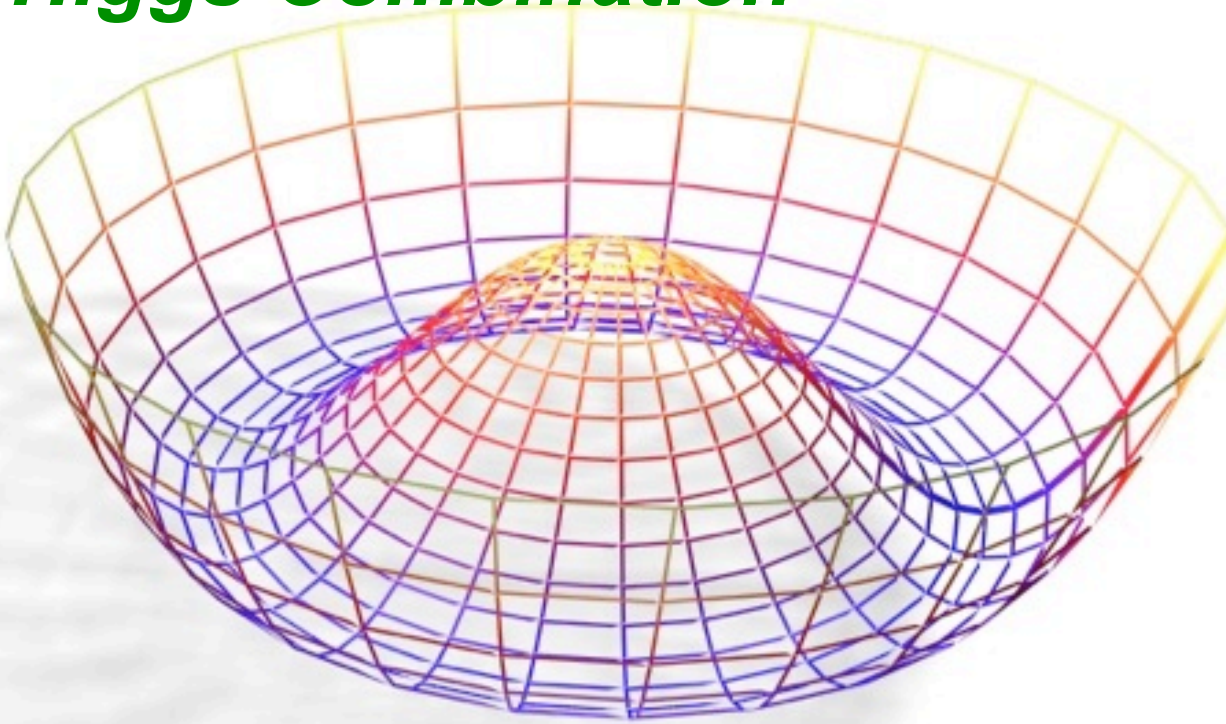
CENTER FOR
COSMOLOGY AND
PARTICLE PHYSICS

RooStats tools used for Top observation and ATLAS+CMS Higgs Combination



Kyle Cranmer,
(NYU)

Thanks to Silas Whiteson for his help

Outline

- Reminder on toy ATLAS+CMS Higgs combination this summer
  - how it was done technically
    - obtaining inputs
    - combining inputs
    - different statistical tests (and test statistics)
  - what we learned and can do better
    - correlated systematics, exposing nuisance parameters
    - separate source of systematic from effect on rates and efficiencies
- The current top observation and cross-section measurement
  - common tool for converting histograms into RooFit models
  - New ToyMC Sampler running on PROOF
- Informal expert working group for combinations

The toy ATLAS+CMS Higgs Combination

Toy ATLAS+CMS Higgs Combination

In June, the ATLAS & CMS Higgs groups agreed to perform a toy combination. Several meetings to agree on details.

- On Thursday, July 1, the inputs were provided by both collaborations
- On Tuesday, July 6 (the last day of the LHC Higgs cross-section workshop) combined results were shown using multiple techniques

http://indico.cern.ch/conferenceDisplay.py?confId=100458

Each experiment wrote down their likelihoods, which were based on number counting in several main measurements and in several control samples.

- Both experiments also implemented these likelihood functions using RooFit's modeling language and saved the result in a Workspace. Likelihood function & Priors were separated.
- Finally, RooStats tools were used for the statistical analysis.

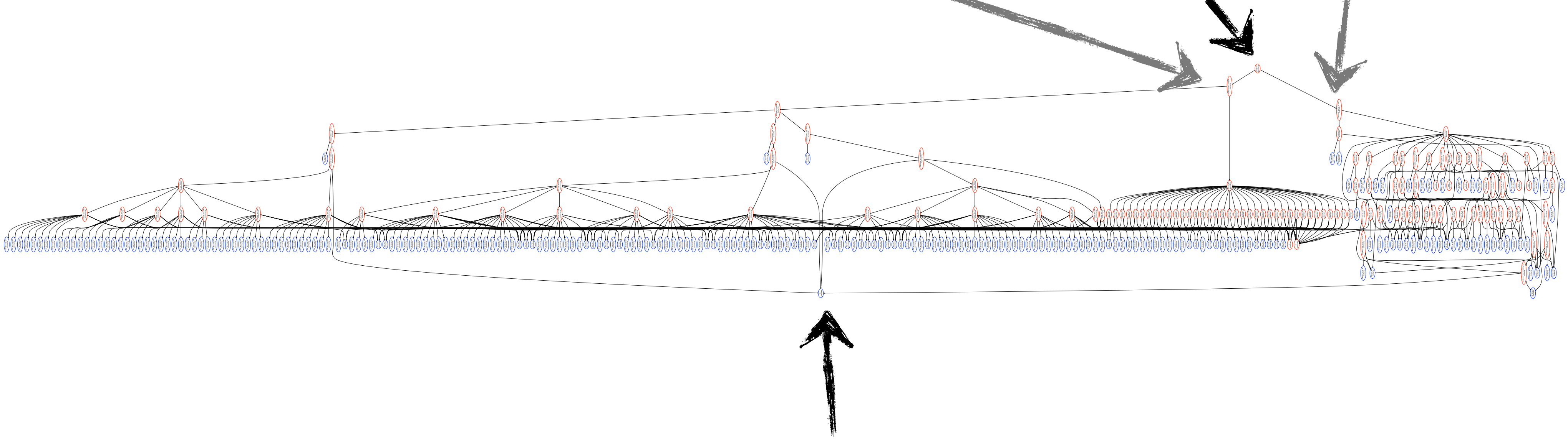
Visualization of the ATLAS+CMS Workspace

The full model has
12 observables and
~50 parameters

top level model

ATLAS part

CMS part



parameter of interest

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

The two approaches

In the ATLAS approach backgrounds were estimated from data using control samples, and extrapolation coefficients between the control and signal regions (called α and β) were assigned systematics.

- different sources of systematic uncertainty were added in quadrature to obtain an overall uncertainty on the α & β
- the systematics were large, and modeled with truncated Gaussians
- The ATLAS analysis was very conservative

In the CMS approach, each source of systematic was identified with a nuisance parameter. The impact on each signal and background was estimated separately.

- auxiliary measurements included as additional terms in the likelihood
- Used log-normal for all systematics
- The CMS analysis used NN's and was much more powerful

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

Basic formalism: one dataset

Start with a **model** for the data, eg. a probability density function for x written $P(x|\mu,\nu)$ that is parametrized by

- parameters of interest: $\mu : m_H, \sigma, ...$
- nuisance parameters: $\nu : b, JES, \epsilon_b, ...$

The **likelihood function** is given by

$$L(\mu,\nu) = \prod_{i \in \mathrm{events}} P(x_i|\mu,\nu)$$

And the **profile likelihood ratio** is given by:

$$\lambda(\mu) = \frac{L(\mu,\hat{\hat{\nu}})}{L(\hat{\mu},\hat{\nu})}$$

Note: for frequentist tests we need more than the likelihood function, we also need the pdf so that we can generate toy data for x

Note: for Bayesian analyses, we also need a prior for all parameters

Note: for Hybrid analyses we need prior for nuisance parameters

Basic formalism: two datasets

With two datasets, the observables may be different and one needs a **model** for each dataset.

- models may have different nuisance parameters, but should share parameters of interest, eg.

$$P_1(x|\mu,\nu) \qquad P_2(y|\mu,\alpha)$$

The **likelihood function** is given by

$$L(\mu,\nu,\alpha) = \prod_{i\in\text{data1}} P_1(x_i|\mu,\nu)\cdot \prod_{i\in\text{data2}} P_2(y_i|\mu,\alpha)$$

In RooFit/RooStats, this type of situation is represented by:

- a "combined dataset" for data1, data2 with a "category label"
- a "simultaneous PDF"
  - keeps track of P_1, P_2 & associated category

x
2.7
1.3

\+

y
5.1
7.2

=

x	y	category
2.7	-	1
1.3	-	1
-	5.1	2
-	7.2	2

Note: if there is only one measurement of *y* for the constraint term (eg. a single number summarizing the auxiliary measurement), it is not necessary to include y in a dataset and produce a RooSimultaneous (which can be cumbersome). In that case a simple RooProduct will suffice.

Providing ingredients via the workspace

To perform the combination, we need these basic ingredients:

- we need the full model (eg. the simultaneous PDF) for each channel
- we need the combined dataset for each channel
- we need to know what the names of the parameters of interest are and what they correspond to for each channel
- we do NOT need to know and we don't want to worry about:
  - structure and complexity of PDF, structure of the combined dataset
  - anything about the nuisance parameters (unless we want to introduce correlations between them in the combination)

This can be achieved for an arbitrary model and dataset with these 5 lines:

- technically we need any custom code, but we don't want to deal with it by hand.

```
RooWorkspace workspace("Example_workspace");
workspace.import(*data);
workspace.import(*pdf);
workspace.importClassCode();
workspace.writeToFile("myWorkspace.root");
```

Within ATLAS, combinations based on this methodology have been performed successfully

Creating the model

In order to communicate the additional "meta information" about the meaning of the parameters, we have introduced a new class called **ModelConfig**

- If the groups want to go "above and beyond" they can also make a ModelConfig and import it into the workspace.

Create a new ModelConfig

Specify workspace that holds pdf, parameters of interest, observables, ...

```
// specify components of model for statistical tools
ModelConfig* modelConfig = new ModelConfig("G(x|mu,1)");
modelConfig->SetWorkspace(*wspace);
modelConfig->SetPdf( *wspace->pdf("normal") );
modelConfig->SetParametersOfInterest( *wspace->set("poi") );
modelConfig->SetObservables( *wspace->set("obs") );
```

```
// Bayesian tools would also need a prior
modelConfig->SetPriorPdf(*wspace->pdf("prior"));

// import modelConfig into workspace too
wspace->import(*modelConfig);
```

now import the modelConfig object into the workspace too (before writing it to the ROOT file)

Add prior for parameters for use in Bayesian methods (kept separate from frequentist part of the pdf)

More documentation here:
http://root.cern.ch/viewcvs/branches/dev/roostats/roofit/roostats/doc/usersguide/RooStats_UsersGuide.pdf

Prototype of a general combination tool

CENTER FOR COSMOLOGY AND PARTICLE PHYSICS

Combination used a variant of the same combination script used in 2009 for ATLAS Higgs combination. This is being abstracted into a RooStats combination utility (more later)

```
void combine(string mass,double& mllSig, double& mlhSig, double& mcombSig){

  TFile llf(("workspace_simul_ll_m"+mass+".root").c_str());
  RooWorkspace* llWS = (RooWorkspace*) llf.Get("Example_workspace");
  TFile lhf(("workspace_simul_lh_m"+mass+".root").c_str());
  RooWorkspace* lhWS = (RooWorkspace*) lhf.Get("Example_workspace");

  // get original models
  RooAbsPdf* llModel = llWS->pdf("simPdf");
  RooAbsPdf* lhModel = lhWS->pdf("simPdf");
  cout << "models = " << llModel << " " << lhModel << endl;

  // should be able to use abs data, but Import(string, data) needs specific on
  RooDataSet* lldata = (RooDataSet*) llWS->data("data");
  RooDataSet* lhdata = (RooDataSet*) lhWS->data("data");

  // add models to new WS with suffix
  RooWorkspace* combWS = new RooWorkspace("combWS");
  combWS->import(*llModel, RenameAllNodes("ll"), RenameAllVariables("ll"));
  combWS->import(*lhModel, RenameAllNodes("lh"), RenameAllVariables("lh"));

  // need to rename type_ll and type_lh back to common type
  combWS->import(*llWS->cat("type"));
  RooCategory* type = combWS->cat("type");
  combWS->factory("EDIT::ll_tmp(simPdf_ll, type_ll=type)");
  combWS->factory("EDIT::lh_tmp(simPdf_lh, type_lh=type)");
  combWS->Print();

  // get new WS back out
  llModel = combWS->pdf("ll_tmp");
  lhModel = combWS->pdf("lh_tmp");

  RooCategory* channel = new RooCategory("channel","channel");
  channel->defineType("ll",0);
  channel->defineType("lh",1);

  RooArgSet vars=*((RooArgSet*)lldata->get()->clone("vars"));
  vars.Print();
  RooRealVar w("w","w",0,10000);
  vars.add(w);
```

```
  RooDataSet* combData = new RooDataSet("combData", "combined dataset",
                                        vars,
                                        Index(*channel),
                                        Import("ll",*lldata),
                                        Import("lh",*lhdata),
                                        WeightVar(w));

  map<string, RooAbsPdf*> pdfMap;
  string llStr="llStr";
  string lhStr="lhStr";
  pdfMap["ll"]=llModel;
  pdfMap["lh"]=lhModel;

  RooSimultaneous* combModel_split
    = new RooSimultaneous("combModel_split", "", pdfMap, *channel);

  combWS->import(*combModel_split);
  channel = combWS->cat("channel"); // get pointer to one in WS

  // here we establish correspondence of common variables
  combWS->import(*llWS->var("mH"));
  combWS->import(*llWS->var("mu"));
  combWS->import(*llWS->var("mTauTau"));
  combWS->factory("EDIT::combModel(combModel_split,mH_ll=mH,mH_lh=mH,\
_ll=mu,mu_lh=mu,mTauTau_lh=mTauTau,mTauTau_ll=mTauTau)");

  RooAbsPdf* combModel = combWS->pdf("combModel");

  // do the combined fit
  combModel->fitTo(*combData, Constrained(), Hesse(kFALSE), Minos(kFALSE));

  // do profile LR
  RooRealVar* mu = combWS->var("mu");
  RooAbsReal* nll = combModel->createNLL(*combData, Constrained());
  RooAbsReal* profile = nll->createProfile(*mu);

  mu->setVal(0);
  cout << "sqrt(-2loglambda(0)) = " << sqrt(2*profile->getVal()) << endl;
```

Statistical tests

After producing a workspace for the combined model and combined data we considered several statistical tests:

- For discovery:
  - profile likeilhood ratio
  - Bayesian/Frequentist Hybrid using Toy MC
    - 3 different test statistics
- For cross-section measurement:
  - profile likelihood ratio
  - Feldman-Cousins (with nuisance parameters)
  - Inverted hypothesis test based on Hybrid Calculator (aka FC-CH)
  - Bayesian MCMC (with BAT interface and native RooStats MCMC tool)

Attacking the same statistical model with ~7 statistical techniques (all in 4 days) is a real triumph!

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 .

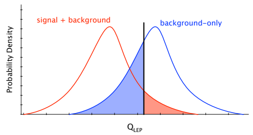
- best fit point is: $\hat{\mu}, \hat{\nu}$
- best fit of nuisance parameters with μ fixed is $\hat{\hat{\nu}}$ (aka “profiled”)

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

Three common test statistics used in the field are:

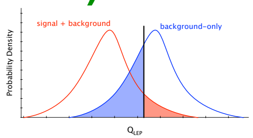
- simple likelihood ratio (used at LEP, nuisance parameters fixed)

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



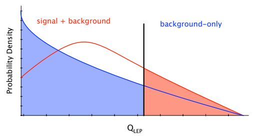
- ratio of profiled likelihoods (used commonly at Tevatron)

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



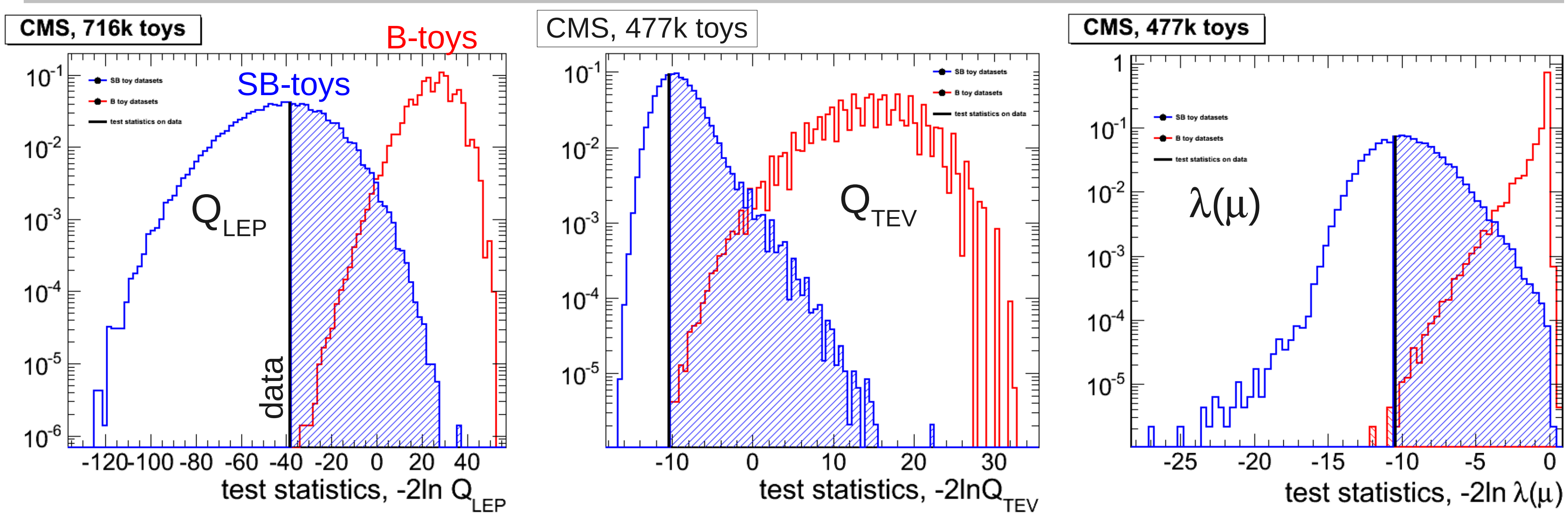
- profile likelihood ratio (related to Wilks’s theorem)

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



Initial results

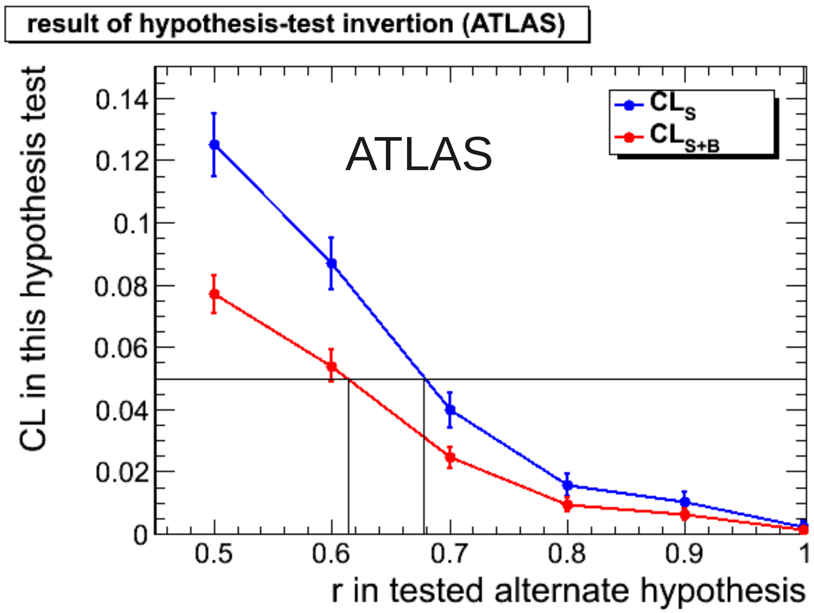
Hybrid Frequentist-Bayesian

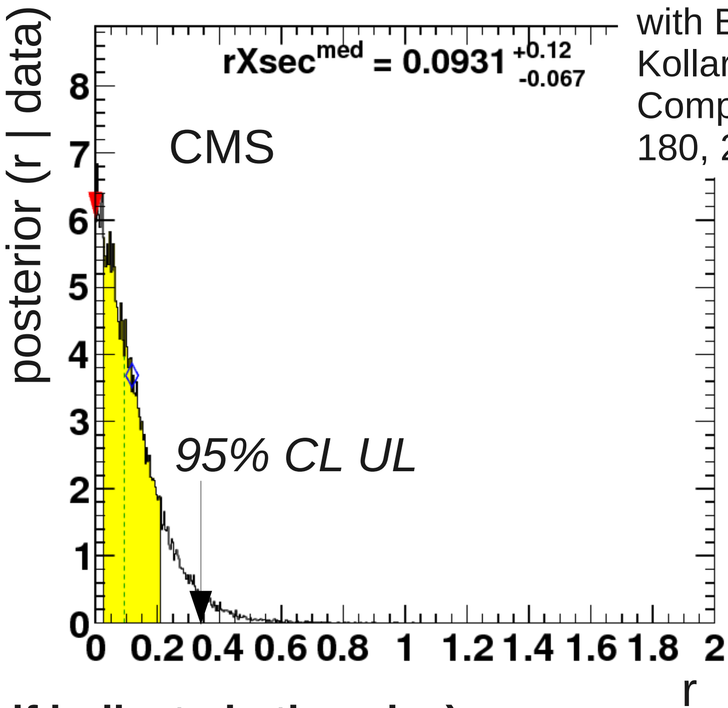


- run for the 3 test statistics mentionned in the previous talk
- need to study the advantage and inconvenient of the 3 approaches

	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 -

Summary of upper limits





with BAT [Caldwell, Kollar, Kroeninger, Comp.Phys.Com. 180, 2197 (2009)]

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

Technical lessons learned

The prototype for the general purpose combination tool is a good start

- sometimes tools that build the individual models can do form the combination internally
  - eg. the SUSY tool shown earlier today
  - eg. the Top tool shown next
- but in other cases we need to combine models coming from different groups that may be modeled in significantly different way

Sometimes the way in which the model is coded can lead to avoidable inefficiencies

- for example the log-normal used by CMS was done as an interpreted "FormulaVar" instead of a compiled class
- sometimes the analytic integral of a PDF is known, but not implemented. This leads to unnecessary numerical integration.

For these reasons, it is useful to build some ATLAS-wide expertise in these areas

The Top Observation & Cross-section measurement

Goal of HistFactory

Many analyses are based on template histograms (ROOT TH1)

- provide a tool that allows one to use RooStats statistical tools without knowing RooFit's data modeling language

In this approach, user provides other templates corresponding to variations of individual systematics

- this is done for each source of systematic and for each signal and background individually
- It is straightforward to provide a combined model for several channels and to identify the same systematic in each channel

The user specifies all of these systematics via an XML file and a compiled command line executable parses the XML file to produce the combined model

- by default, it also runs a profile likelihood analysis on the parameters of interest

The factory

For each background 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>
```

Example xml files

Center for Cosmology and Particle Physics

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

- Can drive parameter settings for “measurements” via XML

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

<Combination OutputFilePrefix="./results/example" Mode="comb" >

  <Input>./config/example_channel.xml</Input>

  <Measurement Name="Example" Lumi="10" LumiRelErr="0.05" BinLow="0" BinHigh="2" Mode="comb">
    <POI>SigXsecOverSM</POI>
    <ParamSetting Const="True">Lumi alpha_syst1</ParamSetting>
  </Measurement>

</Combination>
```

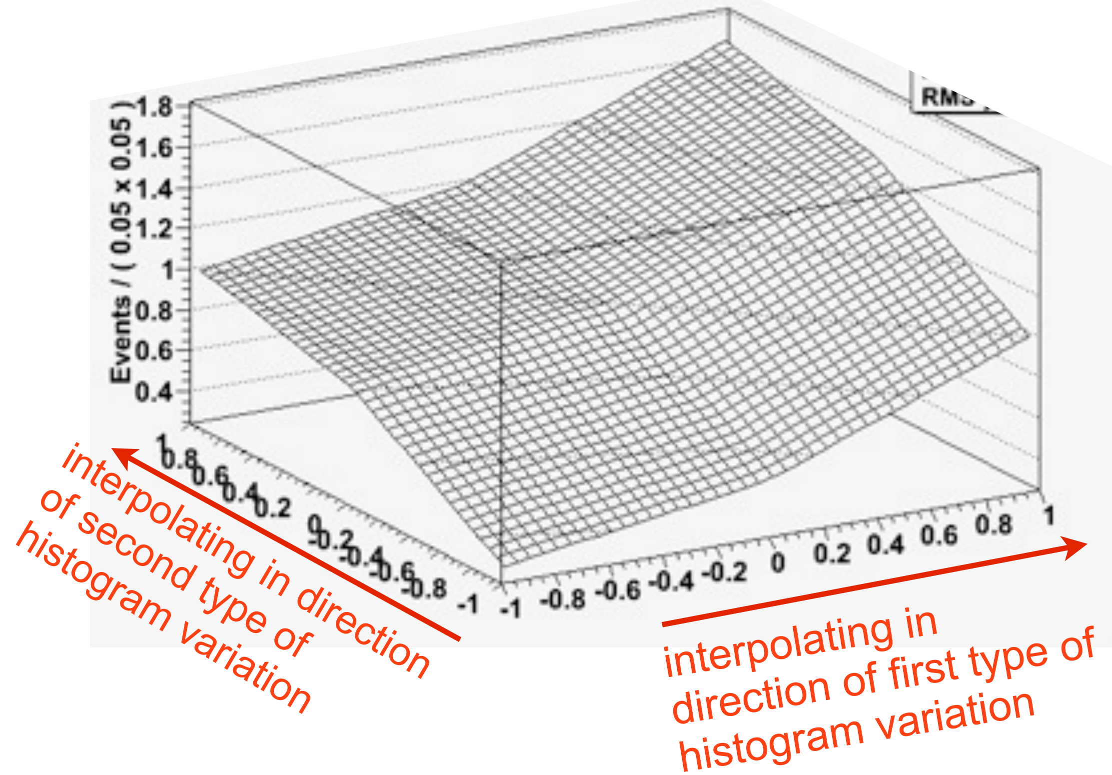
```
<!DOCTYPE Channel  SYSTEM 'Config.dtd'>

  <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>
```

Parametrized templates (eg. systematics)

Given nominal histograms and +/- variations for each source of systematic α_j produce a family of predictions parametrized by α_j with linear interpolation:

$$\varepsilon_{jk}(\alpha_j) = \begin{cases} \tilde{\varepsilon}_{jk} + \alpha_j(\varepsilon_{jk}(\alpha_j^+) - \tilde{\varepsilon}_{jk}) & \text{if } \alpha_j > 0 \\ \tilde{\varepsilon}_{jk} & \text{if } \alpha_j = 0 \\ \tilde{\varepsilon}_{jk} - \alpha_j(\varepsilon_{jk}(\alpha_j^-) - \tilde{\varepsilon}_{jk}) & \text{if } \alpha_j < 0\,. \end{cases}$$



Now expectation is:

$$N_k^{exp} = \mathscr{L}\sigma_k \prod_j \tilde{\varepsilon}_{jk} \frac{\varepsilon_{jk}(\alpha_j)}{\tilde{\varepsilon}_{jk}} = \tilde{N}_k^{exp} \prod_j \frac{\varepsilon_{jk}(\alpha_j)}{\tilde{\varepsilon}_{jk}}$$

Then, constrain α_j with Normal

$$L(\sigma_{sig}, \mathscr{L}, \alpha_j) = Pois(N^{obs}|N^{exp}_{tot}) \times Gaus(\tilde{\mathscr{L}}|\mathscr{L}, \sigma_{\mathscr{L}}) \times \prod_j Gaus(\tilde{\alpha}_j = 0|\,\alpha_j, \Delta_{\alpha_j} = 1)\,.$$

Generalization for multiple bins and multiple channels:

$$L(\sigma_{sig}, \mathscr{L}, \alpha_j) = \prod_{l\in\{ee,\mu\mu,e\mu\}} \left\{ \prod_{i\in bins} \left[Pois(N_i^{obs}|N_{i,tot}^{exp})\, Gaus(\tilde{\mathscr{L}}|\mathscr{L}, \sigma_{\mathscr{L}}) \prod_{j\in syst} Gaus(0|\alpha_j, 1) \right] \right\}$$

Support for different constraints

For large uncertainties, truncated Gaussian are a bad choice

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

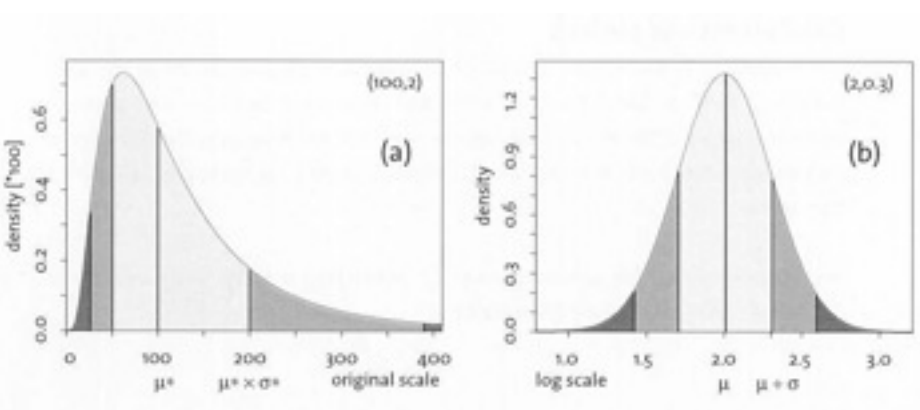
Gamma (reasonable for systematics constrained from measurements)

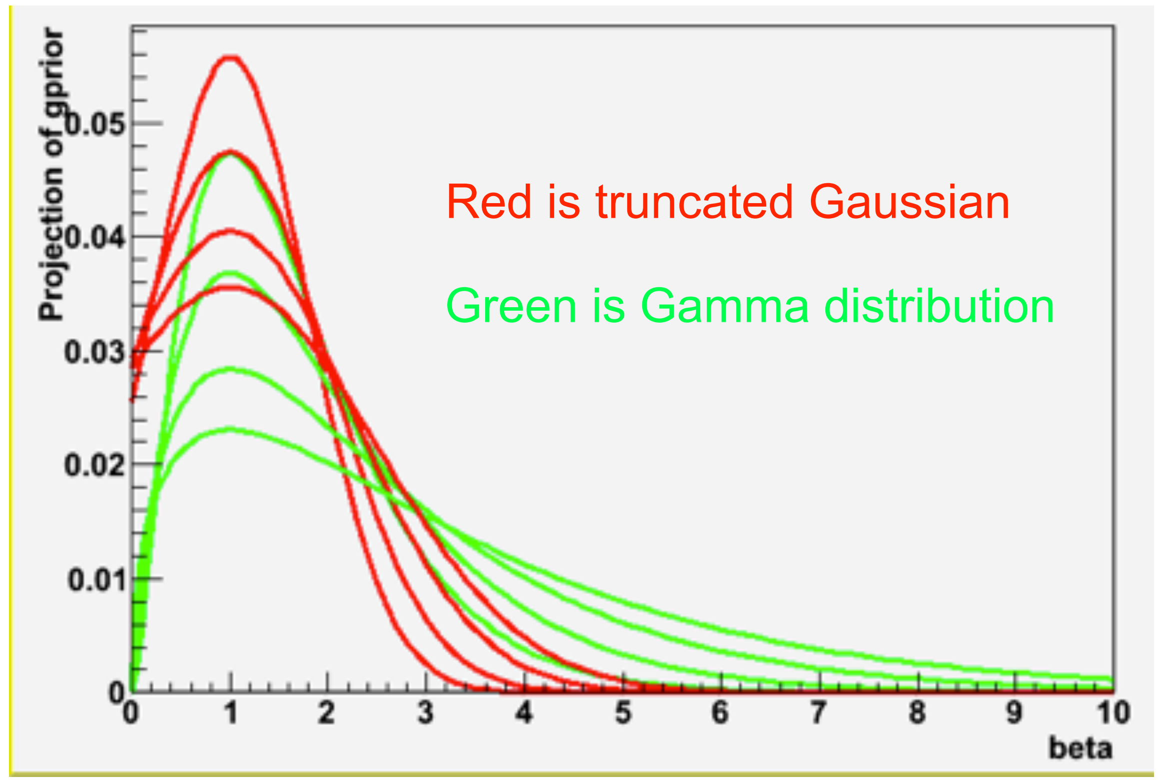
- longer tail, good behavior near 0, natural choice if auxiliary is based on counting

Log-normal is another popular choice

Tool was modified so that one can specify what form of constraint one would like for each source of systematic

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





bkg scale factor

l+jets Input table

Center for Cosmology and Particle Physics

(Michele Pinamonti Sept. 8)

```
Code            Shape          Intr.Unc Description
------          ---------      -------- ------------
LUMI            Gaussian       0.11     Luminosity uncertainty
ZXSEC           Gaussian       1.00     Z+jets cross section uncertainty
MMSYS           Gaussian       0.30     QCD Matrix method lepton+jets correlated systematic uncertainty  <-- input
MMSTATE         ScaledPoisson  0.40     Statistical uncertainty on QCD Matrix method for e+jets  <-- input needed
MMSTATM         ScaledPoisson  0.40     Statistical uncertainty on QCD Matrix method for mu+jets  <-- input needed
BTAGEFF         Gaussian       0.15     B-tagging efficiency uncertainty
CTAGEFF         Gaussian       0.30     C-tagging efficiency uncertainty
LTAGEFF         Gaussian       0.50     L-tagging efficiency uncertainty
STXS            Gaussian       0.10     Single top cross section uncertainty
JES             Gaussian       0.10     Jet Energy Scale uncertainty
W4BG            Gaussian       0.35     W+4 jets uncertainty from Berends Giele Scaling
WHFRAC          Gaussian       1.00     Uncertainty on fraction W+HF(4jet)/W+AF(4jet)
WCFRAC          Gaussian       1.00     Uncertainty on fraction W+c+3jets

Channel: e+jets
Component       Yield   Systematics (name=value)
SIGNAL          14.94
Qcd multije     0.9        MMSYS=  0.30       MMSTATE= 0.40  <-- input needed from Note V group
W+HF jets       0.82       BTAGEFF= 0.05      CTAGEFF= 0.04  LTAGEFF= 0.02   WHFRAC= 1.00   W4BG= 0.47
W+c+jets        0.218      CTAGEFF= 0.17      LTAGEFF= 0.04  WCFRAC= 1.00    W4BG= 0.47
W+LF jets       0.30       LTAGEFF= 0.41      W4BG= 0.47
Z+jets          0.15       ZXSEC= 1.00        LUMI= 0.11     JES= 0.20
single top      0.69       STXS= 0.10         LUMI= 0.11     JES= 0.19   BTAGEFF= 0.13

Channel: mu+jets
Component       Yield   Systematics (name=value)
SIGNAL          15.91
Qcd multije     1.2        MMSYS=  0.30       MMSTATM= 0.40  <-- input needed from Note V group
W+HF jets       0.99       BTAGEFF= 0.02      CTAGEFF= 0.05  LTAGEFF= 0.03   WHFRAC= 1.00   W4BG= 0.34
W+c+jets        0.23       CTAGEFF= 0.12      LTAGEFF= 0.03  WCFRAC= 1.00    W4BG= 0.34
W+LF jets       0.34       LTAGEFF= 0.53      W4BG= 0.34
Z+jets          0.08       ZXSEC= 1.00        LUMI= 0.11     JES= 0.25
single top      0.70       STXS= 0.10         LUMI= 0.11     JES= 0.14   BTAGEFF= 0.10
```

Note: No systematics on signal, needed for xs measurement

Updated Tool

The combination tool has been updated to support Gaussian and Gamma distributions for constraints

- Thanks to Dominique for XML support
- link to tool in SVN
- **next**: add support for uniform constraints, which may be useful for theoretical uncertainties. A few hours work.

The table that was sent was converted to XML

- ejets_2010, mjets_2010 should be checked
- Ideally, this file gets updated directly.
- The top_ljets_2010.xml file has several variations on systematics

To run it:

- make; ./bin/LikelihoodRatioFit config/top_ljets_2010.xml

This tool only does profile likelihood

- **next:** run toys using same model

root / Physics / Top / Software / HistFactory

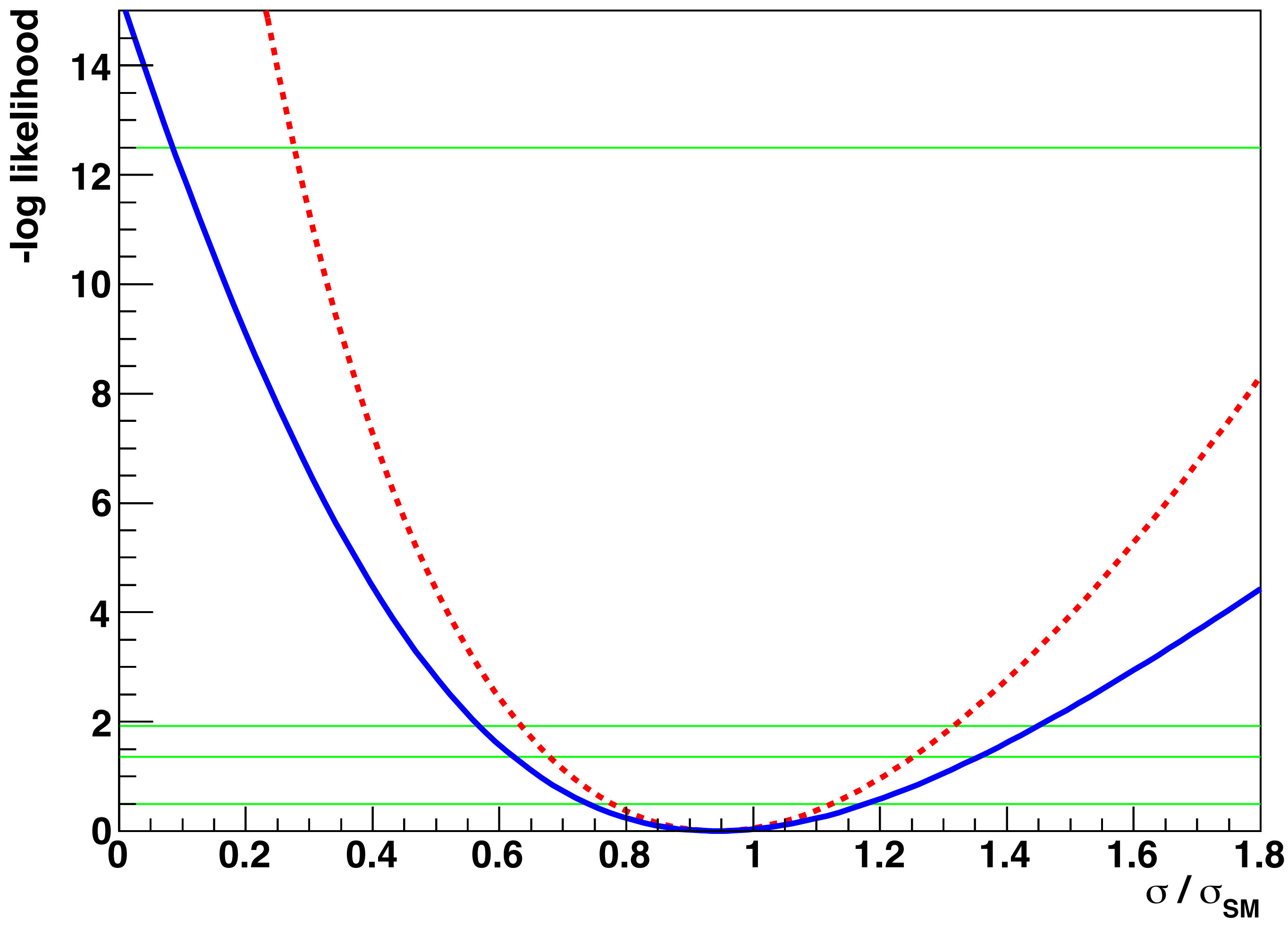
Name	Size	Rev	Age
../			
.settings		28190	4 months
bin		47897	9 hours
config		47961	15 seconds
Combination.xml	8.8 KB	42027	6 weeks
Config.dtd	5.2 KB	47737	32 hours
ee.xml	3.2 KB	42027	6 weeks
ejets.xml	3.5 KB	29524	4 months
ejets_2010.xml	2.6 KB	47961	2 minutes
emu.xml	4.0 KB	42027	6 weeks
example.xml	1.0 KB	47960	68 minutes
example_2.xml	1.1 KB	47960	68 minutes
example_channel2.xml	1.0 KB	47960	68 minutes
example_channel.xml	0.8 KB	47960	68 minutes
FiveChannel.xml	1.0 KB	29524	4 months
mjets_2010.xml	2.7 KB	47961	2 minutes
mujets.xml	3.5 KB	29524	4 months
mumu.xml	3.1 KB	42027	6 weeks
top_ljets_2010.xml	8.9 KB	47961	2 minutes
zee.xml	1.7 KB	42027	6 weeks
zmumu.xml	1.7 KB	42027	6 weeks
data		47960	66 minutes
include		47764	29 hours
lib		28190	4 months
results		28190	4 months
src		47961	15 seconds
tools		29524	4 months
.cproject	15.3 KB	28190	4 months
.project	2.2 KB	28190	4 months
ChangeLog	5.8 KB	47764	29 hours
Makefile	4.2 KB	29525	4 months
README	6.8 KB	42027	6 weeks
setupRoot526_lxplus.sh	150 bytes	28190	4 months

5channel, gamma on WHFRAC

Center for
Cosmology and
Particle Physics

Using a gamma and **80%** uncertainty on WHFRAC >5 sigma

- only if we do combination with dileptons



Toy results from tool

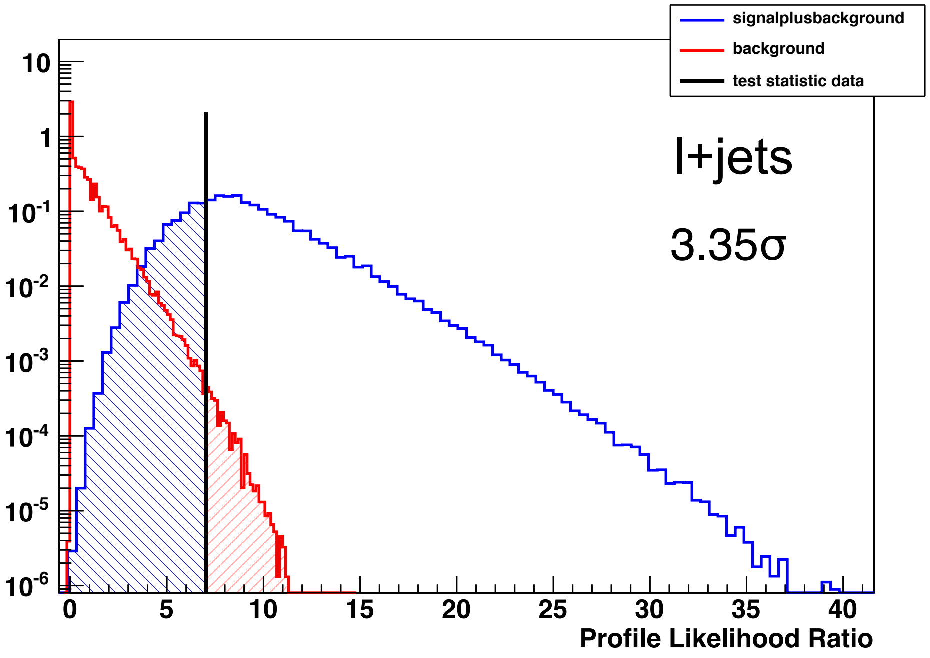
Center for
Cosmology and
Particle Physics

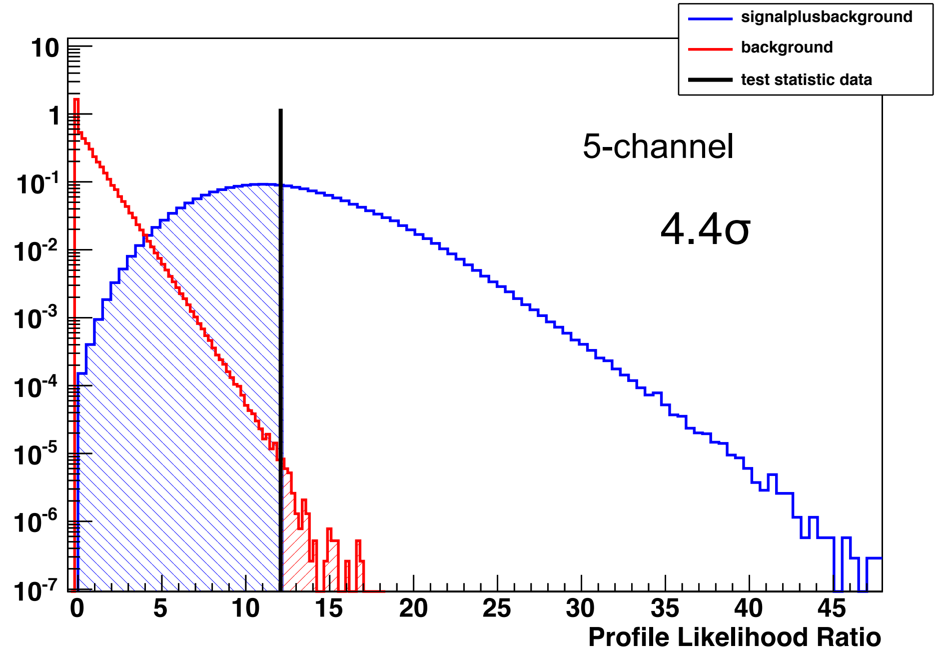
Uses same model produced from combination tool

- New ToyMC Sampler runs on PROOF
- throws millions of toys, does full fit on 50 parameters for each toy
  - (~1day on 30 computers)
- p-values from toys are ~2x higher than expected from profile
  - typically ~0.1 or 0.2σ effect

Now importance sampling is also implemented,

- allows for 1000x speed increase! Still being tested in detail





Conclusions

RooStats tools are in pretty good shape, still developing

- documentation is growing, but need to improve examples

Technically, the approach allows individual groups to model their channels as they wish, but the combination often imposes additional constraints:

- in particular, correlating systematics requires that the individual sources of systematics are explicitly exposed.

Will be useful to have an informal group of expert users to aid in combinations and to avoid inefficient implementation of models

- please contact me if you are interested.
- this group will be doing technical work