Modeling and Statistical Analysis for Higgs Physics at the Large Hadron Collider

Sven Kreiss
with many ideas from Kyle Cranmer

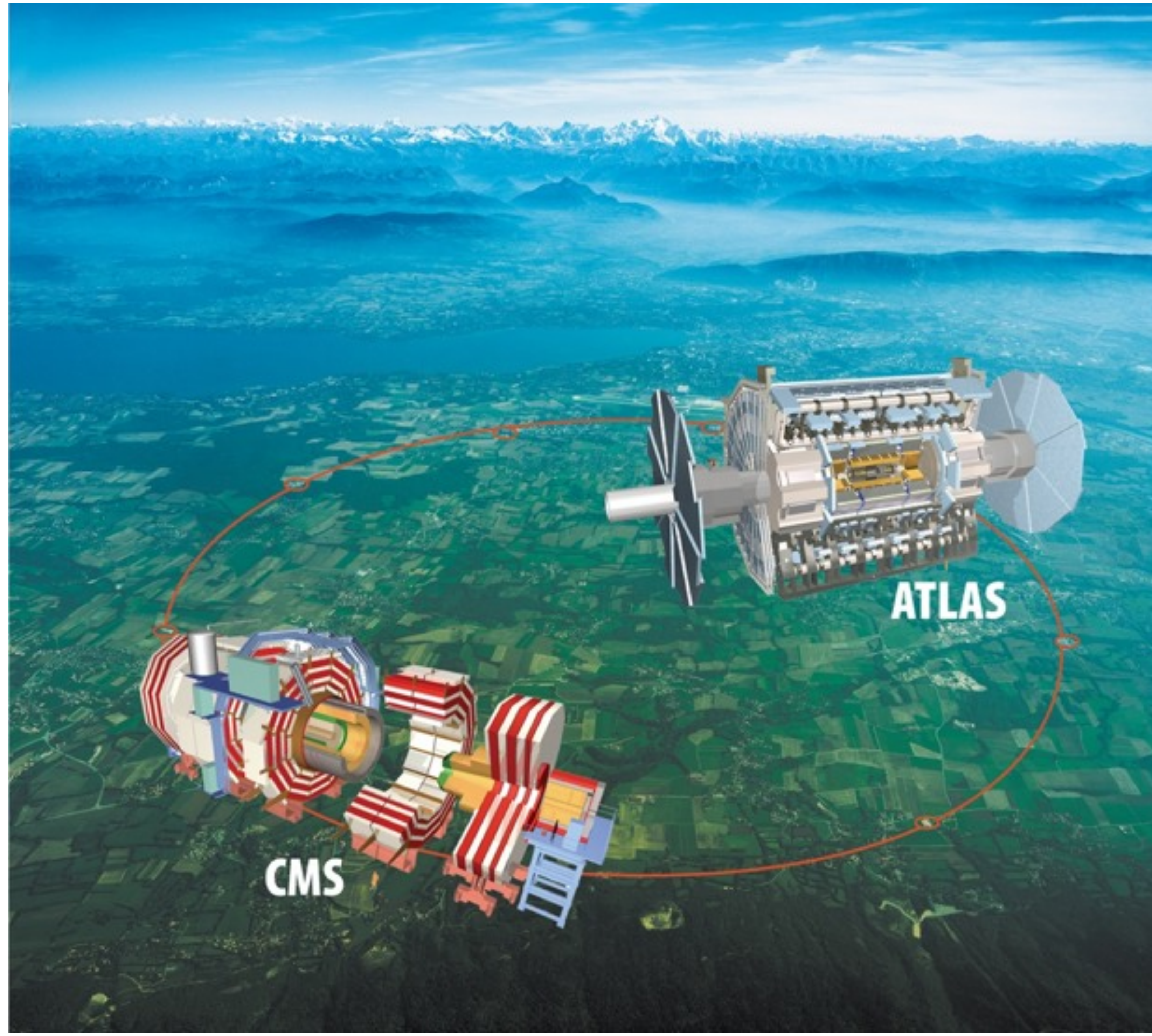
SAMSI-MADAI conference:
Knowledge Extraction via Comparison of Complex Computational Models to Massive Data Sets

29 July 2013

LHC, PARTICLE PHYSICS, HIGGS BOSON

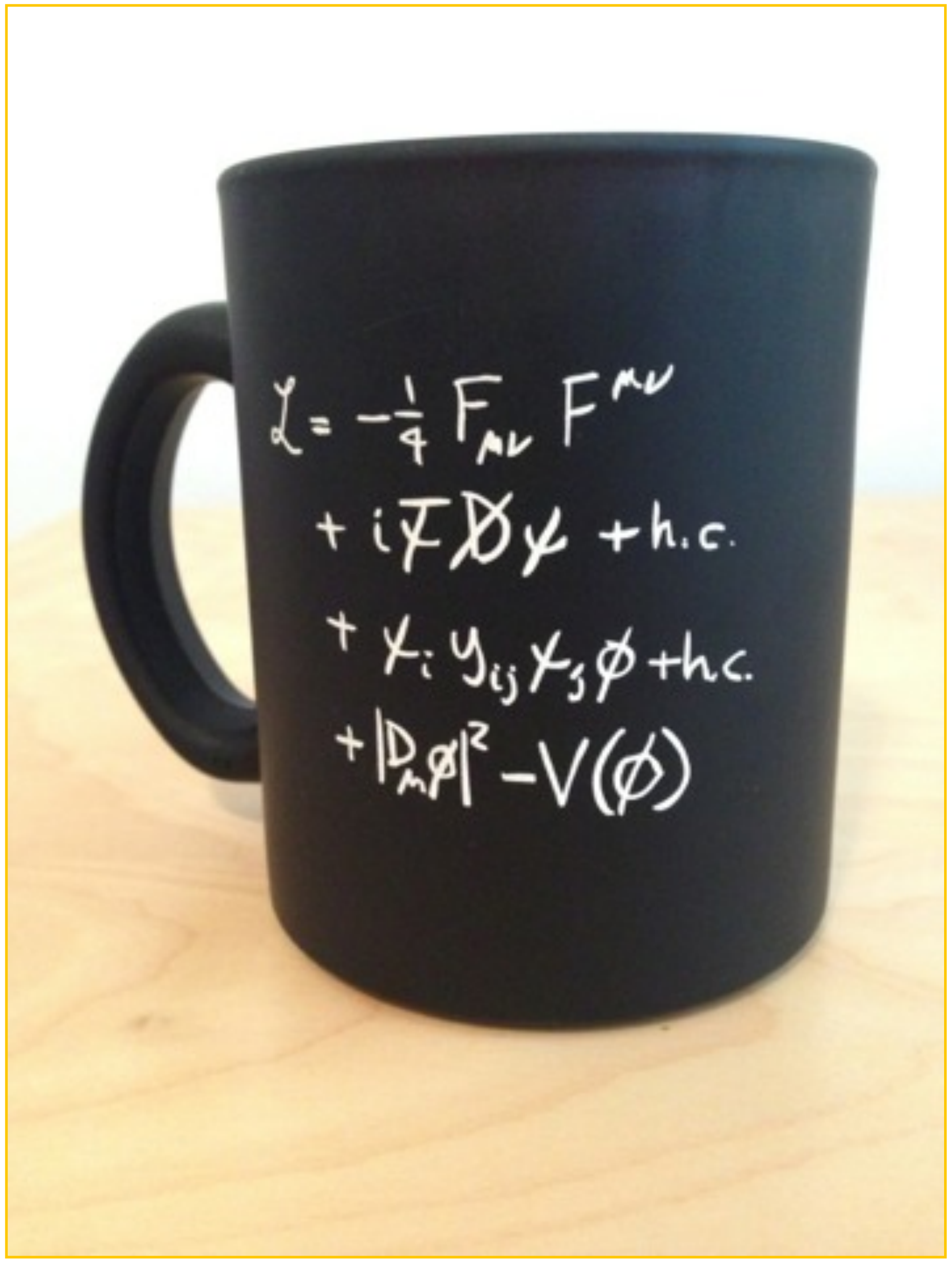
LHC

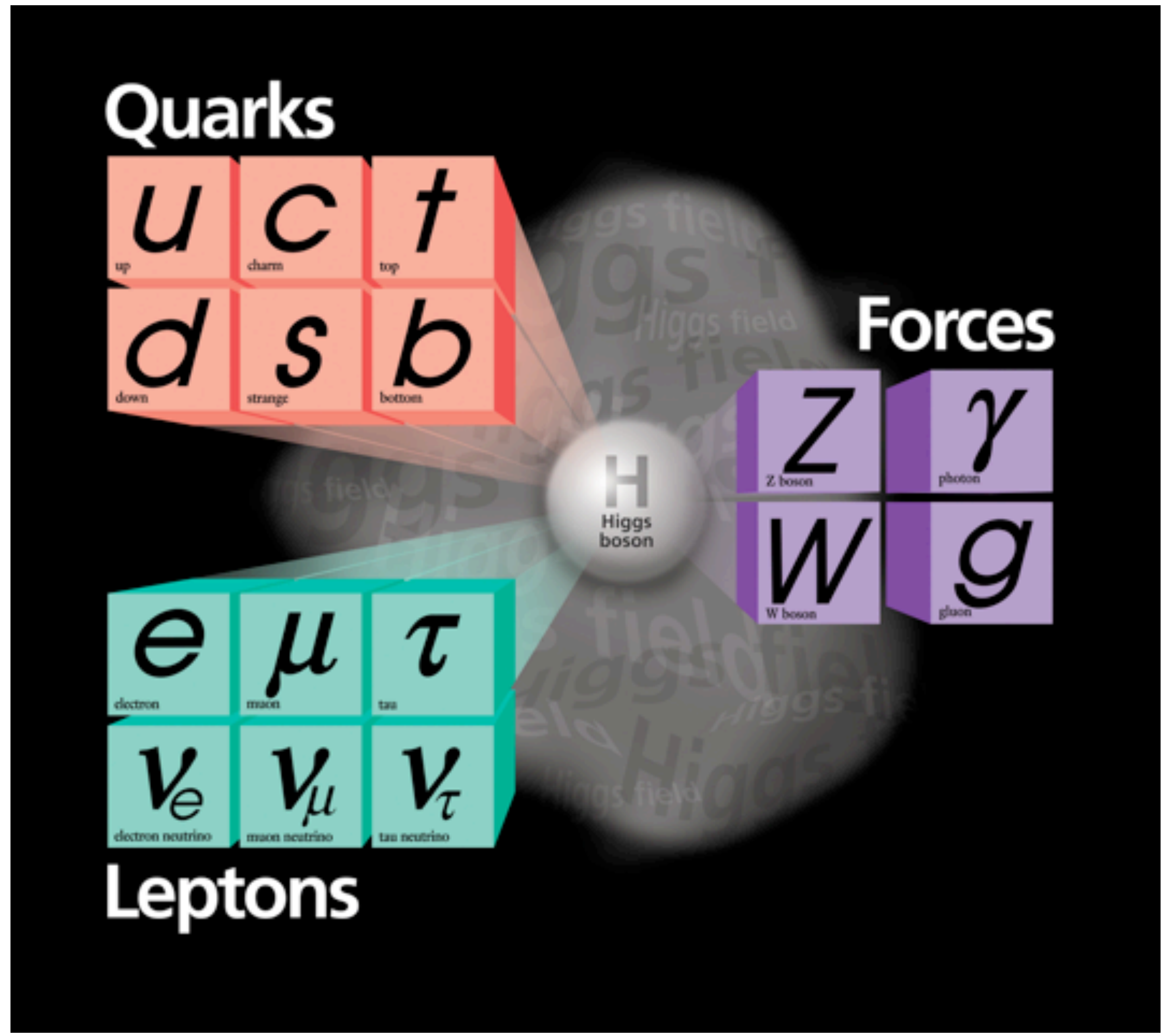
- Geneva, Switzerland
- pp collider with 17 miles circumference
- two multi-purpose detectors: ATLAS, CMS
- Every second, 20million collisions happen at each experiment.
- At ATLAS, 100million detector cells are read out and the 20million events are filtered in real-time down to 400 events per second.



Particle Physics and the Higgs Boson

1964: Mechanism for electroweak symmetry breaking proposed. Higgs boson is manifestation of the underlying scalar field. It was later used to build “The Standard Model of Particle Physics”. Every particle that this model predicted was found: e.g. W, Z and top.



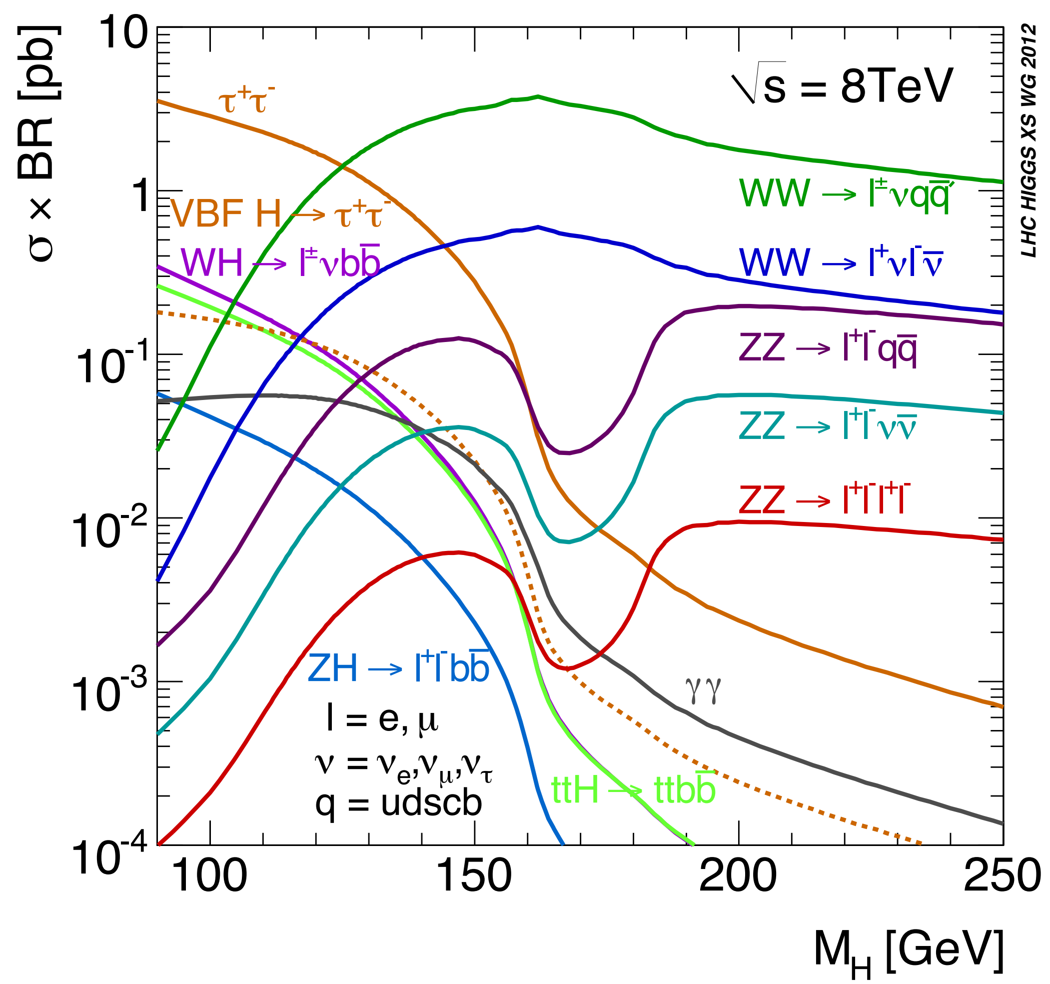


The Higgs Boson and the Need For a Combination

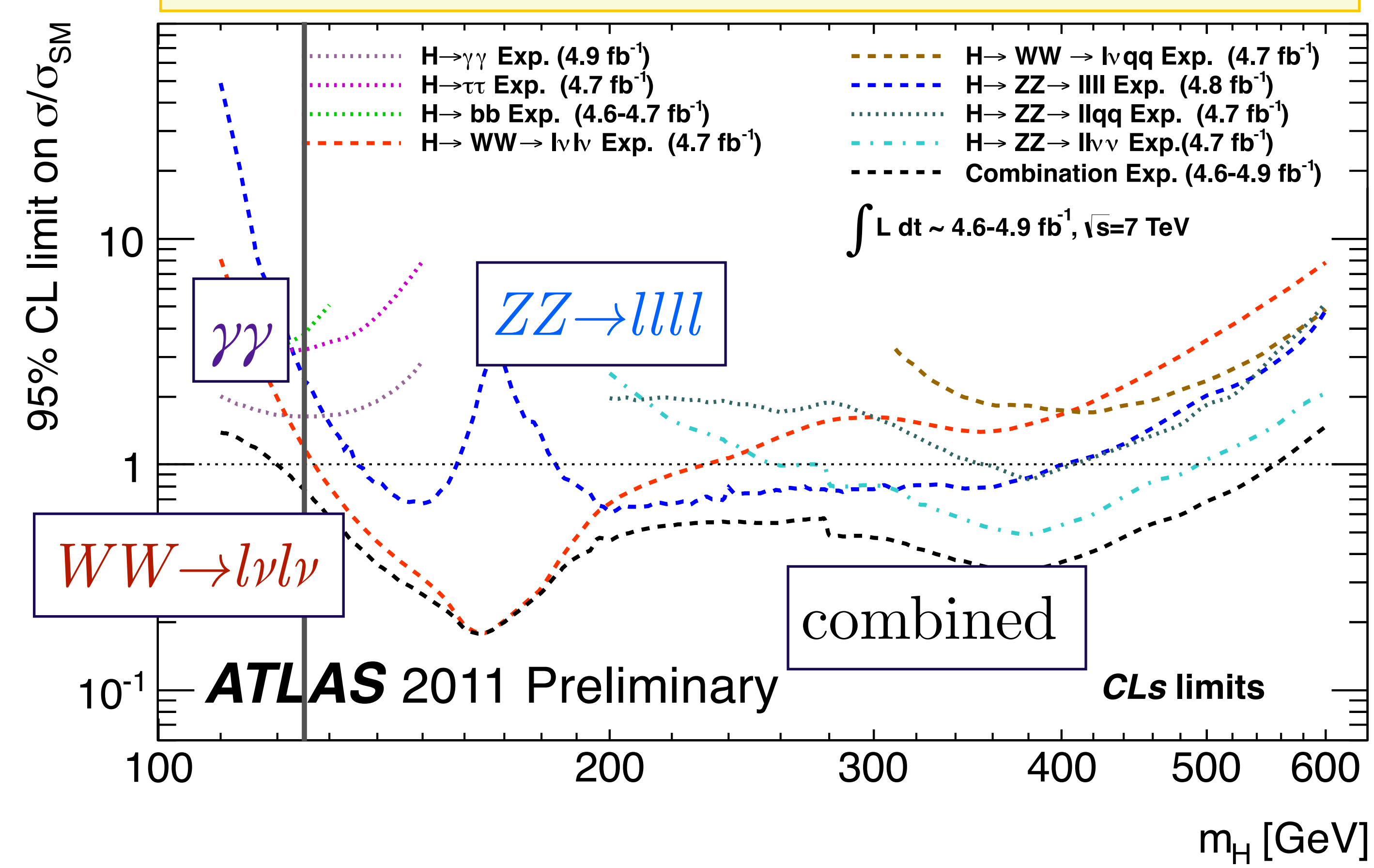
Understanding electroweak symmetry breaking is one of the primary objectives of the LHC. A Higgs boson was recently discovered in the search for the Standard Model Higgs boson. Now attention turns to studying its properties.

There are various options for production and decay of the Higgs. Their importance depends on m_H.
For the currently interesting region, the production through gluon-gluon-Fusion and Vector-Boson-Fusion and the decays to $\gamma\gamma$, $ZZ \rightarrow llll$ and $WW \rightarrow l\nu l\nu$ contribute the most.

Backgrounds (processes that are already understood) are huge. After many months of continuous running both in 2011 and 2012, we found ~60 candidate Higgs events that decayed to four leptons.



Expected upper limits are related to the “weight” of each channel in the combination: [ATLAS-CONF-2012-019]



Inputs

Combination of 15 channels of which 12 are used for the test of a Higgs with m_H = 126 GeV. To enhance sensitivity, a further division into ~100 sub-channels is done.

Table 6: Summary of the individual channels entering the combination. The transition points between separately optimised m_H regions are indicated where applicable. In channels sensitive to associated production of the Higgs boson, V indicates a W or Z boson. The symbols $\otimes$ and $\oplus$ represent direct products and sums over sets of selection requirements, respectively.

	Higgs Boson Decay	Subsequent Decay	Sub-Channels	m_H Range [GeV]	$\int$ L dt [fb^{-1}]	Ref.
			2011 $\sqrt{s}$ =7 TeV			
2011	$H \to ZZ^{(*)}$	4ℓ	$\{4e, 2e2\mu, 2\mu2e, 4\mu\}$	110–600	4.8	[87]
		$\ell\ell\nu\nu$	$\{ee, \mu\mu\} \otimes$ {low, high pile-up}	200–280–600	4.7	[125]
		$\ell\ell qq$	{b-tagged, untagged}	200–300–600	4.7	[126]
	$H \to \gamma\gamma$	–	10 categories $\{p_{\mathrm{Tt}} \otimes \eta_\gamma \otimes$ conversion$\} \oplus$ {2-jet}	110–150	4.8	[127]
	$H \to WW^{(*)}$	$\ell\nu\ell\nu$	$\{ee, e\mu/\mu e, \mu\mu\} \otimes$ {0-jet, 1-jet, 2-jet} $\otimes$ {low, high pile-up}	110–200–300–600	4.7	[106]
		$\ell\nu qq'$	$\{e, \mu\} \otimes$ {0-jet, 1-jet, 2-jet}	300–600	4.7	[128]
	$H \to \tau\tau$	$\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$	$\{e\mu\} \otimes$ {0-jet} $\oplus \{\ell\ell\} \otimes$ {1-jet, 2-jet, VH}	110–150	4.7	[129]
		$\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$	$\{e, \mu\} \otimes$ {0-jet} $\otimes \{E_{\mathrm{T}}^{\mathrm{miss}} < 20$ GeV, $E_{\mathrm{T}}^{\mathrm{miss}} \geq 20$ GeV$\}$ $\oplus \{e, \mu\} \otimes$ {1-jet} $\oplus \{\ell\} \otimes$ {2-jet}	110–150	4.7	
		$\tau_{\mathrm{had}}\tau_{\mathrm{had}}$	{1-jet}	110–150	4.7	
	$VH \to Vbb$	$Z \to \nu\nu$	$E_{\mathrm{T}}^{\mathrm{miss}} \in \{120-160, 160-200, \geq 200$ GeV$\}$	110–130	4.6	[130]
		$W \to \ell\nu$	$p_{\mathrm{T}}^{W} \in \{< 50, 50-100, 100-200, \geq 200$ GeV$\}$	110–130	4.7	
		$Z \to \ell\ell$	$p_{\mathrm{T}}^{Z} \in \{< 50, 50-100, 100-200, \geq 200$ GeV$\}$	110–130	4.7	
			2012 $\sqrt{s}$ =8 TeV			
2012	$H \to ZZ^{(*)}$	4ℓ	$\{4e, 2e2\mu, 2\mu2e, 4\mu\}$	110–600	5.8	[87]
	$H \to \gamma\gamma$	–	10 categories $\{p_{\mathrm{Tt}} \otimes \eta_\gamma \otimes$ conversion$\} \oplus$ {2-jet}	110–150	5.9	[127]
	$H \to WW^{(*)}$	$e\nu\mu\nu$	$\{e\mu, \mu e\} \otimes$ {0-jet, 1-jet, 2-jet}	110–200	5.8	[131]

ONE OF THE HIGGS SEARCHES

Counting events.

Underlying physical process is statistical and not deterministic: $P(i \to f) = \frac{|\langle f|i\rangle|^2}{\langle f|f\rangle \langle i|i\rangle}$

But we can only measure the event rate:

$$\textit{rate} = \textit{flux} \cdot \textit{cross section} \cdot \textit{efficiency} \cdot \textit{acceptance}$$

flux	cross section	efficiency	acceptance
provided by LHC	physics	detector performance	analysis performance

LHC Model Building Introduction

All analyses are counting events. The distribution and number of events is taken into account in the “marked Poisson model”:

$$\mathcal{P}(\{x_1 \ldots x_n\}|\nu) = \underbrace{\mathrm{Pois}(n|\nu)}_{\text{Probability of observing } n \text{ events for } \nu \text{ expected}} \prod_{e=1}^{n} f(x_e)$$

In the context of the search for a signal S with strength μ on top of a background B:

$$\mathcal{P}(\{x_1 \ldots x_n\}|\mu) = \mathrm{Pois}(n|\mu S + B) \left[\underbrace{\prod_{e=1}^{n} \frac{\mu S f_{\mathrm{S}}(x_e) + B f_{\mathrm{B}}(x_e)}{\mu S + B}}_{\text{weighted sum of signal and background PDFs evaluated at all observed events}} \right]$$

- ➡ $f(x)$, $f_S(x)$ and $f_B(x)$ are probability density functions (PDFs).
- ➡ For most of our models, the PDFs and data are provided in binned form.

Template Model

Most Higgs searches are based on Histogram templates.

Still the same Lagrangian from the coffee cup.

Complicated event generators.

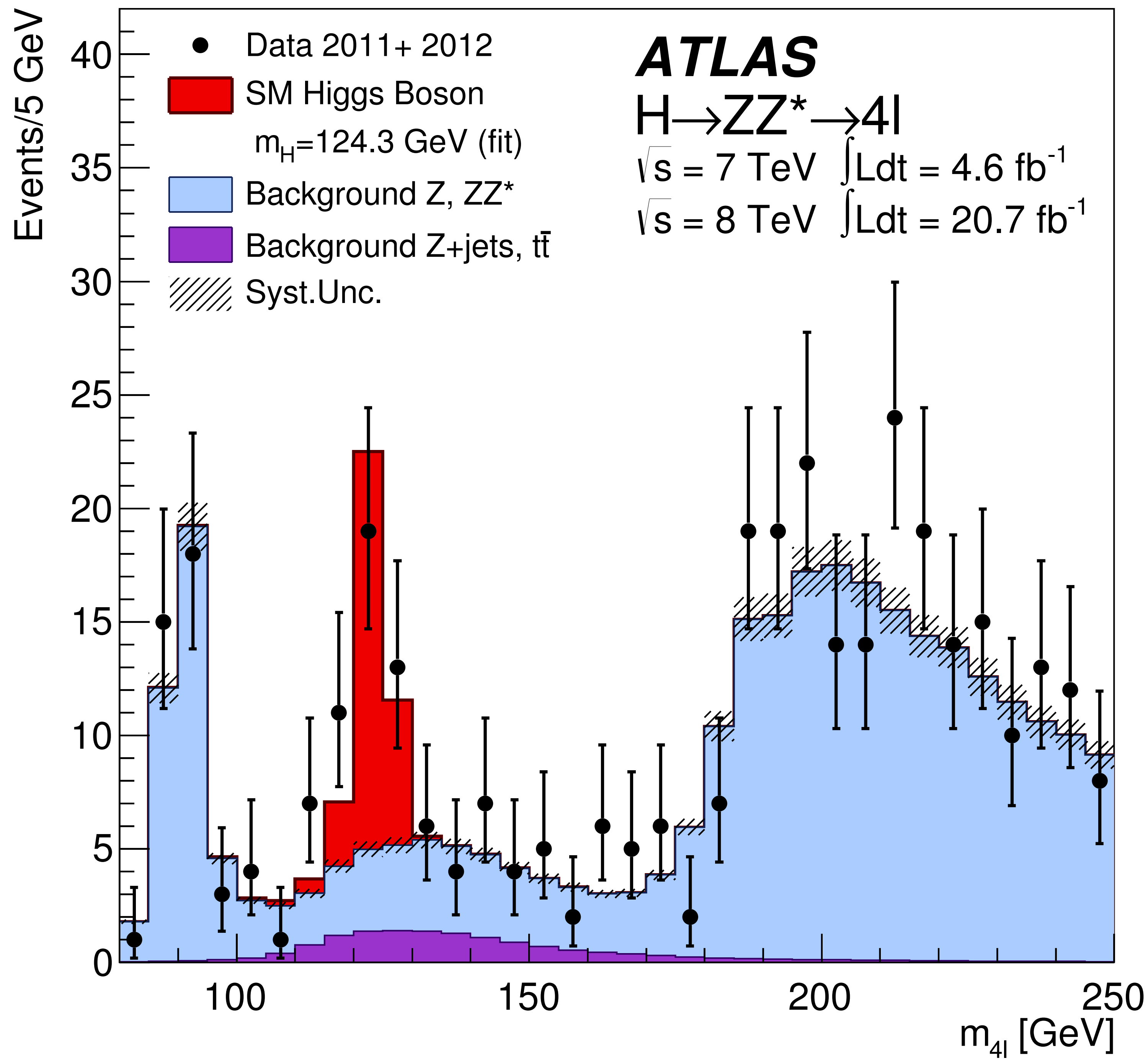
Expensive simulation of the detector (~20min per event).

Allows testing the hypothesis without the red signal histogram (μ=0) against the hypothesis with the red signal histogram ($\mu=\hat{\mu}$).

Generative model.

Can measure μ.

Systematic effects in yield and shape are included (more later).



Measuring the Higgs Mass

The mass resolution in the H→ZZ*→4l is too good for histogram based signal templates. Need a statistical model independent of bin boundaries. It is difficult to converge on an ad-hoc analytic function.

Background shapes are reasonably smooth and can be further smoothened by bin-to-bin linear interpolation or kernel density estimates.

Signal shapes are generated at fixed m_H hypotheses.

toy model, no ATLAS data

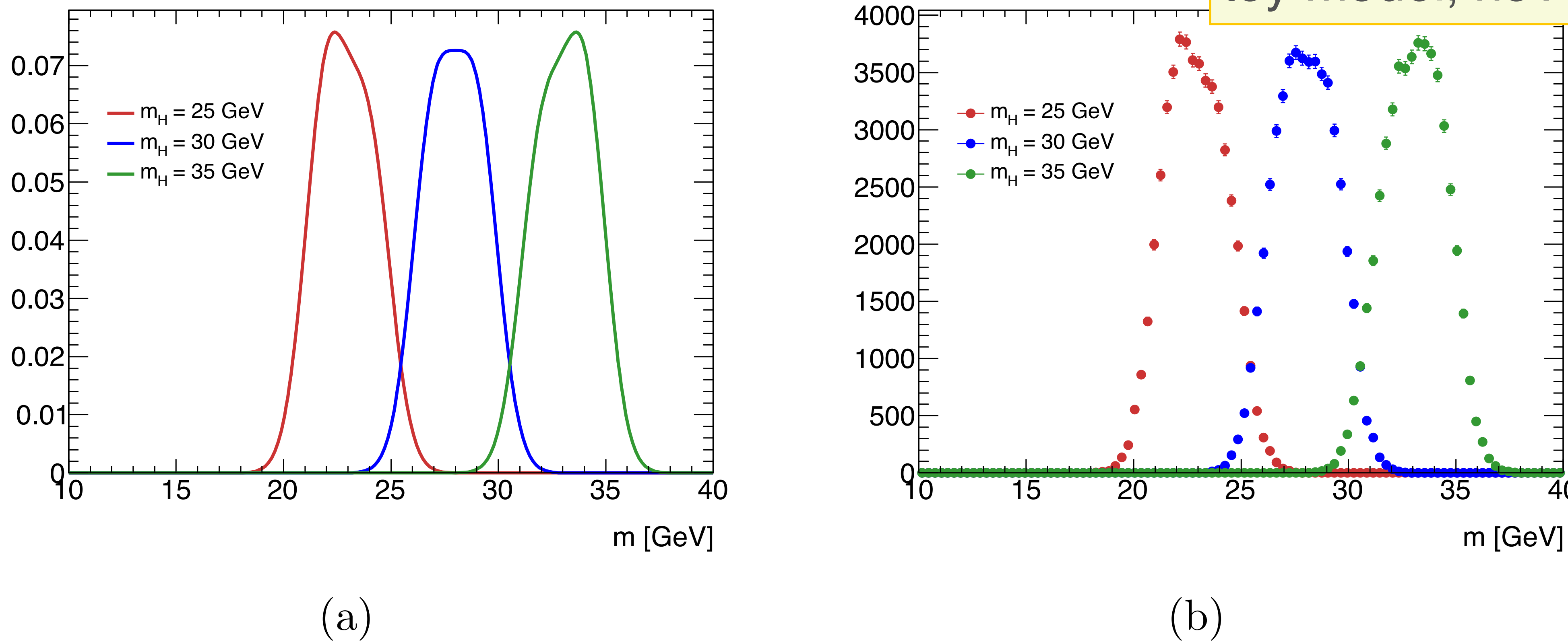
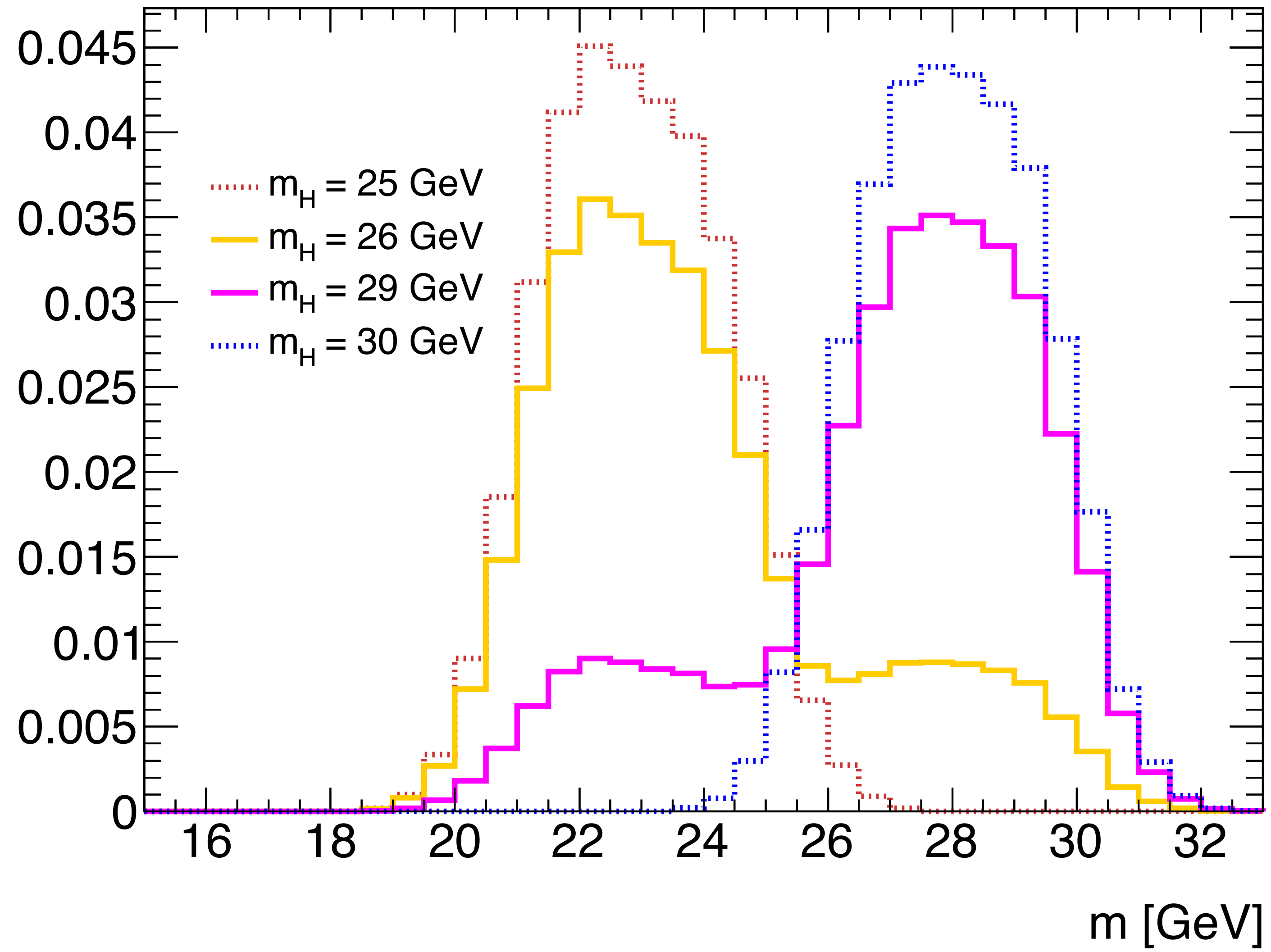


Figure 1: Illustration of the example model that is used in this note. There are three MC samples available with $m_H = 25$, 30 and 35 GeV. All three samples have a slightly different shape. (a) True model. (b) Generated MC samples from the true model.

Vertical Interpolation: Not an Option for a High Resolution Channel

From physics, want: Shapes move left and right when changing m_H.

Vertical Interpolation: Bin heights move up and down creating two peaks at the two initial positions. Those are unphysical shapes.



There are other interpolation algorithms, which can in principle be built directly into the model.

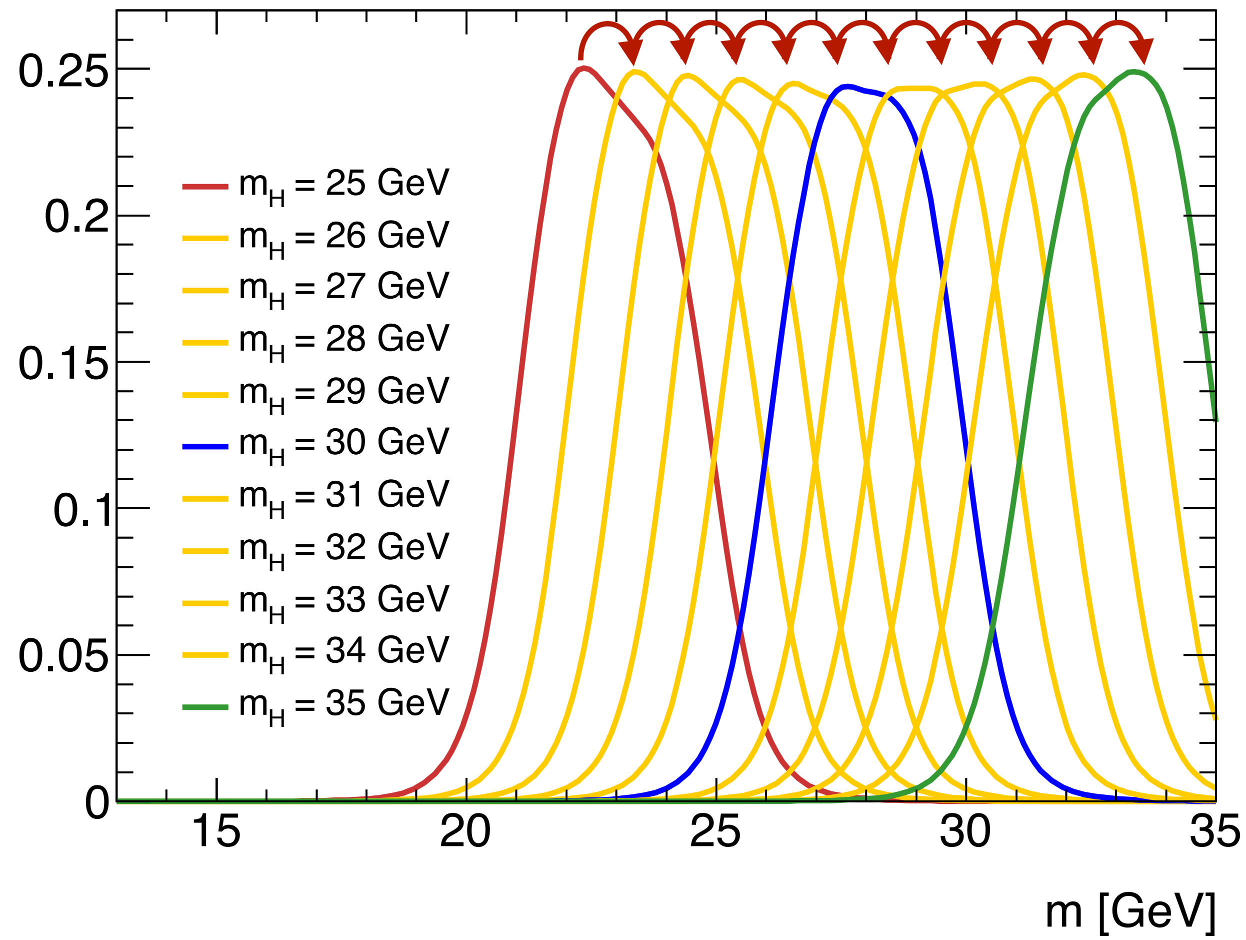
Combining Kernel Density Estimates and B-splines

Adaptive width Kernel Density Estimates: KEYS [Cranmer, 2001]

With KEYS and without the discreteness of bin boundaries, can make use of:

$$f(m_{\text{inv}}|m_H + \Delta) \approx f(m_{\text{inv}} + \Delta|m_H)$$

and shift shapes continuously.



Using KEYS PDFs and B-spline weights:

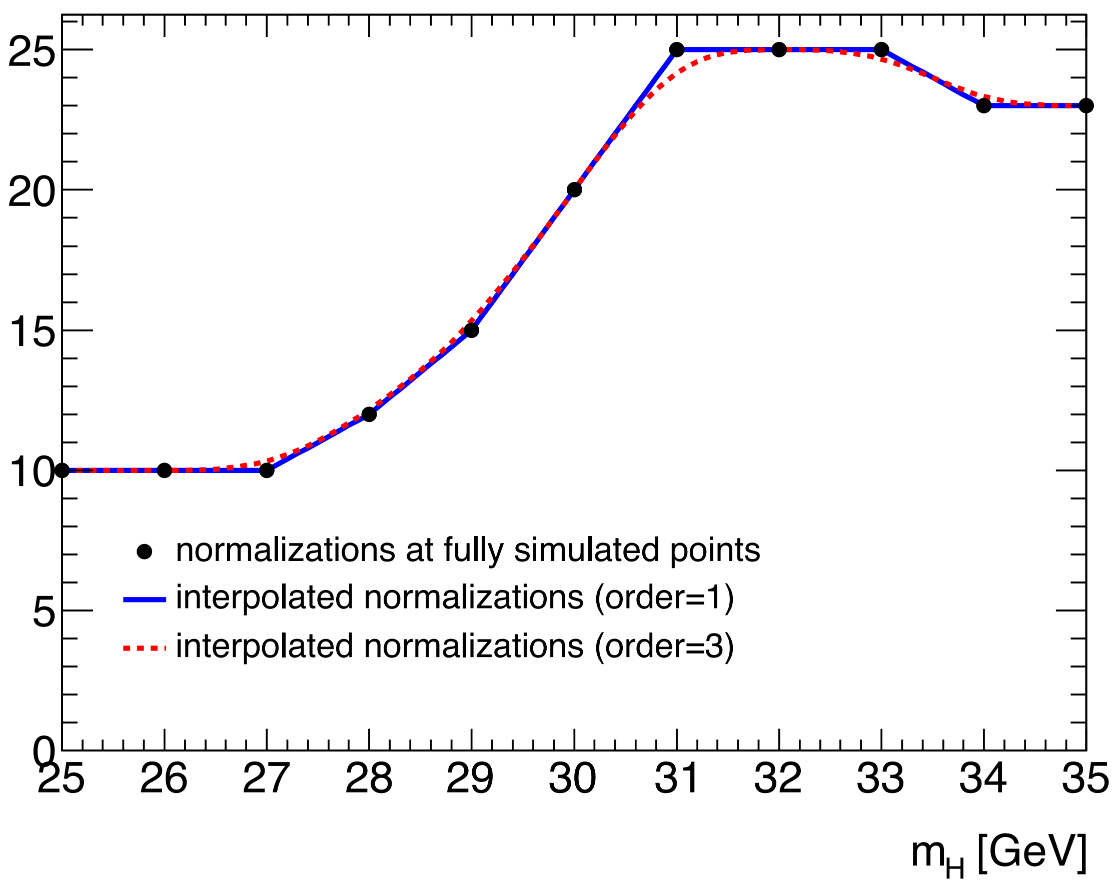
To obtain the signal shape at m_H, all f_j are first shifted by $\Delta m_{\text{inv},j} = m_H - m_j$ and then interpolated according to

$$f_{\text{total}}(m_{\text{inv}} \mid m_H) = \sum_j w_j(m_H)\; f_j(m_{\text{inv}} \mid m_H - m_j)$$

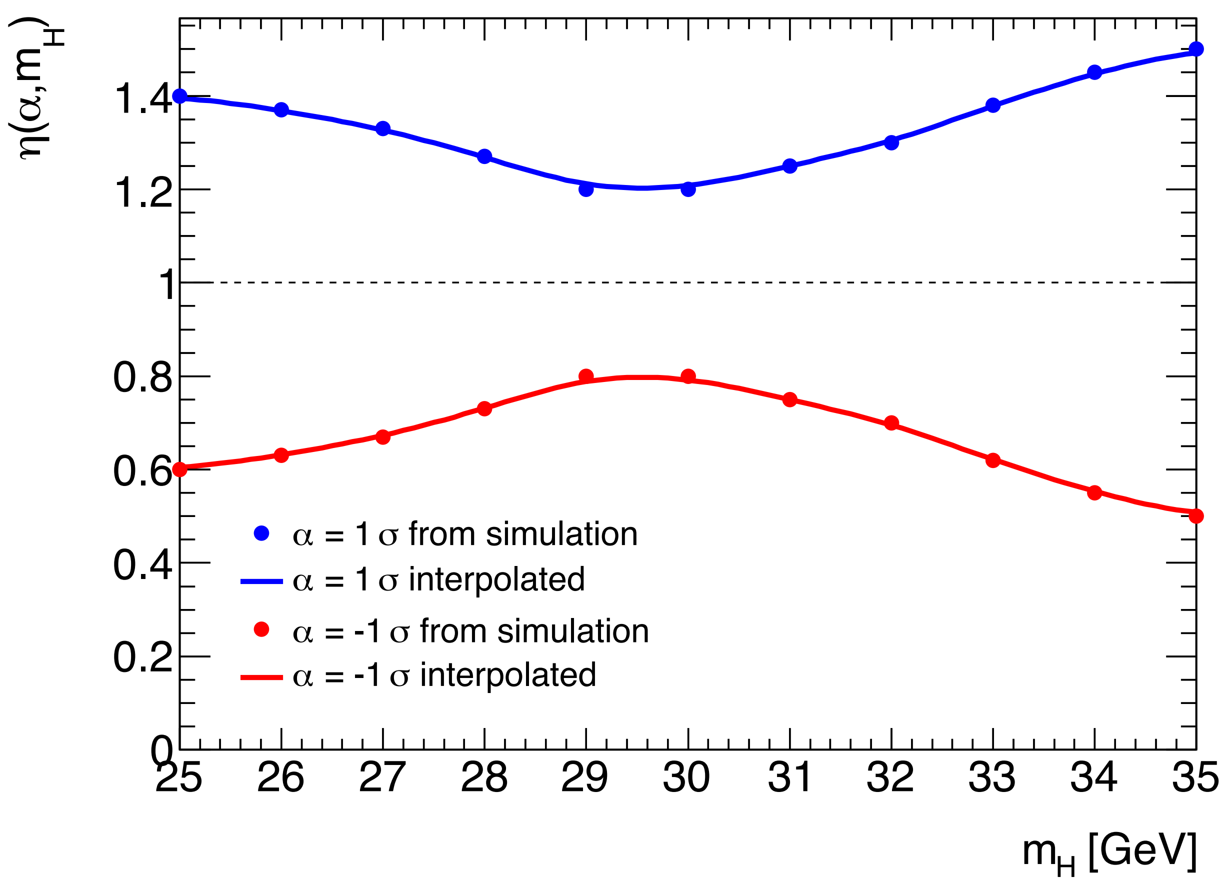
where the coefficients $w_j(m_H)$ are B-spline basis functions.

Continuous Parametrizations for Normalization and Systematics

B-splines interpolate signal normalizations in m_H.

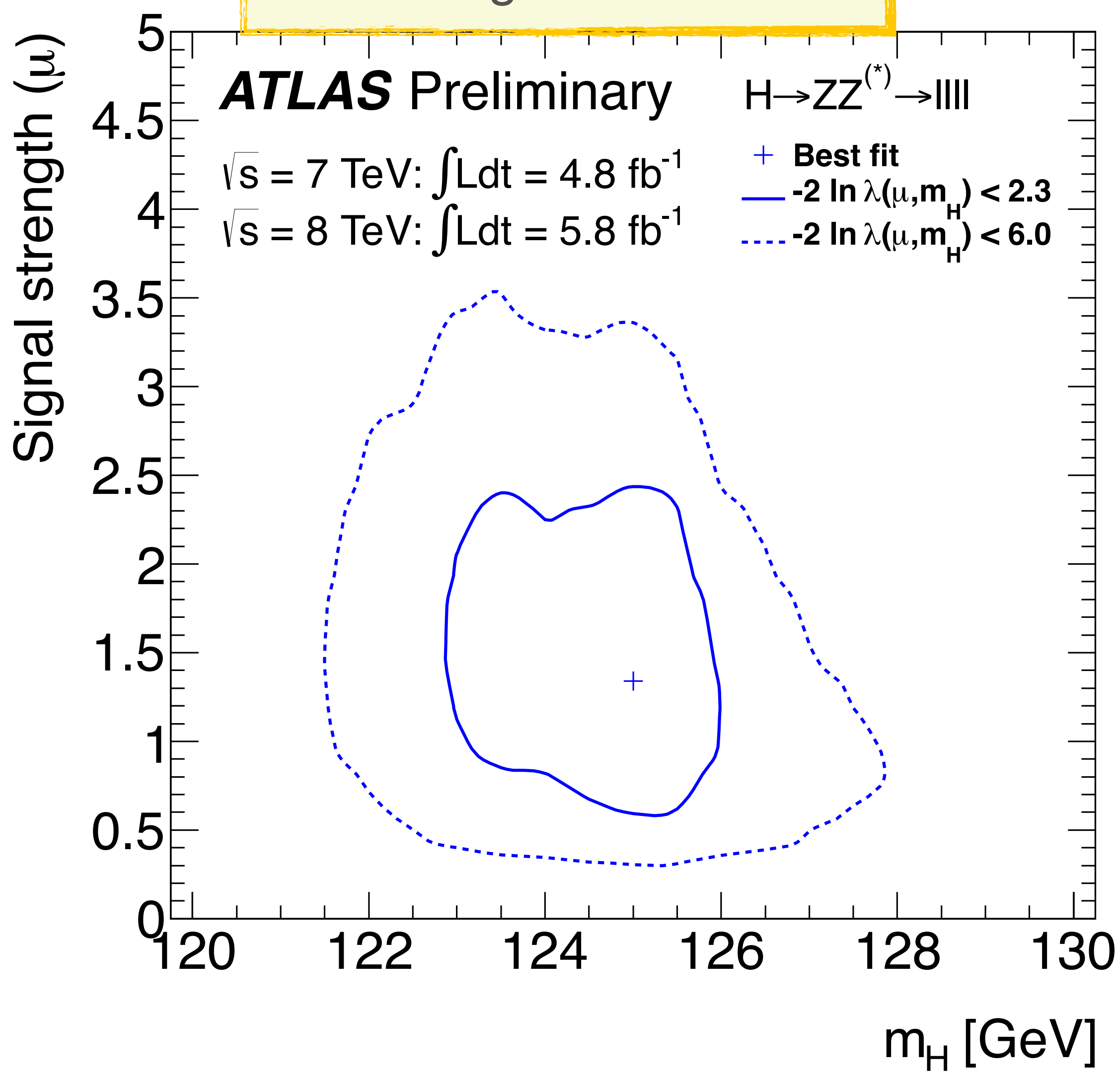


One "dynamic" B-spline interpolates response function for systematics in m_H.

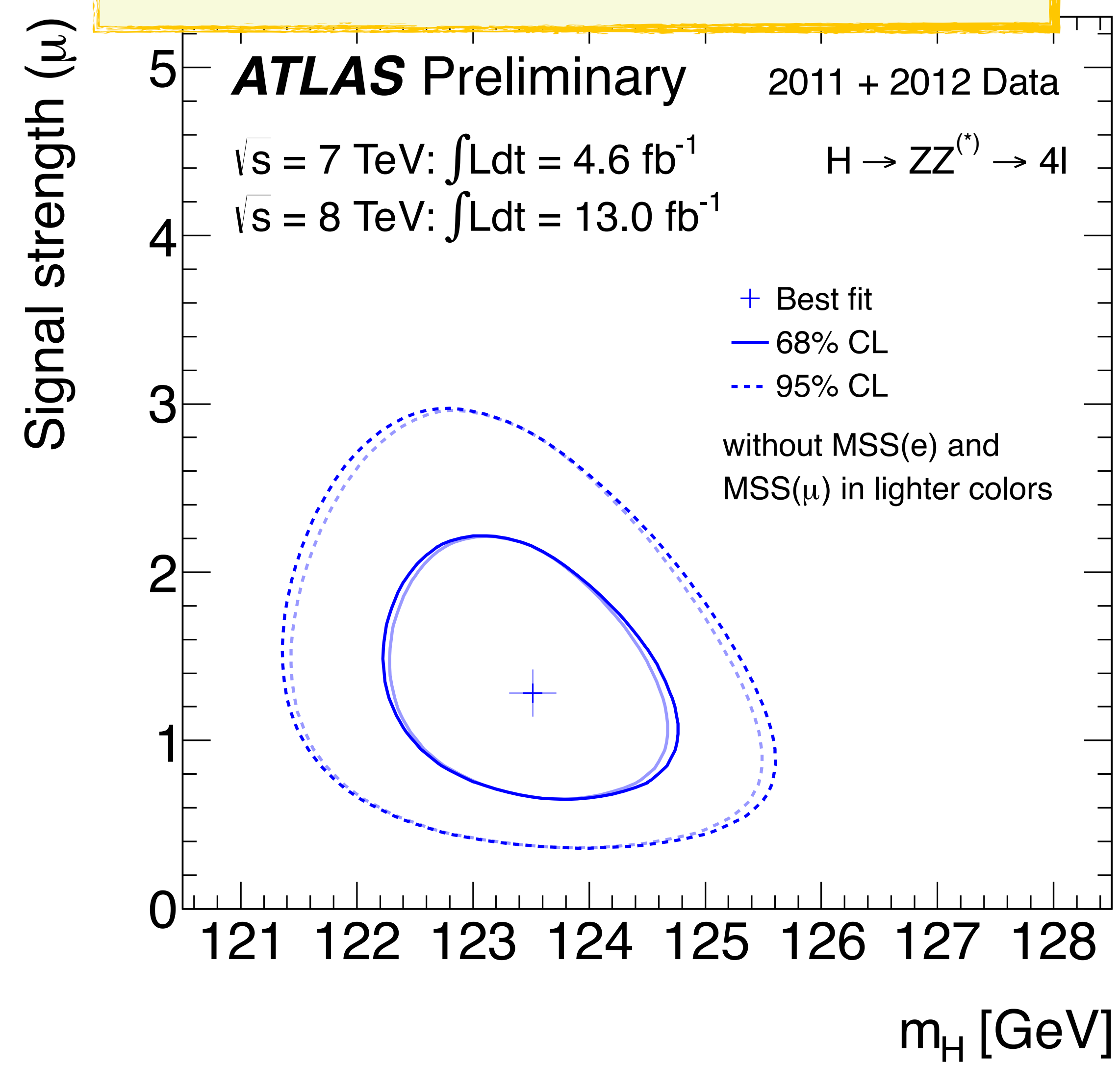


One of the Results

Histogram based



KEYS and B-splines, more data



COLLABORATIVE STATISTICAL MODELING

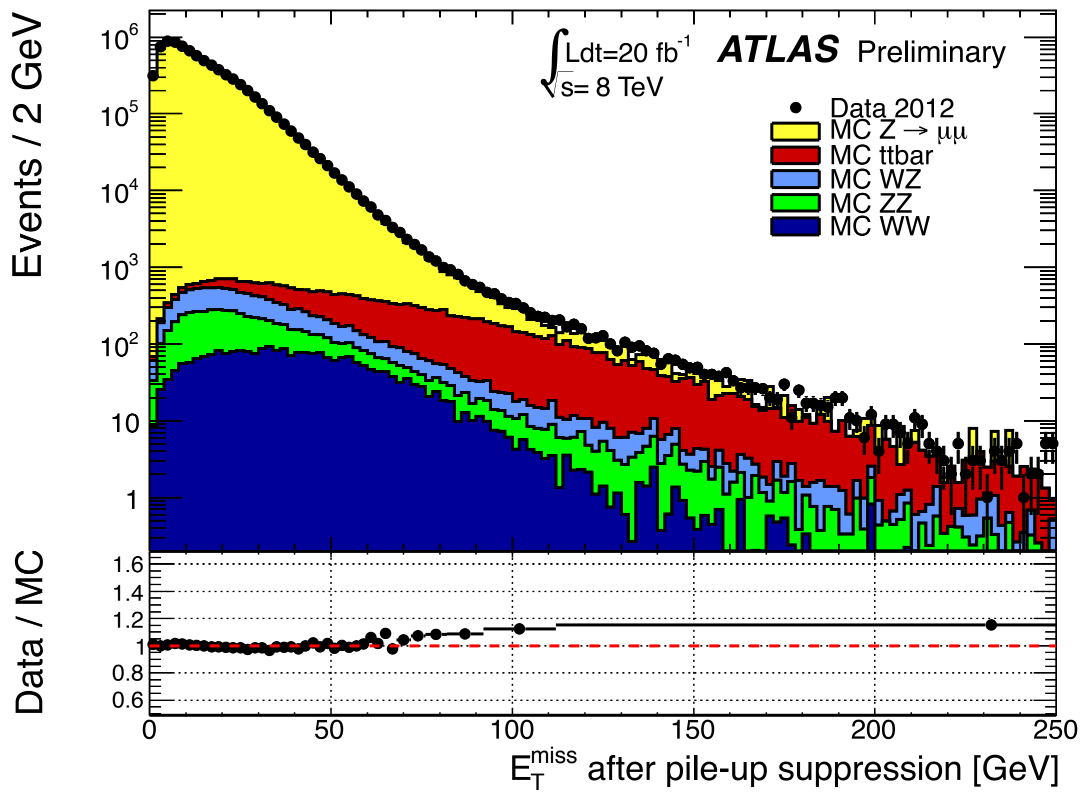
ATLAS Overview: ~3000 People

ATLAS celebrating the discovery

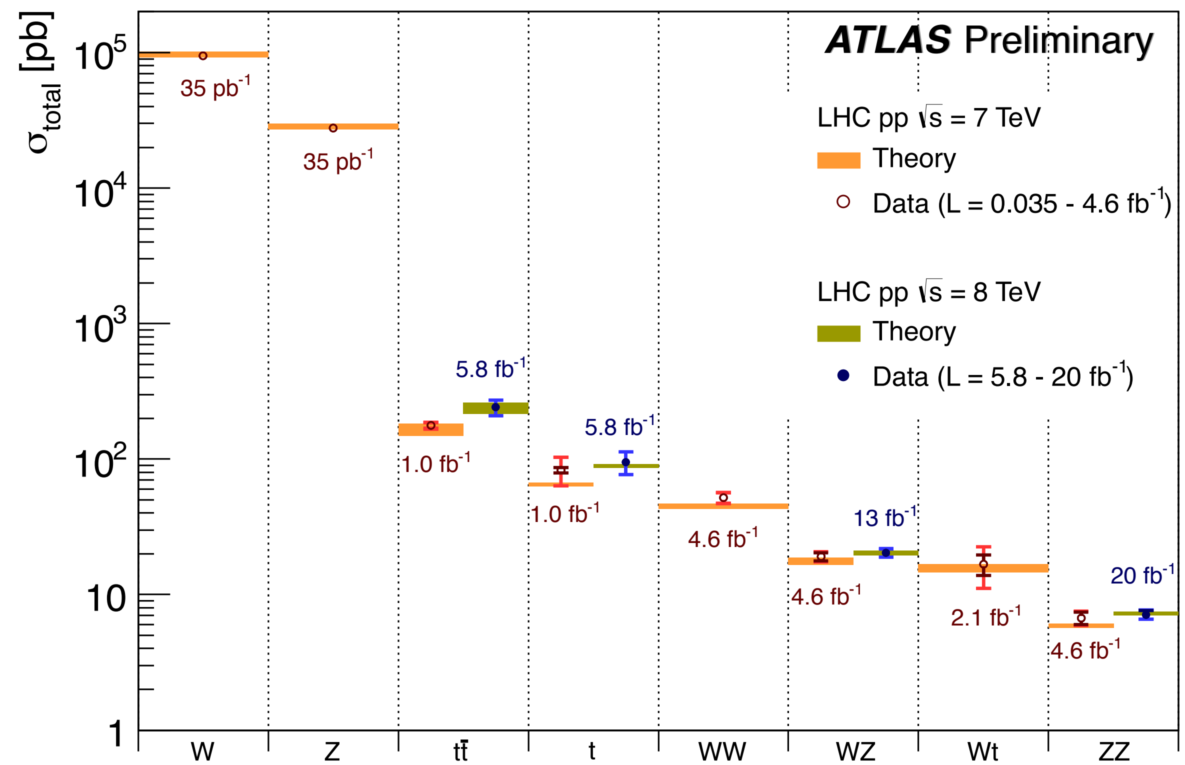
Performance groups: e/gamma, Flavor tagging, Jet/EtMiss, Tau, Muon, Tracking, Simulation
Trigger and Luminosity groups
Physics groups: SM, B, Top, Higgs, SUSY, Exotics, Heavy Ion, Monte Carlo

Likelihoods from Domain Experts

We have experts for counting many things to very high precision.



The sum of all simulations (colored histograms) agrees very well with the observed data (black dots) across a few orders of magnitude (notice log-scale).

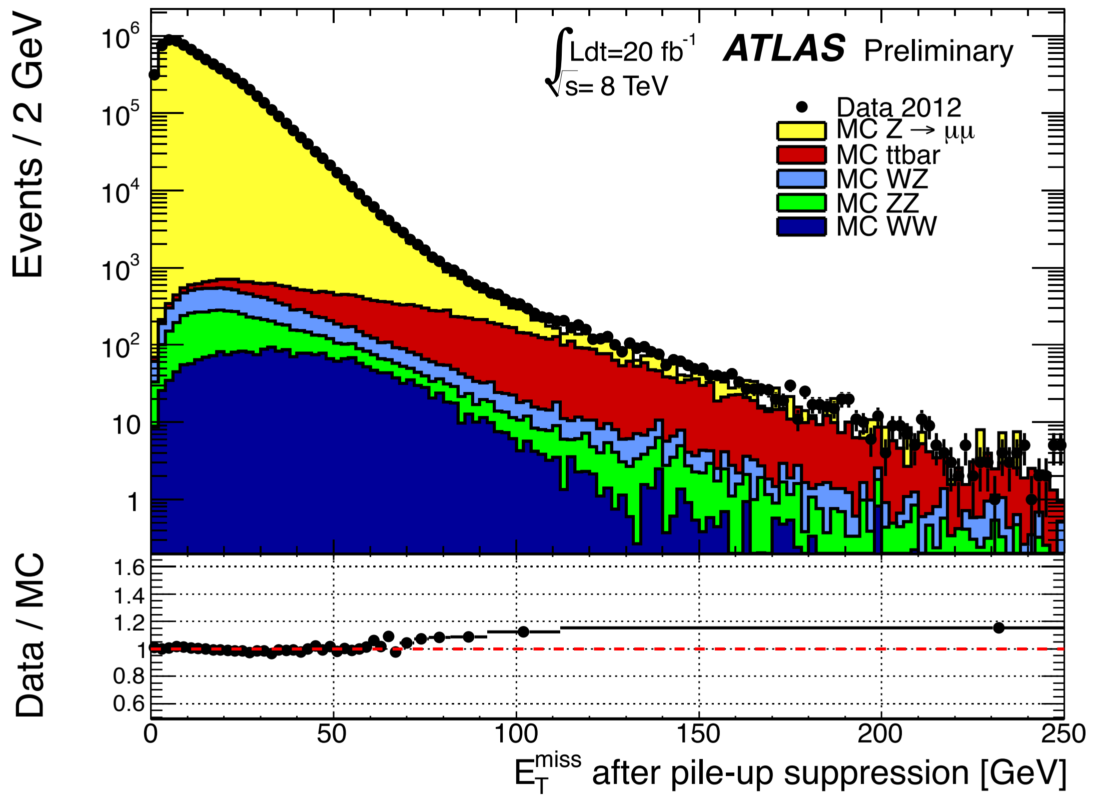


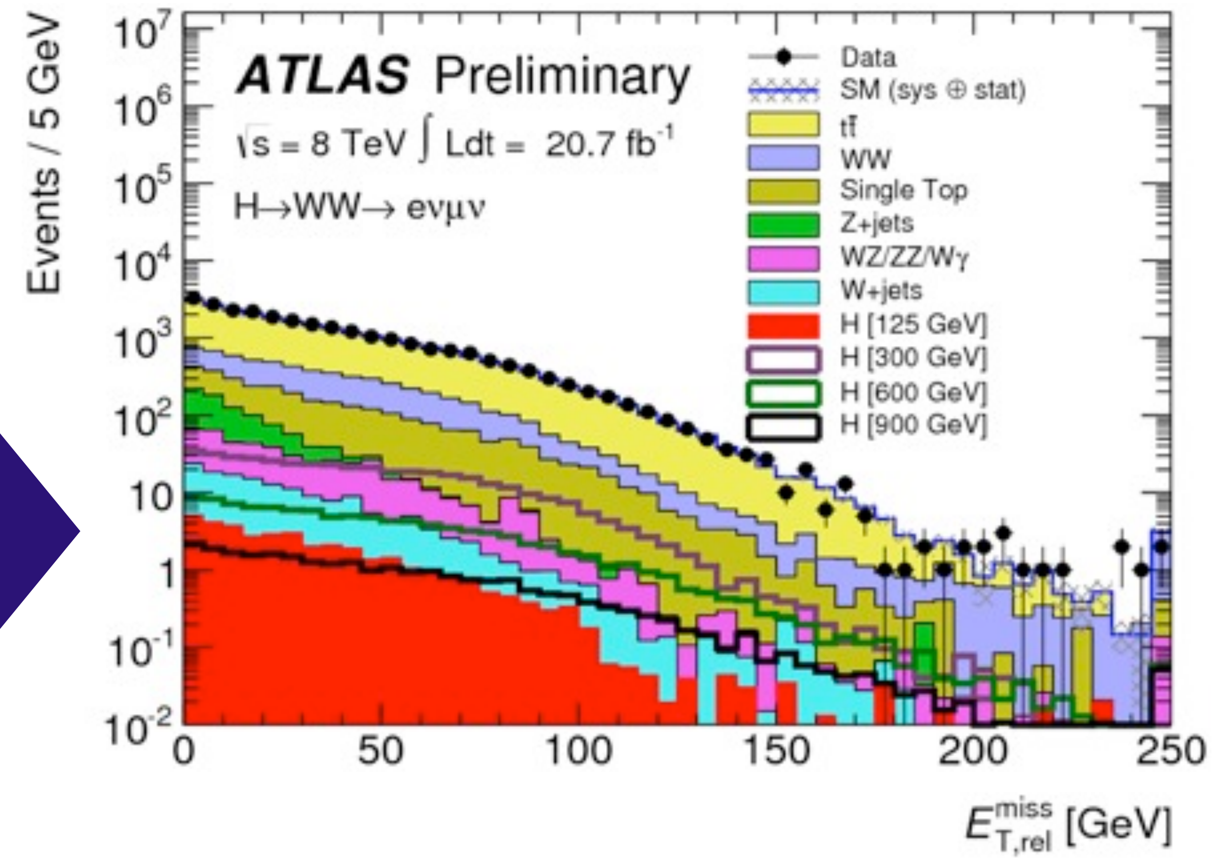
Many Standard Model groups perform precise measurements of the known Standard Model processes which are then shared and used by other Physics groups.

Need to share full Likelihoods: serialize, write to disk and store.

Statistically Independent Measurements with the same Instrument

The Higgs group uses results and their uncertainties from performance groups.





missing energy in performance group

model using missing energy in Higgs group

Correlated Systematics: both measurements involve for example the muon momentum as measured by the muon spectrometer. The same physics object (here, a muon) has the same systematic in the momentum measurement irrespective of whether the process that is studied is $Z \rightarrow \mu\mu$ or $H \rightarrow WW^* \rightarrow e\nu\mu\nu$.

The on/off Problem

(arxiv:0702156v4, HybridInstructional.C and references)

From the paper by Cousins, Linneman, Tucker:

This HEP prototype problem has an exact analog in gamma ray astronomy (GRA), upon which we base our notational subscripts "on" and "off". The observation of n_{on} photons when a telescope is pointing at a potential source ("on-source") includes both background and the source, while the observation of n_{off} photons with the telescope pointing at a source-free direction nearby ("off-source") is the subsidiary measurement. In both the HEP and GRA examples, we let the parameter τ denote the ratio of the expected means of

Often used for counting experiment:

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

$\pi(b)$ is a Bayesian prior obtained from the n_{off} measurement. It cannot be used in pure Frequentist methods.

Alternatively, one can introduce a sideband measurement (which is also how knowledge about b was obtained in reality):

$$\underbrace{P(n_{\text{on}}, n_{\text{off}}|s, b)}_{\text{joint model}} = \underbrace{\text{Pois}(n_{\text{on}}|s+b)}_{\text{main measurement}} \underbrace{\text{Pois}(n_{\text{off}}|\tau b)}_{\text{sideband}}$$

Using the Same Telescope: Correlating Systematics in the on/off Problem

Extending the previous idea for *b*, other nuisance parameters can be incorporated in the same way:

$$P(n_{\text{on}}, n_{\text{off}}, \text{obs}_\theta | s, b, \theta) = \text{Pois}(n_{\text{on}} | s + b + \eta_1(\theta)) \; \text{Pois}(n_{\text{off}} | \tau b + \eta_2(\theta)) \; \text{Gaus}(\text{obs}_\theta | \theta, 1)$$

"global observable"

response functions η

constraint

Example: electric noise is reducing the photon count. A correction is applied using a fudge factor, which introduces a systematic uncertainty.

Solution: from a calibrated source, the fudge factor and it's uncertainty is determined. With the above prescription, the systematic uncertainty is implemented such that the amount by which the fudge factor is corrected is the same in both the on and the off measurement.

➡ The situation is similar at the LHC. The set of objects that are measured (jets, electron, photons, muons, ...) are always the same. The systematics need to be linked between measurements.

Big Model, not Big Data

Combining the individual measurements and correlating systematics between measurements leads to a “big model problem”, not a “big data problem”.

We don’t do

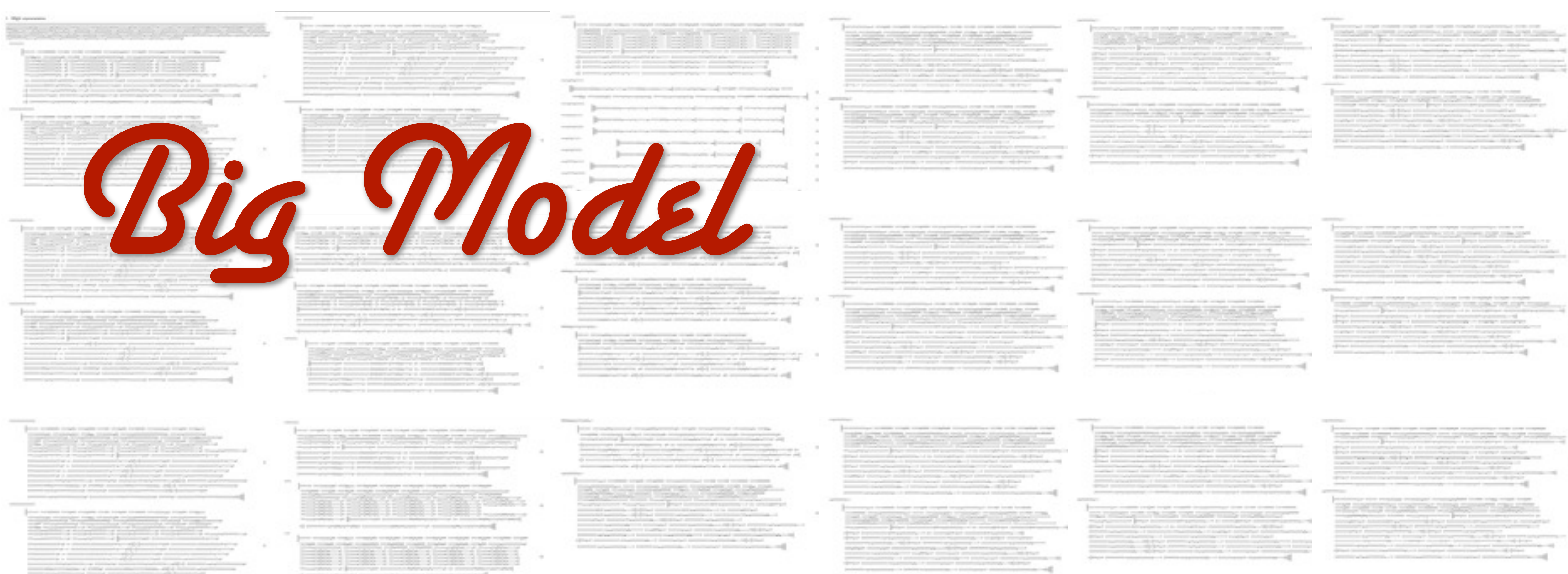
BIG DATA

$$L(\omega, S) = \frac{1}{|S|} \sum_{(x,y)\in S} (y - \omega^T x)^2$$

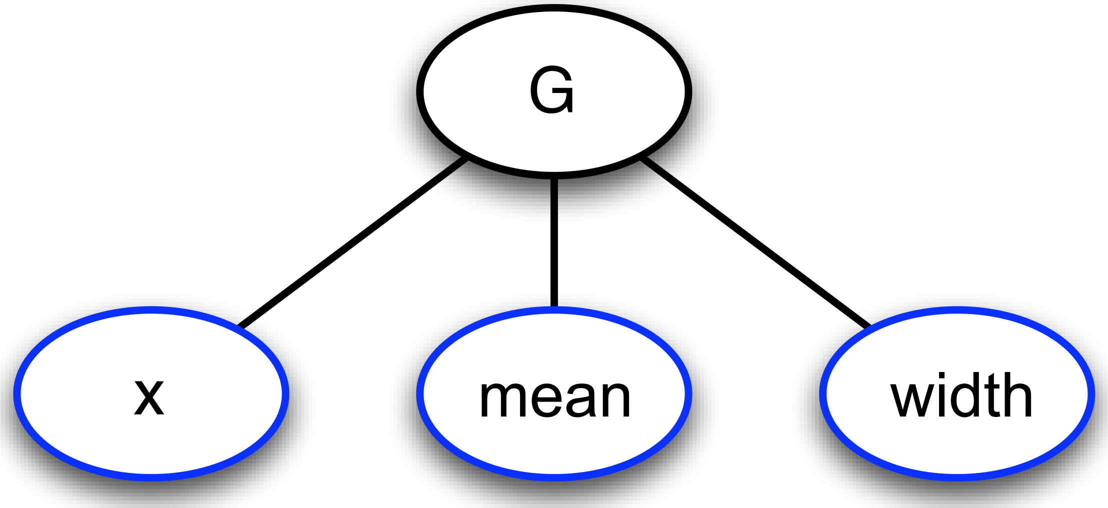
$6\text{x}10^7$ features in RCV1 dataset

We do

L =

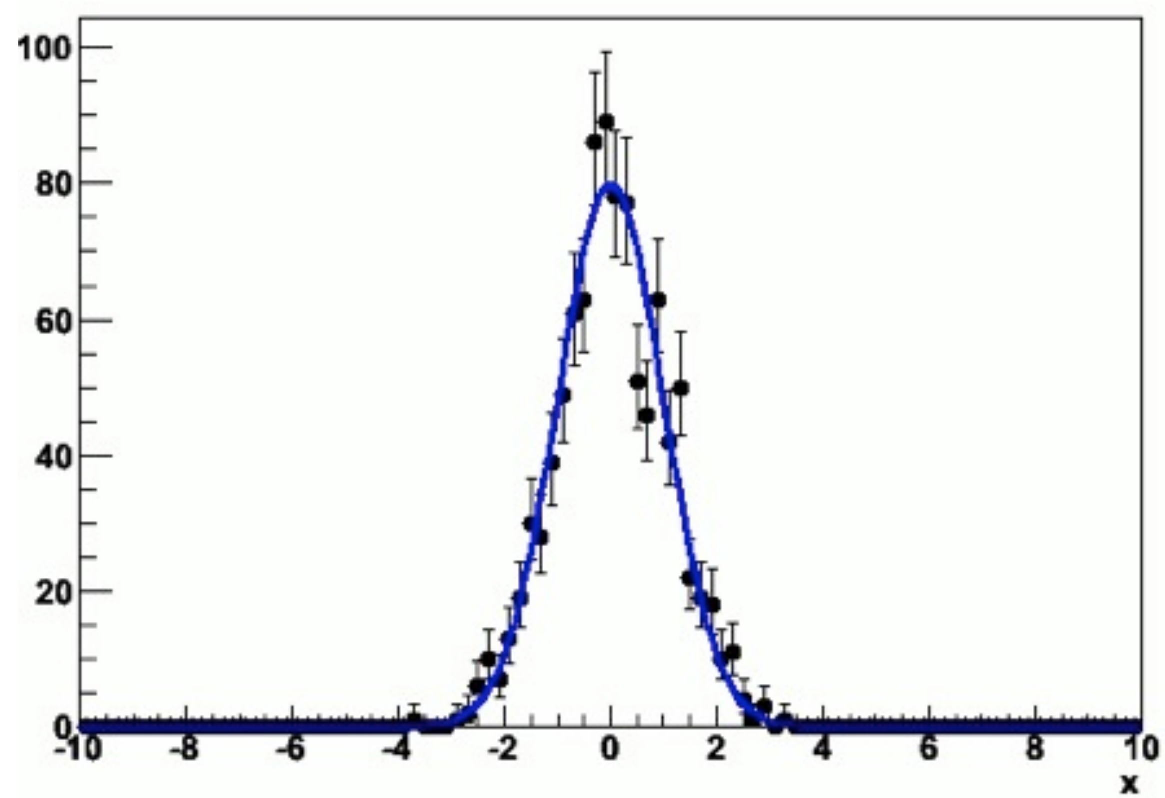


Model: Gaussian example

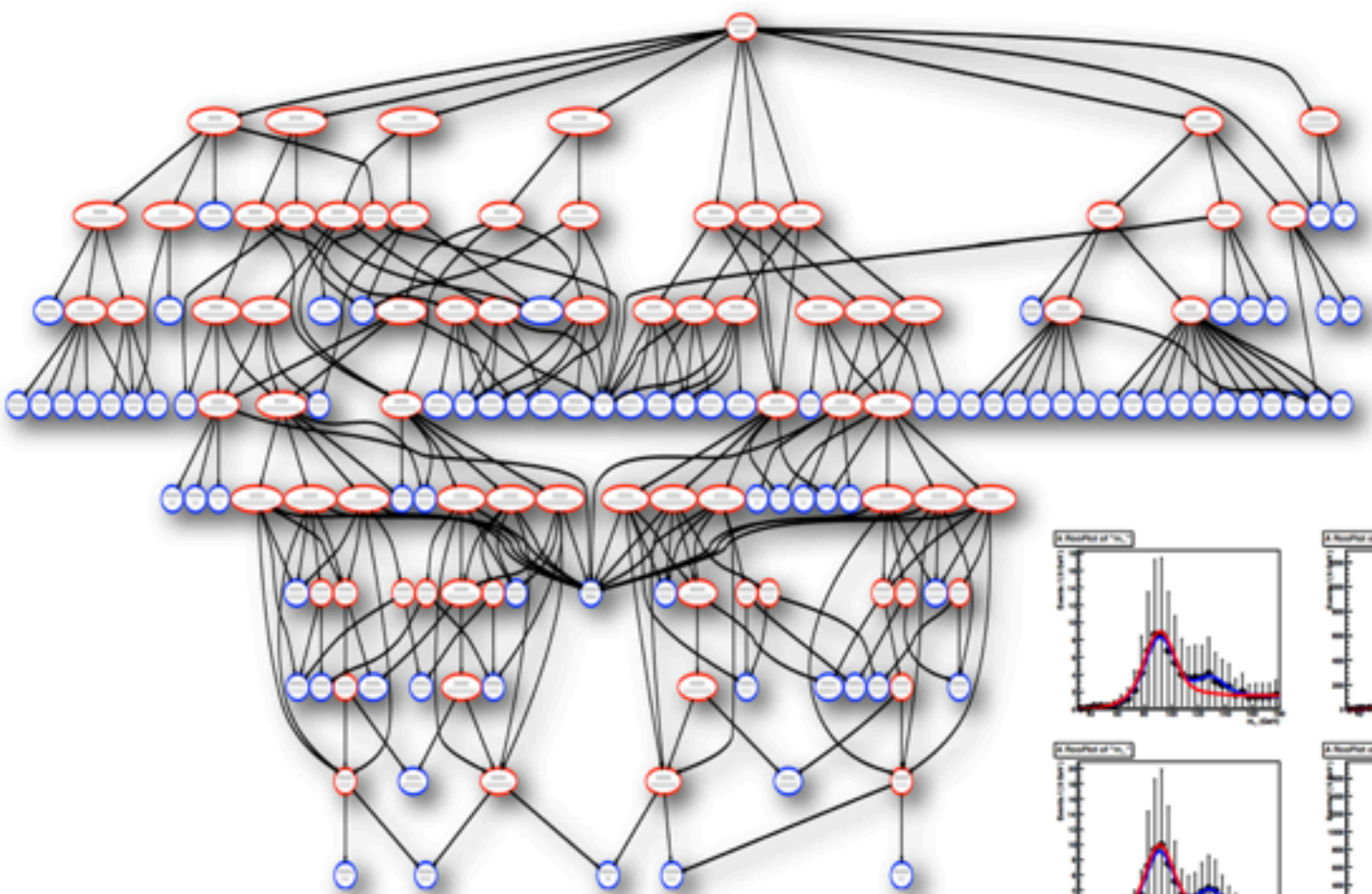


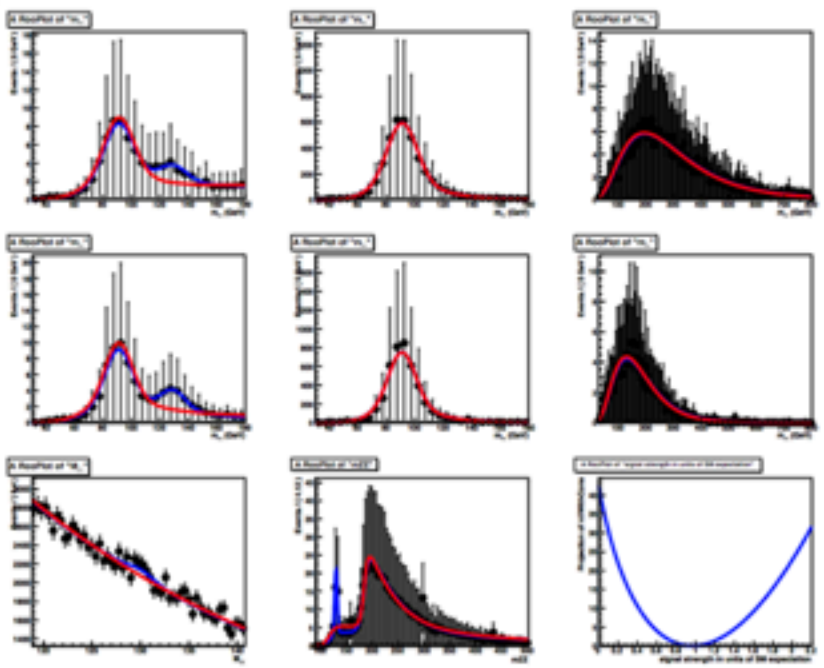
Top node: output value

Nodes below: input variables



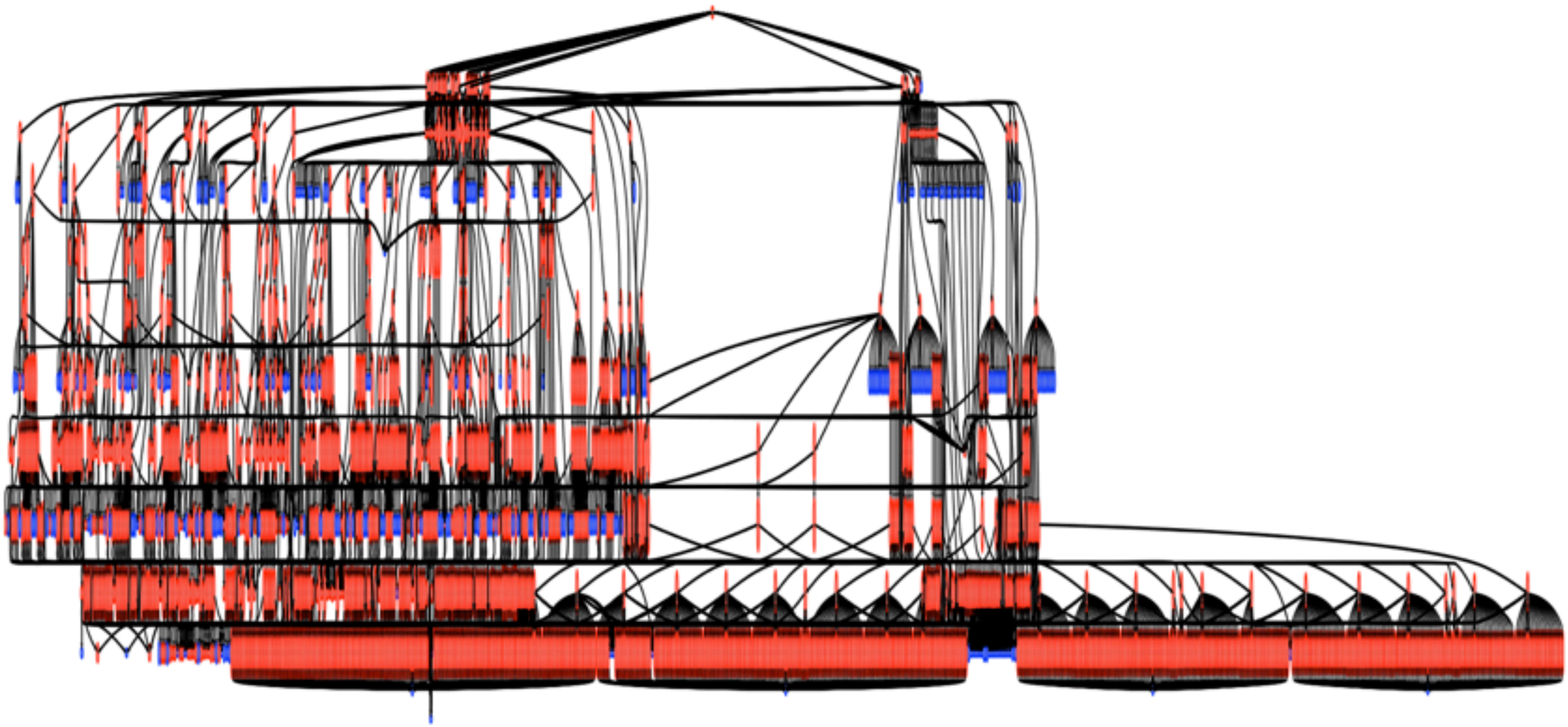
Model: Simple Measurement



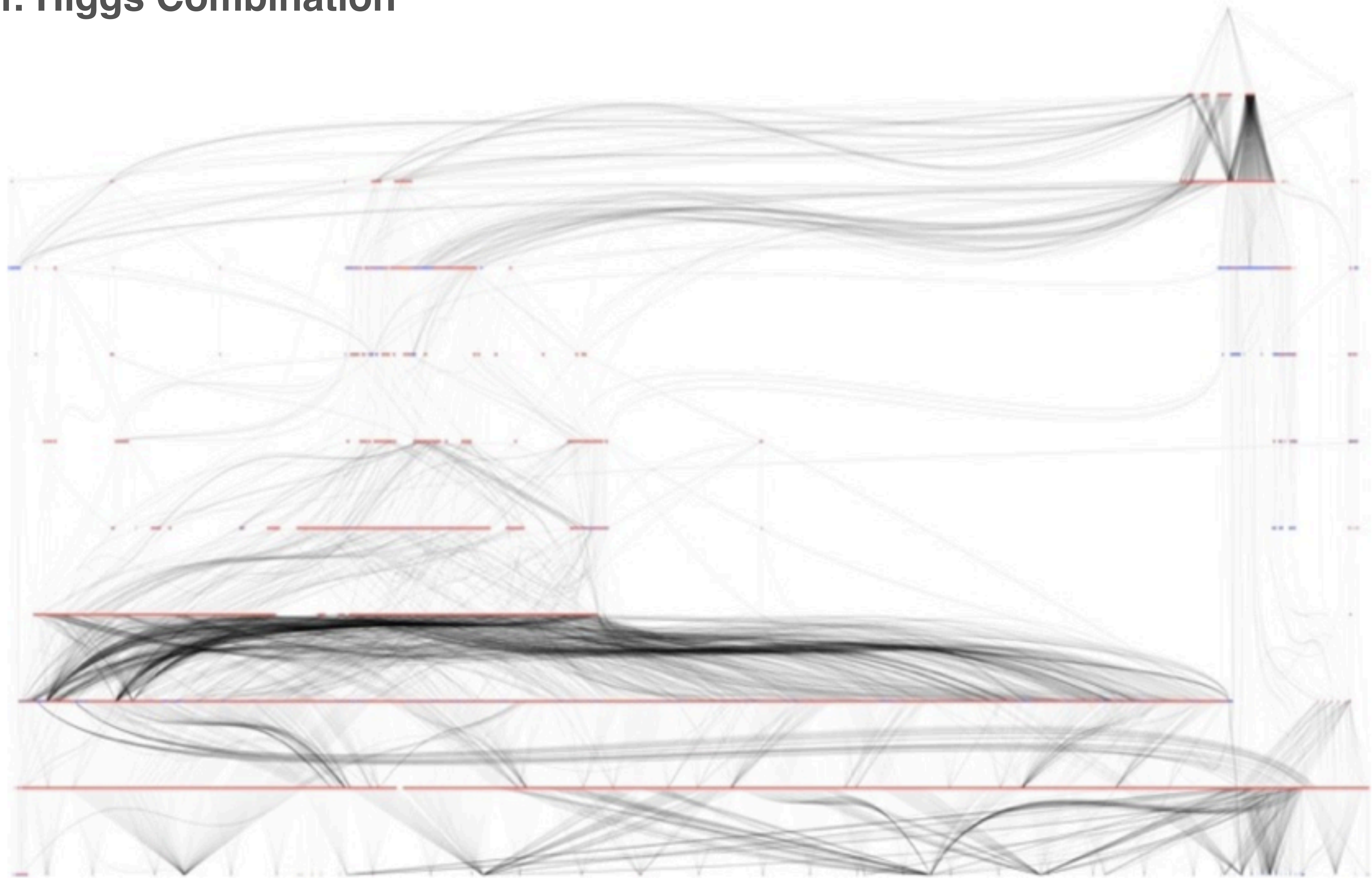


-ln(L) profile for the parameter of interest

Model: Test Higgs Combination



Model: Higgs Combination



Our models now contain about 800 parameters, where ~500 are associated to statistical fluctuations in the Monte Carlo templates.

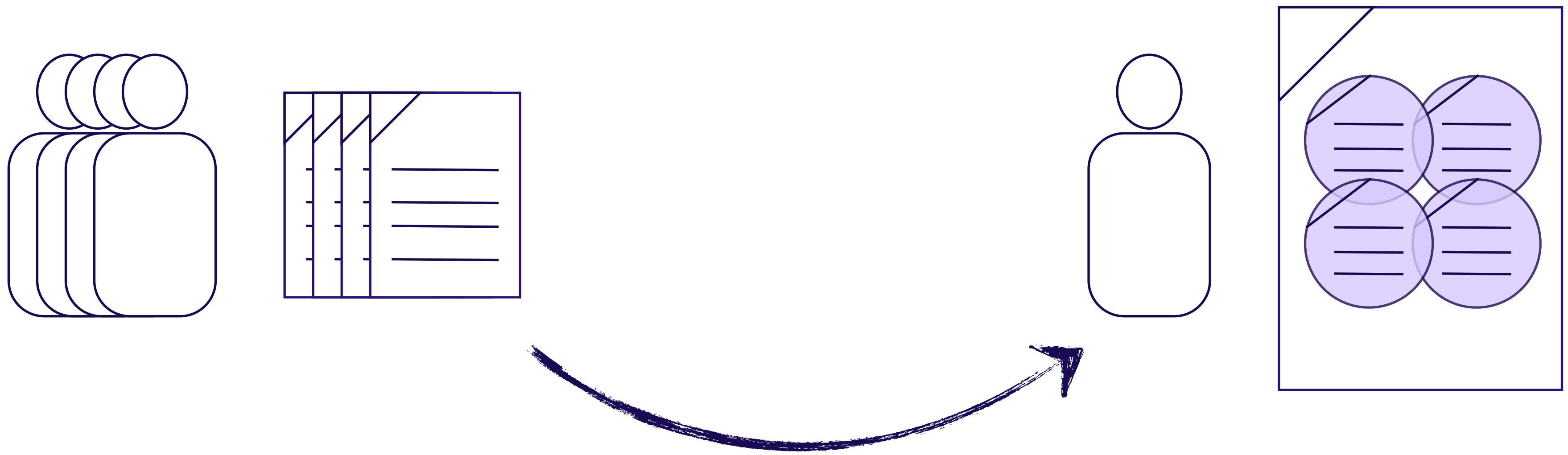
They contain about 20,000 nodes.

STATISTICAL METHODS, BAYESIAN TECHNIQUES

One Model, Many Methods

one subgroup per decay mode:
$\gamma\gamma$, ZZ, WW, ...

combined model includes
common parameters



subgroups provide digital form of their
model to combination group

For every model, we can run various statistical tests and measurements.

Upper Limits are produced with three different methods that are to a large extent technically independent: using asymptotic formulas of the Likelihood, Frequentist Ensemble Tests and Bayesian MCMC

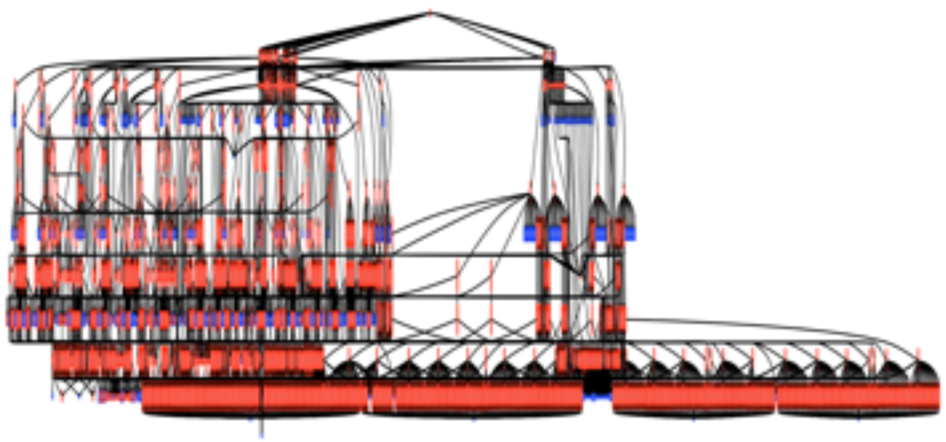
➡ Ensemble tests rely on Likelihood minimizations which can fail, MCMC does not.

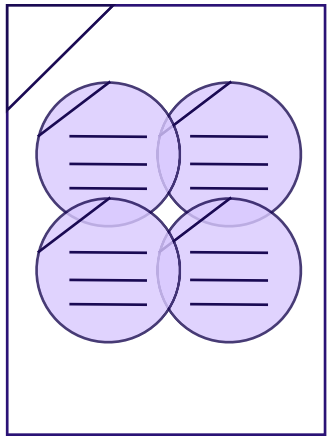
Discovery Significances are produced with Asymptotics and Ensemble Tests.

Underlying technology is based on RooFit and RooStats: https://twiki.cern.ch/twiki/bin/view/RooStats

Separate Model and Method

Serializable model. Stored in a binary file that includes a descriptor object for the model (ModelConfig) and the observed data.





Model files are stored and shared on a shared filesystem (for example AFS). Various groups can run their statistical analysis on these models.

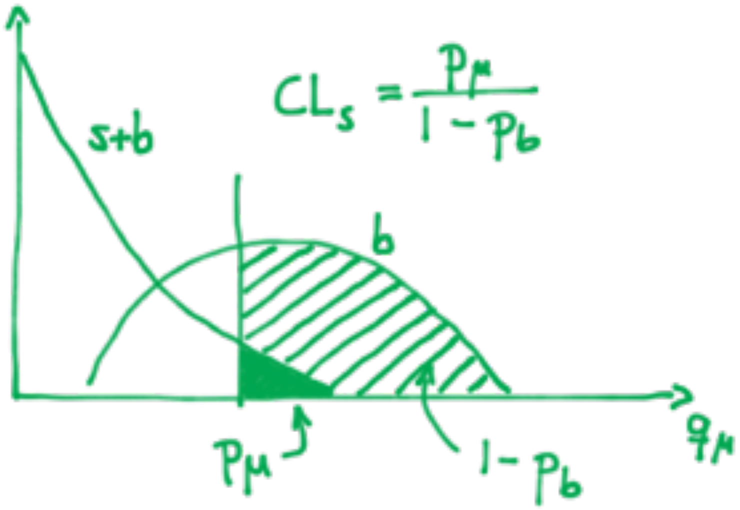
➡ A few standard methods come with RooStats:

```
double StandardFrequentistDiscovery(  infile, workspaceName, modelConfigName, dataName );
void   StandardHypoTestInvDemo(       infile, workspaceName, modelConfigName, dataName );
void   StandardProfileLikelihoodDemo( infile, workspaceName, modelConfigName, dataName );
void   StandardBayesianMCMCDemo(      infile, workspaceName, modelConfigName, dataName );
void   StandardBayesianNumericalDemo( infile, workspaceName, modelConfigName, dataName );
```

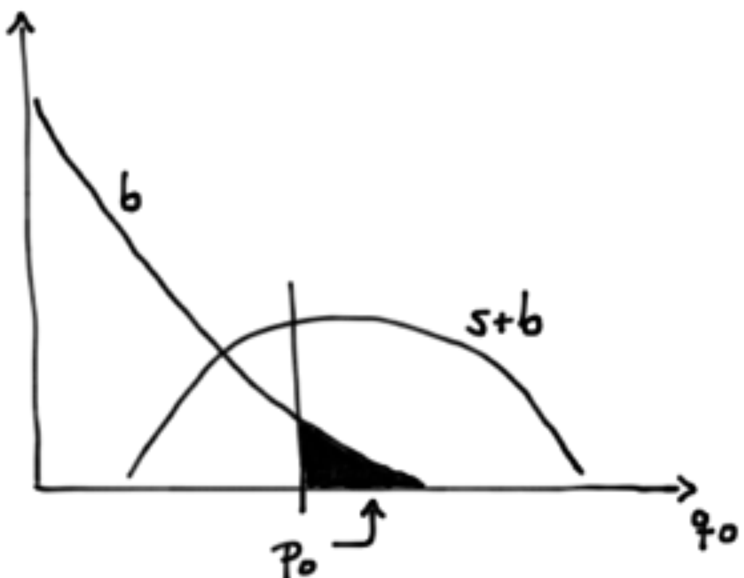
... and more

Statistics for Discovery

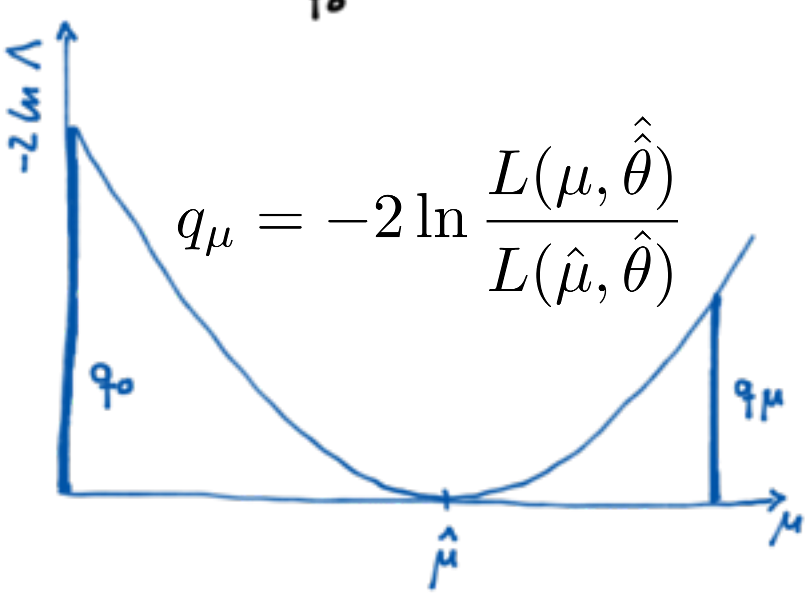
Follow LHC-HCG Combination Procedures



CL_s to test signal hypothesis

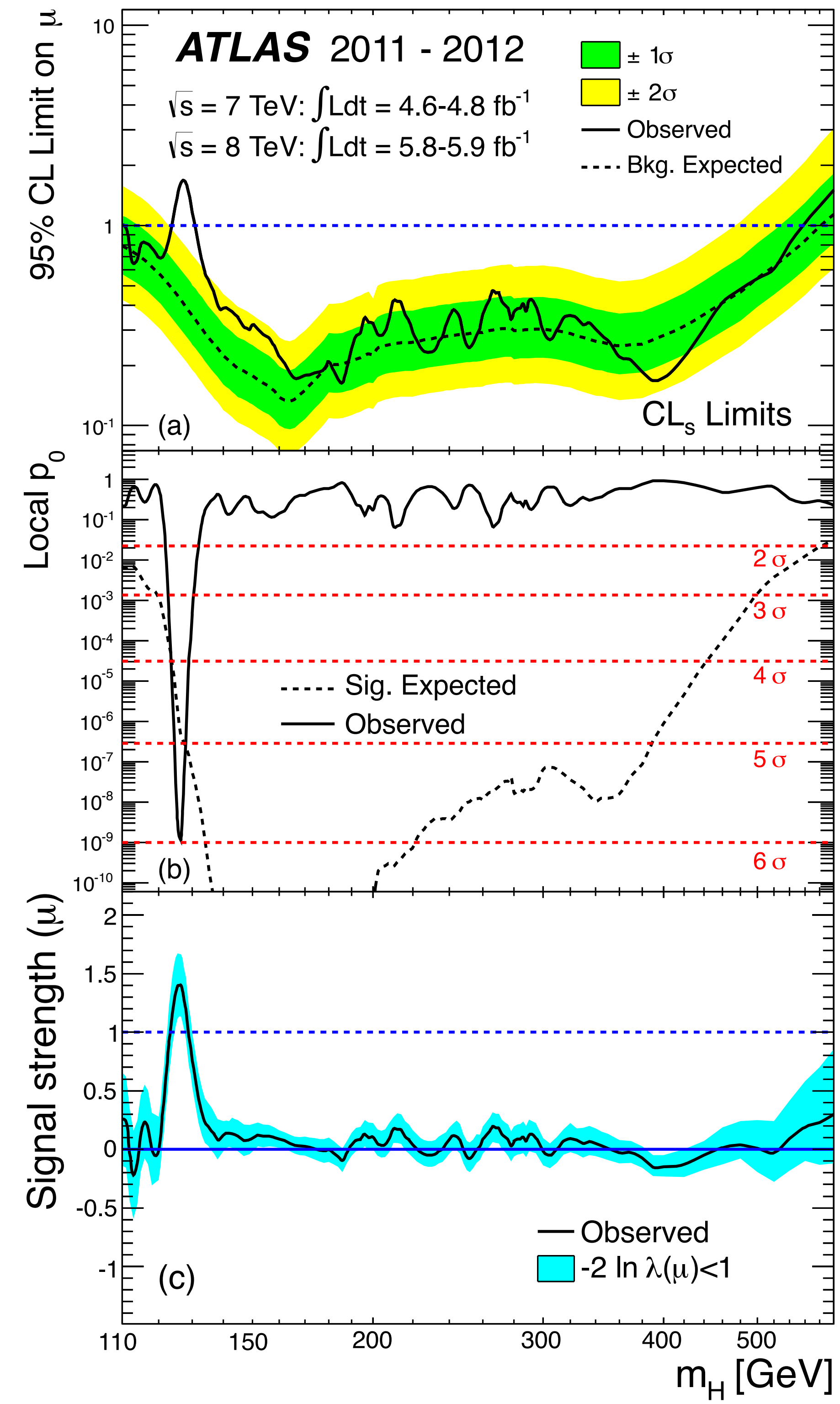


p_0 to test background hypothesis



$\hat{\mu}$ to estimate signal strength

$$\mu = \frac{\sigma}{\sigma_{SM}}\frac{B}{B_{SM}}$$



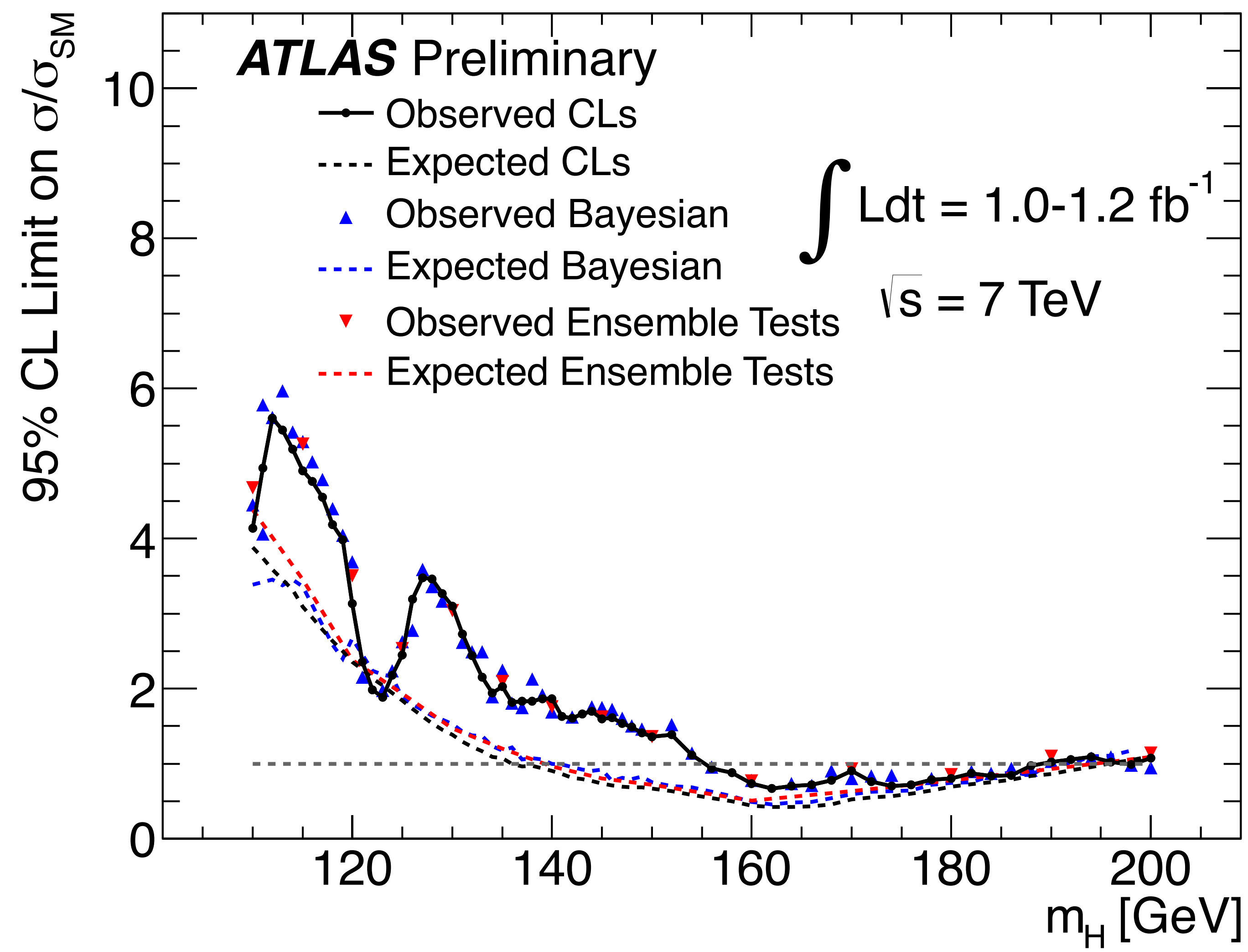
Cross Checks with MultiNest and MCMC

MultiNest:

- ➡ multi-modal ellipsoidal sampling
- ➡ ellipsoidal approximation to posterior contour becomes exponentially worse with increasing number of parameters
- ➡ problems with >60 dimensions

MCMC can do fairly large number of dimensions if a sequential proposal function is used.

“Modified Frequentist upper limits” [Read, 2000] for Gaussian and Poisson models are expected to agree numerically between Frequentist, Likelihood and Bayesian methods. They also agree well for very large Higgs models as shown by ATLAS on the right.



Priors for Bayesian Studies

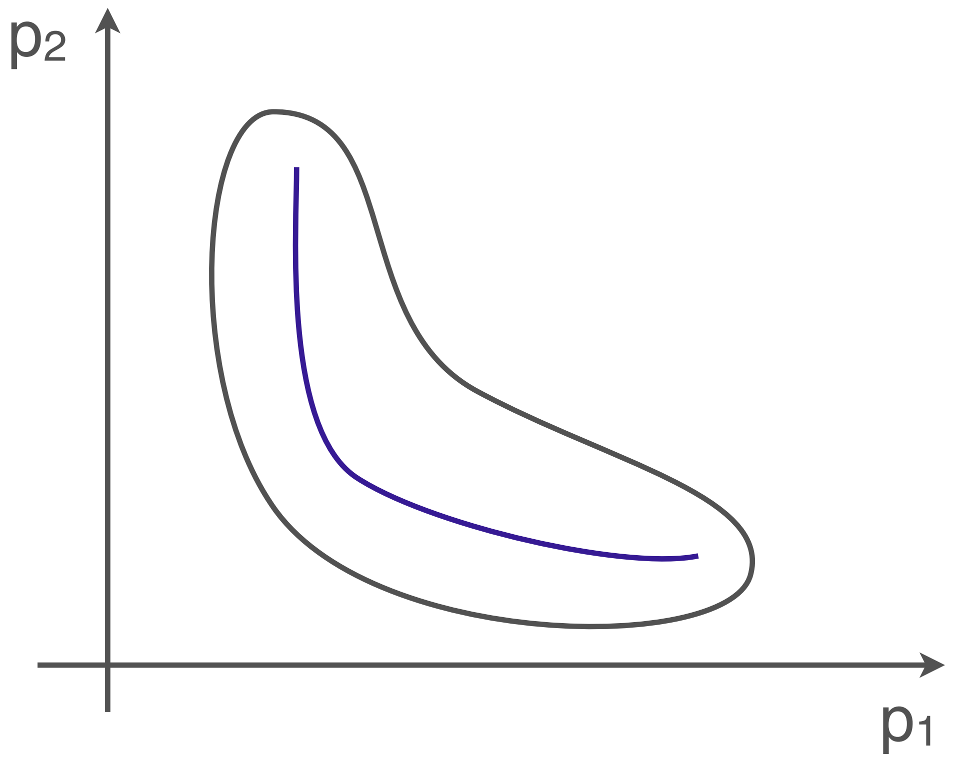
Nuisance Parameters: “prior”s are the auxiliary measurements. The priors of the auxiliary measurement, “ur-prior”s, are usually taken as flat. It was shown that the result is fairly insensitive to their exact form given the additional information from the auxiliary measurement.

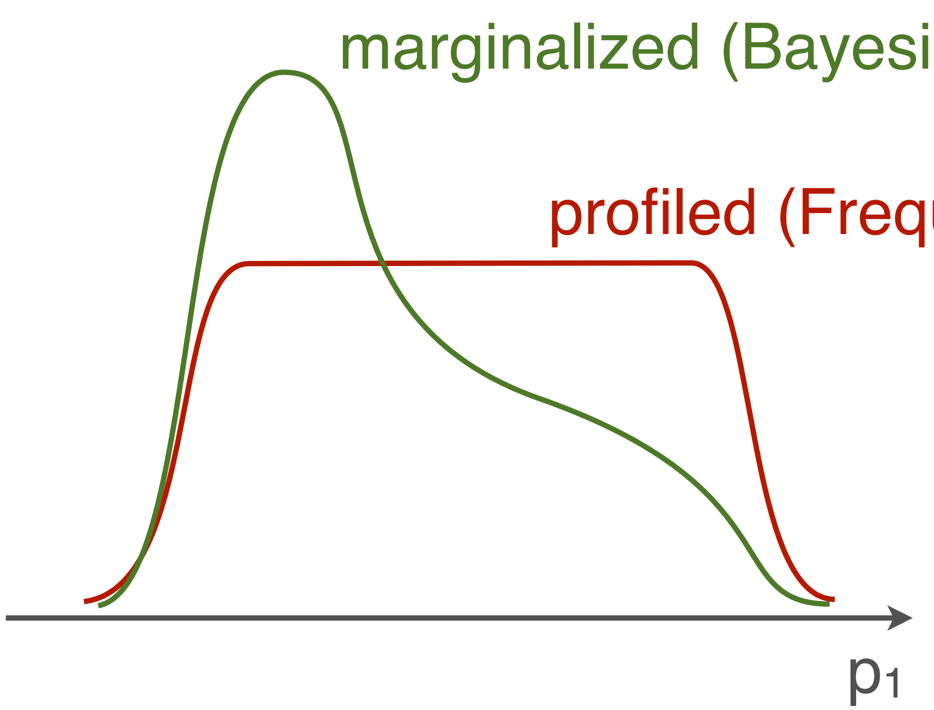
Discovery: one POI which is signal strength μ.
In the cross checks, we assumed flat priors in signal strength μ. For simpler but similar models, we calculated numerically the Jeffreys Prior which was fairly similar to the flat prior.
There are different subjective views, whether the “best” subjective prior is flat in the experimentally motivated parameter μ or in the theoretically motivated coupling scale factors that scale like $\sqrt{\mu}$.

Property studies: many POI, for example κ_V and κ_F.
In addition to the above problem, so called “volume effects” can lead to interesting and sometimes counterintuitive properties.

Difference between Frequentist and Bayesian methods for Coupling Studies

We wanted to study the “volume effect” in banana-type Likelihood contours.





There is no scale on these sketches to derive intervals. When constructing credibility and confidence intervals in practice, we found tighter limits using marginalization.

Conclusion

This works for us: Invest effort into problem specific Likelihood model. Distribute work to domain experts. Combine result into one model.

Showed an improvement of template models beyond simple histograms that is used for the mass measurement of the Higgs boson.

Separate model and method. Conceptual and technical issues can arise with any one technique. Comparison of various methods for the same model is an important check.