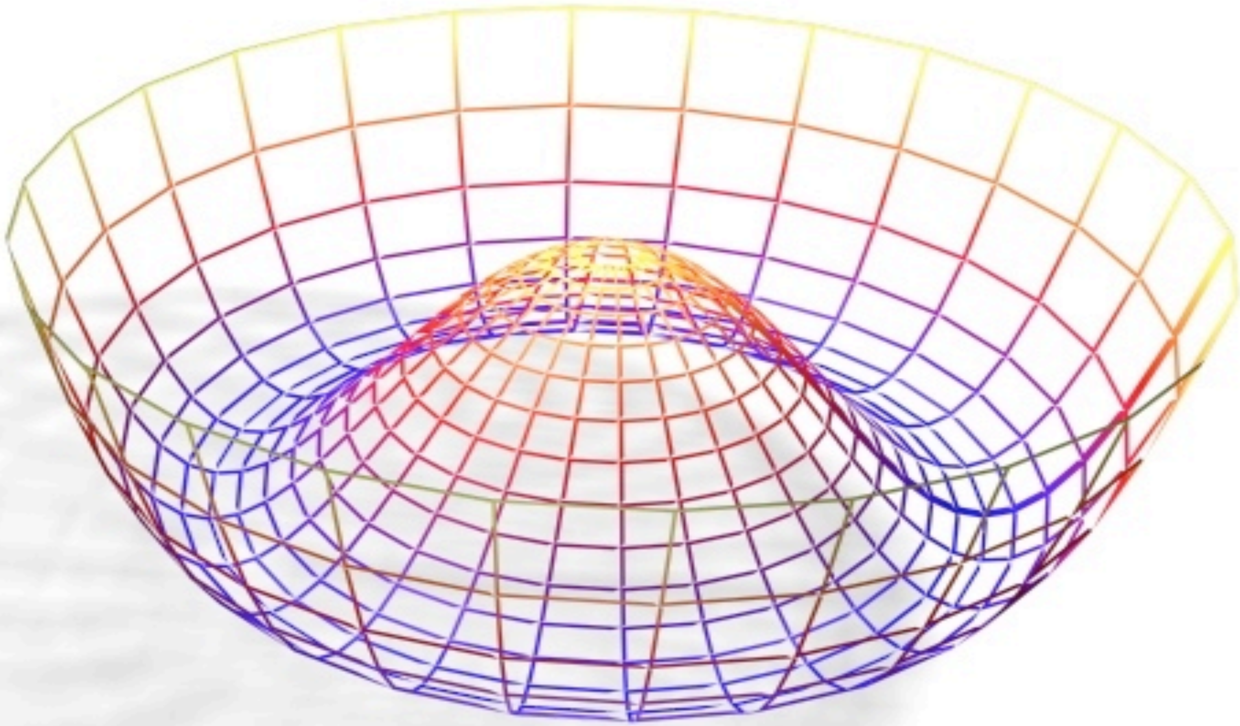
Center for
Cosmology and
Particle Physics

The ATLAS+CMS Higgs Combination: where physics and statistics meet



Kyle Cranmer,

on behalf of the ATLAS & CMS Higgs and Statistics working groups

Thanks to Gregory Schott, Lorenzo Moneta, Wouter Verkerke, Eilam Gross, Glen Cowan, Bob Cousins, Andrey Korytov, Louis Lyons, Luc Demortier, Chiara Mariotti, Vivek Sharma, Guillelmo Gomez-Ceballos, Hao Liu, Bill Quayle, William John Murray, Ketevi A. Assamagan, Francesco Polci, Konstantinos Nikolopoulos, Mingshui Chen

Goals

The goals of this talk are to:

- outline the connection among
  - the analysis strategy,
  - statistical technique,
  - and necessary inputs from theoretical community
- introduce statistical concepts necessary to
  - appreciate why we are using multiple statistical techniques
  - and understand the terminology used in Gregory's talk
- communicate some lesson's learned to focus future discussions of the x-sec group.

Preliminaries

The cross-section measurement is inferred from the relation

$$n_{exp} = \epsilon \, A \, L \, \sigma Br(H \to X) + n_{bkg}$$

expected = efficiency * acceptance * luminosity * cross-section * Branching Ratio + num background

- In the most naive form, the cross-section measurement just comes from solving for σ using N_{obs} and assuming everything else is known perfectly.

The situation quickly becomes more complicated as we take into account uncertainties each of the quantities.

- **statistical uncertainties** from finite event counts in data or Monte Carlo
- **systematic uncertainties** in the simulation and knowledge of our detectors
- **theoretical uncertainties** in cross-sections and distributions

Accurately reporting our inference on the cross-section requires careful treatment of each of these uncertainties, and a coherent story of physics knowledge, statistical methodology, and any necessary assumptions.

Background Estimation

A huge amount of effort is going into advancing signal and background generation (Monte Carlo) tools.

- We should make sure those efforts are well coordinated with the analysis strategies being used

Many backgrounds are estimated from **data-driven techniques**, where

- the total cross-section for the process is typically **not used**,
- the uncertainty on the luminosity **does not enter** (unless in some indirect way),
- BUT the **distributions** of the background from Monte Carlo **is used** to validate the shape used when fitting a **sideband** to determine a factor that is used to **extrapolate** from the control measurement to the signal-like region. [link]
- Most commonly used for backgrounds with fakes or when tails populate the signal-like region

Rare processes and very signal-like backgrounds often **rely on Monte Carlo** predictions, thus

- the total cross-section, the luminosity, and their uncertainties **enter directly** into the cross-section measurement.

Of course, the signal relies on Monte Carlo predictions, and we want to take special care to **separate** the effect of **different sources** of uncertainty on the measurement.

Three Schools of Statistical Inference

When dealing with complex problems there is no `optimal' statistical technique in general. There are several approaches for dealing with systematics, and they highlight the differences in the major 'schools' of statistical inference:

- Classical (Probability data given theory):
  - define probability as a limiting frequency of outcomes
  - insist on (or at least strive for) 'coverage': meaning that the confidence intervals cover the true value with at least a pre-specified rate, eg. 95%
  - violate "likelihood principle"
- Bayesian (Probability theory given data):
  - define probability as a degree of believe
  - do not insist on 'coverage'
  - require priors based on subjective degree of belief or some formal rule
  - obey likelihood principle
- Likelihood-based (the middle road, MINUIT/MINOS):
  - use frequentist definition of probability
  - do not insist on coverage, but rely on approximate asymptotic properties
  - obey likelihood principle

"When methods agree, asymptotic nirvana, when they don't we learn something."

Systematics, Nuisance Parameters, and Constraints

Effect of systematics is parametrized by one or more "**nuisance parameters**" denoted ν. Some examples

- rate for jet to fake an electron, reconstruction efficiency, ...
- extrapolation factor from a control region into the signal-like region
- cross-section for a background estimated from Monte Carlo

Sometimes these nuisance parameters are **constrained** by the data.

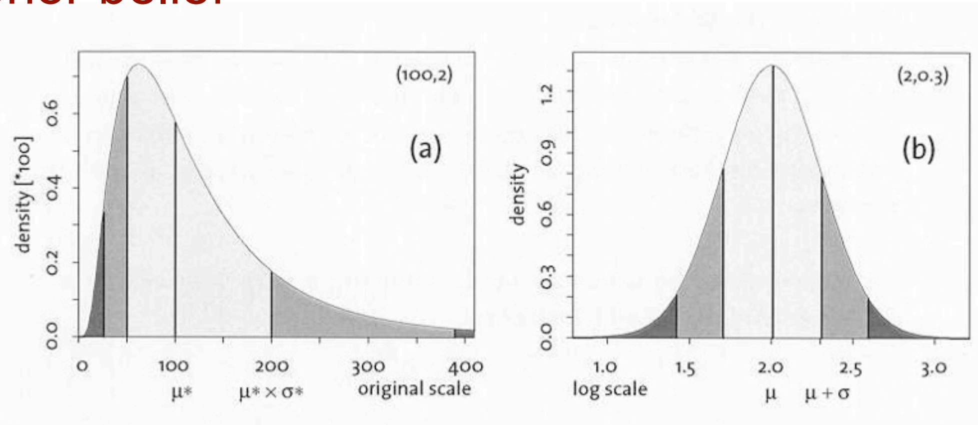
- by auxiliary measurements on control samples performed in a previous step
- or by **simultaneously** considering the signal-like region and control samples
  - probability model of the auxiliary measurement summarized in a **constraint term**

And sometimes the constraints are summarized as a Bayesian probability density

- may be **posterior** from auxiliary measurement, but will never escape original **prior**
- or may just be a subjective summary of prior belief

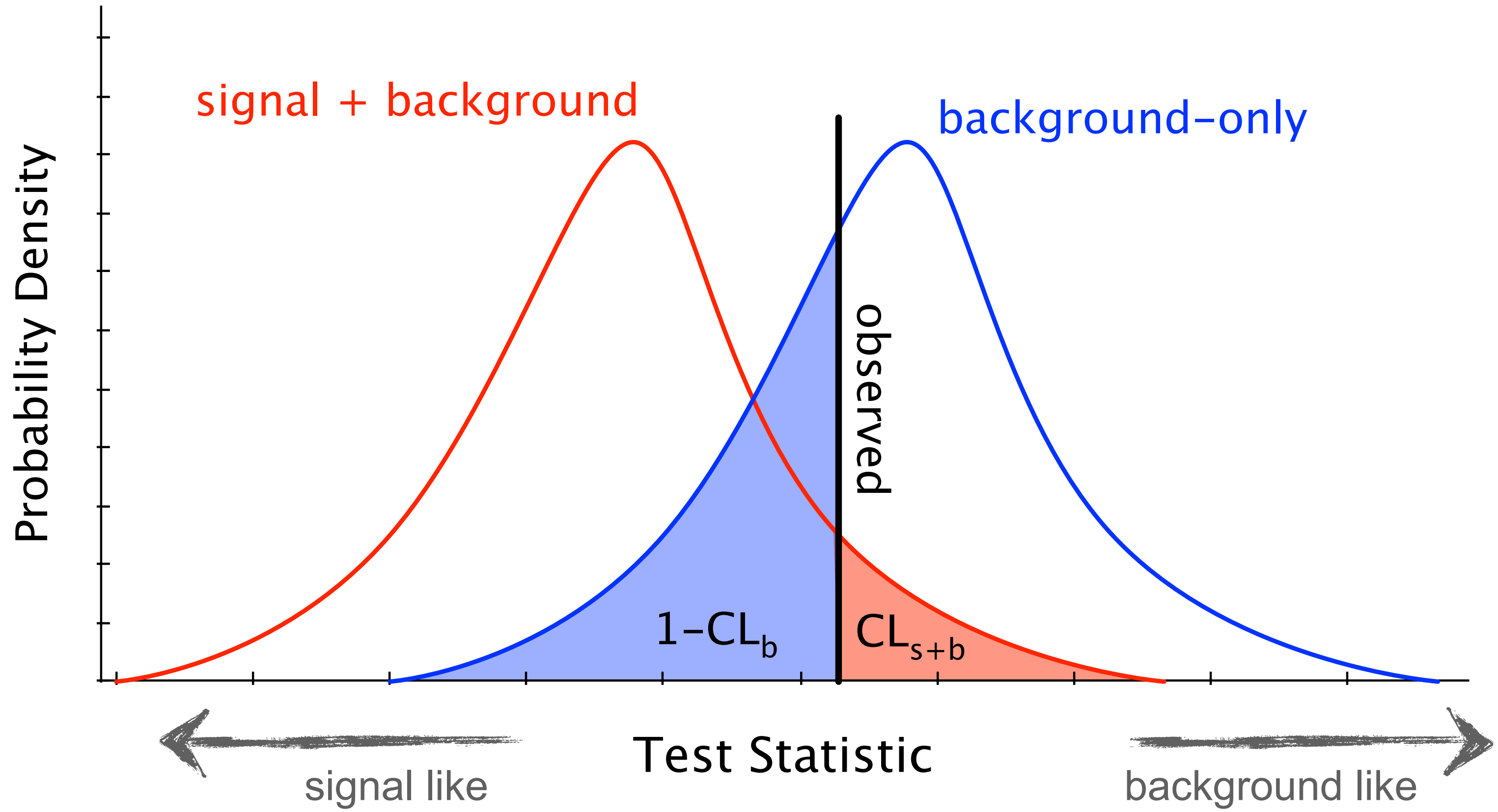
Common forms for the constraint terms are:

PDF	Prior	Posterior
Gaussian	uniform	Gaussian
Poisson	uniform	Gamma
Log-normal	reference	Log-Normal



The Test Statistic and its distribution

To get a feel for the different approaches, consider this schematic diagram



The "**test statistic**" is a single number that quantifies the entire experiment, it could just be number of events observed, but often its more sophisticated, like a likelihood ratio. What test statistic do we choose?

And how do we build the **distribution**? Usually "toy Monte Carlo", but what about the uncertainties... what do we do with the nuisance parameters?

Three common test statistics

We express cross-section as $\mu = \sigma/\sigma_{SM}$ for convenience.

Effect of systematics is parametrized by one or more “nuisance parameters” denoted ν .

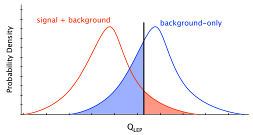
- best fit point is: $\hat{\mu}, \hat{\nu}$
- best fit of nuisance parameters with μ fixed is $\hat{\hat{\nu}}$ (aka “profiled”)

In principle, s+b and b-only models can have different parametrizations

Three common test statistics used in the field are:

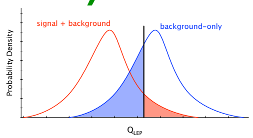
- simple likelihood ratio (used at LEP, nuisance parameters fixed)

$$Q_{LEP} = L_{s+b}(\mu = 1)/L_b(\mu = 0)$$



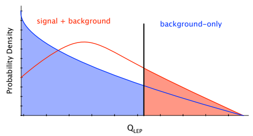
- ratio of profiled likelihoods (used commonly at Tevatron)

$$Q_{TEV} = L_{s+b}(\mu = 1, \hat{\hat{\nu}})/L_b(\mu = 0, \hat{\hat{\nu}}')$$



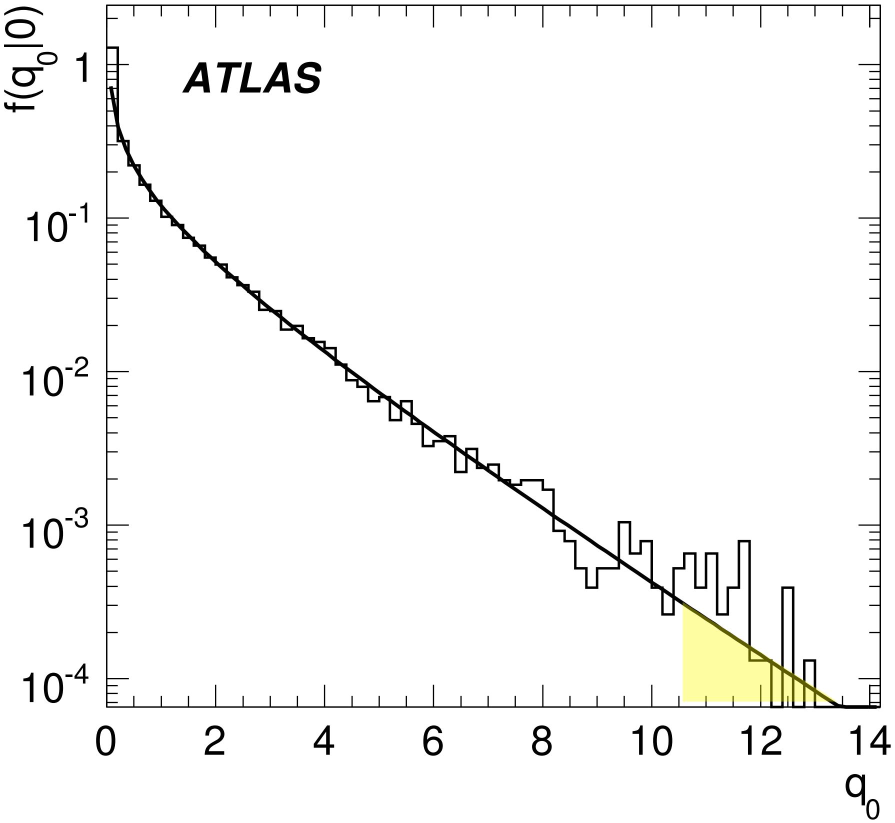
- profile likelihood ratio (related to Wilks’s theorem)

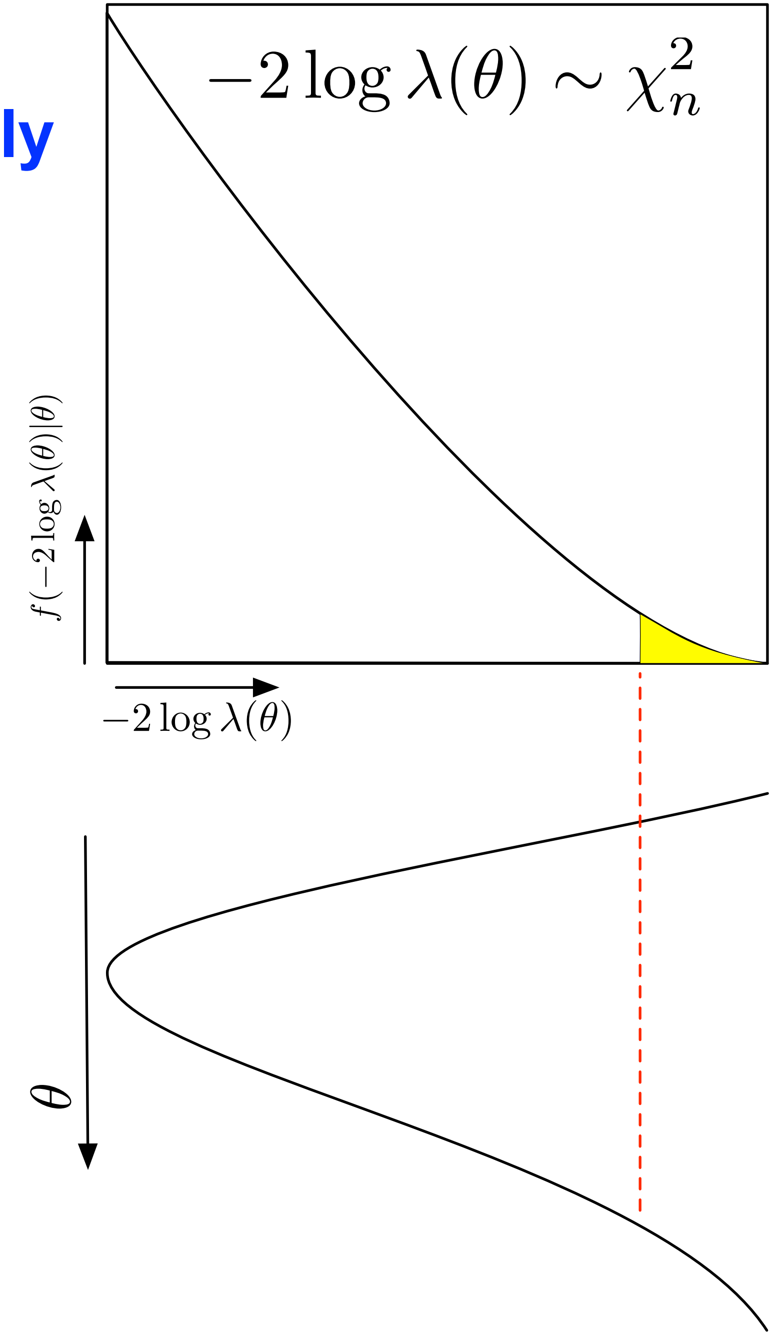
$$\lambda(\mu) = L_{s+b}(\mu, \hat{\hat{\nu}})/L_{s+b}(\hat{\mu}, \hat{\nu})$$



Likelihood-based Intervals (Wilks's theorem)

Wilks's theorem tells us how the profile likelihood ratio is distributed **asymptotically**

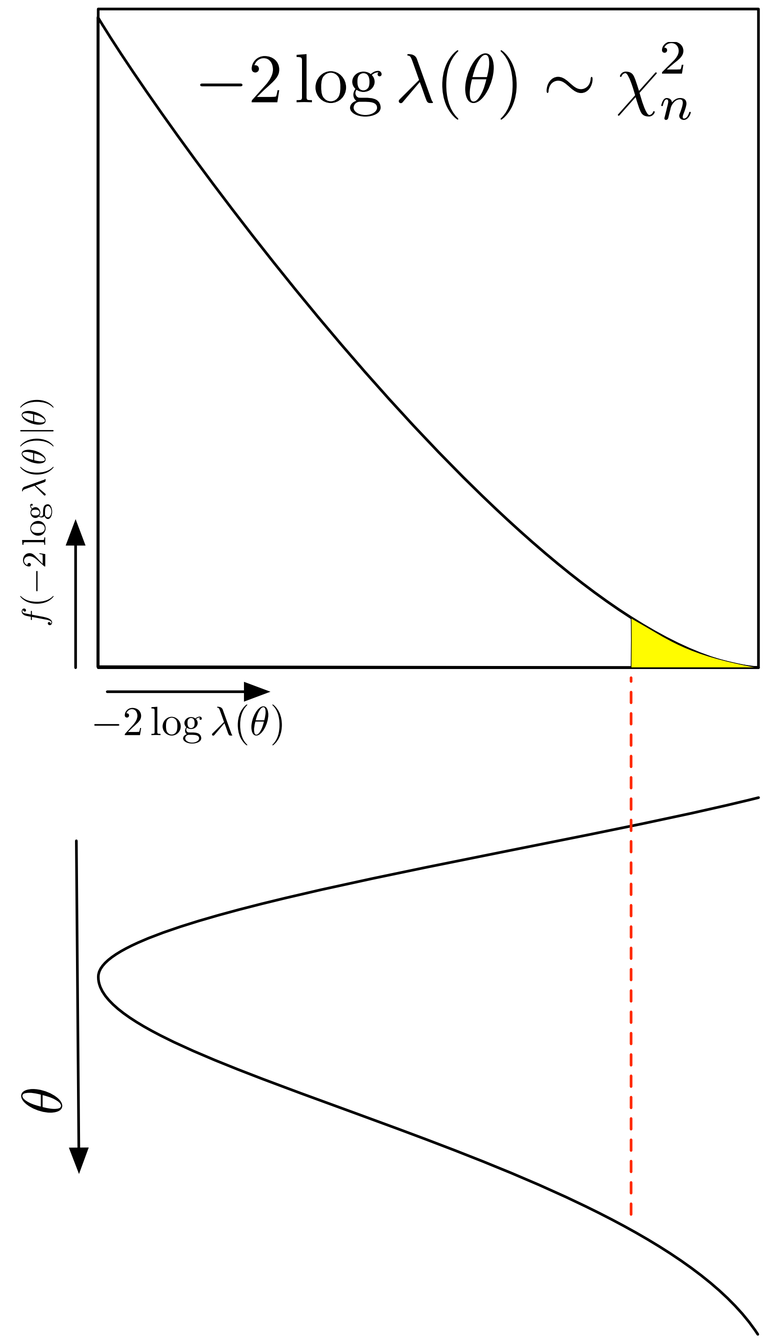




We can go immediately to the cutoff value of the profile likelihood ratio

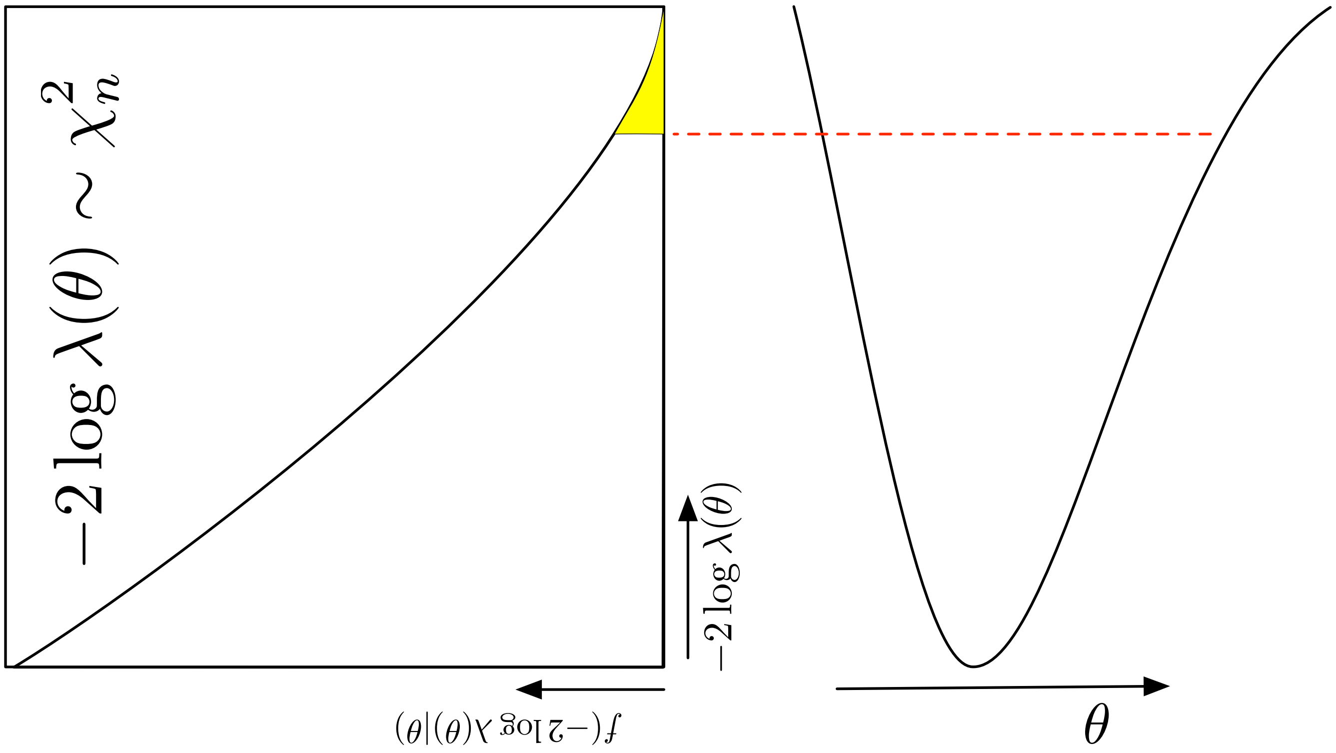
Likelihood-based Intervals

Center for Cosmology and Particle Physics



We can go immediately to the cutoff value of the profile likelihood ratio

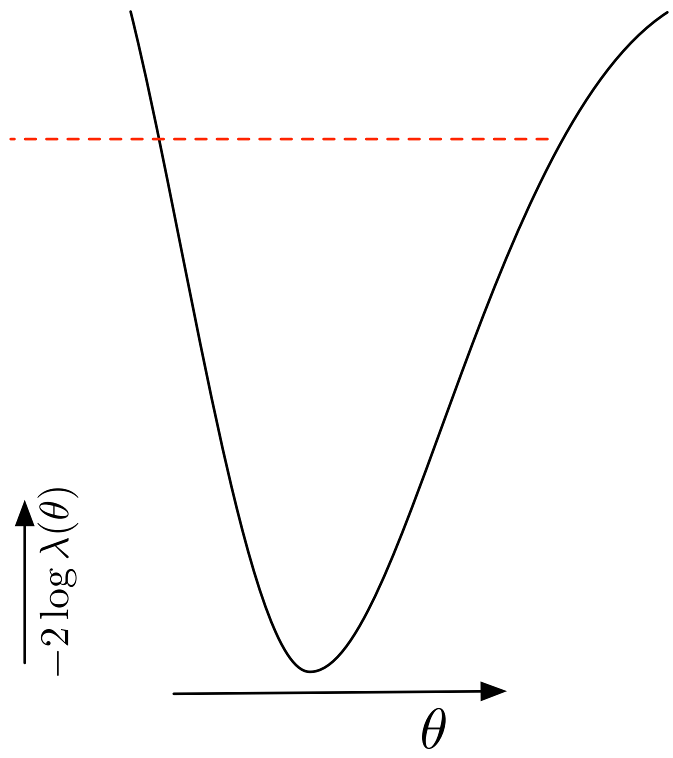
Likelihood-based Intervals



And typically we only show the likelihood curve and don’t even bother with the implicit (asymptotic) distribution

We can go immediately to the cutoff value of the profile likelihood ratio

Likelihood-based Intervals



And typically we only show the likelihood curve and don't even bother with the implicit (asymptotic) distribution

We can go immediately to the cutoff value of the profile likelihood ratio

A Fully Frequentist Method

While Wilks's theorem is very convenient, it does not guarantee that the limits cover the true value 95% of the time.

- with few events, the asymptotic distribution can be a poor approximation
- in order to guarantee "coverage" one must calibrate the cutoff value using pseudo-experiments
  - this is the basis of the **Feldman-Cousins** technique
- in this context, taking the systematics into account means that the limit covers for every value of the nuisance parameters
  - dealing with systematics was beyond the scope of the original Feldman-Cousins paper, and previously this was considered impractical
  - but there is a generalization that works with nuisance parameters

Because this approach is very new to the field, usually systematics have been dealt with in a Bayesian way...

Bayesian-Frequentist Hybrid Solutions

Goal of Bayesian-frequentist hybrid solutions is to provide a frequentist treatment of the main measurement, while eliminating nuisance parameters (deal with systematics) with an intuitive Bayesian technique.

$$P(n_{\rm on}|s) = \int db\, {\rm Pois}(n_{\rm on}|s+b)\, \pi(b),$$

"Principled" version:

- clearly state prior $\eta(b)$; identify control samples, sidebands, etc. and base prior on

$$\pi(b) = P(b|n_{\rm off}) = \frac{P(n_{\rm off}|b)\eta(b)}{\int db P(n_{\rm off}|b)\eta(b)}.$$

"Ad-hoc" version:

- unable or unwilling to justify $\pi(b)$, so go straight to some distribution
  - often the case for real systematic uncertainty (eg. MC generators, different background estimation techniques, etc.)

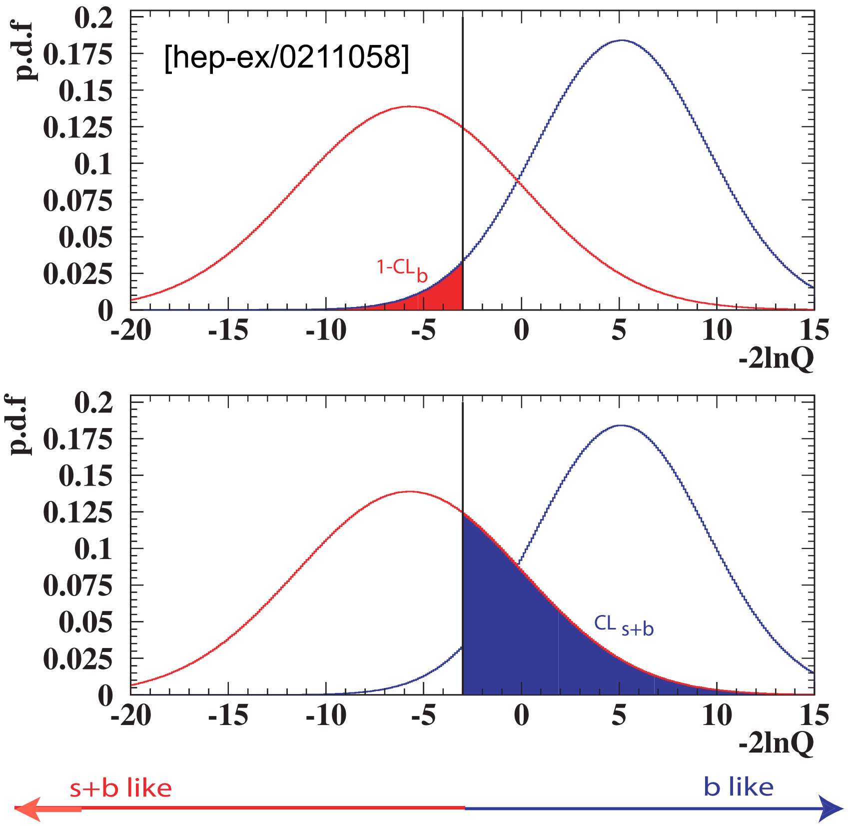
CL_b, CL_{s+b} and the curious CL_s

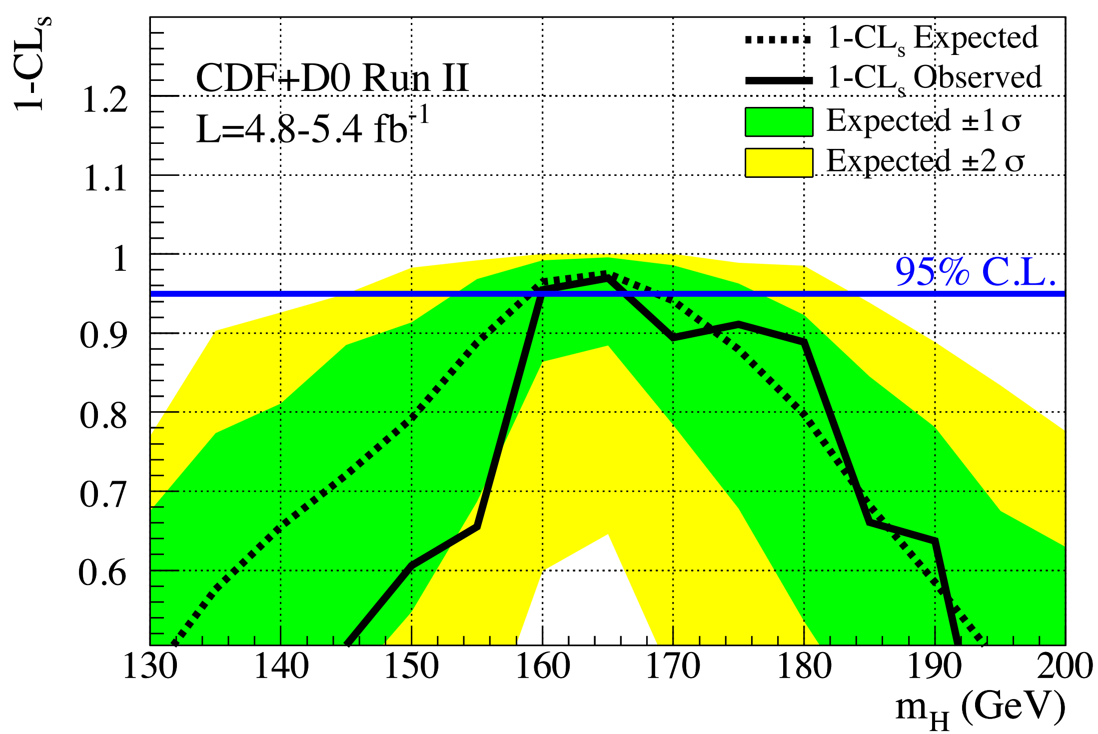
Claims of discovery are based on the p-value of the Null, in physics called $1-CL_b$

Standard frequentist limits are based on CL_{s+b}, but they have the undesired feature that they can exclude a region where we have no sensitivity.

- to address this, the LEP experiments introduced "CLs" = CL_{s+b}/CL_b, which builds in some conservatism.
  - Note: CLs is a ratio, it's not a statistical technique (or a normal probability)

Also used at Tevatron; however, most statistics experts really don't like CL_s.

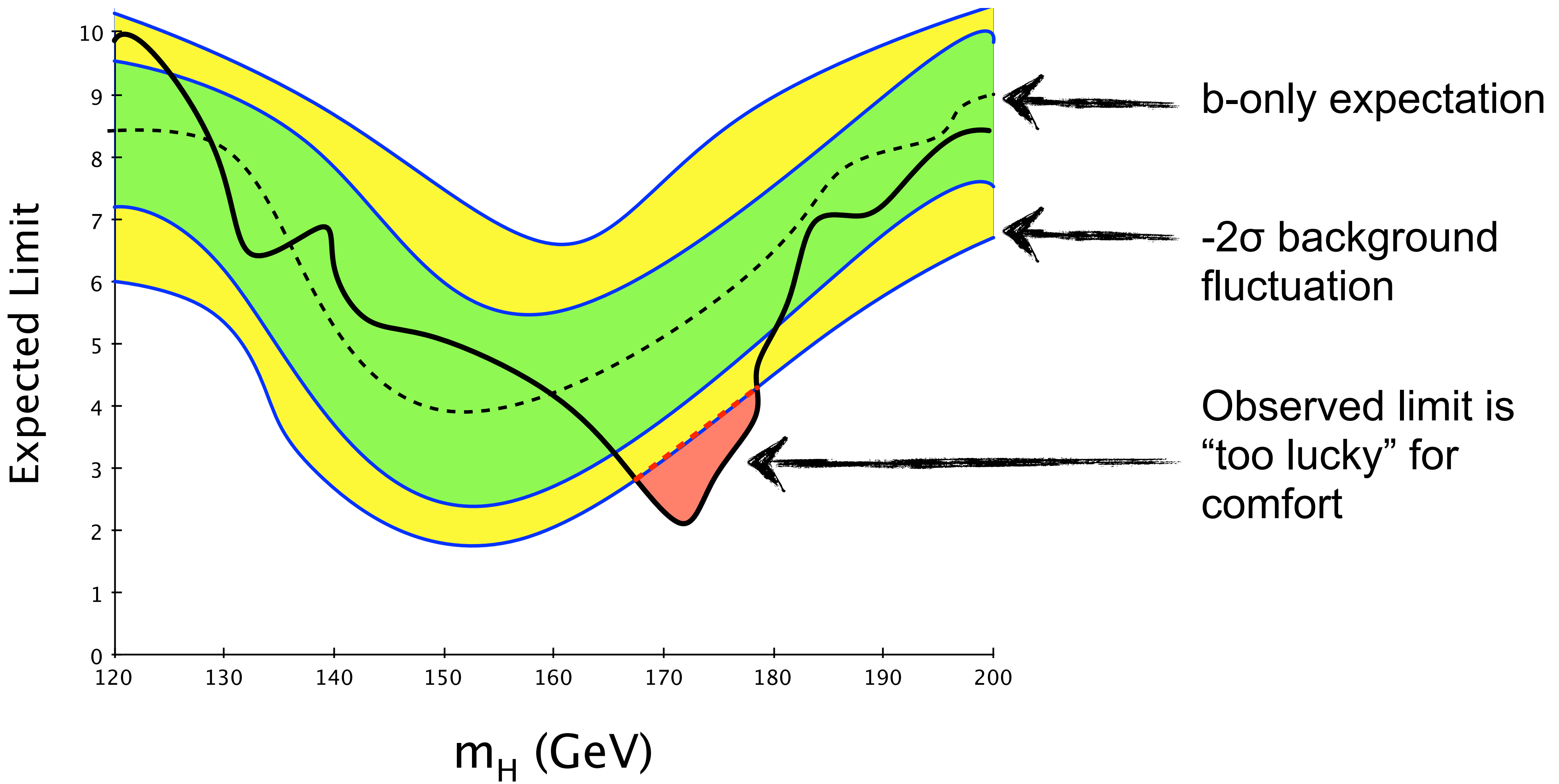




“Power-Constrained” limits

The ATLAS+CMS statistics committees are looking into a different way to avoid setting limits where we have no sensitivity

- don’t quote limit below some threshold defined by an N-σ downward fluctuation of b-only pseudo-experiments



Bayesian Techniques

The goal of Bayesian techniques is to answer "probability of theory given data" based on a posterior

$$P(\mu) \propto \int d\nu \; L(\mu,\nu)\pi(\mu,\nu)$$

This inevitably involves a prior $\pi(\mu)$, which is not predicted by any theory.

- "the heart of scientific Bayesianism" is to assess the sensitivity to the choice of prior

In problems with several nuisance parameters, the integration (called marginalization) is challenging

- one of the most useful techniques is called Markov Chain Monte Carlo or just MCMC

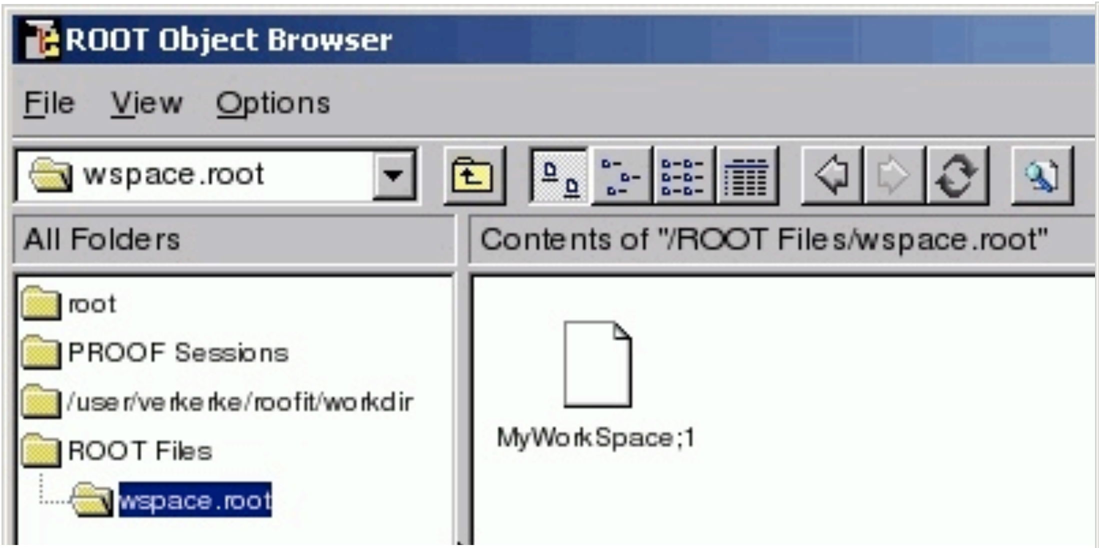
The RooStats Approach

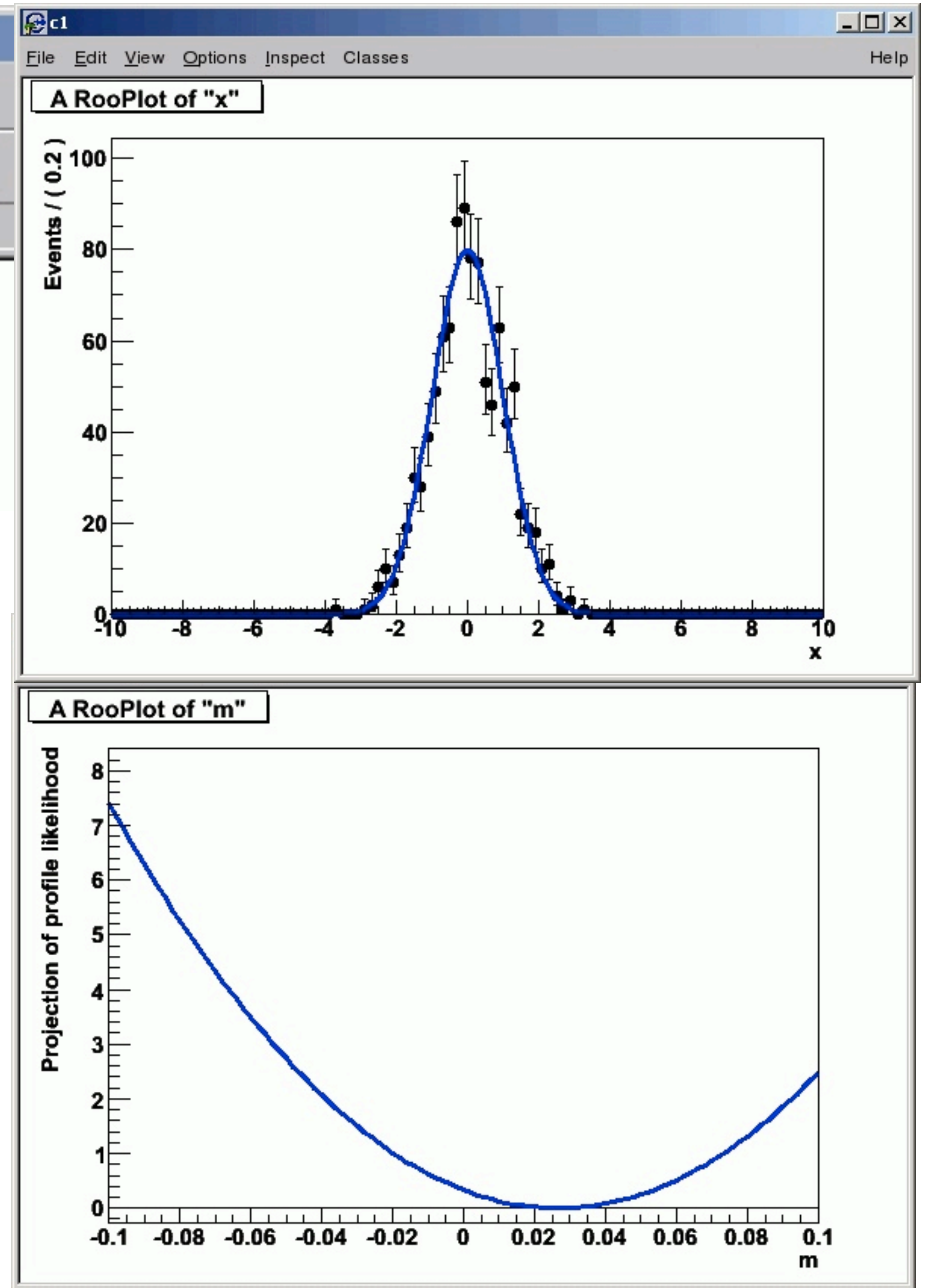
There are a few core principles to the RooStats approach for combinations

- Create one probability model relating the data and the parameters, then try multiple statistical techniques for the same problem
  - the probability model is what codifies the analysis strategy. It's hard, and we should only have to do this once. Here we use RooFit as a data modeling language.
  - we don't want to compare apples and oranges,
  - RooStats has at least one implementation of the major Frequentist, Bayesian, and Likelihood-based techniques. Most work with arbitrary probability models.
    - Gregory will show results from Profile Likelihood, Hybrid, Feldman-Cousins, and MCMC
- Communicate the entire probability model digitally
  - The RooFit/RooStats workspace allows one to capture the entire probability model (both the likelihood function and the ability to generate toy data) and is an ideal tool for combinations. [an exciting possibility for publications!]
  - Avoid having those doing combination interpreting/recoding/modifying probability model
- Make technical aspect of combinations part as easy as possible so that we can focus on physics decisions
  - With the workspace technology, the combination is technically very easy. The combiner is insulated from the complexity of the models.
  - "With great power comes great responsibility"...

The RooFit/RooStats workspace

CENTER FOR
COSMOLOGY AND
PARTICLE PHYSICS





RooFit's Workspace now provides the ability to save in a ROOT file the full probability model, any priors you might need, and the minimal data necessary to reproduce likelihood function.

Need this for combinations, exciting potential for publishing results.

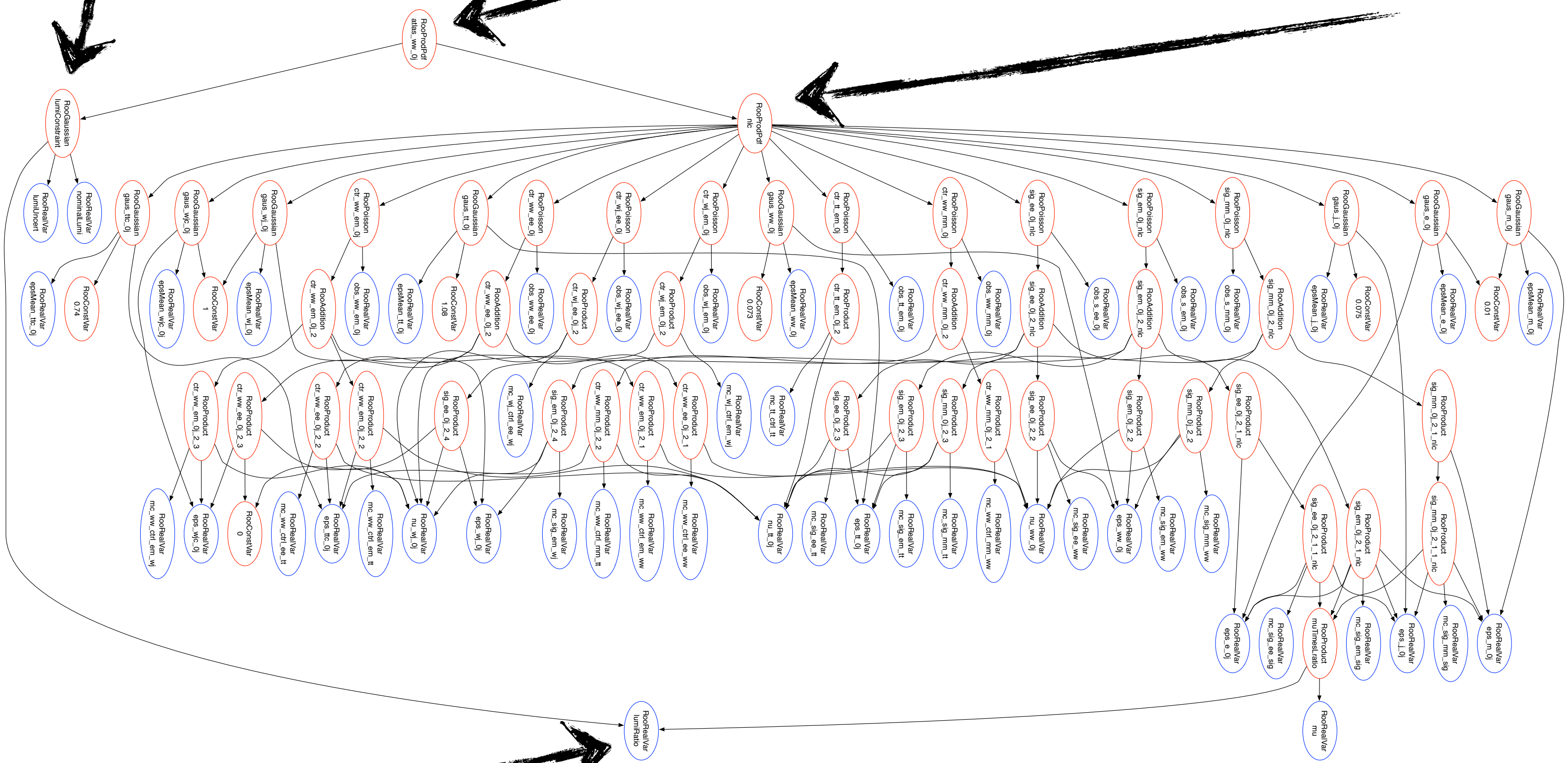
Visualization of the ATLAS Workspace

Center for Cosmology and Particle Physics

lumi constraint
(to be replaced by correlated term)

top level model

ATLAS part



luminosity parameter
(to be included in correlated term)

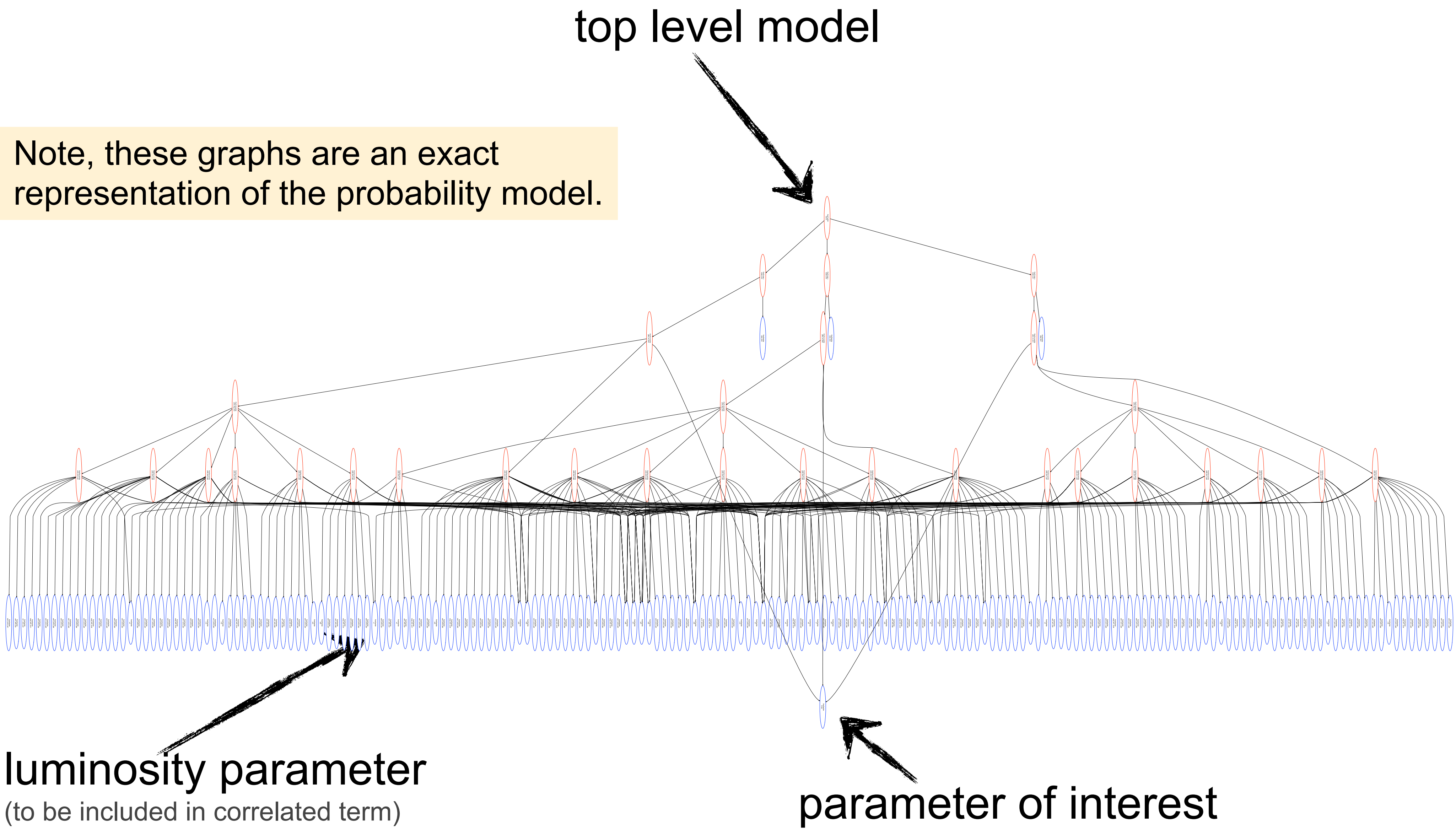
parameter of interest

$$\mu = \frac{\sigma BR}{\sigma_{SM} BR_{SM}}$$

Recall, ATLAS alone is a combination of ee, eμ, μμ channels together with 6 control regions

Visualization of the CMS Workspace

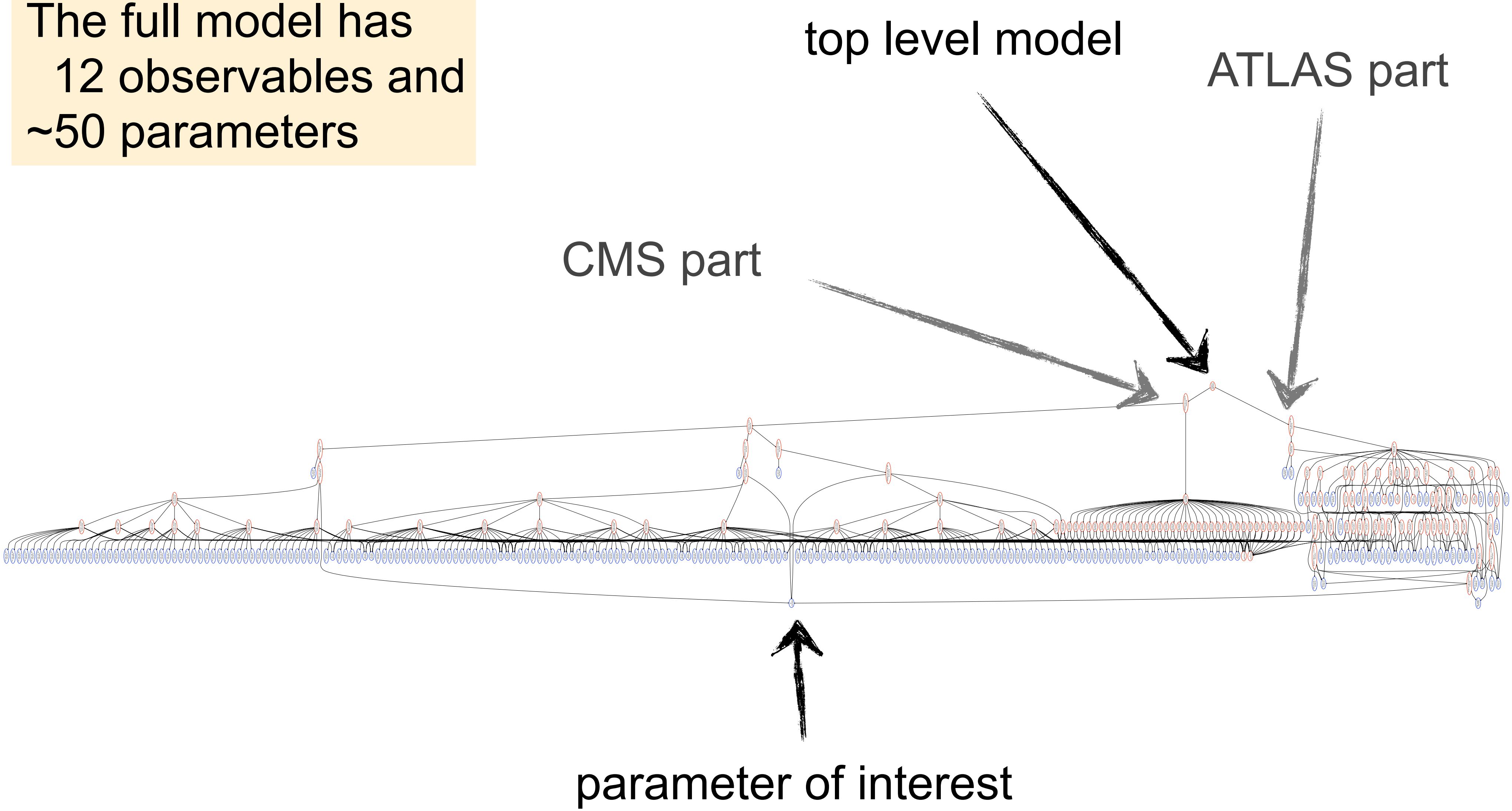
Note, these graphs are an exact representation of the probability model.



$$\mu = \frac{\sigma BR}{\sigma_{SM} BR_{SM}}$$

Visualization of the ATLAS+CMS Workspace

The full model has
12 observables and
~50 parameters



$$\mu = \frac{\sigma BR}{\sigma_{SM} BR_{SM}}$$

Some lessons learned

In general, this combination has been a great success

- in our first meeting we were already discussing correlated systematics between ATLAS and CMS

We need to identify each of the backgrounds estimated from theory, because their theoretical uncertainties in the are correlated between experiments

We need to separate and individually parametrize the effect of individual systematics

- the ability to correlate across experiments (and for different channels within the same experiment) requires the ability to relate parameters in the model in a consistent way
- this means parametrizing the **effect** of an uncertainty in terms of variations in the **source** of the uncertainty
  - **simple example:** if electron identification efficiency were 1% higher, it has a different effect on the ee,eμ, and μμ channels [done]
  - **complicated example**: if jet energy scale is 5% higher, the expected yield of some backgrounds go up, while others go down [not done]
  - **complicated example:** the qg, qQ, and gg parts uncertainties in the parton density functions affect different processes in a different way, lumping them all together may be missing some essential physics. [not done]

The current modeling is not sufficiently detailed, and does not always separate or expose the individual sources of a systematics. Often several effects are "added in quadrature"

- this group should be thinking about what physics we want to be sure the model captures

Most of the LEP Higgs results were presented as exclusions in the space of some theoretical model

- SM Higgs mass assuming SM branching ratios and production rates
- exclusions in the $(m_A, \tan\beta)$ plane assuming some SUSY scenario

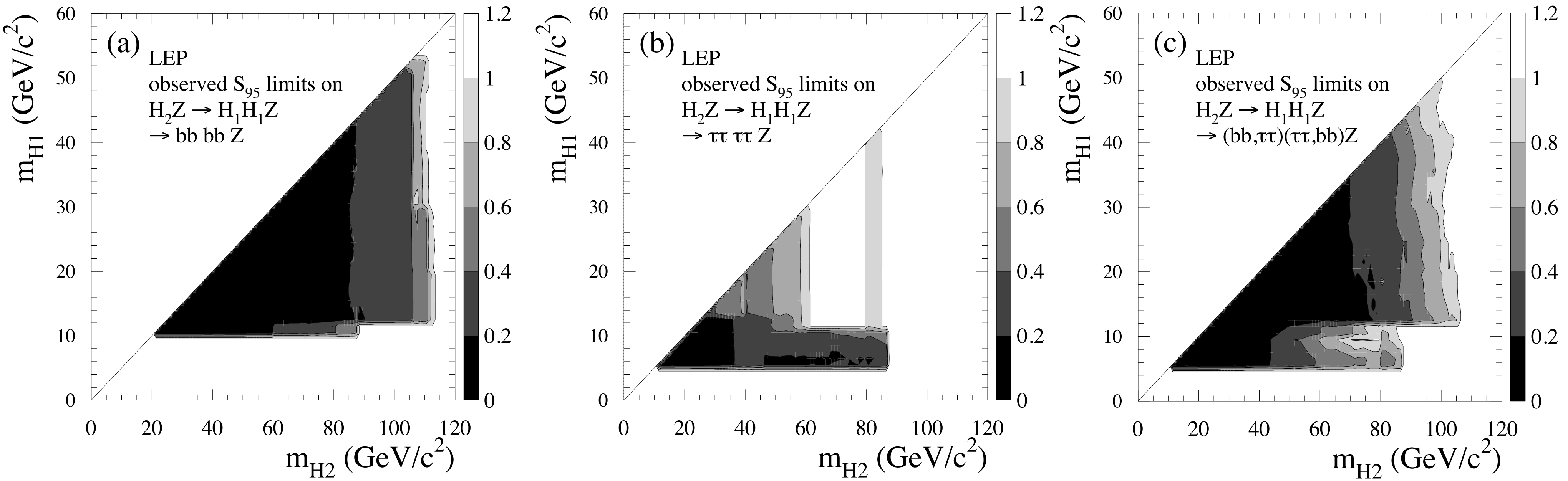
Recently, I've been considering non-standard models, and the most useful results are presented in terms of limits on cross section X branching ratio

- in terms of experimentally observable masses

But... when combining different decays or production modes, assumptions are made on the relative branching ratios and cross-sections -- not ideal.

(factor x SM cross section that corresponds to 95% exclusion)

Eur.Phys.J. C33 (2004) [hep-ex/0602042]



Conclusions

We have made great progress...

- in the last few years developing methodology and building a set of tools able to cope with the complexity of LHC physics analysis
  - great advances for the field through the PhyStat conference series
  - the RooStats project brings high-level statistical tools in each of the schools of statistical inference built on top of RooFit's data modeling language
- in the last few weeks bringing together realistic toy models from ATLAS and CMS to combine
  - Grégory is about to show the impressive progress that we made in the last few days once the ATLAS and CMS workspaces became available

We have much to do...

- the tools are pretty good, but they need more work and more validation
- the current modeling needs to further separate individual sources of uncertainty and expose parameters in order to consistently introduce correlations between channels and across experiments
  - this is not statistical issue, this is a physics issue that will require thoughtful consideration and planning for the next round of combinations