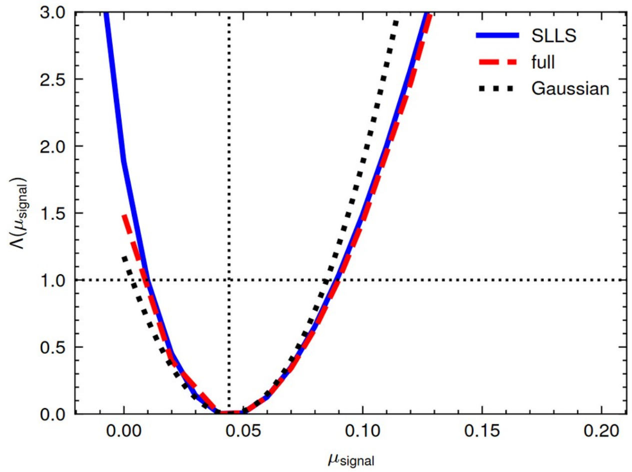
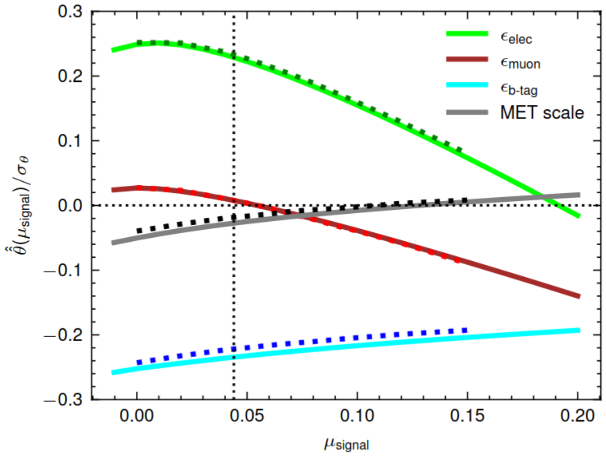
Simplified likelihoods with linearized systematics

JHEP04(2023)084

Nicolas Berger (CERN & LAPP)





Publishing experimental results

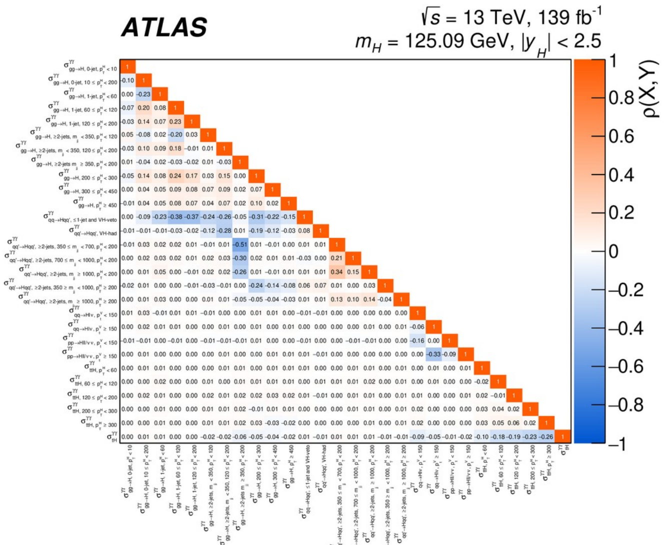
CERN-EP-2022-094 (accepted by JHEP)

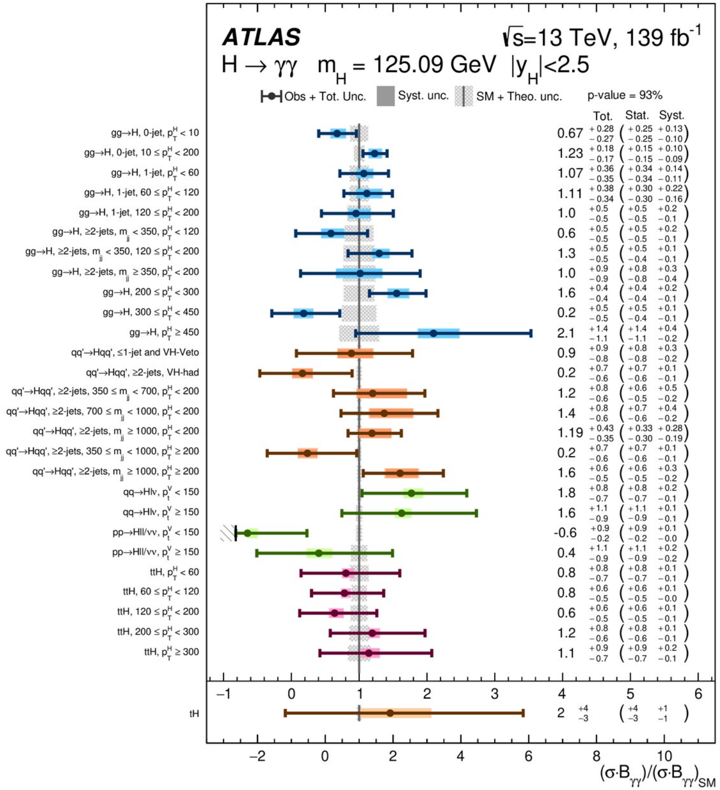
Example: ATLAS full-Run 2 H→γγ "couplings" analysis : measure σ_i in various kinematic regions ("STXS")

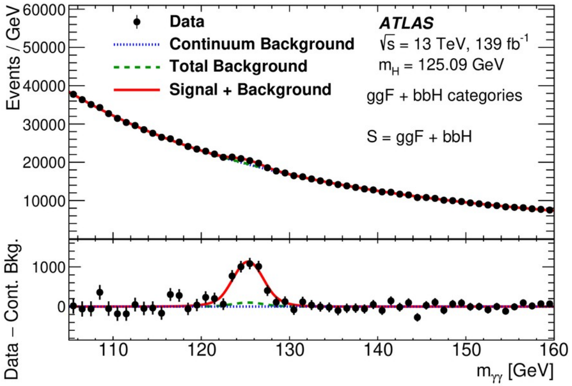
- Some high-stats, syst-dominated regions (e.g. targeting gg→ H);
- Also low-stats regions for rare mode (pp→tH, high-p_T gg→H)

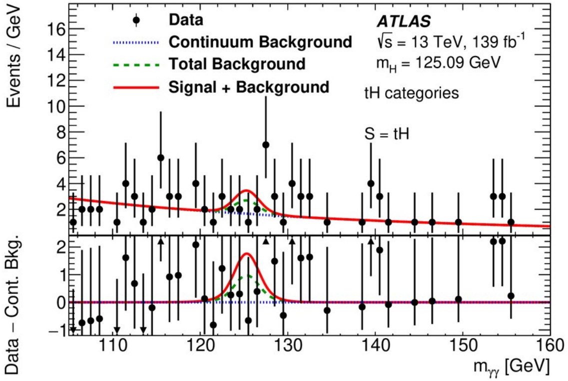
What we report:

- Best-fit values + uncertainties
- Correlation matrix









As seen from the collaboration

Results obtained from a statistical model describing the full measurement:

- Signal normalization (σ_i) + systematic uncertainties (nuisance parameters θ_k)
- Background normalization and $m_{\gamma\gamma}$ shapes (more θ_k)

W. Verkerke, SOS 2014

Get results from:

$$\textbf{model (PDF)}: \; p(n_a; \sigma_i, \theta_k)$$

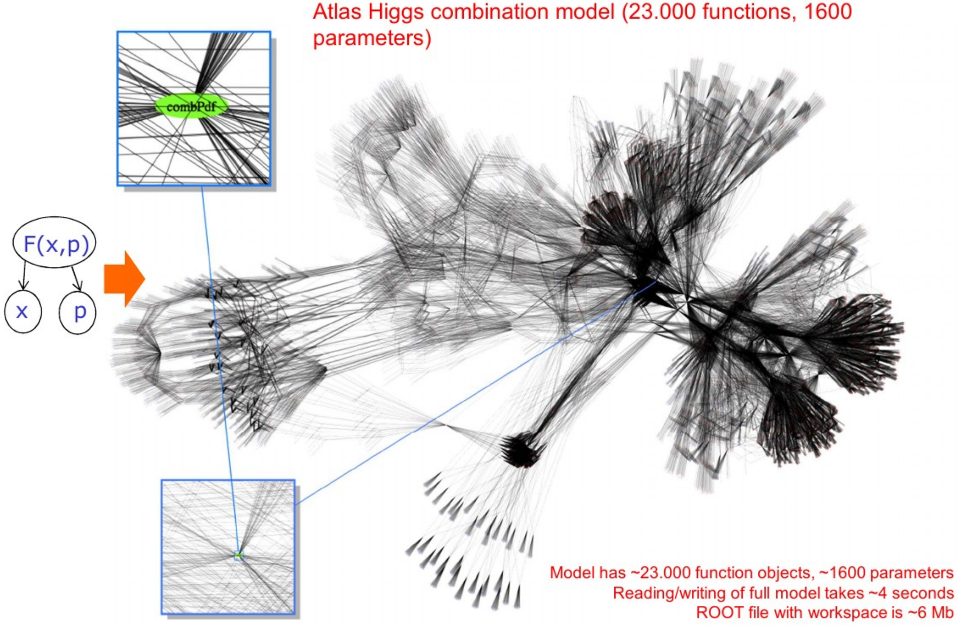
$$\textbf{observed data}: \; n_a^{\text{obs}}$$

Implemented within ROOT as a "workspace"

~O(few 10s) parameters of interest (POIs) σ_i

~O(few 1000s) nuisance parameters (NPs) θ_k

Few hours/days per fit



Likelihoods in HEP results

Usual description of LHC measurements: measurement PDF, a.k.a the **likelihood** L with

- Parameters of interest (μ)
- Nuisance parameters (θ).

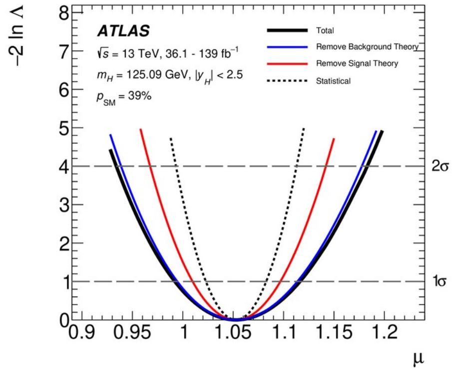
$$L(\mu, \theta)$$

The nuisance parameters describe systematics and other "nuisances" fit to data.

Profile likelihood : uses "profiled" values $\hat{\hat{\theta}}(\mu)$ of the NPs = best-fit values at given POI values.

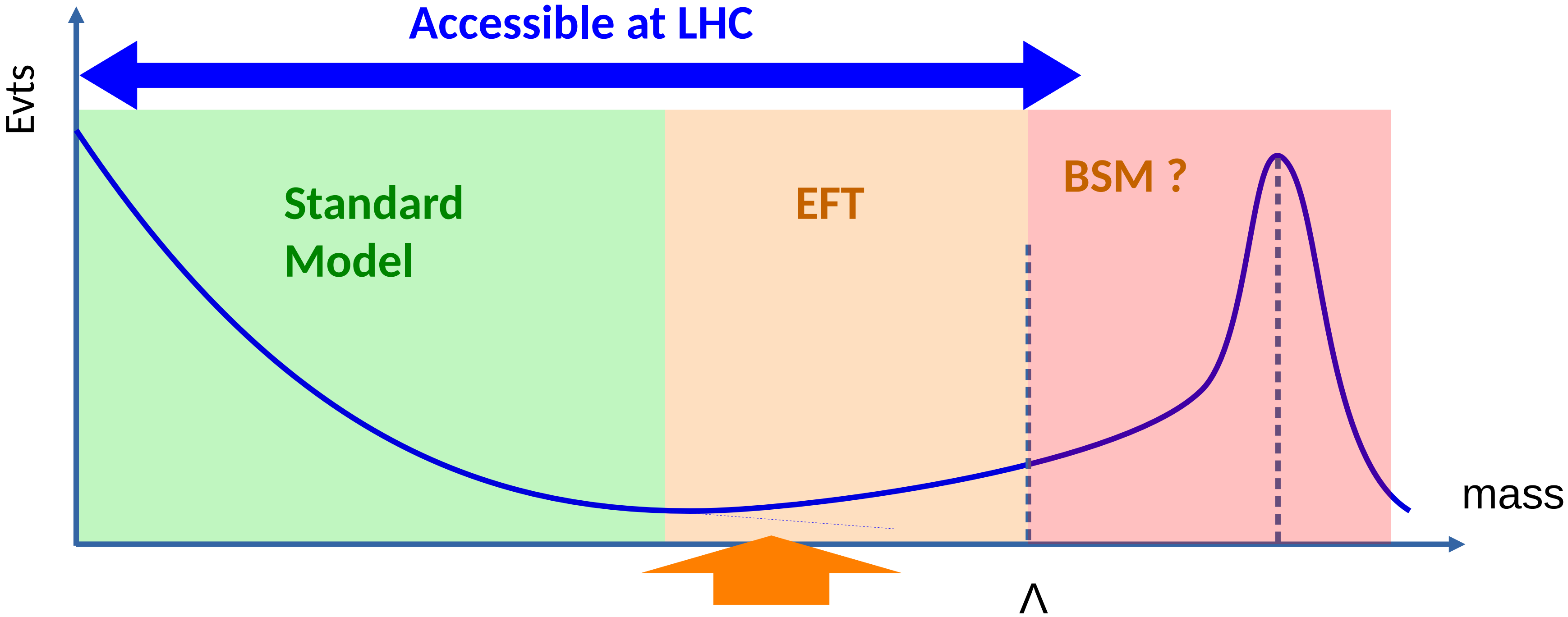
$$\Lambda(\mu) = -2 \log \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})}$$

Used compute e.g. a confidence interval:
same for upper limits, etc.



Reuse and reinterpretation

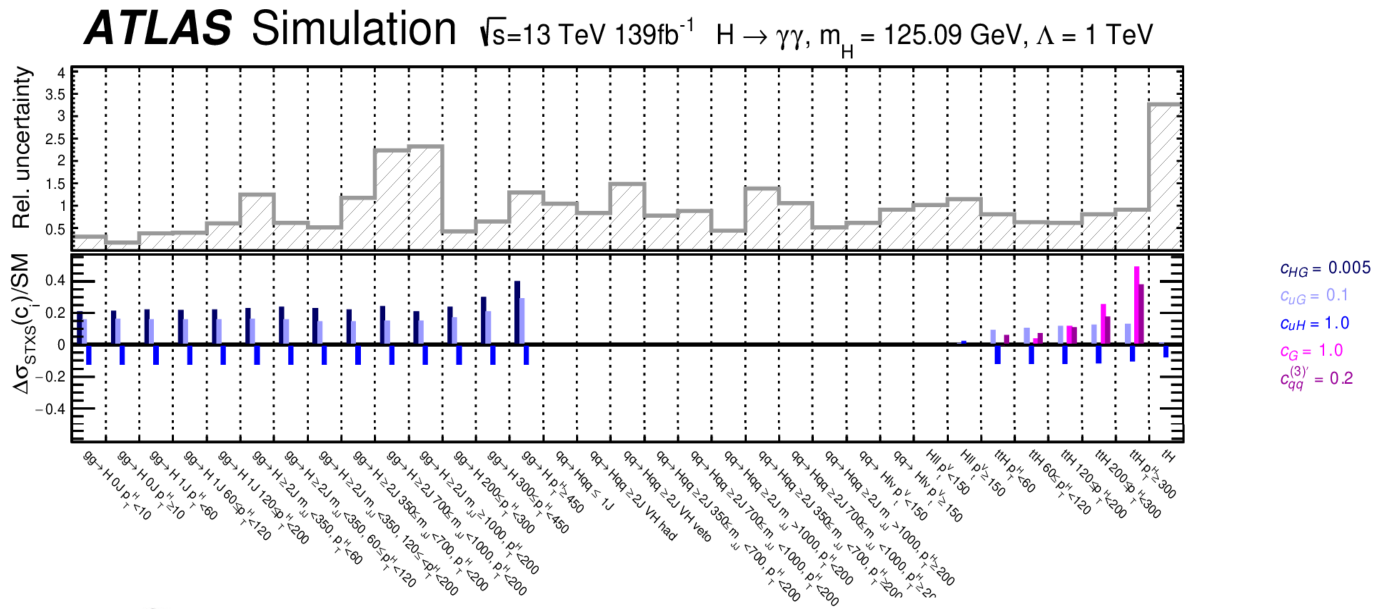
Ultimate goal: use σ_i to constrain new physics models e.g. EFT



$$\mathscr{L}_{EFT} = \mathscr{L}_{SM} + \frac{c^{(5)}}{\Lambda} O^{(5)} + \sum_i \frac{c_i^{(6)}}{\Lambda^2} O_i^{(6)} + \sum_i \frac{c_i^{(7)}}{\Lambda^3} O_i^{(7)} + \sum_i \frac{c_i^{(8)}}{\Lambda^4} O_i^{(8)} + \ldots$$

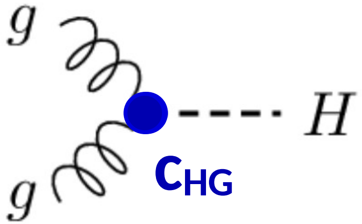
Reuse and reinterpretation

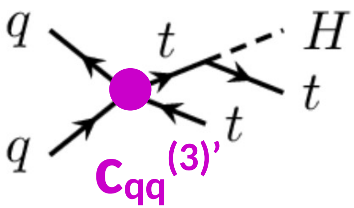
Ultimate goal: use σ_i to constrain new physics models e.g. EFT



EFT constraints from

- **Precision measurements** in High-stats, syst-dominated regions
- **High-pT regions** with large EFT effects and low stats.

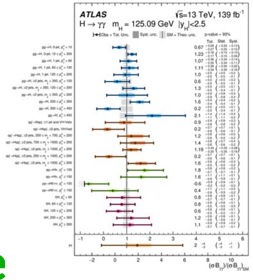




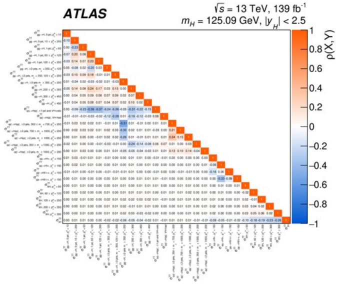
Reuse and reinterpretation

Ultimate goal: use σ_i to constrain new physics models e.g. EFT

- **Using published results**



+



Gaussian model

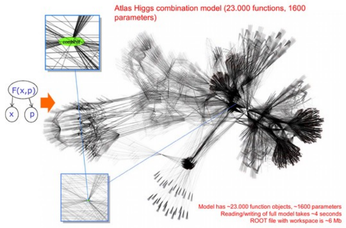
$$\longrightarrow G\left(\boldsymbol{n}_a^{\mathbf{obs}};\boldsymbol{\sigma}_i\right) \longrightarrow G\left(\boldsymbol{c}_i^{\mathbf{EFT}}\right)$$

⊕ **Simple, publicly available**

⊖ Gaussian approximation only (no low stats!)

⊖ No nuisance parameters: cannot correlate systematics (e.g. when including CMS)

- **Using full workspace**



Full model

$$\longrightarrow p\left(\boldsymbol{n}_a^{\mathbf{obs}};\boldsymbol{\sigma}_i,\boldsymbol{\theta}_k\right) \longrightarrow L\left(\boldsymbol{c}_i^{\mathbf{EFT}}\right)$$

⊕ **Exact model, including non-Gaussian effects; All NPs: can correlate systematics**

⊖ Model not always accessible outside the collaboration

⊖ Long fits times!

Likelihood publication: some history

(from Sabine Kraml)

- **2000 First PHYSTAT workshop** [CERN 2000-005]

  Unanimous agreement that particle physicists should publish likelihood functions, given their fundamental importance in extracting quantitative results from experimental data.

- **2012 Les Houches** Recommendations **for the Presentation of LHC Results**

  Recommendation 3b: When feasible, provide a mathematical description of the final likelihood function in which experimental data and parameters are clearly distinguished, [....].

  Recommendation 3c: Additionally provide a digitized implementation of the likelihood that is consistent with the mathematical description.

- **2020:** Reinterpretation of LHC Results for New Physics: Status and Recommendations after Run 2 [**SciPost Phys. 9, 022 (2020)**]

- **2021:** White paper **on publishing statistical models** →

Publishing statistical models: Getting the most out of particle physics experiments

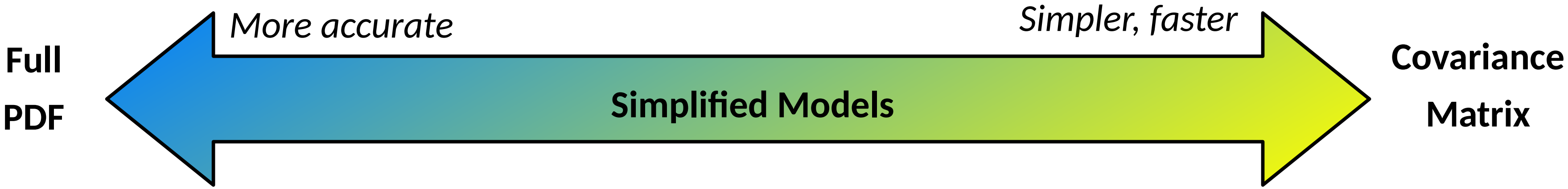
Kyle Cranmer[1*], Sabine Kraml[2‡], Harrison B. Prosper[3§] (editors), Philip Bechtle[4], Florian U. Bernlochner[4], Itay M. Bloch[5], Enzo Canonero[6], Marcin Chrzaszcz[7], Andrea Coccaro[8], Jan Conrad[9], Glen Cowan [10], Matthew Feickert[11], Nahuel Ferreiro Iachellini[12,13] Andrew Fowlie[14], Lukas Heinrich[15], Alexander Held[1], Thomas Kuhr[13,16], Anders Kvellestad[17], Maeve Madigan[18], Farvah Mahmoudi [15,19], Knut Dundas Morå[20], Mark S. Neubauer[11], Maurizio Pierini[15], Juan Rojo[8], Sezen Sekmen[22], Luca Silvestrini[23], Veronica Sanz[24,25], Giordon Stark[26], Riccardo Torre[8], Robert Thorne[27], Wolfgang Waltenberger[28], Nicholas Wardle[29], Jonas Wittbrodt[30]

The statistical models used to derive the results of experimental analyses are of incredible scientific value and are essential information for analysis preservation and reuse. In this paper, we make the scientific case for systematically publishing the full statistical models and discuss the technical developments that make this practical. By means of a variety of physics cases — including parton distribution functions, Higgs boson measurements, effective field theory interpretations, direct searches for new physics, heavy flavor physics, direct dark matter detection, world averages, and beyond the Standard Model global fits — we illustrate how detailed information on the statistical modelling can enhance the short- and long-term impact of experimental results.

Current Situation

- **Full likelihood** publication is gathering steam (pyhf, ROOT)
  - ⊕ **Support for non-Gaussian effects,** both from small yields (Poisson PDFs) and systematics.
  - ⊕ **Independent NPs for systematics**: can properly correlate systs (not always trivial in practice!)
  - ⊖ **Models sometimes quite large**: difficult to handle and long to evaluate (few min - few hours)
  - ⊖ **More difficult to tackle unbinned models** (needs arbitrary PDFs, requires e.g. roofit…)

- **Simplified likelihoods** provide intermediate solutions. Many flavors:
  - ATLAS SUSY SLs [ATL-PHYS-PUB-2021-038] : Poisson PDFs, 1 NP for systs.
  - Simplify [JHEP04(2019)064] : Poisson PDFs, Keep all POIs, 1 NP per bin with quadratic impact
  - DNNLikelihood [Eur. Phys. J. C 80, 664 (2020)]: train a DNN to approximate the likelihood function



Simplified Likelihoods using Linearized Systematics

JHEP04(2023)084

Requirements:

- **Describe non-Gaussian effects** from small event counts (Poisson behavior)
- **Preserve all POIs** ⇒ allow reinterpretation through reparameterization
- **Preserve all NPs** ⇒ allow correlation of systematic uncertainties

A particular use-case: SMEFT interpretations

- Reparameterize cross-section measurements using EFT Wilson coefficients: $\sigma_i = f(c_\alpha)$
- Important constraints from both
  - High-mass tail regions (e.g. pp → VV, ll) ⇒ need Poisson description
  - Syst. dominated precision measurements (e.g. W,Z, top, Higgs) ⇒ need accurate syst. treatment

How can we simplify ??

1. **Keep all NPs/systematics but at linear order only.**
2. **Assume all systematics are Gaussian.** (common assumption even for full likelihoods)

Starting point: the HistFactory description

Binned likelihood form, with parameters of interest (**μ**) and nuisance parameters (**θ**) :

$$L(\mu, \theta) = \prod_{c=1}^{n_\text{channels}} \prod_{b=b_c^\text{first}}^{b_c^\text{last}} \textbf{Pois}\left(n_b^\text{obs} ; \sum_{s=1}^{n_\text{samples}} N_{s,b}^\text{exp}(\mu, \theta) \right) \prod_{p=1}^{n_\text{syst NPs}} C(a_p ; \theta_p)$$

Product over channels (independent measurement regions)

Product over channel bins

Poisson PDF in each bin

Observed bin yield

Expected bin yield, function of both POIs and NPs.

- Several possible forms: linear, exponential, etc.
- Implements correlations between bins

NP constraints (from auxiliary measurements a)

Constrained nuisance parameters describe systematic uncertainties

Simplified Likelihoods with Linearized Systematics

→ Consider **NP effects** at **linear order only**.

→ Consider only **Gaussian constraints**

→ Keep **full description** of **bin counting** (Poisson PDF) and **POIs** (μ)

Exact treatment of the POI μ

NP dependence at linear order only

$$N_{sb}^{\exp}(\mu,\theta) = N_{sb}^{\text{nom}}(\mu)\left(1 + \sum_p \Delta_{sbp}\theta_p\right)$$

Impact coefficients

$$L(\mu,\theta) = \prod_{c=1}^{n_{\text{channels}}} \prod_{b=b_c^{\text{first}}}^{b_c^{\text{last}}} \text{Pois}\left(n_b^{\text{obs}}; \sum_{s=1}^{n_{\text{samples}}} N_{sb}^{\exp}(\mu,\theta)\right) \prod_{p=1}^{n_{\text{non-free NPs}}} C(a_p;\theta_p)$$

$$L(\mu,\theta) = \prod_{c=1}^{n_{\text{channels}}} \prod_{b=b_c^{\text{first}}}^{b_c^{\text{last}}} \text{Pois}\left(n_b^{\text{obs}}; \sum_{s=1}^{n_{\text{samples}}} N_{sb}^{\text{nom}}(\mu)\left(1 + \sum_p \Delta_{sbp}\theta_p\right)\right) G(\theta_p^{\text{obs}}; \theta_p, \Gamma^{-1})$$

Linear NP impacts

Gaussian constraints

Simplified Likelihoods with Linearized Systematics

JHEP04(2023)084

→ Consider **NP effects** at **linear order only**.

→ Consider only **Gaussian constraints**

→ Keep **full description** of **bin counting** (Poisson PDF) and **POIs** (μ)

P, Q matrices built from **N**^{nom}(μ, θ) and **Δ**

Benefit: fast profiling!

- Minimization wrt NPs is a simple matrix operation

$$\hat{\hat{\theta}}(\mu) = [\Gamma + P(\mu)]^{-1}[\Gamma\ \theta^{obs} - Q(\mu)]$$

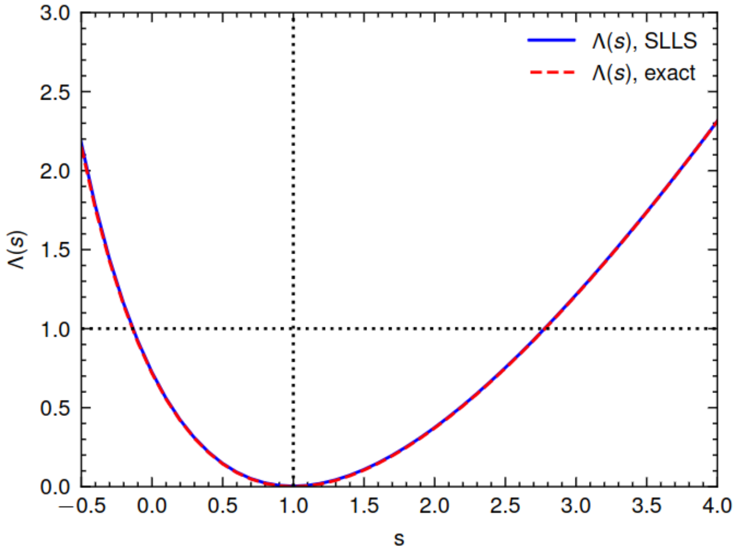
- **Obtain the profile likelihood Λ(μ) more quickly than with the full likelihood.** POIs are treated exactly (non-linear minimization, as in full likelihood). Typically $N_{POIs} \ll N_{NPs}$...
- All NPs retained: can correlate everything across measurements as for full likelihoods
- Usually an excellent approximation
  - Searches have small systematics ⇒ OK to linearize
  - Precision measurements often in Gaussian regime ⇒ well described by linear systematics.
- Cannot describe asymmetric and non-Gaussian uncertainties

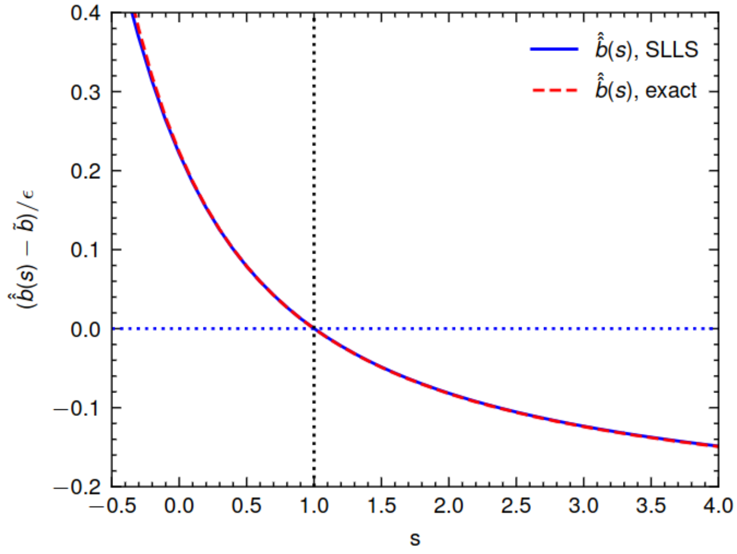
Simple example

JHEP04(2023)084

Simple S+B counting experiment, B = 1 ± 0.25, observe n=2

Describe the uncertain background using an NP: $L(s,b) = \mathrm{Pois}(2, s+b)\, G(1; b, 0.25)$





In this (simple!) case, can compute everything in closed form

SLLS gives very precise estimation of Λ(s) and $\hat{\hat{b}}$(s)

⇒ Describes both Poisson effects and systematics.

$$\hat{\hat{b}}(s) = \frac{1}{2}\left[\sqrt{(s + \tilde{b} - \tilde{b}^2\epsilon^2)^2 + 4\tilde{b}^2\epsilon^2 n} - (s - \tilde{b} + \tilde{b}^2\epsilon^2)\right]$$

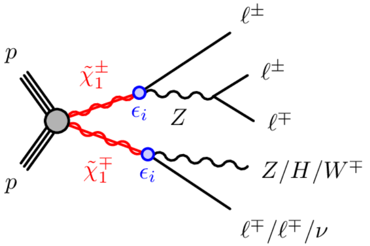
$$\Lambda(s) = 2(s - \hat{s} + \hat{\hat{b}}(s) - \hat{b}) - 2n\log\left(\frac{s + \hat{\hat{b}}(s)}{\hat{s} + \hat{b}}\right),$$

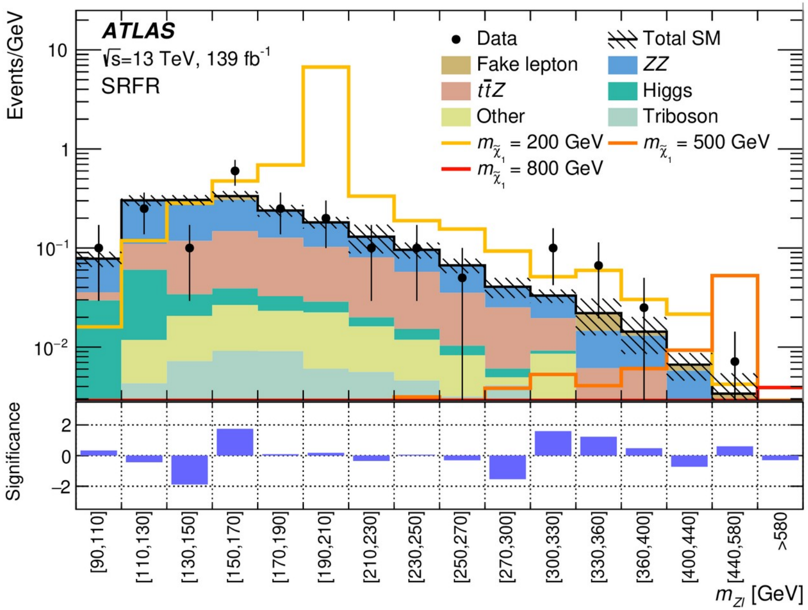
Not-so-simple Example: ATLAS SUSY search in trilepton states

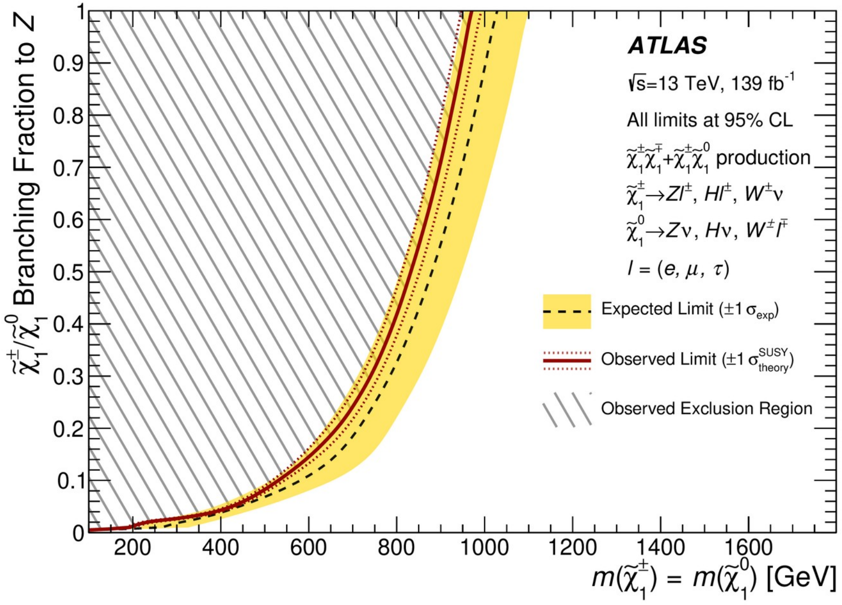
ATLAS search for charginos in trilepton final states (χ+ → Z(ll)l) from Phys. Rev. D 103, (2021) 112003

– 3 signal regions (3l, 4l, full-reco (FR)), binned in m_{Zl}.

– 3 control regions for main backgrounds (CRWZ, CRZZ, CRttZ)

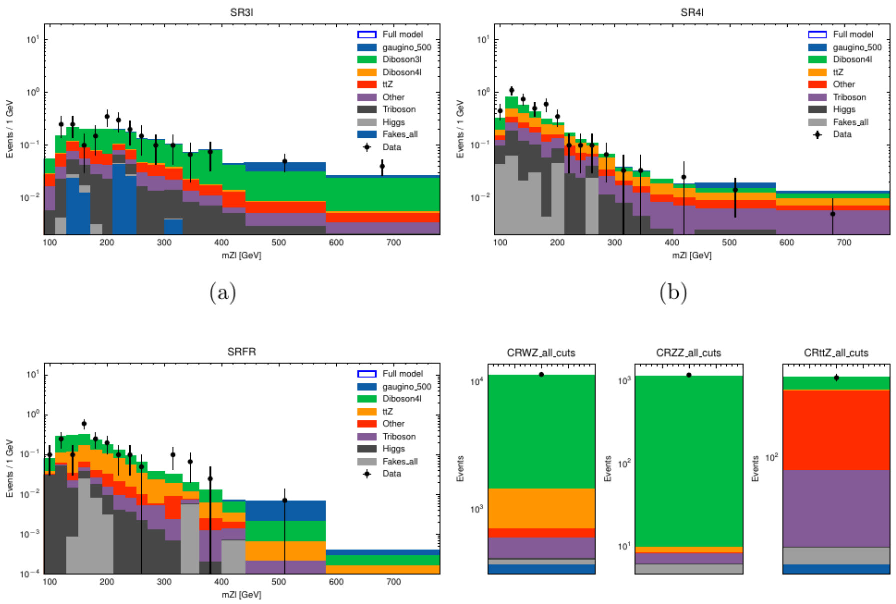


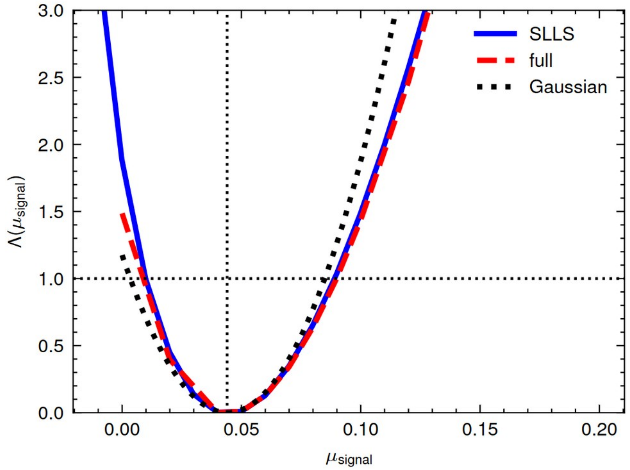


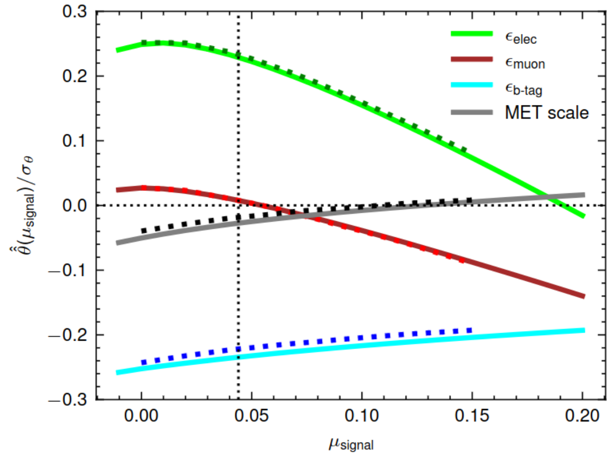


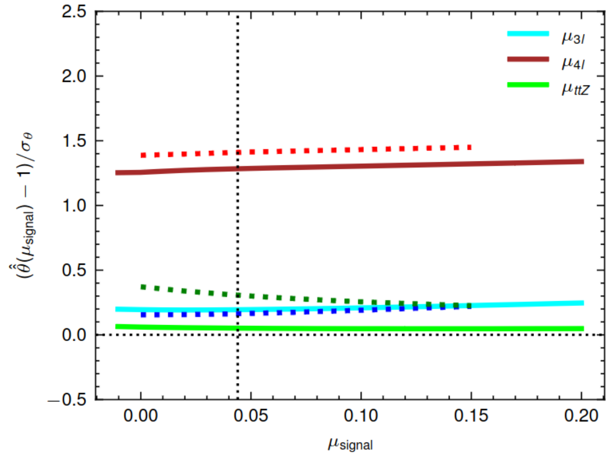
ATLAS SUSY search in trilepton states

- Use the pyhf workspace description of the analysis likelihood, available on HEPdata
- Produce a SLLS linearized likelihood using an automatic script.
- Check profile likelihood scan and profiled values for various model points
- <1s per fit on a laptop, O(1000) faster than full L









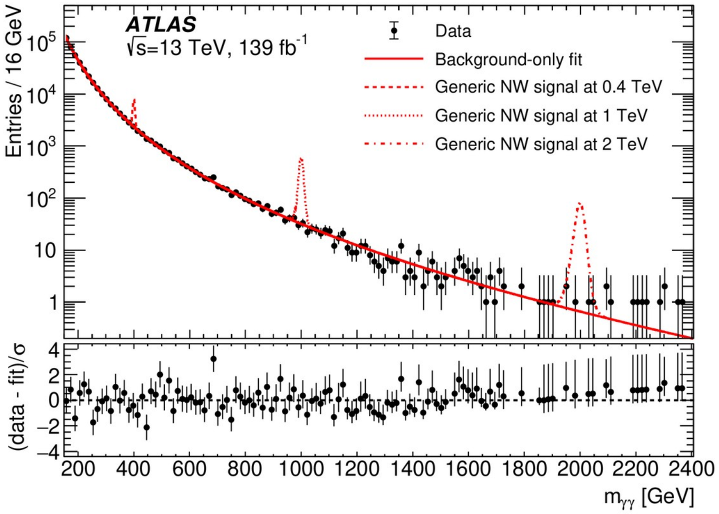
Unbinned models

Phys. Lett. B 822 (2021) 136651

Models described in previous slides are ***binned*** : just counting events in bins.

Some analyses with smooth backgrounds (H/X→γγ, H→μμ, X→jj, …) typically use ***unbinned*** modeling instead → Describe the shape of a continuous observable

Difficult problem: need to implement **all** the PDFs required to model signal and background.



→ Can describe the unbinned distribution as a set of very fine bins, and go back to a binned description

→ Large number of bins required, but feasible for simplified models.

Unbinned example: toy H→γγ measurement

JHEP04(2023)084

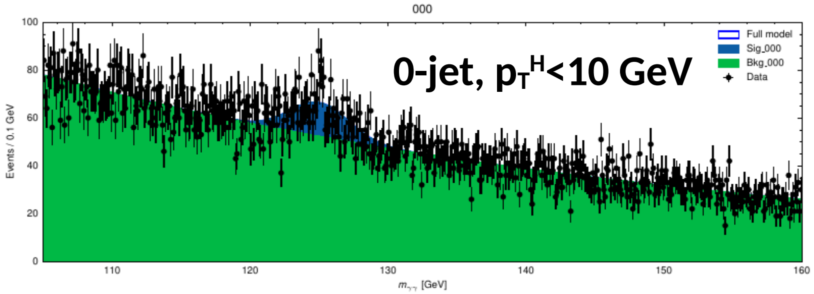
Build a toy unbinned likelihood from the ggF and VBF regions of the ATLAS H→γγ measurement in 2207.00348.

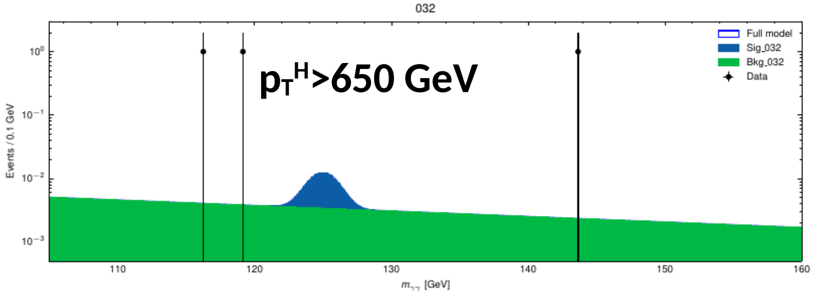
Full model: unbinned measurement of μ over $105 < m_{\gamma\gamma} < 160$ GeV, in 33 measurement regions.

SLLS: discretize into bins of 0.1 GeV($\ll \sigma_H \sim$ 1-2 GeV)

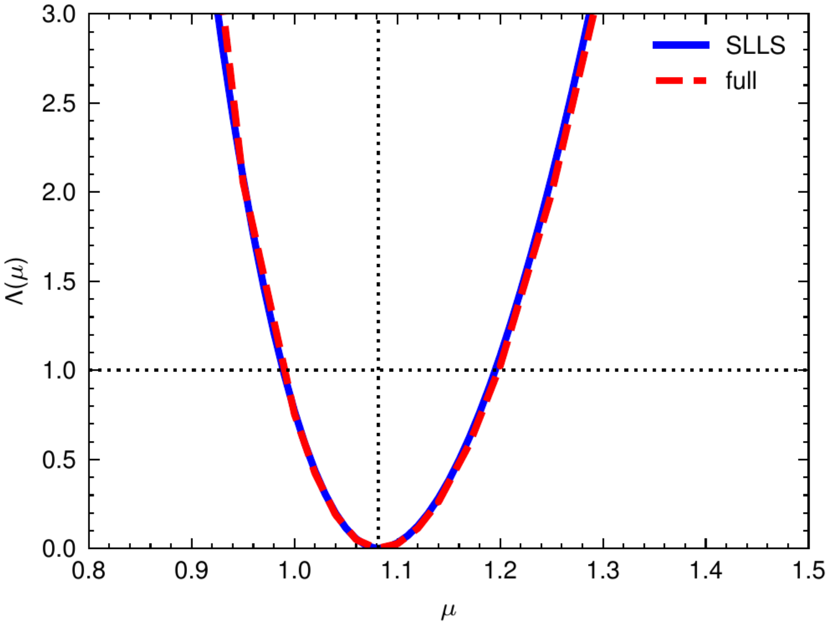
⇒ 18150 bins in total (!) but still rather fast:

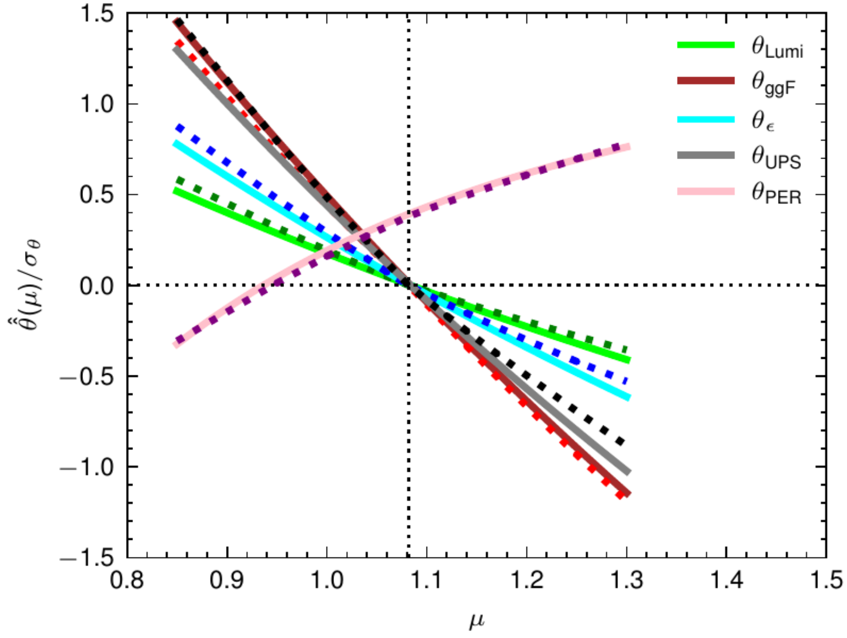
- ~50 ms per fit with fixed μ
- ~1s per fit for floating μ.





Unbinned example: toy H→γγ measurement





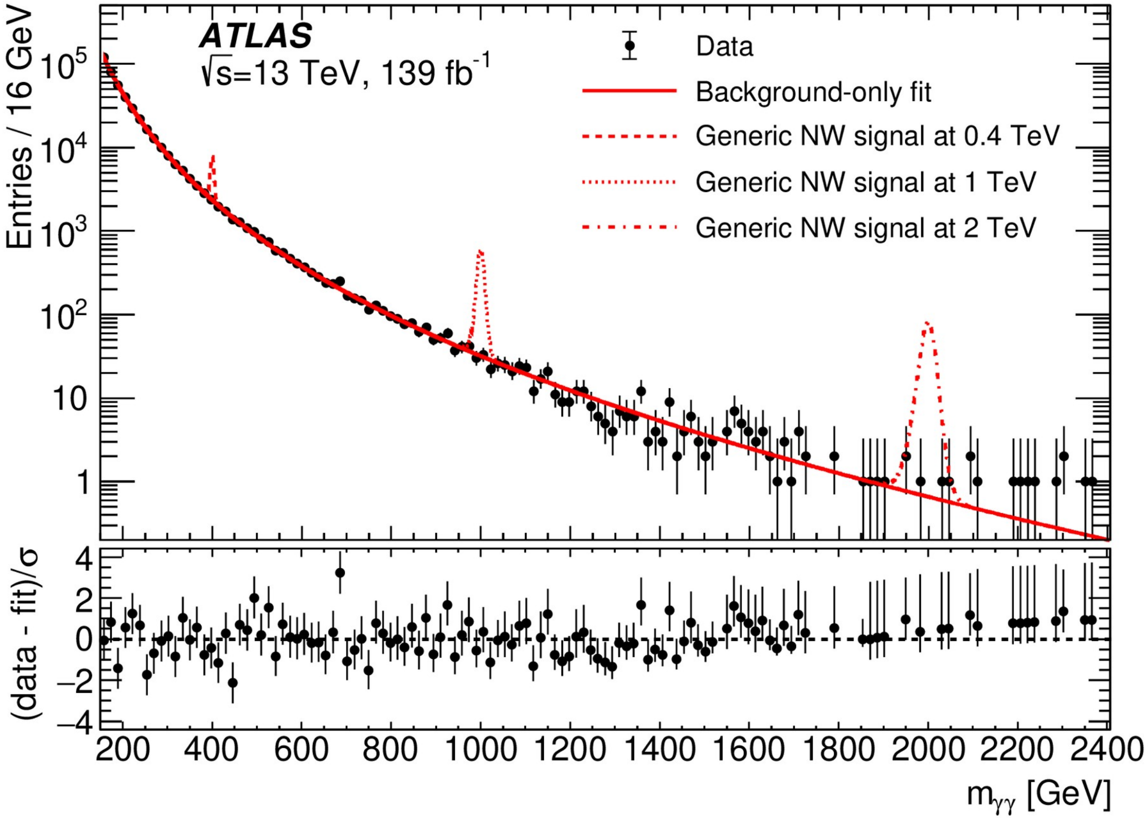
Profile likelihood scan

Profiled values of syst NPs

Unbinned case: toys for X→γγ

Phys. Lett. B 822 (2021) 136651

Results (e.g upper limits) often computed using "**asymptotic formulas**" assuming Gaussian behavior (see e.g. Eur. Phys. J. C71 (2011) 1554) but not valid In tails of distributions with low event counts.

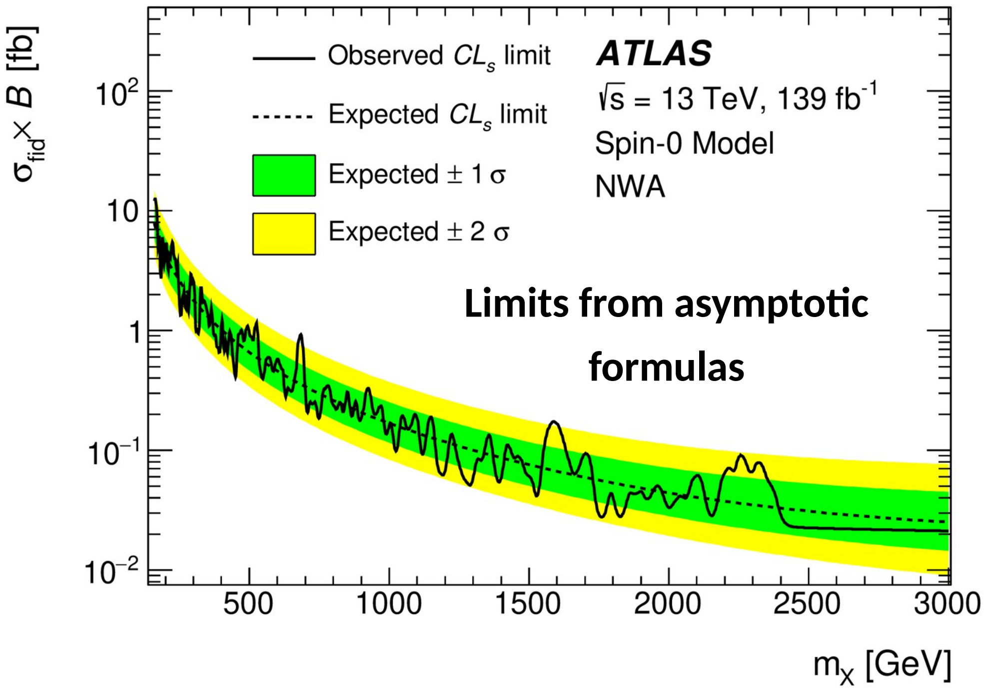


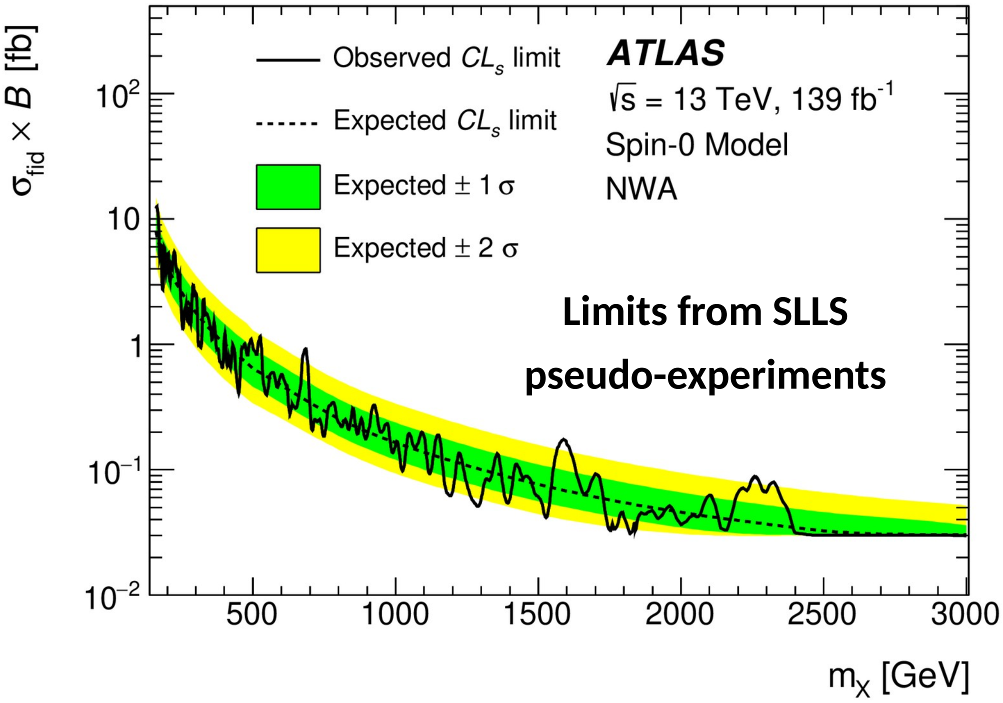
Unbinned case: toys for X→γγ

Phys. Lett. B 822 (2021) 136651

Results (e.g upper limits) often computed using “**asymptotic formulas**” assuming Gaussian behavior (see e.g. Eur. Phys. J. C71 (2011) 1554) but not valid In tails of distributions with low event counts.

⇒ Can use SLLS models to generate and fit pseudo-experiments ⇒ avoid relying on asymptotics.

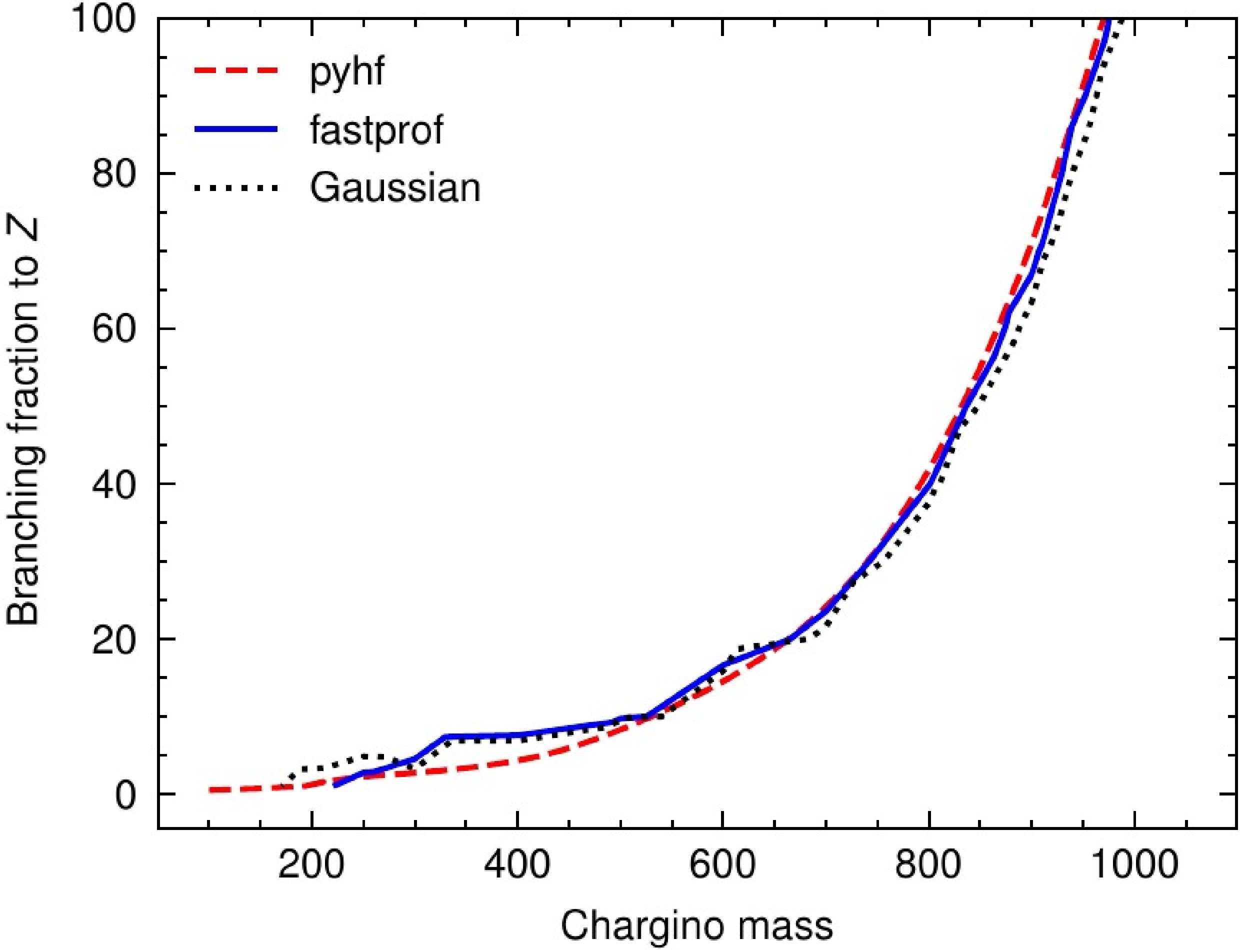




Conclusion

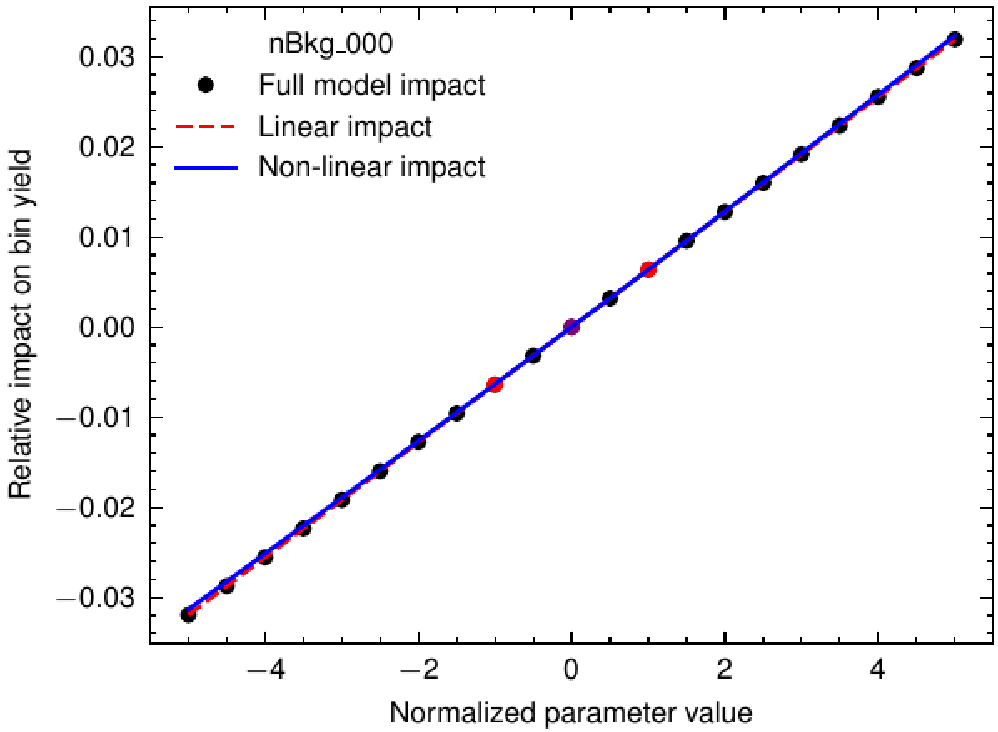
- **Simplified likelihoods** are a critical ingredient for accurate real-world reinterpretation/reuse
- **SLLS likelihoods** provide a simplified description that retains key aspects:
  - Poisson description of event counts
  - all parameters of interest, all nuisance parameters
  - More details in JHEP04(2023)084
  - Python implementation available in github.
- **Linearized NP impacts** allow profiling through matrix algebra, which is very fast .
  - ~1s per full fit, O(10 ms) for profiling NPs.
  - Typically model setup times are longer (few seconds to load all the coefficients)
- **Other simplified approaches available:**
  - ATLAS SUSY SLs [ATL-PHYS-PUB-2021-038] : Poisson PDFs, 1 NP for systs.
  - Simplify [JHEP04(2019)064] : Poisson PDFs, Keep all POIs, 1 NP per bin with quadratic impact
  - DNNLikelihood [Eur. Phys. J. C 80, 664 (2020)]: train a DNN to approximate the likelihood.
- Hopefully all useful for further likelihood publications and reuse!

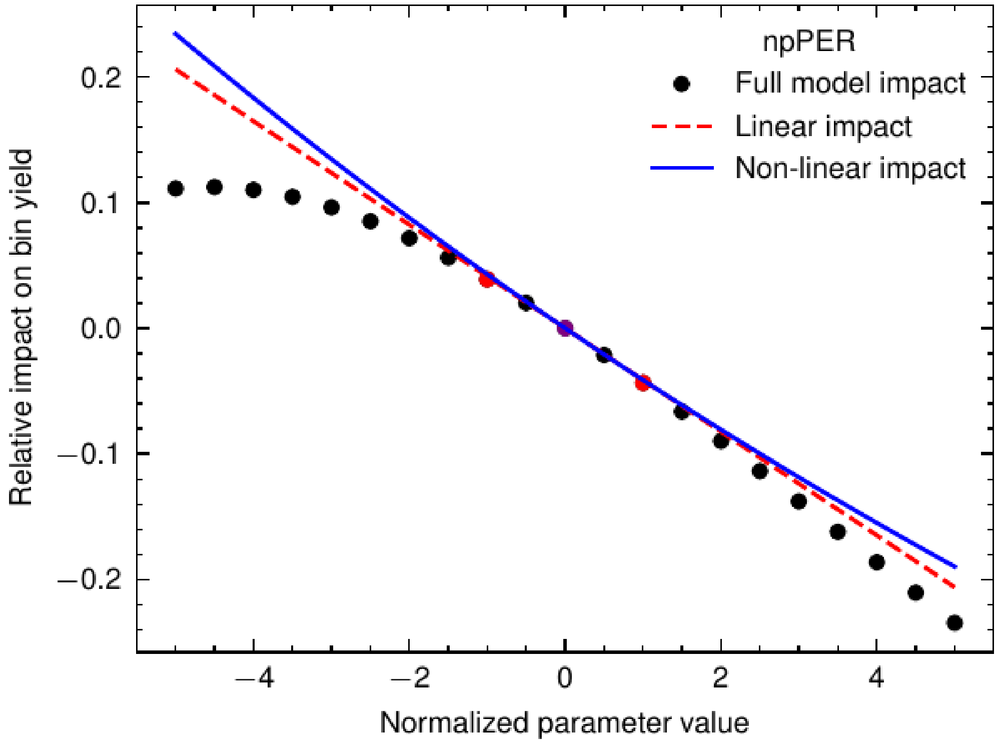
Backup



100
80
60
40
20
0
Branching fraction to Z
200
400
600
800
1000
Chargino mass
pyhf
fastprof
Gaussian

Non-linearities





X→γγ mass spectrum

Phys. Lett. B 822 (2021) 136651