



Implications of LHC results for TeV-scale physics:

Discussion about data interpretation

A. Read (U. Oslo)

31.08.2011 - CERN

Topics (**not** latest Higgs results)

- Milli-review of search stats (can we skip??)
- Possible search outcomes
- CLs, Bayes, PCL
- “We need more than limit plots”
- LEE
- Combining indirect and direct (non)evidence
- Data-driven thoughts
- When is your favorite model “excluded”?!

Search statistics

- p_o – p-value used to reject background hypothesis (in favor of an alternative hypothesis with **an unknown signal strength**)
- p_b – p-value used to reject background hypothesis (in favor of a completely specified alternate hypothesis), also called $1-CL_b$
- p_μ – p-value used to reject the signal+**background** hypothesis, also called CL_{s+b}
- CL_s – pseudo-("modified frequentist") p-value used to reject the signal hypothesis

1 ~~Two~~ slides on statistics: limits

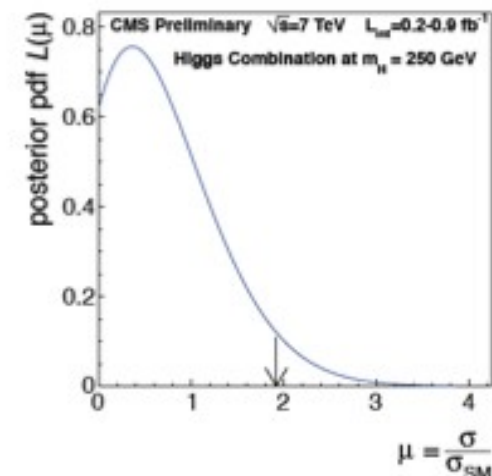
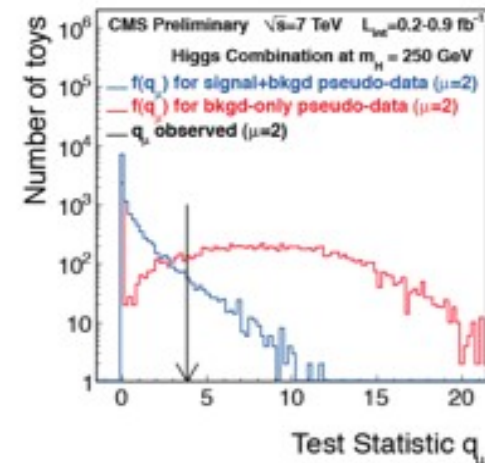
Pursue both frequentist and Bayesian paradigms, which allows us to validate robustness of results...

- CL_s (exact formulation as agreed with ATLAS, details are in backup)

$$CL_s = \frac{P(q_\mu \geq q_\mu^{obs} | \mu s(\hat{\theta}_\mu^{obs}) + b(\hat{\theta}_\mu^{obs}))}{P(q_\mu \geq q_\mu^{obs} | b(\hat{\theta}_0^{obs}))} = 0.05$$

- Bayesian (with flat prior on signal strength)

$$\int_0^{\mu_{95\%CL}} p(\mu | \text{data}) d\mu = 0.95$$

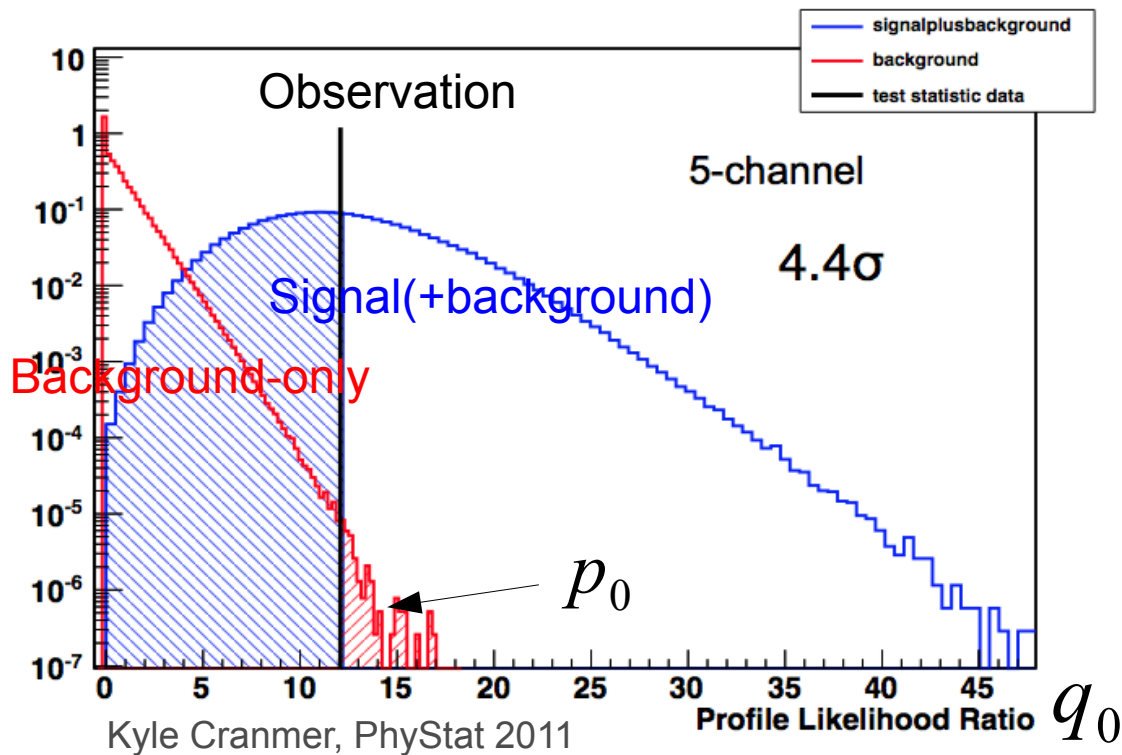


LLR test statistics

	Test statistic	Test statistic	Nuisance parameters	Pseudo-experiments
LEP	$-2 \ln \frac{L(\mu, \tilde{\theta})}{L(0, \tilde{\theta})}$	Simple LR	Fixed by MC	Nuisance parameters randomized about MC
Tevatron	$-2 \ln \frac{L(\mu, \hat{\hat{\theta}})}{L(0, \hat{\hat{\theta}})}$	Ratio of profiled likelihoods	Extracted from priors	Nuisance parameters randomized from priors
LHC	$-2 \ln \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\hat{\theta}})}$	Profile likelihood ratio	Profiled (fit to data)	New nuisance parameters fitted for each pseudo-exp.

LHC sampling of test statistic is frequentist, LEP and Tevatron Bayes-frequentist hybrid. CL_s can be used together with any of these – must be specified! No longer sufficient to write e.g. “the CL_s method was used”.

p_0 – reject b-only hypothesis (“discover”)

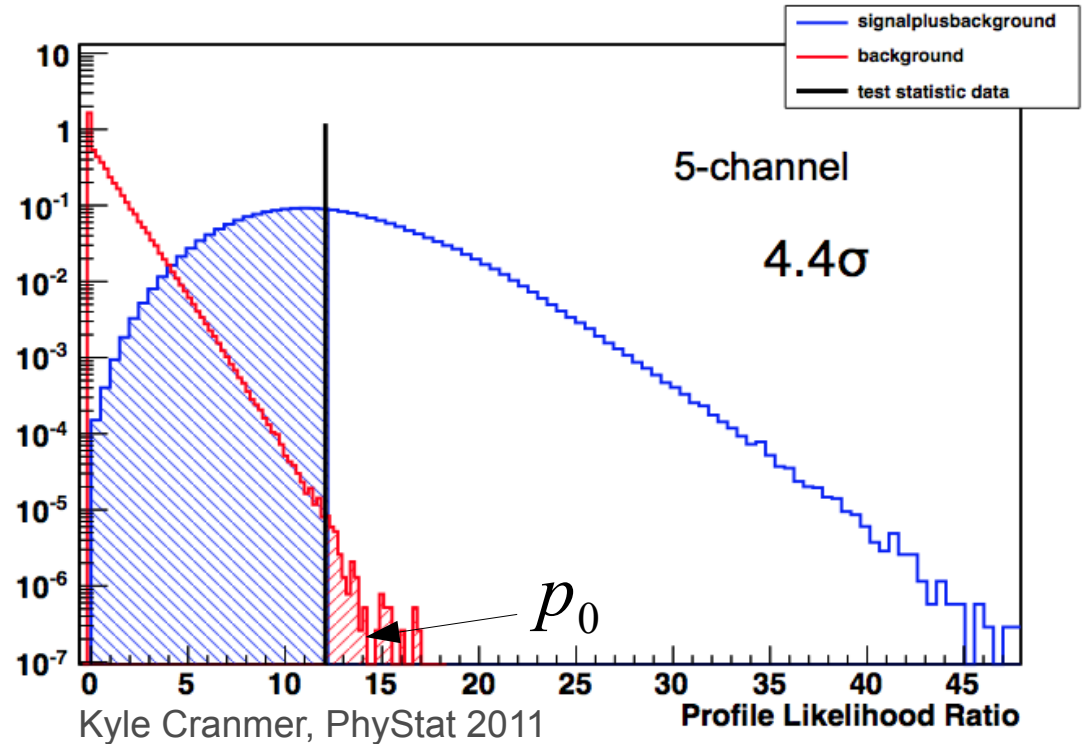
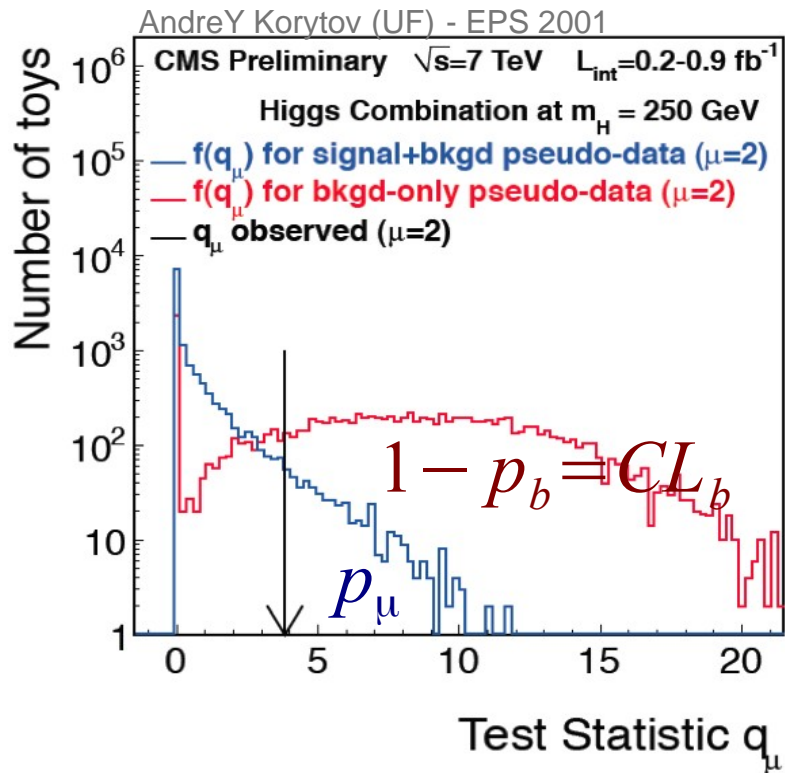


- p_0 gives significance of background rejection, e.g. 5σ for $p_0 = 2.8 \times 10^{-7}$

$$p_0 = P(q_0 \geq q_0^{obs})$$

$$q_0 = -2 \ln \frac{L(\mu = 0, \hat{\theta}_0)}{L(\hat{\mu}, \hat{\theta})}, \hat{\mu} \geq 0$$

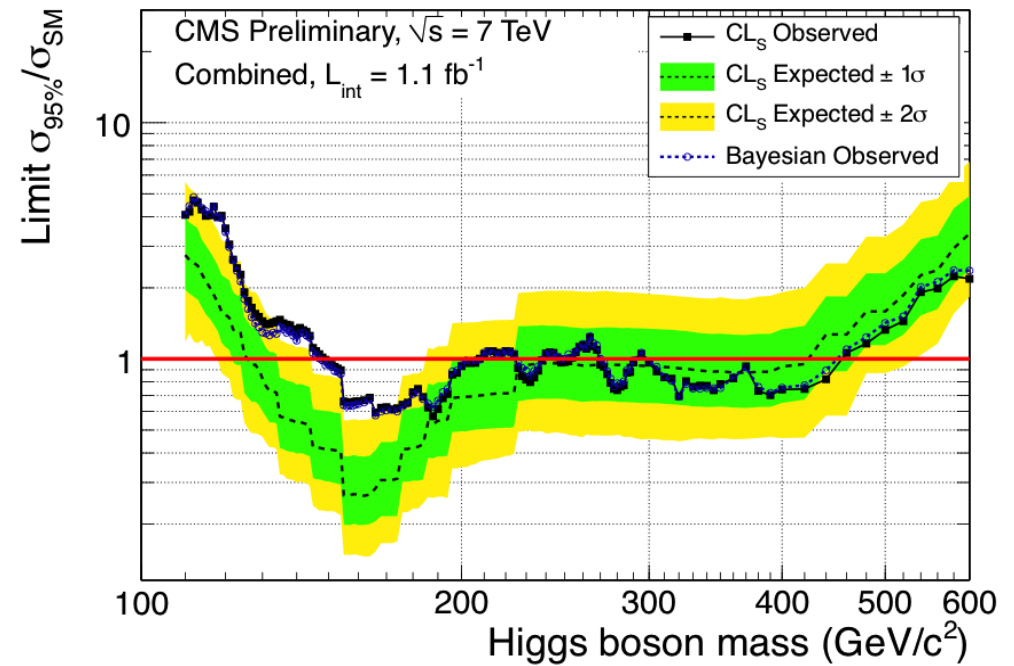
p_μ - reject s+b hypothesis (exclude)



- A small $p_\mu = p_{\mu+b} = CL_{s+b}$ rejects the signal+background hypothesis
- $CL_s = p_\mu / (1 - p_b)$ - penalty to p_μ for lucky b-fluct.

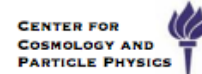
Difference between CLs and Bayes (uniform prior)?

- For most searches, not much!
- For the moment CLs is interim solution for frequentist(-based) searches at LHC



PLR (CL_{s+b}) also needs “protection”

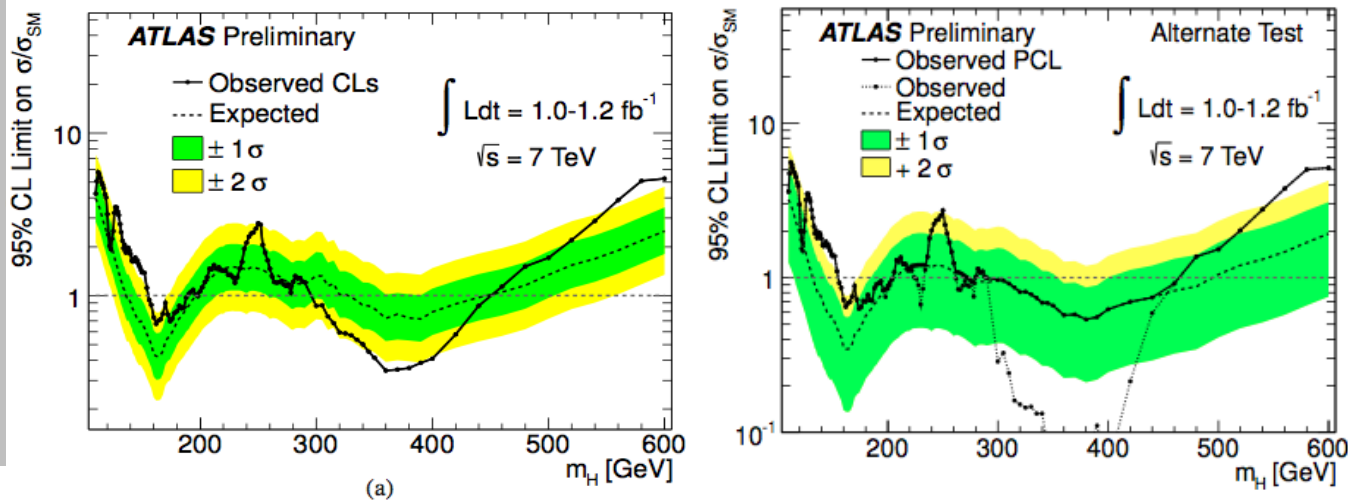
Results with PCL



While CLs is well a established technique in our field, it is considered a non-standard procedure by statistician mixing notions of power and coverage

- it intentionally over-covers to protect against setting limits beyond the experiments sensitivity due to downward fluctuations

An alternative approach (PCL) is based on purely frequentist CLs+b together with a “power-constraint” at the experiments sensitivity achieves the same protection without mixing the notions of coverage and power



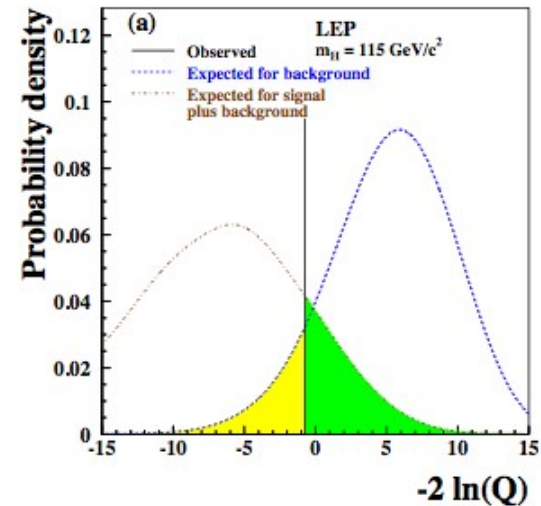
20

Kyle Cranmer (NYU) EPS-HEP 2011 - ATLAS Higgs Combination

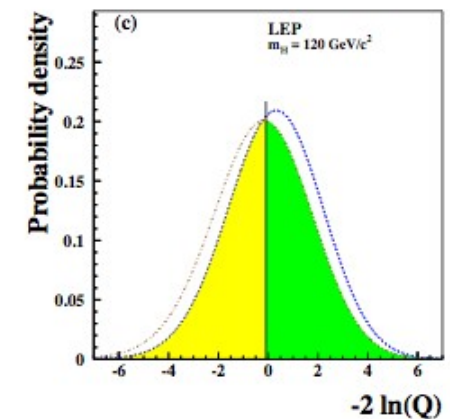
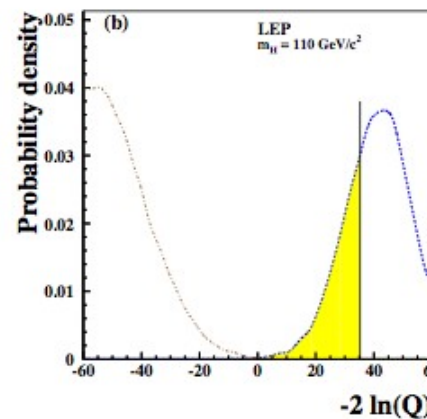
- CLs intuitive solution with connection to Bayes
- PCL IS simpler
- All “protections” break freq. Coverage
- CLs+b and deficits...

Outcomes of a (model-dependent) search

- Not enough sensitivity -- b and s+b distributions look like each other
- Plenty of sensitivity:
 - obs outcome looks like it came from the b-only distribution and not the s+b
 - obs outcome looks like it came from the s+b distribution and not b-only
 - obs outcome looks like it came from neither distribution:
 - Excess is too big for a signal.
 - Excess is too small for the proposed signal but too big to be just background
 - Deficit so big b-only hypothesis looks wrong (mismodeled).
- *Non-negligible probability to **not** get yes/no answer!!!*



(LEP-style LLR for illustration purposes)



Beyond limits

Appendix B - Results Summary

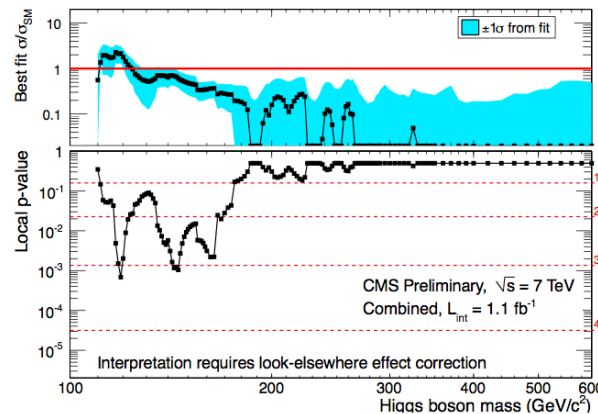
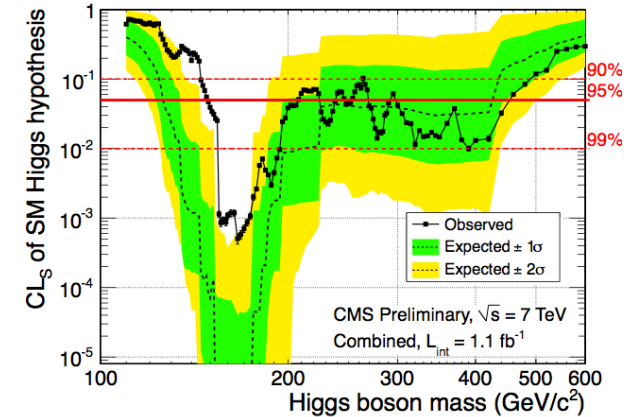
Table 4: The observed and expected CL_s , the observed and expected 95% CL upper limit on the Standard Model Higgs boson production cross section normalized to the Standard Model values using the CL_s method, and the p_0 for a set of Higgs boson mass hypotheses.

m_H (GeV)	CL_s		CL_s limits				p-values $p_{\mu=0}$
	Obs.	Exp.	Obs.	-1 σ	Median	+1 σ	
110	0.473	0.558	3.11	2.6	3.62	5.03	0.5
111	0.609	0.543	3.76	2.48	3.45	4.8	0.412
112	0.678	0.526	4.1	2.4	3.33	4.63	0.281
113	0.649	0.504	3.88	2.27	3.16	4.39	0.284
114	0.626	0.483	3.67	2.16	3	4.17	0.289
115	0.586	0.454	3.4	2	2.78	3.86	0.308
116	0.574	0.428	3.31	1.9	2.63	3.67	0.299
117	0.584	0.403	3.23	1.8	2.5	3.47	0.258

Appendix B - Results Summary

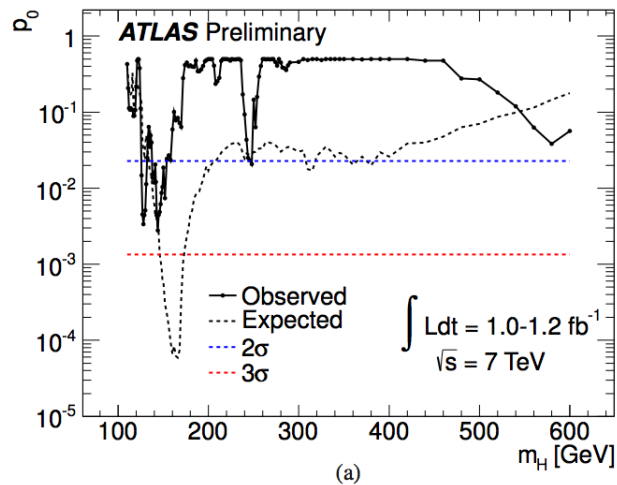
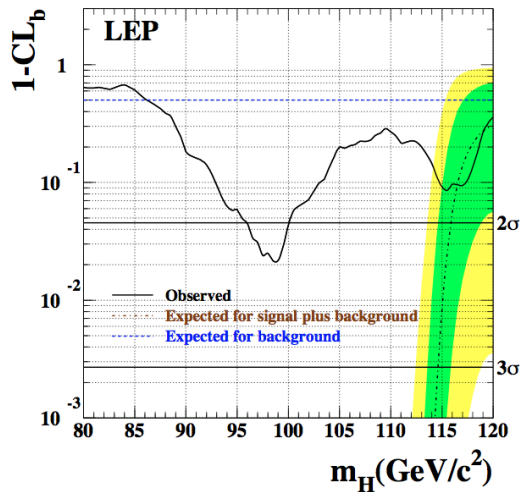
Table 3: The observed and expected CL_s and PCL limits on the cross sections of a Standard Model Higgs boson production normalized to the Standard Model values for a set of Higgs boson mass hypotheses. The CL_s numbers should be taken as the final result. The p_0 -values and $-2 \times \ln \frac{\mathcal{L}(1,\hat{\theta})}{\mathcal{L}(0,\hat{\theta})}$ are also given.

m_H (GeV)	CL_s limits				PCL limits		$-2 \times \ln \frac{\mathcal{L}(1,\hat{\theta})}{\mathcal{L}(0,\hat{\theta})}$	p-values $p_{\mu=0}$
	Obs.	-1 σ	Median	+1 σ	Obs.	Median		
110	4.2	2.9	4.0	5.5	3.6	3.2	0.1	0.43
111	5.1	2.7	3.8	5.3	4.8	3.1	-0.6	0.21
112	5.8	2.6	3.6	5.1	5.6	3.0	-1.1	0.11
113	5.6	2.5	3.5	4.9	5.4	2.9	-1.2	0.11



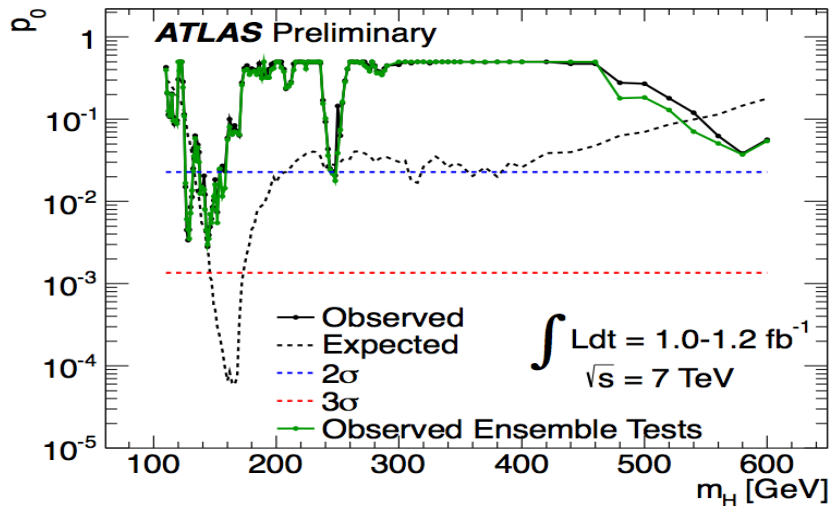
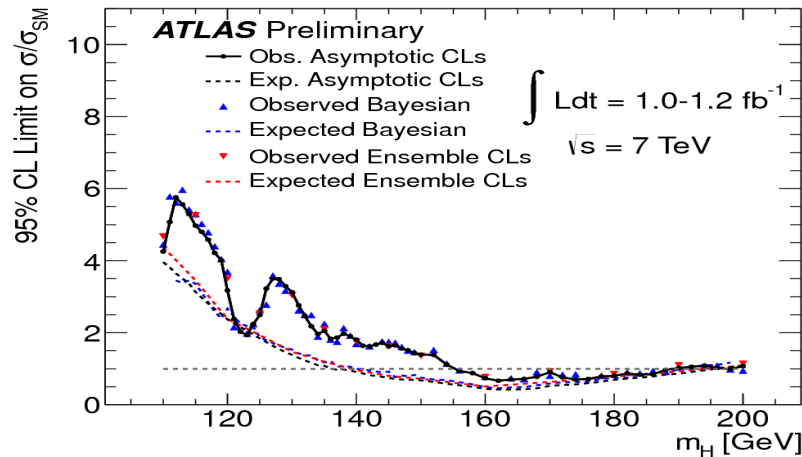
- Ultimate is to publish RooStats workspace
 - This will take time...
- In meantime, p-values, CLs, μ_r , etc...

p_b vs. p_0



- p_0 not fixed to signal strength – good asymptotic properties
- p_b more direct model interpretation, also show deficits
- p_b is T. Junk's favorite, close to “all in one”

Asymptotic expressions validated



- Potential vast savings in CPU time (ATLAS used at least 250 kh for this combination)
- Only need to check stability of single fit to data
- Quick response: test new models

Trial factors for HEP (LEE)

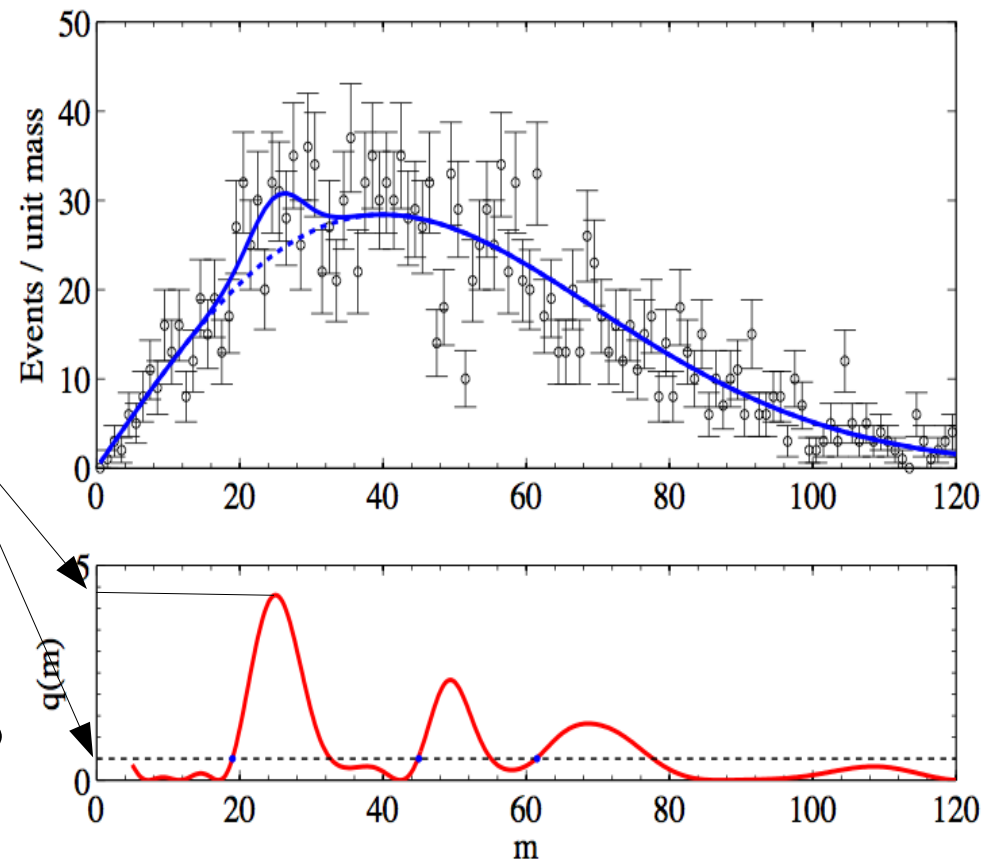
- Gross and Vitells, EPJ C70 (2010) 525-530

- $\langle N_{q_{test}} \rangle = \langle N_{q_{ref}} \rangle e^{-(q_{test} - q_{ref})/2}$

- $p_0^{global} \leq \langle N_{q_{test}} \rangle + p_0^{local}$

- Approximations best above $q_{test} \sim 9$ (3σ)

- Where is “elsewhere”?



Where is “elsewhere”?

- All HEP searches? Since when? 1970? This year?
- All LEP+Tevatron+LHC searches?
- All LHC searches?
- Only “my” search?
- Only the new data in previously unexcluded regions?
- My neighbors' data where I have a peak?
- No easy answer!
- Should all p0-plots have trial factors applied??
- \Rightarrow “5sigma debate” prolonged

Combining precision measurements and searches

- SM Higgs, SUSY models
 - Observables: cosmology, EW, flavor, g_{μ}^{-2}
 - Direct searches
- Searches so far tell where **not** to look...
 - Penalty in chi-squared seems fair, but wrt bg-hyp?
 - Help to localize minimum with exclusion/discovery tests?
 - No principle of “maximum significance”
 - Need rather e.g. $L(m_H)/L(\hat{m}_H)$

Data-driven worries

- In the extreme, we might take all shapes and normalizations from data to estimate backgrounds.
 - Are we sure signal is always “somewhere else”? That we miss nothing?
- Avoid data-driven biases
 - Killing 3σ (we should have a few)
 - Sculpting peaks
- Should exercise on small SM signals (e.g. single top at Tevatron), e.g. SM diboson production at LHC

To combine or not?

- Quality control
- Rapid sample doubling
- Effort
- Milestone
- Approximate combinations

From exclusion to measurement

- Fail to exclude the signal model (with sufficient sensitivity)
- Exclude (“5 sigma”) the background “in the direction” of the model
 - Without (LHC-style) or maybe with (LEP-style) the rate constrained by the model
- Consistency between channels, experiments, accelerators, data and the model
- Maximize likelihood (rate, mass, etc), constrained, unconstrained,...

Exclude SM Higgs?

- Suppose LHC+Tevatron+LEP excludes SM Higgs at $>95\%$ CL below 6-800 GeV?
 - At $>99\%$ CL? At 99.9% CL?
- We may be there sooner than we like...

Summary

- CLs is still around
- Freq. systematics treatment much improved since LEP (but results \sim Bayes!)
- LEE a major headache but we know more
- Limits-users want more information
 - The derivative is small but positive
- Start to see more specific tests of $\mu=1$ (SM)
- We would like more exercise of the discovery machinery!

Additional material

Will we ever reach consensus? Do we need to?

- Don't agree:
 - Bayesian or frequentist
 - Bayesian priors
 - 1-sided vs. Unified freq. confidence intervals
 - “One plot to rule them all”
- Agree:
 - Useful to compare Bayes, freq., likelihood results
 - All purists hate CLs
 - Treatment of deficits with vanishing signal sensitivity is necessary but difficult
 - Treatment of systematics in hypothesis tests improved since LEP

Is "5 σ " meaningful?

- Historically introduced to compensate for Look-elsewhere effect (LEE)/trial factor in HEP
 - Too many 3 σ effects came and went...
- Statisticians say nobody can estimate such tiny probabilities reliably.
- Proper accounting of trial factor AND 5 σ could be considered excessive.
- Could *consider* as criteria some function of:
 - Local 5 σ i.e. no trial factor
 - 3 σ after trial factor (*this* search, *other* searches...)

