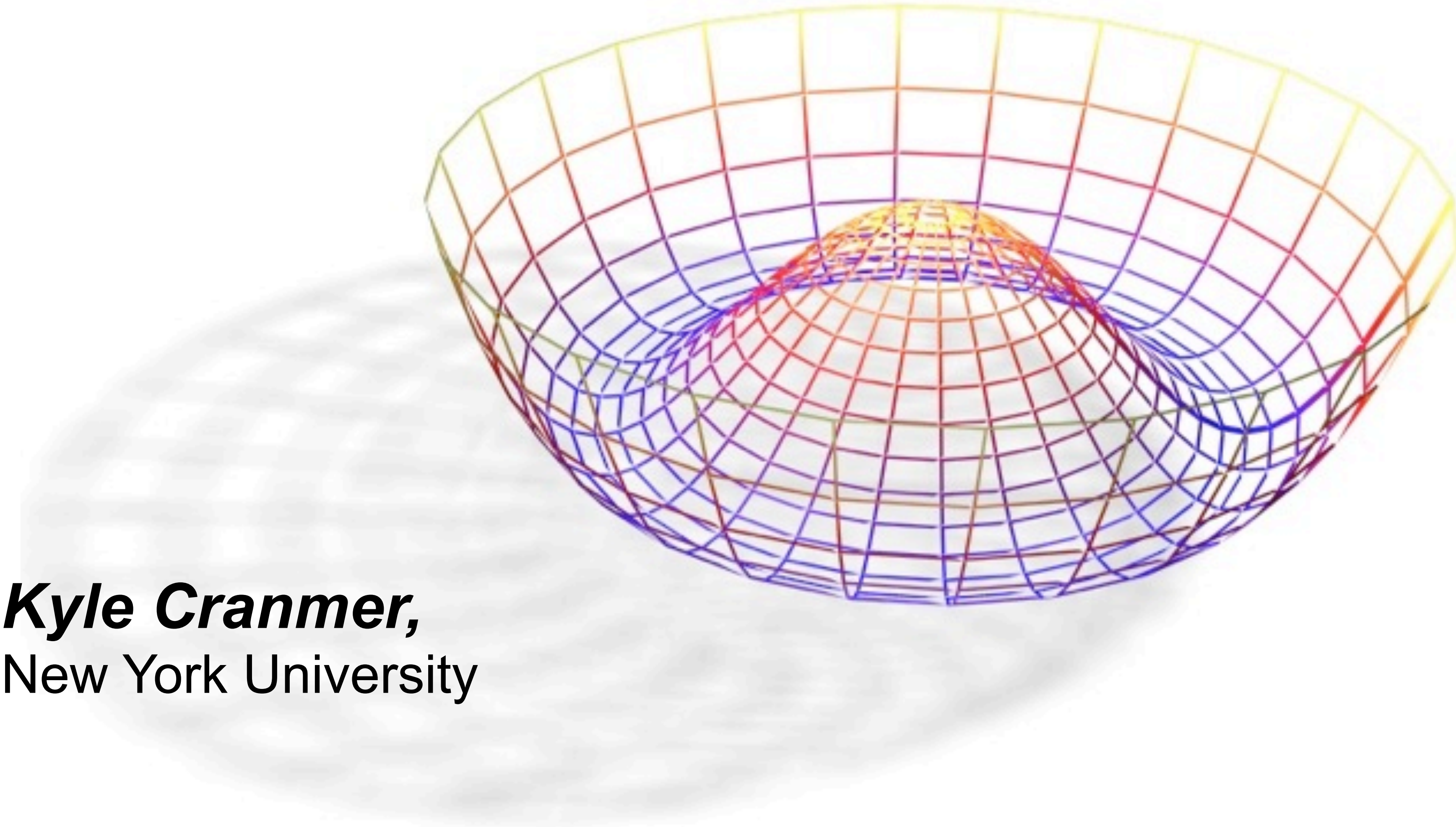
Center for
Cosmology and
Particle Physics

Review of the 2010 combination: experience gained and the lessons learned



Kyle Cranmer,
New York University

The 2010 ATLAS+CMS Higgs Combination Exercise

The exercise was based on "toy" data and models, though realistic in complexity

- An intense effort between in June 2010, toy results shown July 6

Initial meetings were mainly focused on

- aligning language, philosophy, strategy, and priorities.
- discussion practical and technical issues

Early on we decided the initial combination would be based on H→WW+0j and that the analyses would be number counting in a few channels

- attempt to provide inputs in a technology neutral way as well as a RooStats workspace format
  - Andrey suggested a tabular format breaking down the effect on each background due to each source of systematic [similar to approach used by a RooStats tool that was in development, but not yet publicly available]
- early discussions on form of constraint terms (Gaussian, gamma, lognormal)
- later discussions on methods, test statistics, etc.

Took ~1 month to prepare and validate inputs

- Four days from the time the inputs were shared to final results!
- Very impressive and encouraging exercise... but still an exercise.

Timeline & Milestones

31 May: kick-off meeting

- http://indico.cern.ch/conferenceDisplay.py?confId=96787
- Andrey Korytov: general remarks (technology independent)
- KC: details for RooStats input format

10 June: Update

- http://indico.cern.ch/conferenceDisplay.py?confId=98055
- Initial ATLAS results with 9 channels (ee,eμ,μμ)⊗ (0,1,2j) ⇒ decide to use only 0j
- Preliminary CMS tables
- Discussions on truncated Gaussian vs. Log-Normal, Gamma.
  - some requests to change ATLAS model parametrization, but deferred to next exercise

~1 Month!

24 June: Update

- http://indico.cern.ch/conferenceDisplay.py?confId=99459
- Inputs fully specified, testing and cross-checks within experiments

1 July: Pre-combination Meeting

- http://indico.cern.ch/conferenceDisplay.py?confId=99935
- Individual experiments have finalized workspaces, after meeting they are shared

6 July: presentation of initial results at Higgs Cross-Section Workshop

- http://indico.cern.ch/conferenceDisplay.py?confId=100458
- Limits and Significance shown using 6 different methods

The H→WW+0j analyses

CENTER FOR COSMOLOGY AND PARTICLE PHYSICS

More details in the supporting talks, including treatment of systematics

- http://indico.cern.ch/conferenceDisplay.py?confId=100458
- Two leptons (e or μ), missing E_T, no hard jets
- Additional selection based on kinematic variables such as Mll, $\Delta\,\varphi_{ll}$, etc.
  - ATLAS analysis was based on straight cuts on these quantities
  - CMS used a multivariate approach, and cuts on the output of the algorithm

Control Regions

- Main backgrounds in H+0j:
  - W+jets: both CMS and ATLAS normalize this using a control sample with loosened lepton selection.
  - ttbar: CMS normalizes this using events with a soft muon. At low luminosity, ATLAS normalizes it using a 1-lepton plus 4 jets ("top box") control sample
  - Continuum WW. Both collaborations normalize this using a control region with large dilepton invariant mass
  - Z+jets: Normalized using Z peak in both collaborations
  - For ATLAS numbers, minor backgrounds like WZ, ZZ, Wbb, Zbb, etc. are lumped in with the other processes

W. Quayle Page 5 ATLAS-CMS Meeting, 06 July 2010

General Remarks on Systematics

- The Likelihood function used by ATLAS includes Poisson terms for the number of events in each control region
  - Systematic error estimates therefore focus on ratios of cross-sections in the signal region and control regions
  - In most cases, various sources of systematic error are added in quadrature and treated with a single nuisance parameter because it's not clear that it's meaningful to assume these individual sources of error are correlated across experiments
- Because CMS is using a multivariate method, systematic errors are separated into two parts:
  - The uncertainty on the extrapolation from the control region to a "preselection" region with no cut on the output of the multivariate algorithm
  - The uncertainty on the efficiency of the cut on the output of the multivariate algorithm

W. Quayle Page 10 ATLAS-CMS Meeting, 06 July 2010

Constraints on Nuisance Parameters

For large uncertainties, a truncated Gaussian is a bad choice for modeling uncertainty

- often lead to optimistic p-values, short tail, bad behavior at 0

For systematics constrained from control samples dominated by statistical uncertainty, a Gamma distribution is a more natural choice [PDF is Poisson for the control sample]

- longer tail, good behavior near 0, natural choice if auxiliary is based on counting

For “factor of 2” notions of uncertainty log-normal is a good choice

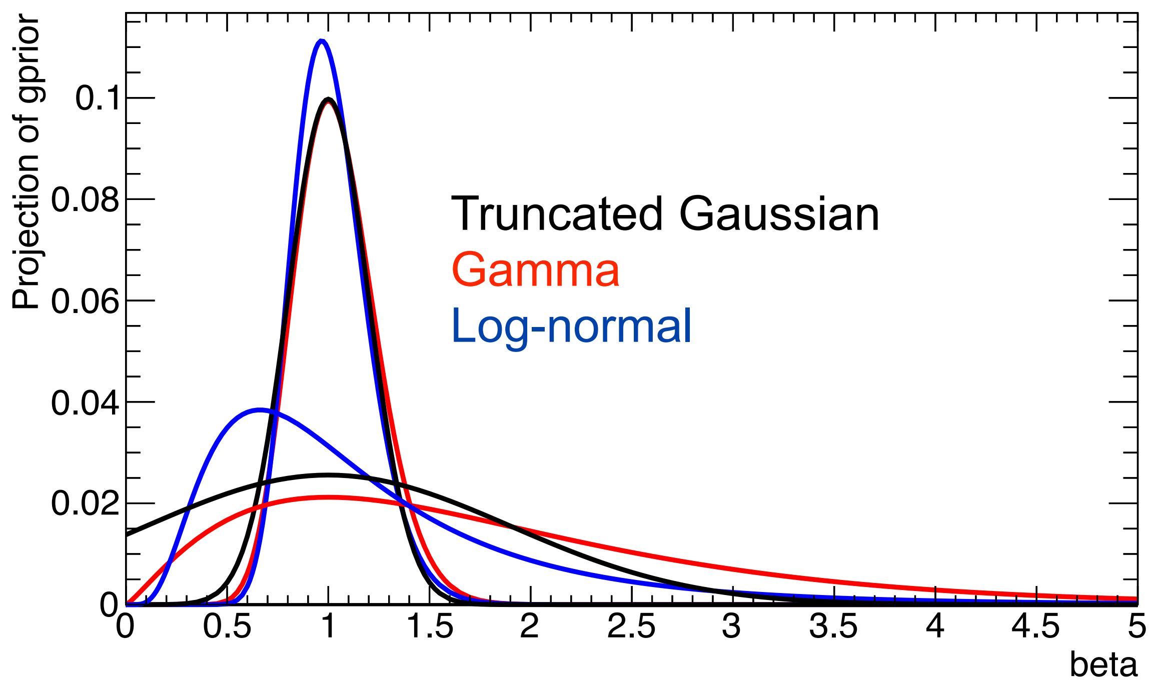
- can have a very long tail for large uncertainties

All of them are approximately Gaussian for small relative uncertainty

None of them are as good as an actual model for the auxiliary measurement, if available

To consistently switch between frequentist, bayesian, and hybrid procedures, need to be clear about prior vs. likelihood function

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



Clarifying our terminology

Center for Cosmology and Particle Physics

When we describe method, we should specify each of the following:

- what is the test statistic
  - simple likelihood ratio (LEP) $Q_{LEP} = L_{s+b}(\mu = 1)/L_b(\mu = 0)$
  - profile likelihood ratio (Wilks) $\lambda(\mu) = L_{s+b}(\mu, \hat{\hat{\nu}})/L_{s+b}(\hat{\mu}, \hat{\nu})$
  - ratio of profiled likelihoods (Tevatron) $Q_{TEV} = L_{s+b}(\mu = 1, \hat{\hat{\nu}})/L_b(\mu = 0, \hat{\hat{\nu}}')$
- how was it sampled:
  - toy MC randomizing nuisance parameters according to $\pi(\nu)$
  - toy MC with nuisance parameters fixed
  - assuming asymptotic distribution (Wilks)
- For limits, what is the condition that defines upper bound?
  - CL_{s+b}, CL_s, power-constrained, something else
    - recall, CL_s is not a method, so let's don't use that term except for this context.
- For Bayesian, what was the prior on the parameter of interest

Tables, Formulae, and Workspaces

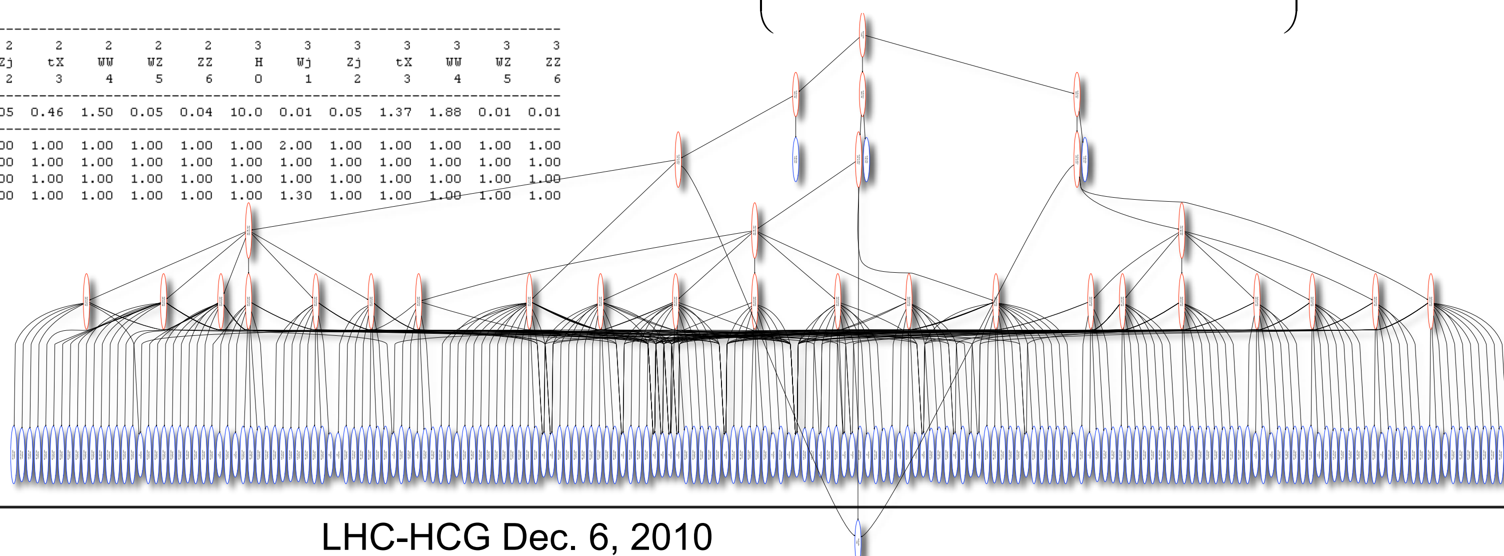
The CMS input:

- cleanly tabulated effect on each background due to each source of systematic
- broke systematics down into uncorrelated subsets
- used lognormal distributions for all systematics
- started with a txt input, defined a mathematical representation, and then prepared the workspace
  - The implementation of model in the workspace used many interpreted strings instead of compiled functions. **Slow to evaluate, and must be numerically integrated for normalization!**

```
RooFormulaVar::yield_bin3_cat2[ formula="@0*@1^@2*@3^@4*@5^@6*@7^@8*@9^@10" ]
```

```
Date: June 22, 2010
Description: HWW-->2l2v, 0jets, cut-and-count for 3 channels: mumu, ee, emu; made-up numbers for a ATLAS+CMS combination exercise
mH    160  Higgs mass hypothesis
comE  7.0   center of mass energy
lumi  1  luminosity in fb-1
------------------------------------------------------------------------------------------------------------------------------------
imax   3  number of channels
jmax   6  number of backgrounds
kmax  37  number of nuisance parameters
------------------------------------------------------------------------------------------------------------------------------------
Observation   15  7  13
------------------------------------------------------------------------------------------------------------------------------------
bin         1     1     1     1     1     1     1     2      2     2     2     2     2     2     3     3     3     3     3     3     3
process     H    Wj    Zj    tX    WW    WZ    ZZ     H     Wj    Zj    tX    WW    WZ    ZZ     H    Wj    Zj    tX    WW    WZ    ZZ
process     0     1     2     3     4     5     6     0      1     2     3     4     5     6     0     1     2     3     4     5     6
------------------------------------------------------------------------------------------------------------------------------------
rate     10.5  0.01  0.05  0.94  3.39  0.01  0.01  5.39  0.01  0.05  0.46  1.50  0.05  0.04  10.0  0.01  0.05  1.37  1.88  0.01  0.01
------------------------------------------------------------------------------------------------------------------------------------
 1  lnN  1.00  2.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  2.00  1.00  1.00  1.00  1.00  1.00
 2  lnN  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  2.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00
 3  lnN  1.00  1.30  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00
 4  lnN  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.30  1.00  1.00  1.00  1.00  1.00  1.00  1.30  1.00  1.00  1.00  1.00  1.00
```

$$L_{b+rs} = \prod_i \left(\frac{\left(\sum_{j=0,1,..} \tilde{n}_{ij} \cdot \kappa_{ijk}^{\theta_k}\right)^{N_i}}{N_i!} \cdot \exp\left(-\sum_{j=0,1,..} \tilde{n}_{ij} \cdot \kappa_{ijk}^{\theta_k}\right) \right) \cdot \prod_k f(\theta_k)$$



Tables, Formulae, and Workspaces

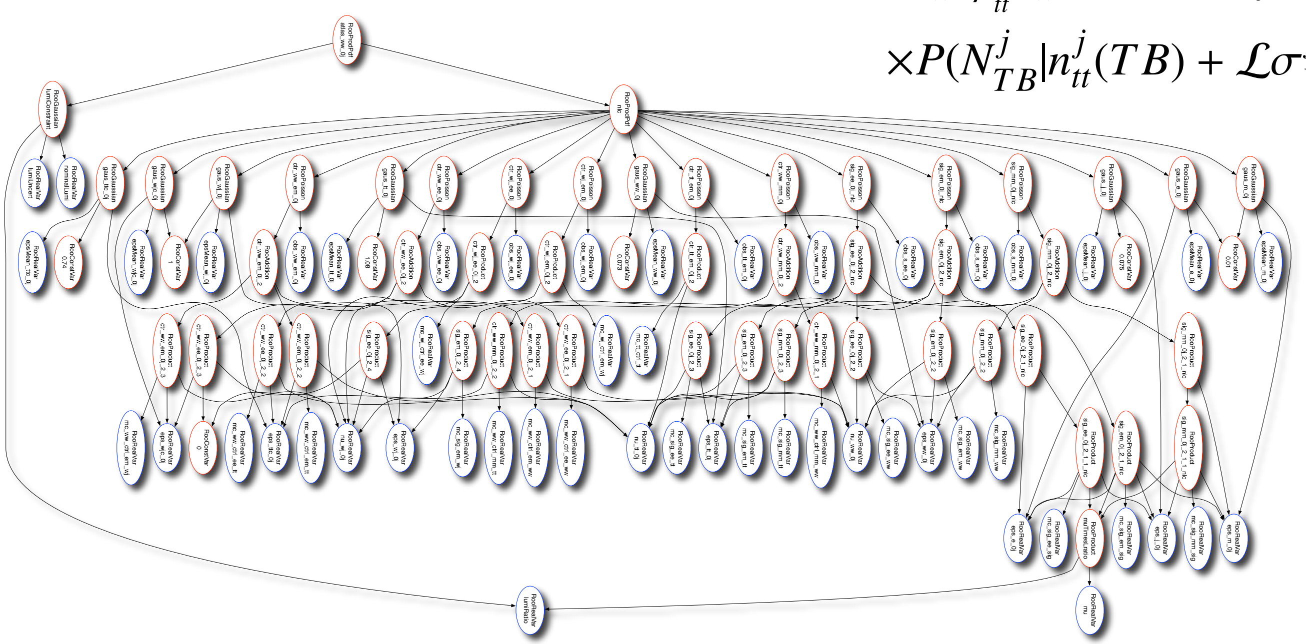
The ATLAS input:

- Poisson terms for statistical variation in control regions
- Uncertainties in extrapolation coefficients treated with truncated Gaussians and individual systematics on extrapolation coefficients were summed in quadrature
  - thus, unable to identify any correlated systematic (eg. theory uncertainty)
- after discussions, decided to use this approach for initial exercise, but the need to evolve parametrization for real combination was recognized.

$$L^{j,e\mu}_{Pois} = P(N^j_{SR}|n^j_s(SR) + \alpha^j_{WW}\nu_{\alpha^j_{WW}} n^j_{WW}(CR) + \alpha^j_{t\bar{t}}\nu_{\alpha^j_{t\bar{t}}} n^j_{t\bar{t}}(TB) + \alpha^j_{W jets}\nu_{\alpha^j_{W jets}} n^j_{W jets}(LL) + \mathcal{L}\sigma^j_{DY}(SR))$$

$$\times P(N^j_{CR}|n^j_s(CR) + n^j_{WW}(CR) + \beta^j_{t\bar{t}}\nu_{\beta^j_{t\bar{t}}} n^j_{t\bar{t}}(TB) + \beta^j_{W jets}\nu_{\beta^j_{W jets}} n^j_{W jets}(LL) + \mathcal{L}\sigma^j_{DY}(CR))$$

$$\times P(N^j_{TB}|n^j_{tt}(TB) + \mathcal{L}\sigma^j_{W jets}(TB)) \times P(N^j_{LL}|n^j_{W jets}(LL))$$



Cross-checks

CENTER FOR COSMOLOGY AND PARTICLE PHYSICS

There were a number of cross-checks performed, though not very systematic

- Clearly an area that will have more attention during the real combination

Within CMS:

- Standalone LandS tool (when equivalent test available)
  - at the time LandS was noted to be much faster (comment later)
- Bayesian results cross-checked using BAT, RooStats by Stefan Schmitz

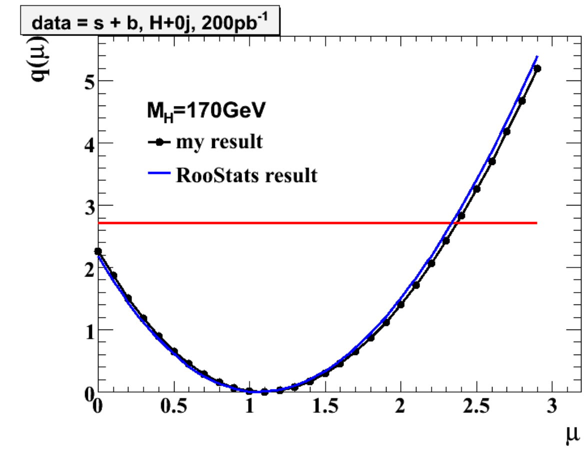
"95%" C.L. exclusion limits on signal strength modifier $r = \sigma/\sigma_{SM}$

Tools	Bayesian	Simple LR (LEP)	Profiled LR (Tevatron)	Profile LR	Profile Likelihood*
RooStats	**0.312±TBD**				**0.218**
LandS**	**0.315±0.001**	**0.290±0.003**		n/a	n/a

** LandS (Limits-and-Significance): a standalone tool used for crosschecks, plan to absorb in RooStats later
https://mschen.web.cern.ch/mschen/LandS/index.html

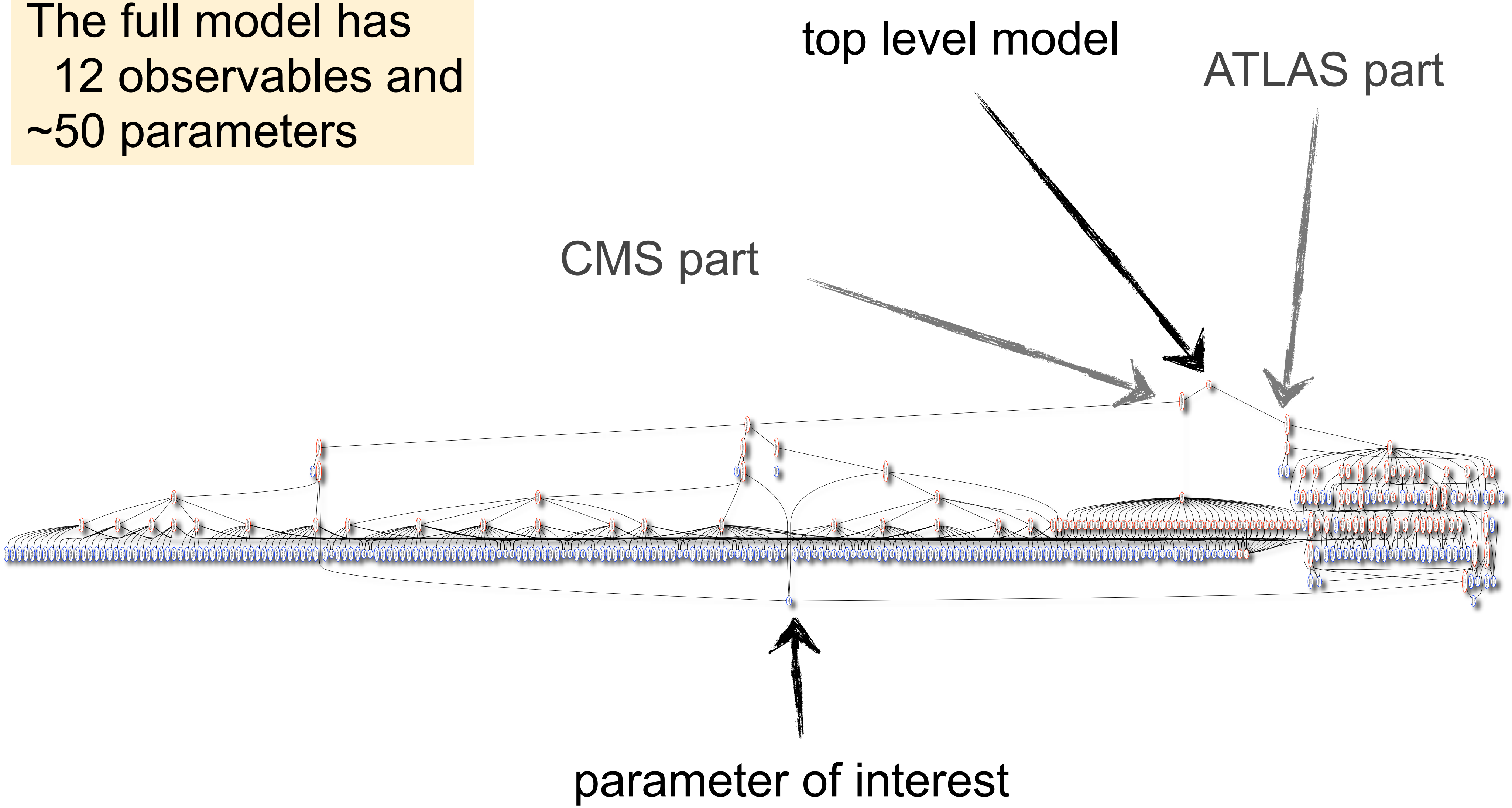
Within ATLAS:

- Standalone implementation of H→WW analysis (H. Liu)
- identical results in-memory & reading from workspace



Visualization of the ATLAS+CMS Workspace

The full model has
12 observables and
~50 parameters



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

Preliminary Results

Center for Cosmology and Particle Physics

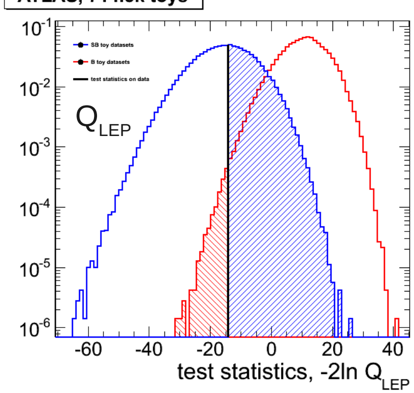
Despite the complexity, we were able to go from inputs to results in 4 days!

- not only did we get results for the combination, we did it with six techniques
- a testament to the power and flexibility of the workspace technology and the RooFit/RooStats tools

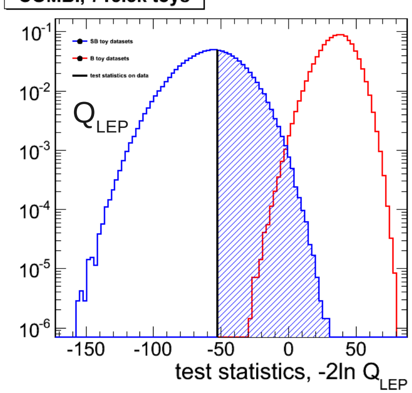
Note, the CMS result was much more powerful. Although a toy, it is loosely representative -- they are using multivariate analyses and systematics uncertainties are not so extreme

Hybrid test statistics distributions

ATLAS, 714.5k toys



COMBI, 715.5k toys



	test statistics	significance (no syst.)	significance (with syst.)
ATLAS	Q_{LEP} Q_{TEV} $\lambda(\mu)$	3.78 -	3.07 ± 0.01 2.8 ± 0.1 -
CMS	Q_{LEP} Q_{TEV} $\lambda(\mu)$	6.22 ± 0.02 - -	4.77 ± 0.02 > 4.6 4.3 ± 0.1
COMBI	Q_{LEP} Q_{TEV} $\lambda(\mu)$	- - -	> 4.6 > 3.5 -

- computing the p-value for significance in this approach is challenging:
  - speed improvements would be useful
  - or use importance sampling techniques
- CMS distribution (and results previous slide) made with a RooFit-independent tool

Grégory Schott - ATLAS-CMS statistics meeting - 01.07.2010

95% CL upper limits: results with systematics (except if indicated otherwise)

technique	test stat	rule	sampling	UL ATLAS	UL CMS	UL COMBI
Feldman-Cousins (no syst.)	$\lambda(\mu)$	CL_{S+B}	toys	0.69 ± 0.05	-	-
Profile LR (Wilks)	$\lambda(\mu)$	CL_{S+B}	asymptotic	0.79	0.28	0.25
Feldman-Cousins++	$\lambda(\mu)$	CL_{S+B}	toys	0.78 ± 0.05	0.26 ± 0.02	0.23 ± 0.02
Hybrid	Q_{LEP}	CL_S	toys	~ 0.68	0.29 ± 0.03 (LandS)	-
Hybrid	Q_{LEP}	CL_{S+B}	toys	~ 0.61	-	-
Bayesian	n/a, flat prior on r		MCMC*	0.72	0.31	0.28

Some lessons learned

In general, this combination was 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

- they are affected by luminosity uncertainty
- their theoretical uncertainties are correlated between experiments
  - **separate production modes:** 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.

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
  - **consistent procedures** are needed for assessing effect of common systematics

Attempt to directly incorporate model for control samples when feasible

- superior to approximating by Gaussian, Gamma, etc. (though often not feasible)

Anticipate and address some technical challenges early on

- for speed: make sure functions and PDFs are compiled, integrals are implemented, etc.
- specify meaningful validation exercises early on

Progress Since July

ROOT developments for 5.28

Center for Cosmology and Particle Physics

Since July, big effort in RooFit/RooStats aimed at ROOT 5.28 production release

- to be released mid-December
- bug fixes, memory leaks, etc.
- validation, tutorial macros, documentation

A few big developments relevant for LHC-HCG

- ToyMC Sampler is now PROOF-enabled
  - Test: probe 5σ in a 3-channel combination with 50 nuisance parameters by using with 30 machines to generate 10 million pseudo-experiments
- ToyMC Sampler now has importance sampling
  - development from Banff workshop. Can lead to 100-1000 speed improvements
- Validation example using the RooStats HybridCalculator for the prototype problem where $Z_{bi}=Z_{\Gamma}$ correspondence is known (results agree to several digits).

ROOT 5.28 will now ship with **HistFactory**, a tool that transforms information in tabular format into a RooFit/RooStats workspace

- command line tool **$ hist2workspace input.xml**
- supports histograms as well as pure number counting
- supports Gaussian, Gamma, Lognormal, Uniform constraints & asymmetric uncertainties
- exports models ready to be used by RooStats tools

The basic template

For each sig & bkg estimate, the expected number of events is modeled as

$$N_{exp} = L\, f\, \epsilon(\alpha)\, \sigma(x;\alpha)$$

- For data-driven estimates, $L=L_0$, the nominal luminosity
- For theory-driven estimates L is an nuisance parameter (constrained)
- f is an overall scaling factor that is left unconstrained
  - these are typically things we measure, like $\mu=\sigma/\sigma_{SM}$
  - can also be a ratio of cross-sections $r=\sigma_{tt}/\sigma_{Z}$ or $r=\sigma_{\mu\mu}/\sigma_{e\mu}$
- $\varepsilon(\alpha)$ is an efficiency or acceptance term assembled from the individual systematics, and there is an α for each source of systematic
- $\sigma(x;\alpha)$ is a histogram for the variable x (in units of cross-section) that interpolates between different variational histograms

By using the same name for the systematic source or scale factor, one can assemble complex combined models that are very general

Example xml files

A 1-channel example, where signal histogram normalization multiplied by “SigXsecOverSM”, which is considered the parameter of interest.

- Nuisance parameters α_j: “Lumi”, “syst1” (sig only), “syst2” (bkg1 only), “syst3” (bkg2 only)

$$N_{exp} = L\, f\, \epsilon(\alpha)\, \sigma(x;\alpha)$$

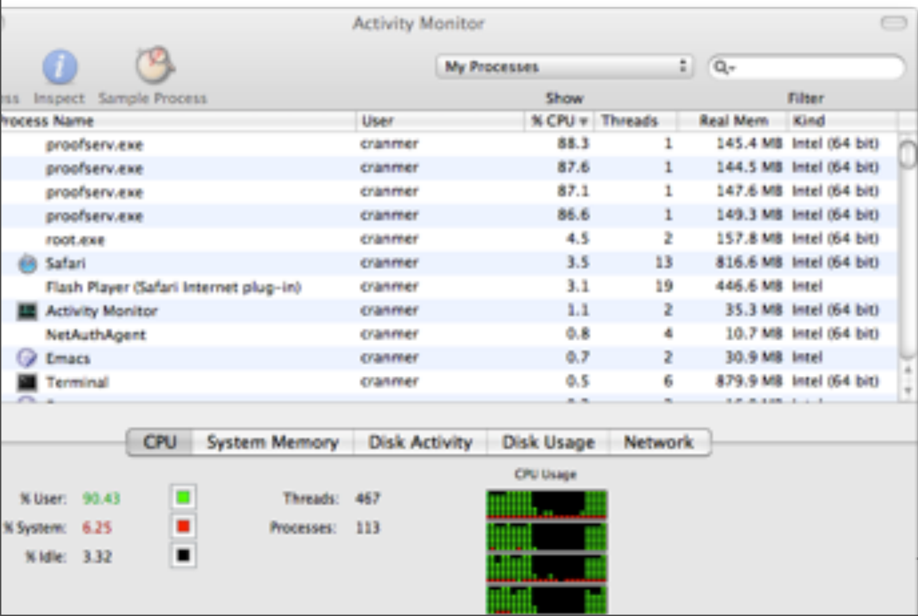
```
<!DOCTYPE Channel  SYSTEM 'Config.dtd'>

  <Channel Name="channel1" InputFile="./data/example.root" HistoName="" >
    <!--<Data Name="data" InputFile="" HistoPath="" HistoName=""/>-->
    <Sample Name="signal" HistoPath="" HistoName="signal">
      <OverallSys Name="syst1" High="1.05" Low="0.95"/>
      <NormFactor Name="SigXsecOverSM" Val="1" Low="0.5" High="1.8" Const="True" />
    </Sample>
    <Sample Name="background1" HistoPath="" NormalizeByTheory="True" HistoName="background1">
      <OverallSys Name="syst2" Low="0.95" High="1.05"/>
    </Sample>
    <Sample Name="background2" HistoPath="" NormalizeByTheory="True" HistoName="background2">
      <OverallSys Name="syst3" Low="0.95" High="1.05"/>
      <!-- <HistoSys Name="syst4" HistoPathHigh="" HistoPathLow="histForSyst4"/>-->
    </Sample>
  </Channel>
```

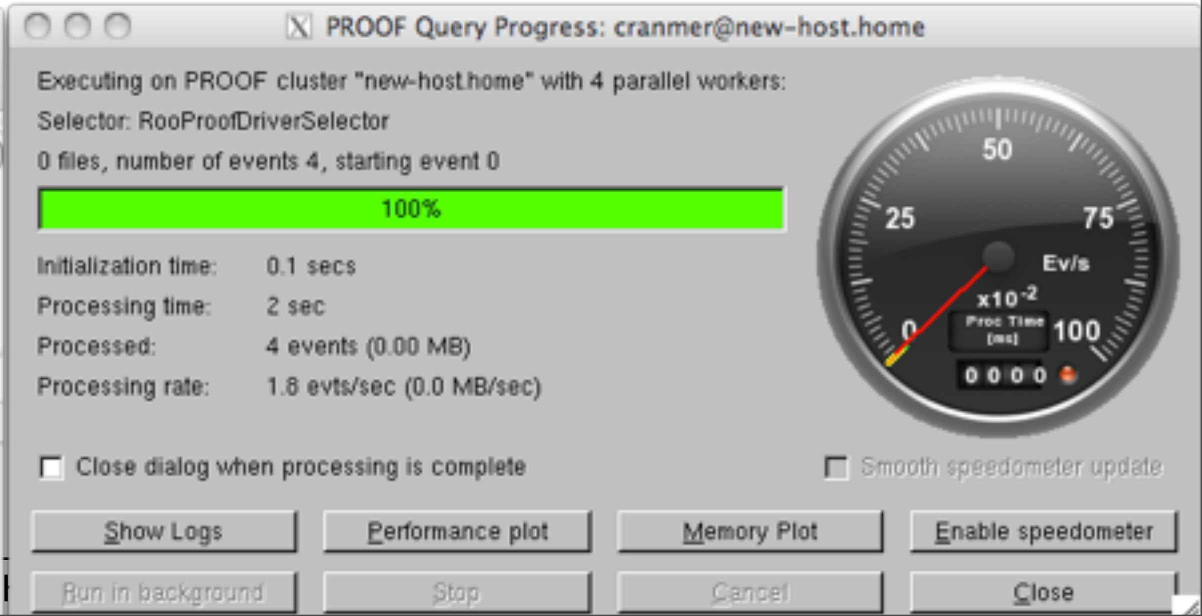
An example session

On my laptop with 4 processors using PROOF-lite

- create model from XML file, then test with FeldmanCousins tool

```
$ hist2workspace config/atlas_example.xml
$ root.exe results/atlas_model.root
root [1] using namespace RooStats
root [2] data = combined->data("simData")
root [5] mc = (ModelConfig*) combined->obj("ModelConfig")
root [6] FeldmanCousins fc(*data,*mc)
root [7] fc.SetConfidenceLevel(0.95)
root [8] fc.UseAdaptiveSampling(true)
root [9] fc.FluctuateNumDataEntries(false)
root [10] ProofConfig pc(*combined, 4, "workers=4");
root [11] toymcsampler = (ToyMCSampler*) fc.GetTestStatSampler();
root [12] toymcsampler->SetProofConfig(&pc);     // enable proof
root [13] interval = fc.GetInterval()
```





Hypothesis Testing

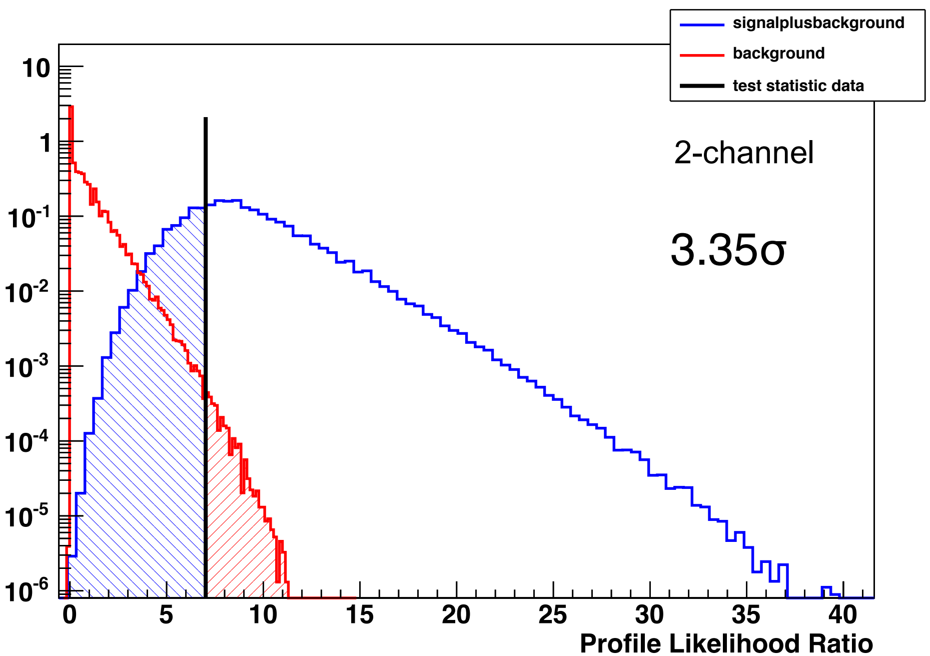
Center for Cosmology and Particle Physics

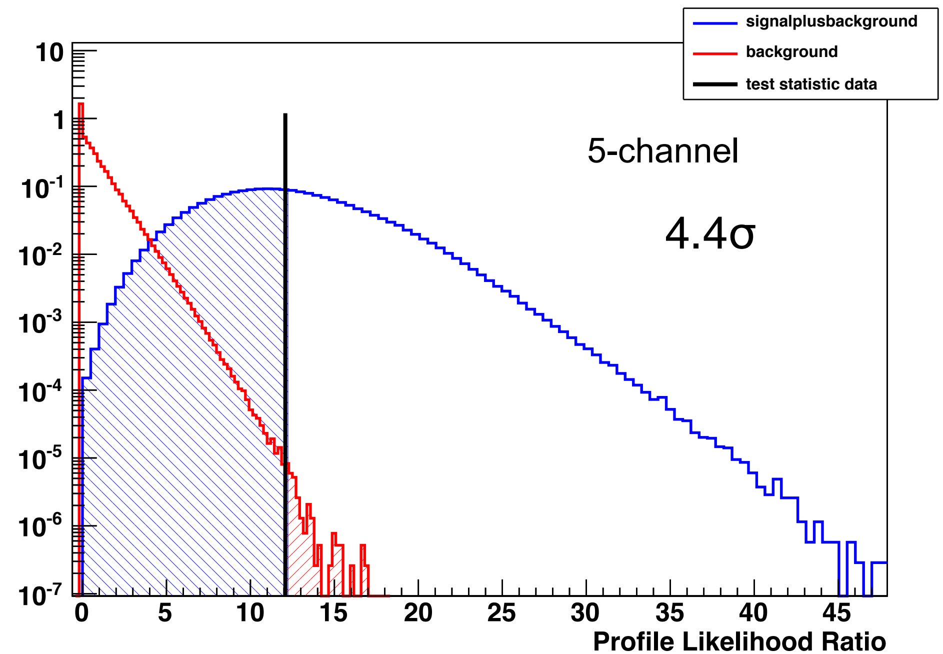
Now on a real PROOF cluster with 30 machines

- real world example throws millions of toys experiments, does full fit on 50 parameters for each toy.
- also supports producing simple shells scripts for use with GRID or batch queues

Now **importance sampling** is also implemented,

- following presentation at Banff with particle physics & statistics experts
- allows for 1000x speed increase!
- Still being tested in detail

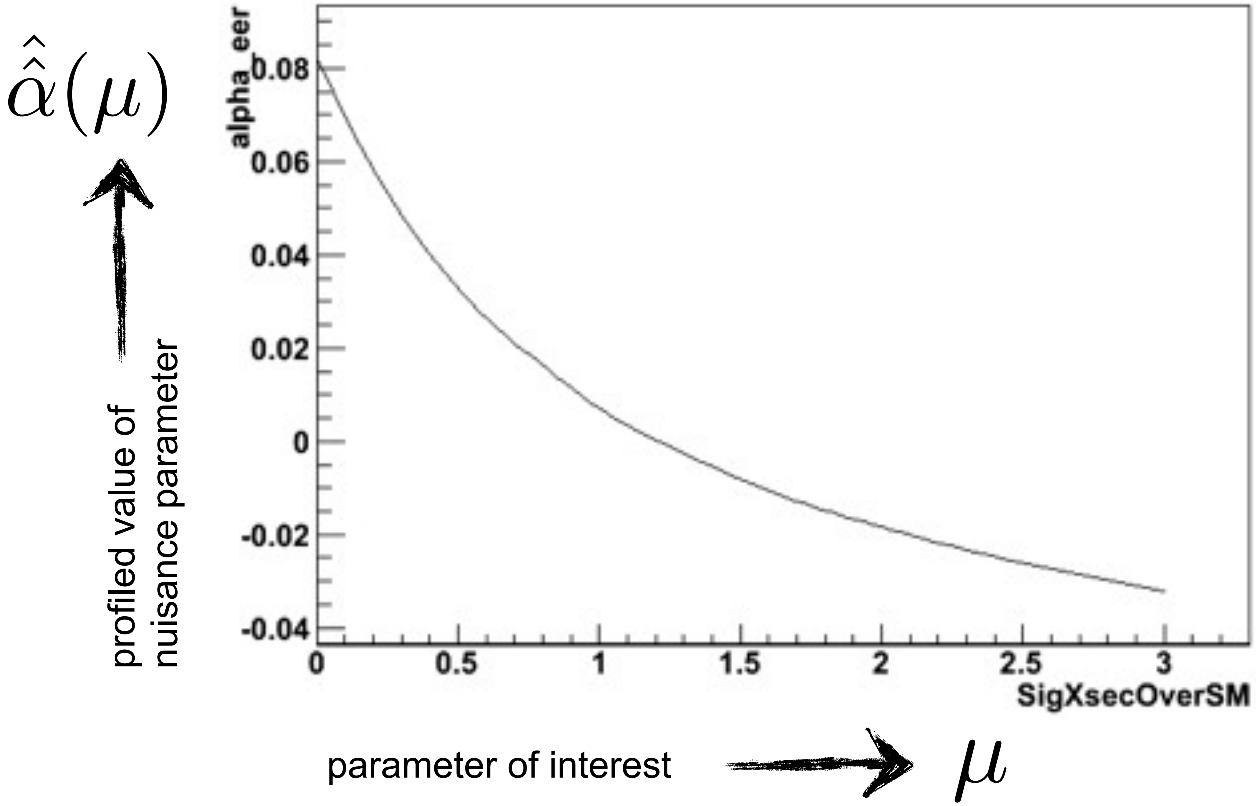




ProfileInspector

Request to be able to inspect the profiling of nuisance parameters ⇒ new tool

```
$ hist2workspace config/top_dilep_2010.xml
$ root.exe results/dilep_2010_combined_dilep_allsys_model.root
root [1] using namespace RooStats;
root [2] ProfileInspector p
root [3] data = *combined->data("simData")
root [4] ModelConfig* mc = combined->obj("ModelConfig”)
root [5] TList* list = p.GetListOfProfilePlots(*data,mc)
root [6] list->At(5)->Draw("al")
```



Summary and Conclusions

ATLAS and CMS completed a Higgs combination exercise in July 2010

- intense effort lasting roughly ~1 month
- results of toy combination were shown on the last day of the Higgs cross-section workshop
- Inputs from ATLAS & CMS were based on $H \rightarrow WW$ +0j
- RooFit/RooStats workspaces were used to communicate and RooStats tools were used for statistical tests

It was a big success in terms of cooperation and technical achievement

- much of the ground work was established, several lessons learned
  - those lessons drove bulk of RooFit/RooStats development effort since July

The timeline proposed is ambitious, but seems realistic given our previous experience.

Good luck to the LHC-HCG in 2011!