University of Colorado
Boulder

HH+met reinterpretation exercise

Bill Ford

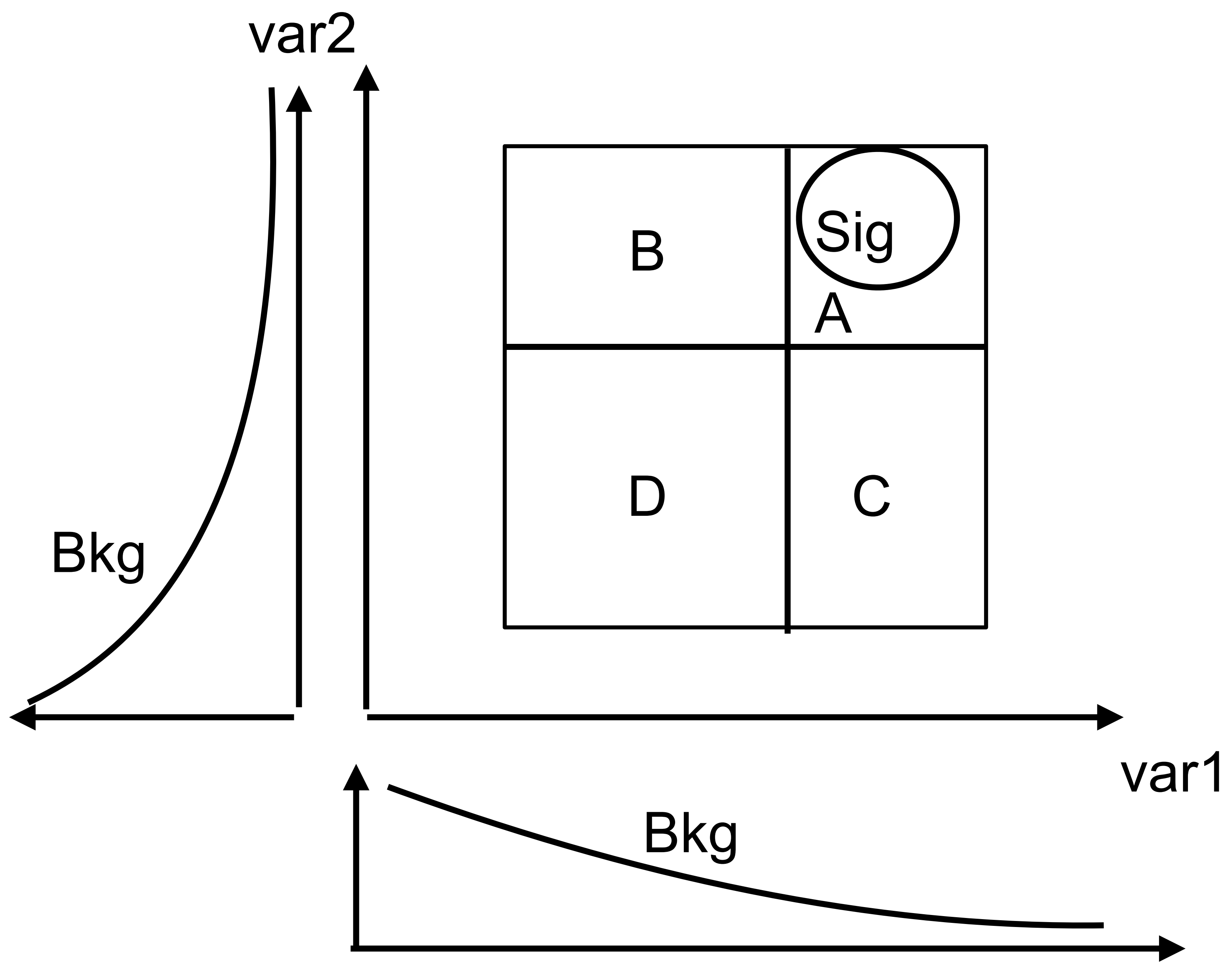
July, 2022

Methods for calculating limits from the data in search experiments

- Higgs combine tool (used in Higgsino search [1], RA2/b [2], …)
- Question: how well can one reproduce our results from HEPData?
- Chisquare fit
- Improved version: simplified likelihood
  - Described by Waltenberger at the June SUSY workshop:
- Results, comparisons
- Observations

- [1] CMS-SUS-20-004
- [2] CMS-SUS-19-006

Data-driven background prediction, “ABCD”



- Predicted N_{bkg} in A = N_B (N_C / N_D)
- All N’s are event counts, so Poisson distributed

Higgs combine tool method

- Inputs are
  - N^{obs}_i in each SR (signal region, A)
  - N^{obs}_i in each SB, CSR, CSB (B, C, D) region for the ABCD estimate of N^{bkg}_i in A
  - N^{sig}_i (MC) in each SR, SB, CSR, CSB (for each (m_{NLSP}, m_{LSP}) scan point)
  - κ_i for each SR, with uncertainty nuisance parameter
  - Uncertainty nuisances for other systematics (scale factors, JEC, …)
- Likelihood is built from
  - Poisson pdfs for N^{obs}_i in all SR, SB, CSR, CSB regions
  - Constraints $N^{bkg} = A = \kappa$ B C / D (rateParams)
  - Gaussian pdfs for κ uncertainties
  - Log-normal pdfs for other nuisances
- The expected yields N^{exp}_i are given by
  - $N^{exp}_i = N^{bkg}_i + \mu\, N^{sig}_{i,}$ where μ (r in combine) is the signal strength

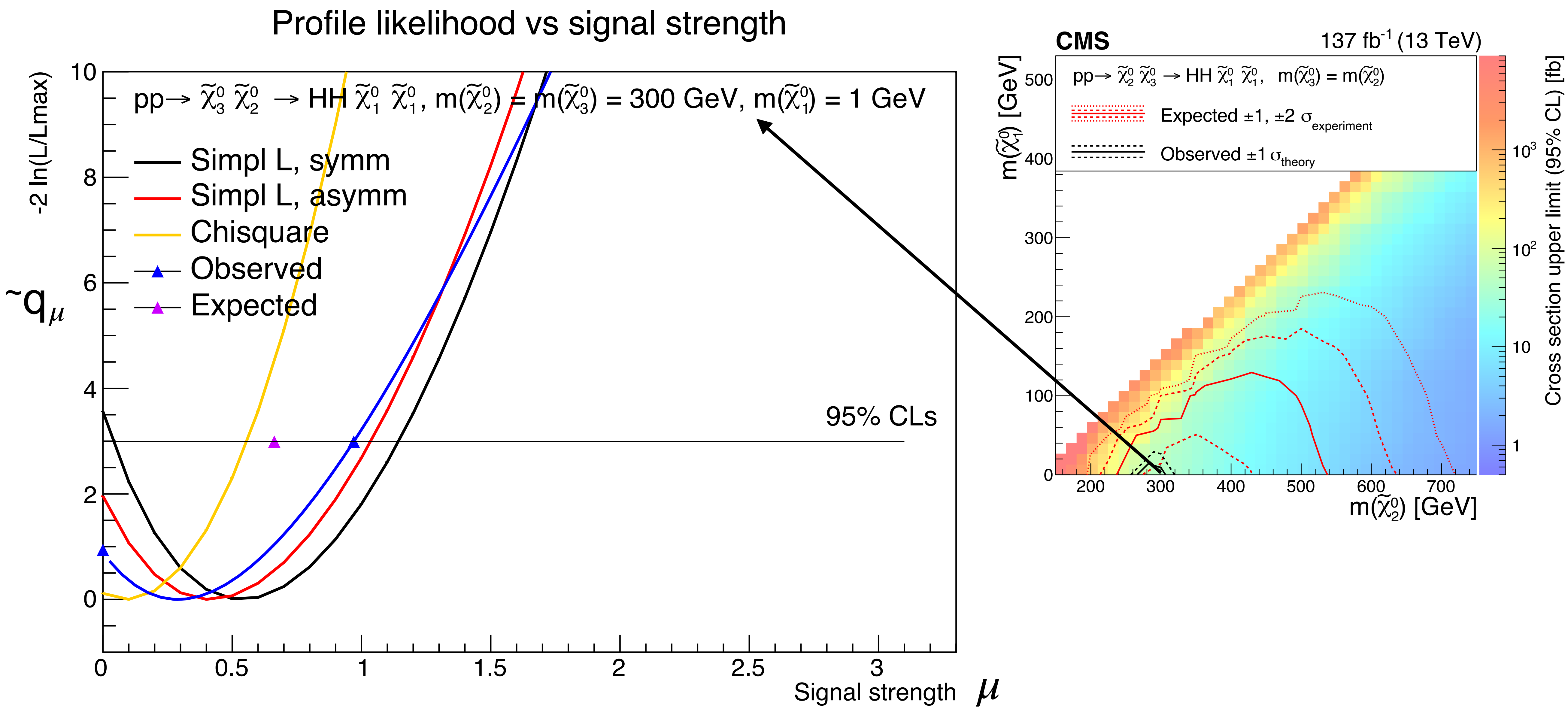
Extraction of limit & profile likelihood

- Running "combine -M AsymptoticLimits" on the datacard for a model scan point gives:
  - 95% CL upper limit from the data
  - 95% CL upper limit expected, with +/- 1 and 2-sigma deviations
- The criterion for 95% CL is that CLs = 0.05
  - CLs = CL_{s+b} / CL_b
  - CL_{s+b} = 1 - $\Phi(\sqrt{\tilde{q}_\mu})$, where $\tilde{q}_\mu$ is the profile likelihood test statistic:

    $$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(data|\mu, \hat{\theta}_\mu)}{\mathcal{L}(data|\hat{\mu}, \hat{\theta})}, \quad 0 \leq \hat{\mu} \leq \mu$$

    , and Φ is the normal cumulative distribution function (cdf)
  - CL_b measured with the Asimov data set (N^{obs} set to $N^{expected}$)
  - Details in CMS-NOTE-2011/005 (joint w ATLAS)
  - Adding —verbose 1 to AsymptoticLimits gives CL_b at limit
- Running "combine -M MultiDimFit --algo grid" gives $\tilde{q}_\mu$ vs μ

Profile likelihood vs μ



- Blue curve from combine tool method
  - blue triangles: significance, 95% CLs limit
    - $\mu < 1$ @ 95% CLs ⇒ this (300, 1) point is (barely) excluded
  - purple triangle: expected limit

Reproducing the result from HEPData

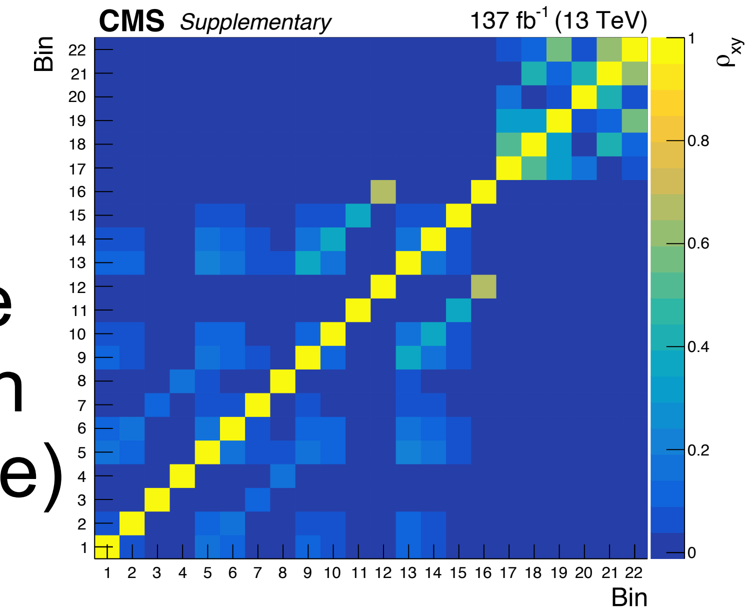
- The (relevant) available information:
- From Tables 4 & 5
  - Observed yields
  - Estimated pre-fit background yields
    - upper and lower uncertainties
- From figs. 11 (TChiHH-G), 13 (TChiHH), & 14 (T5HH)
  - Cross sections vs scan point (m_{NLSP}, m_{LSP})
    - Observed 95% CL upper limit, +/- 1 sigma
    - Expected 95% CL upper limit, +/- 1, 2 sigma
    - Theory (vs m_{NLSP})
- Covariance matrix
- Efficiency vs scan point and SR bin
- Significance vs scan point

From HEPData

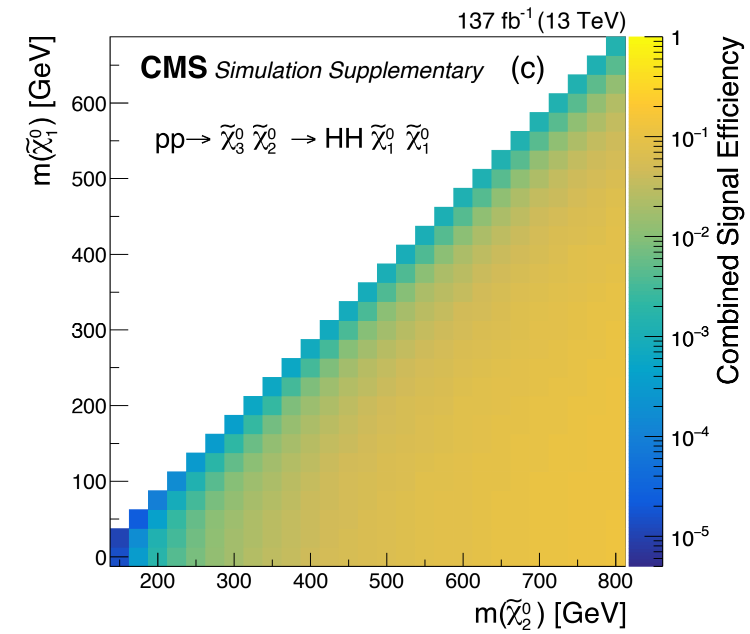
Bin	$\Delta R_{\max}$	$N_{\rm b}$	$p_{\rm T}^{\rm miss}$ [GeV]	κ	$N_{\rm SR}^{\rm pred}$	$N_{\rm SR}^{\rm fit}$	$N_{\rm SR}^{\rm obs}$
1	1.1–2.2	3	150–200	$1.09 \pm 0.04 \pm 0.02$	161^{+14}_{-13}	$149.7^{+8.9}_{-8.5}$	138
2			200–300	$0.92 \pm 0.04 \pm 0.02$	$90.4^{+9.7}_{-9.0}$	$91.5^{+6.9}_{-6.5}$	91
3			300–400	$0.94 \pm 0.09 \pm 0.01$	$11.5^{+3.4}_{-2.7}$	$12.8^{+2.6}_{-2.2}$	14
4			$>$400	$0.98^{+0.19}_{-0.16} \pm 0.02$	$2.8^{+2.3}_{-1.4}$	$2.8^{+1.4}_{-1.0}$	3
5		4	150–200	$1.13 \pm 0.09 \pm 0.08$	$53.5^{+8.8}_{-7.8}$	$54.1^{+5.6}_{-5.2}$	54
6			200–300	$0.96 \pm 0.07 \pm 0.07$	$28.3^{+5.6}_{-4.8}$	$33.2^{+4.2}_{-3.9}$	38
7			300–400	$0.89^{+0.16}_{-0.15} \pm 0.05$	$2.6^{+1.5}_{-1.1}$	$3.2^{+1.3}_{-1.0}$	4
8			$>$400	$0.92^{+0.27}_{-0.22} \pm 0.07$	$2.6^{+2.4}_{-1.4}$	$1.27^{+0.98}_{-0.63}$	0
9	$<$1.1	3	150–200	$1.05^{+0.18}_{-0.15} \pm 0.12$	$5.1^{+1.6}_{-1.3}$	$5.9^{+1.4}_{-1.2}$	8
10			200–300	$1.04^{+0.14}_{-0.13} \pm 0.11$	$2.17^{+0.79}_{-0.60}$	$2.31^{+0.73}_{-0.57}$	2
11			300–400	$0.72^{+0.33}_{-0.22} \pm 0.08$	$0.06^{+0.11}_{-0.04}$	$0.72^{+0.53}_{-0.33}$	4
12			$>$400	$1.24^{+0.67}_{-0.45} \pm 0.10$	$0.89^{+1.42}_{-0.60}$	$0.52^{+0.65}_{-0.35}$	0
13		4	150–200	$1.26^{+0.21}_{-0.20} \pm 0.23$	$2.68^{+1.06}_{-0.79}$	$2.58^{+0.85}_{-0.67}$	1
14			200–300	$1.21^{+0.22}_{-0.21} \pm 0.22$	$1.26^{+0.62}_{-0.44}$	$1.62^{+0.65}_{-0.48}$	3
15			300–400	$2.35^{+0.88}_{-0.72} \pm 0.34$	$0.42^{+0.61}_{-0.27}$	$1.16^{+0.87}_{-0.55}$	1
16			$>$400	$0.94^{+0.53}_{-0.36} \pm 0.13$	$0.67^{+1.10}_{-0.46}$	$0.78^{+0.76}_{-0.43}$	1

Bin	$N_{\rm H}$	$p_{\rm T}^{\rm miss}$ [GeV]	$N_{\rm SR,\,tot}^{\rm pred}$	$f_{p{\rm Tmiss}}$	$N_{\rm SR}^{\rm pred}$	$N_{\rm SR}^{\rm fit}$	$N_{\rm SR}^{\rm obs}$
17	1	300–500	42.6 ± 4.2	0.789 ± 0.030	$33.6^{+6.1}_{-5.2}$	$37.0^{+4.2}_{-4.0}$	42
18		500–700		0.172 ± 0.028	$7.3^{+2.0}_{-1.6}$	$7.2^{+1.5}_{-1.3}$	6
19		$>$700		0.039 ± 0.014	$1.65^{+1.04}_{-0.66}$	$1.50^{+0.75}_{-0.53}$	1
20	2	300–500	5.1 ± 1.0	0.789 ± 0.030	$4.0^{+1.5}_{-1.1}$	$4.0^{+1.2}_{-1.0}$	4
21		500–700		0.172 ± 0.028	$0.88^{+0.40}_{-0.28}$	$0.74^{+0.29}_{-0.21}$	0
22		$>$700		0.039 ± 0.014	$0.20^{+0.21}_{-0.10}$	$0.14^{+0.13}_{-0.07}$	0

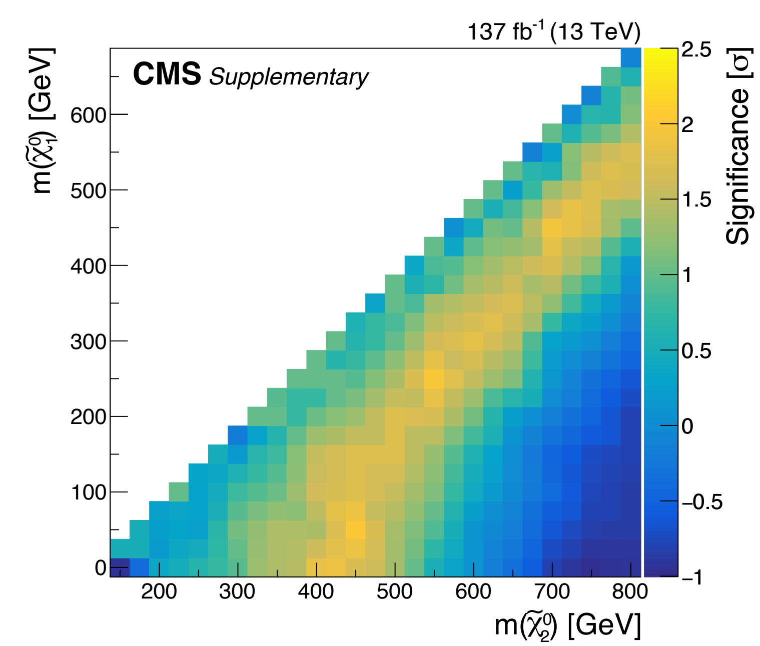
covariance (correlation shown here)



efficiency (by bin available)



significance



- HEPData tables downloadable as csv, yaml, …

Chisquare method

- Simplest approximation: build the likelihood as a product of Gaussians.
- Then minimize χ^2 = - 2 ln(L):
- Define the residual Δ_i for each bin i:

$$\Delta_i \equiv N_i^{\rm obs} - N_i^{\rm pred},$$
$$N_i^{\rm pred} \equiv N_i^{\rm bkg} + \mu N_i^{\rm sig},$$
$$N_i^{\rm sig} = \epsilon_i \mathcal{B}^2(H \to b\bar{b})\sigma\mathcal{L}$$

  - μ is the signal strength
  - ε, $\boldsymbol{\sigma}$, and $\mathcal{L}$ are the efficiency, cross section, and luminosity
- Given the covariance matrix V, we find

$$\chi^2 = \Delta_i V_{ij}^{-1} \Delta_j,$$

  - $$\to \sum \Delta_i^2/\sigma_i^2$$ in the limit of uncorrelated bins.

Accounting for signal uncertainty in chisq

- The covariance matrix is given for the background.
- In the chisq formulation we need to add the contribution of the signal uncertainty:

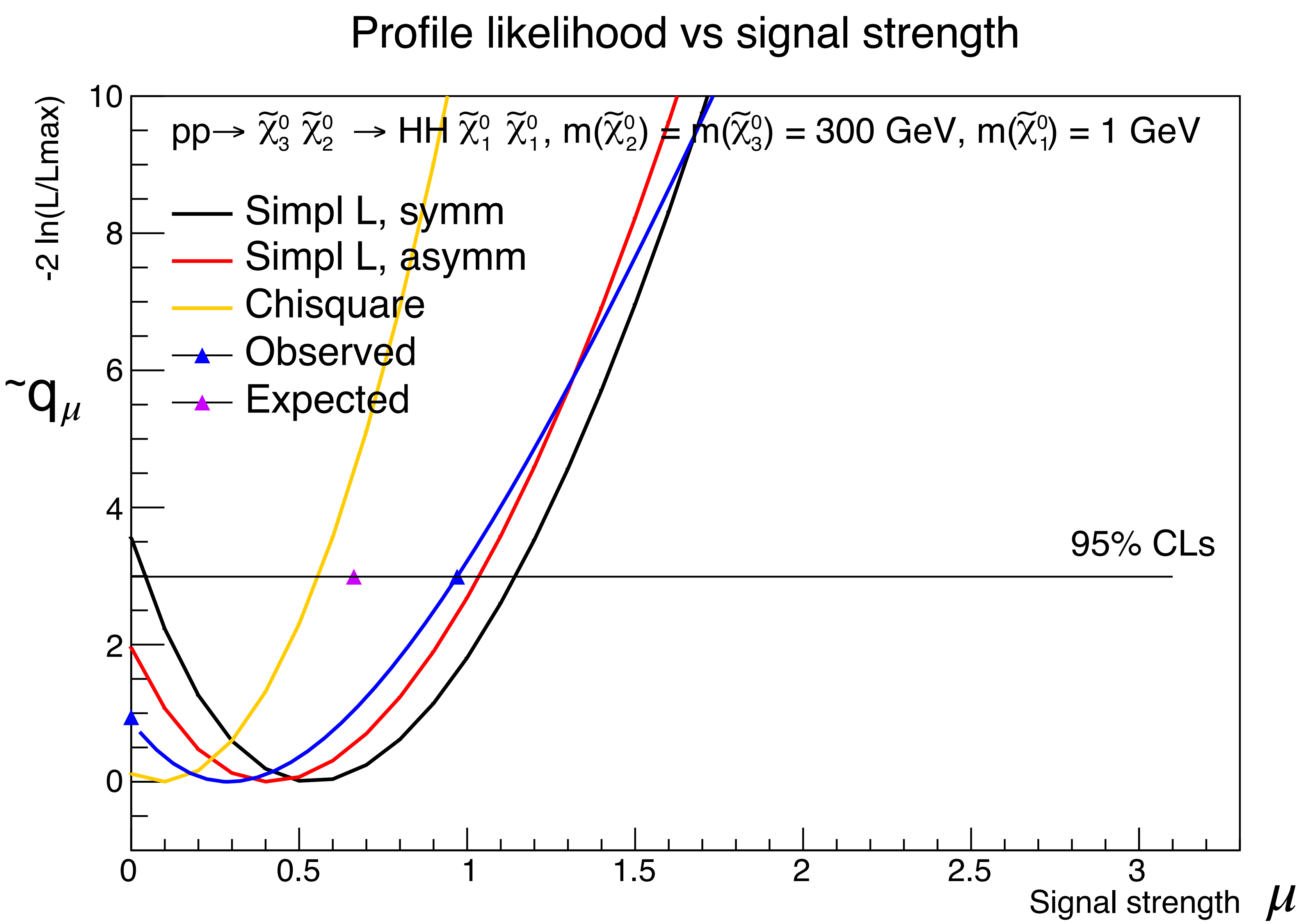
$$V = V^{\rm bkg} + {\rm diag}(N^{\rm obs})$$

  - N^{obs} being the square of the Poisson uncertainty $\sqrt{N^{obs}}$
- The resulting function $\chi^2(\mu)$ is a parabola, with minimum at the best-fit value of μ, μ_0.
- The square root of $\chi^2(\mu)$ - $\chi^2(\mu_0)$ at any other value of μ gives the Z-value (number of standard deviations of a Gaussian pdf).

Limitations of the chisq method

- All uncertainties assumed Gaussian
  - and symmetric
  - not so good for small yields
  - V^{bkg} reflects the combined effects of all sources (CR yields, scale factors, JECs, etc., whereas some contributions are statistical (sideband yields).
  - V^{obs} represents Poisson stat uncertainties with a Gaussian approximation
- Signal uncertainties are not accounted for.
- The contribution of each bin to chisq represents the consistency of background and observation in that bin.
  - But the real question is, given N^{bkg}, what's the probability of observing N^{obs}?
  - E.g., the bin 11 contribution before squaring is (very nearly) $(4 - 0)/\sqrt{4}$, which is 2 sigma, vs the detailed study giving 3.3 sigma local significance.

Chisquare method for HH+met



- Orange curve from the χ^2 method
  - underestimates μ_0
  - underestimates the high-side uncertainty

Simplified likelihood method

- Feedback from phenomenology team (Wolfgang Waltenberger): need efficiencies by SR bin
  - We now provide these in HEPData (and public twiki)
- Wolfgang also spoke at the June SUSY workshop
- Introduced the simplified likelihood method, SM:
  - https://arxiv.org/abs/1809.05548, https://link.springer.com/article/10.1007/JHEP04(2019)064
- Reproduces one of RA2/b Run 2 limits quite well
- Improvements over chisq method:
  - Poisson pdfs for the observed yields
  - Option to include asymmetric background uncertainties:

Building the simplified likelihood

- The predicted yield in bin i is

$$N_i^{\text{pred}} \equiv N_i^{\text{bkg}} + \mu N_i^{\text{sig}},$$
$$N_i^{\text{bkg}} = a_i + b_i\theta_i + c_i\theta_i^2$$

  - a_i is the central value of the bkg prediction
  - θ_i is a nuisance parameter drawn from a unit Gaussian
  - b_i is the effective sigma of the bkg uncertainty, $\sqrt{V_{ii}}$ in the limit of symmetric uncertainties
  - c_i gives the asymmetry of the bkg uncertainty
- The simplified likelihood is

$$L_S(\mu,\theta) \propto \prod_i \text{Pois}(N_i^{\text{obs}}|N_i^{\text{pred}}(\mu)) \exp\left(-\tfrac{1}{2}\theta_i\rho_{ij}^{-1}\theta_j\right)$$

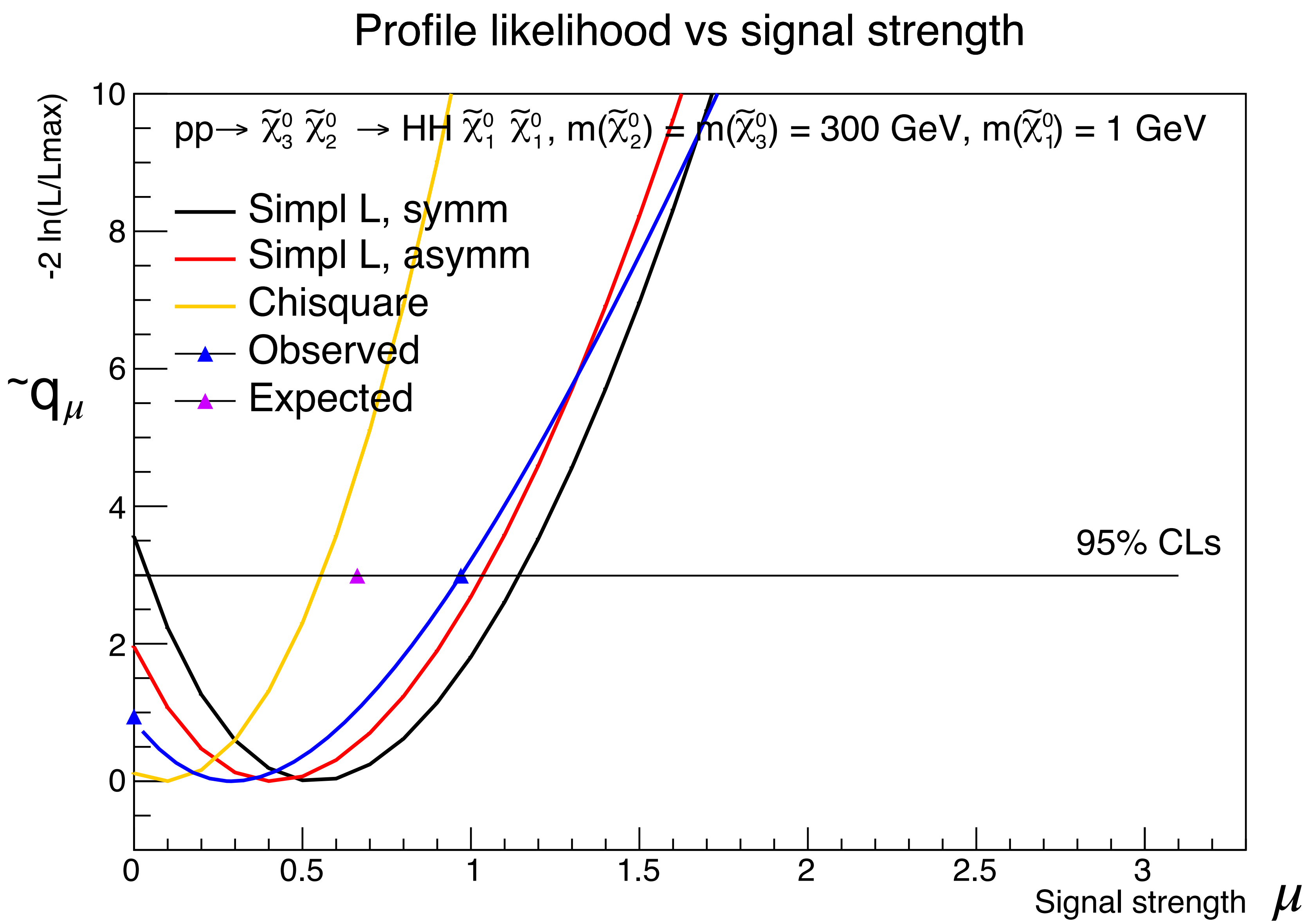
  - where $\rho \rightarrow$ correlation matrix in the symmetric limit
- The code, and example input, are in gitHub:
  - https://gitlab.cern.ch/SimplifiedLikelihood/SLtools

SL: asymmetric bkg uncertainties

- The covariance matrix gives second moments, i.e., sigma^2, on the diagonal, and correlations, on off-diagonal elements
- To incorporate asymmetric uncertainties, SL uses the diagonal elements of the 3rd moment m_3 of the background nuisances.
  - Maybe these could be extracted from combine tool?
- Compute m_3 from a bifurcated Gaussian using the asymmetric uncertainties $\sigma_{1,2}$:

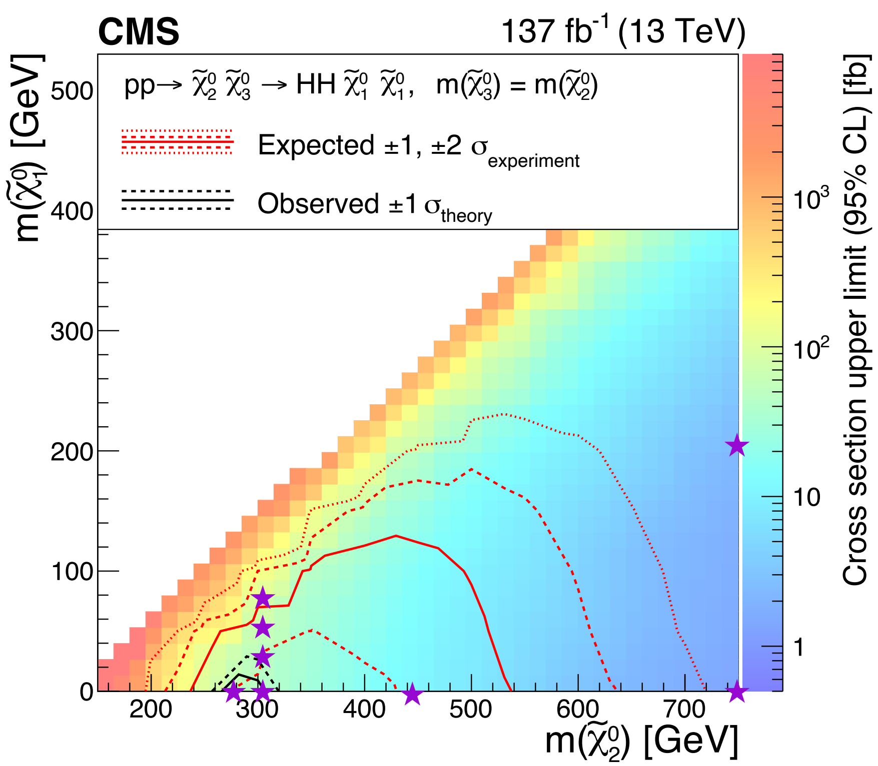
$$m_3 = \frac{2}{\sigma_1 + \sigma_2}\left[\sigma_1 \int_{-\infty}^{0} x^3 G(x;0,\sigma_1)dx + \sigma_2 \int_{0}^{+\infty} x^3 G(x;0,\sigma_2)dx\right]$$

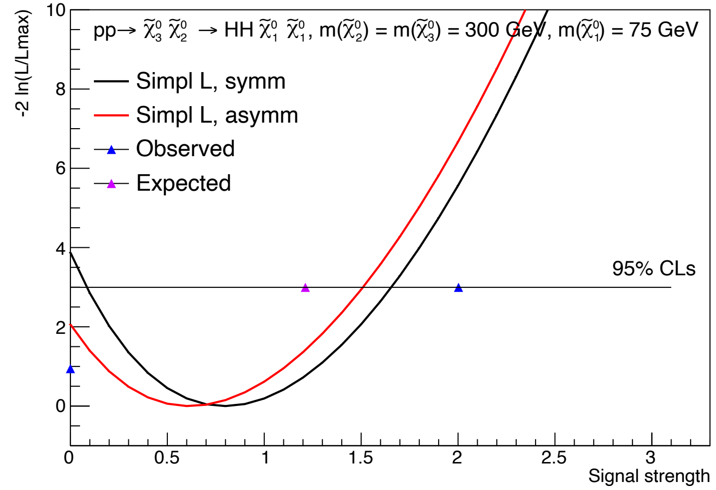
Compare SL with full L

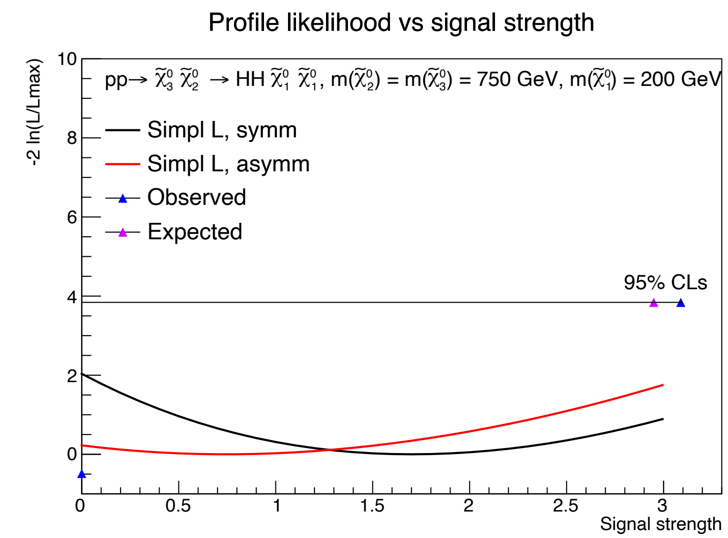


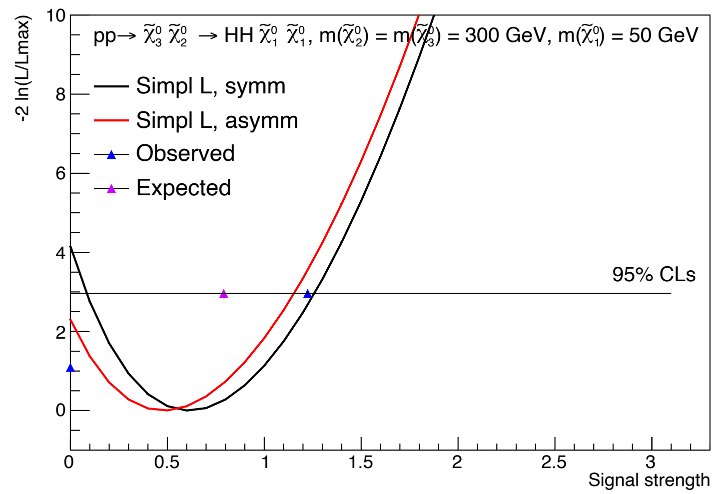
- The red (black) curve is SL with(out) asymmetric uncertainties
- SL treats N^{obs} as Poisson,
- N^{bkg} as (asymmetric) Gaussian
  - doesn't fully account for Poisson fluctuations of low-stats CR yields

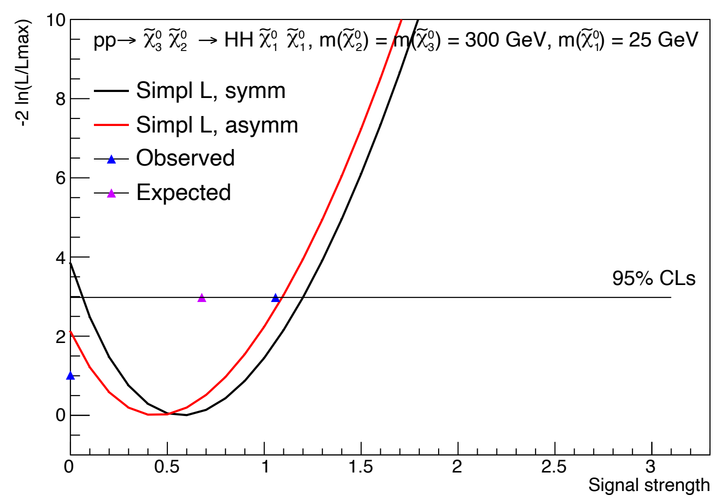
L & SL vs scan pt

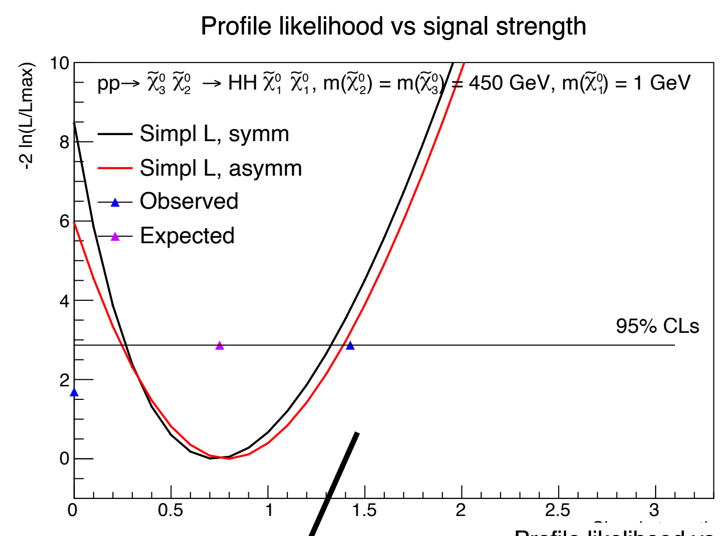


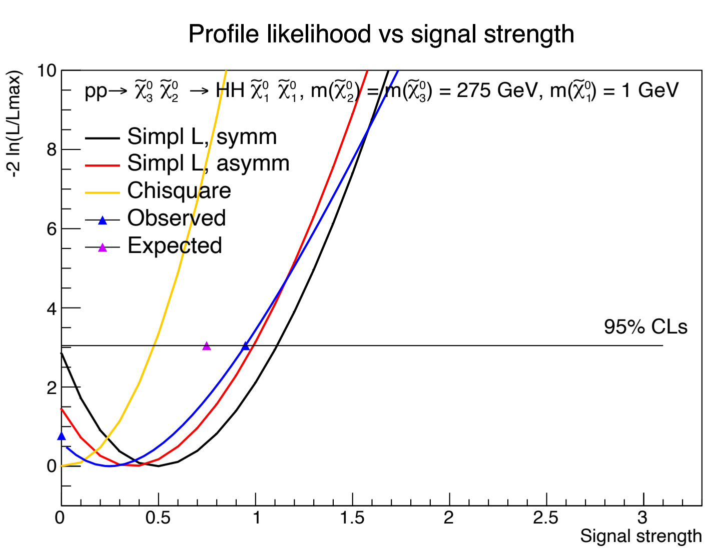




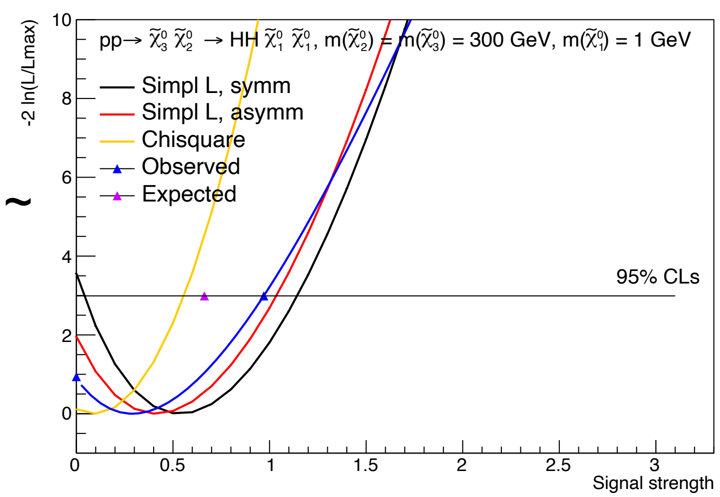


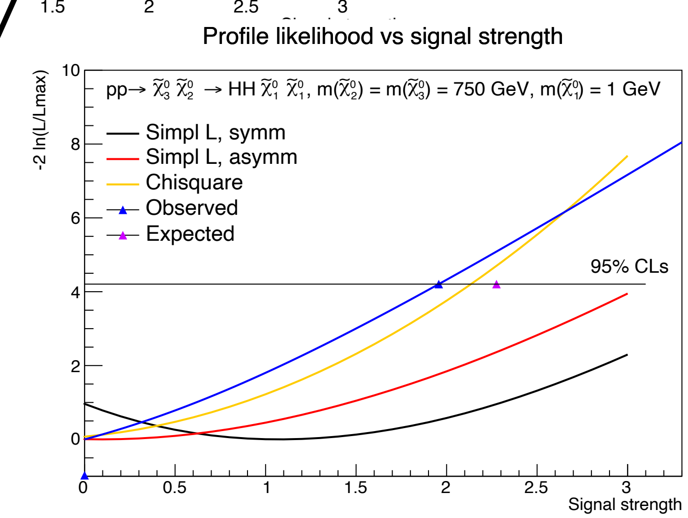






~





Some observations

- Phenomenologists have rather sophisticated tools for incorporating published data from experiments
- For the HH+met and RA2/b papers, the provided information is adequate for reasonably closely reproducing the results
- How can one test other models with our HH+met data?
  - Reweight bin efficiencies?
    - The b quark content would have to be the same, else sorting of the model into 3b, 4b, or 1H, 2H bins would be impossible
    - Sorting the model into resolved/boosted, maybe with a cut on ΔR between the H daughter quarks?
    - MET OK
    - ΔR_{max} OK?
- For some ATLAS papers, “likelihood” is provided.
  - What would this entail?

Update: combine with lnN nuisances

- Jaebak presented a fit made with combine, but with
  - backgrounds entered directly from the paper
  - bkg uncertainties from the paper, entered into the datacard as (asymmetric) lnN nuisances
  - no correlations
- I've modified this approach by adding the bkg correlations
  - Same lnN nuisances, but multiplied by the correlation matrix elements to extend to the off-diagonal terms
  - Results shown as green plots in the profile plots (following page)
- Compared with Waltenberger's simplified likelihood
  - these do much better estimating the significance
  - not necessarily better in predicting the limits

L & SLs vs scan point