Usage of full likelihoods in reinterpretation studies

Jack Y. Araz

Thomas Jefferson National Accelerator Facility

pyhf Users and Developers Workshop 2023
December 5th, 2023

Jefferson Lab
EXPLORING THE NATURE OF MATTER

U.S. DEPARTMENT OF ENERGY

MCnet

OLD DOMINION UNIVERSITY

Outline

- Introduction
- Other approaches: beyond pyhf
- Combining statistical models
- Computational challenges of full statistical models
- Conclusion & future directions

Introduction

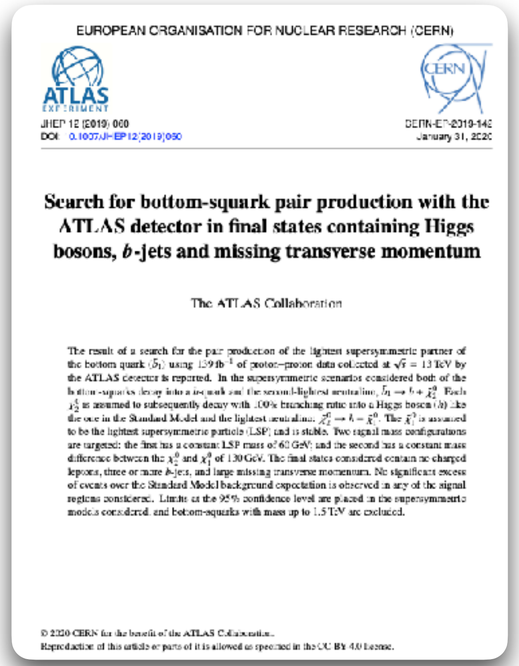
Why designing & recasting is important?

- Exploiting the full potential of LHC (for new physics)
  - *Designing* new analyses (based on MC simulations)
  - *Recasting* LHC analyses (The LHC legacy)
- Data preservation in HEP is mandatory
  - Going beyond raw data via analyses
- Related tools need to be supported by the entire community
  - Both theorists & experimentalists
- Universal recasting tool

Les Houches Recommendations (EPJC '12)

Reinterpretation Forum Report (SciPost '20)

(Re)interpretation of an analysis



EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

ATLAS EXPERIMENT

CERN

JHEP 12 (2019) 060
DOI: 10.1007/JHEP12(2019)060

CERN-EP-2019-142
January 31, 2020

Search for bottom-squark pair production with the ATLAS detector in final states containing Higgs bosons, *b*-jets and missing transverse momentum

The ATLAS Collaboration

The result of a search for the pair production of the lightest supersymmetric partner of the bottom quark ($\tilde{b}_1$) using 139 fb^{-1} of proton–proton data collected at $\sqrt{s} = 13$ TeV by the ATLAS detector is reported. In the supersymmetric scenarios considered both of the bottom-squarks decay into a bottom and the second-lightest neutralino, $\tilde{b}_1 \to b + \tilde{\chi}_2^0$. Each $\tilde{\chi}_2^0$ is assumed to subsequently decay with 100% branching ratio into a Higgs boson (h) like the one in the Standard Model and the lightest neutralino, $\tilde{\chi}_2^0 \to h + \tilde{\chi}_1^0$. The $\tilde{\chi}_1^0$ is assumed to be the lightest supersymmetric particle (LSP) and is stable. Two signal mass configurations are targeted: the first has a constant LSP mass of 60 GeV and the second has a constant mass difference between the $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$ of 130 GeV. The final states considered contain no charged leptons, three or more b-jets, and large missing transverse momentum. No significant excess of events over the Standard Model background expectation is observed in any of the signal regions considered. Limits at the 95% confidence level are placed in the supersymmetric models considered, and bottom-squarks with mass up to 1.5 TeV are excluded.

Statistical Models

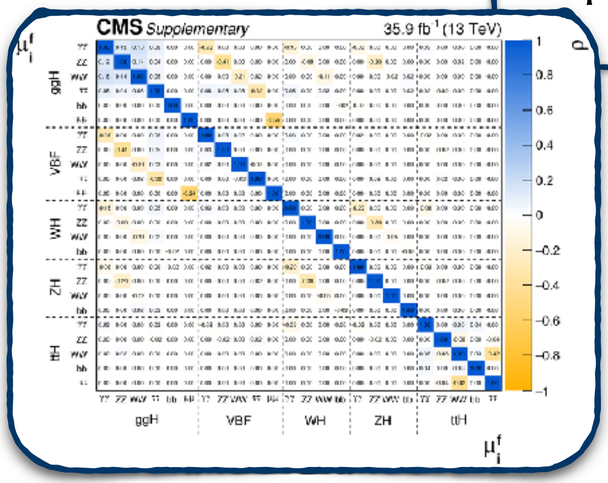
Why important?

ATLAS SUSY and Exotics workshop
S. Kraml '20

- The mathematical description of the analysis is provided within its statistical model.
- A likelihood enables the standard statistical approaches to extract information.
- *i.e.* how reasonably aligned the theoretical predictions with the experimental observations?

Les Houches Recommendations (EPJC '12)

Simplified likelihoods from CMS



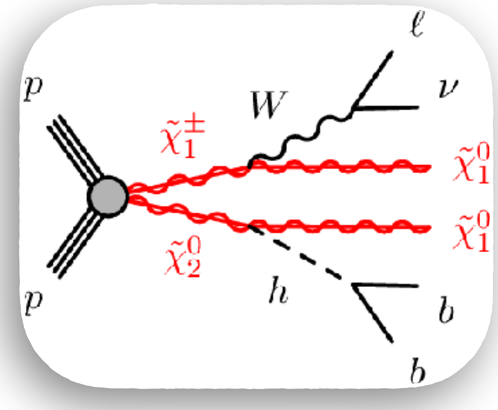
CMS-NOTE-2017-001

Full likelihoods from ATLAS

pyhf
∂ifferentiable
ℒikelihoods

ATL-PHYS-PUB-2019-029

Full likelihoods in action



ATLAS-SUSY-19-08

HEPData

See Graeme's talk

- Wino-like electroweakinos
- ATLAS shares HistFactory like json files to form full profile likelihoods.
- Each file includes detailed information on backgrounds and corresponding nuisance parameters.

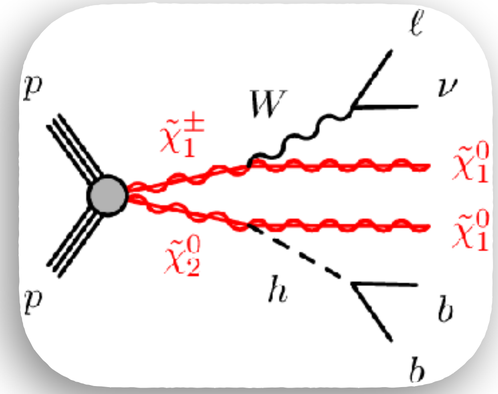
gz File

Archive of full likelihoods in the HistFactory JSON format described in CERN-EP-2019-188. For each signal point the background-only model is found in the file named BkgOnly.json. All jsonpatches are contained in the file patchset.json. Each patch is identified in patchset.json by the metadata field "name": "C1N2_Wh_hbb_[m1]_[m2]" where m1 is the mass of both the lightest chargino and the next-to-lightest neutralino (which are assumed to be nearly mass degenerate) and m2 is the mass of the lightest neutralino.

Download

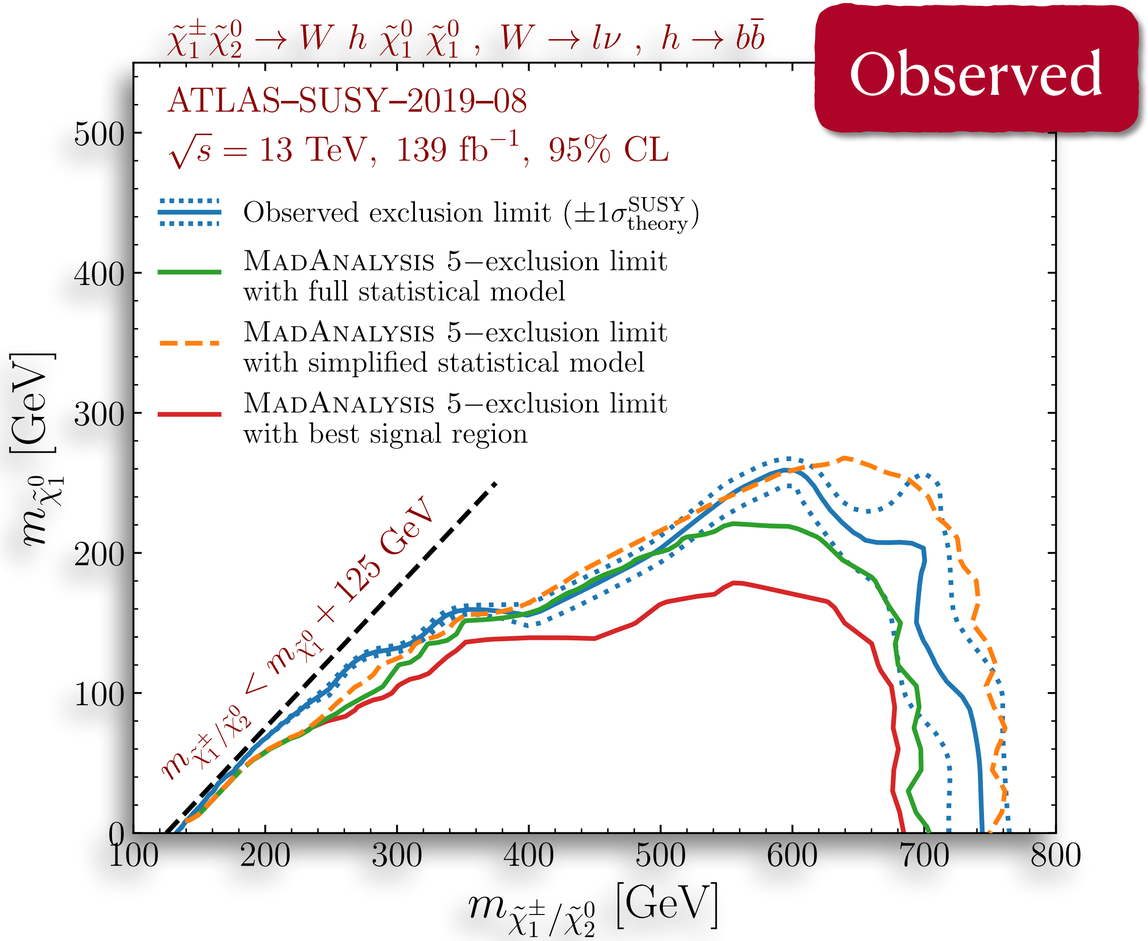
Full likelihoods in action

Alguero, JYA, Fuks, Kraml; SciPost '23



- Wino-like electroweakinos
- ATLAS shares HistFactory-like JSON files to form full profile likelihoods.
- The difference between green and blue is due to the lack of CR/VR!
- Simplified likelihoods are achieved by compressing all the information into a single nuisance parameter.

eschanet/simplify

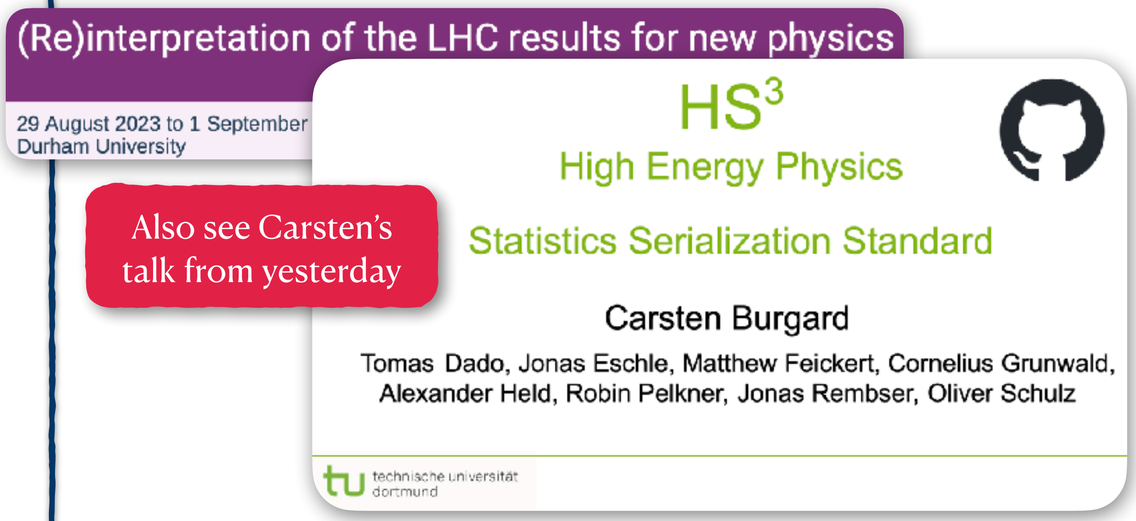


Jefferson Lab EXPLORING THE NATURE OF MATTER

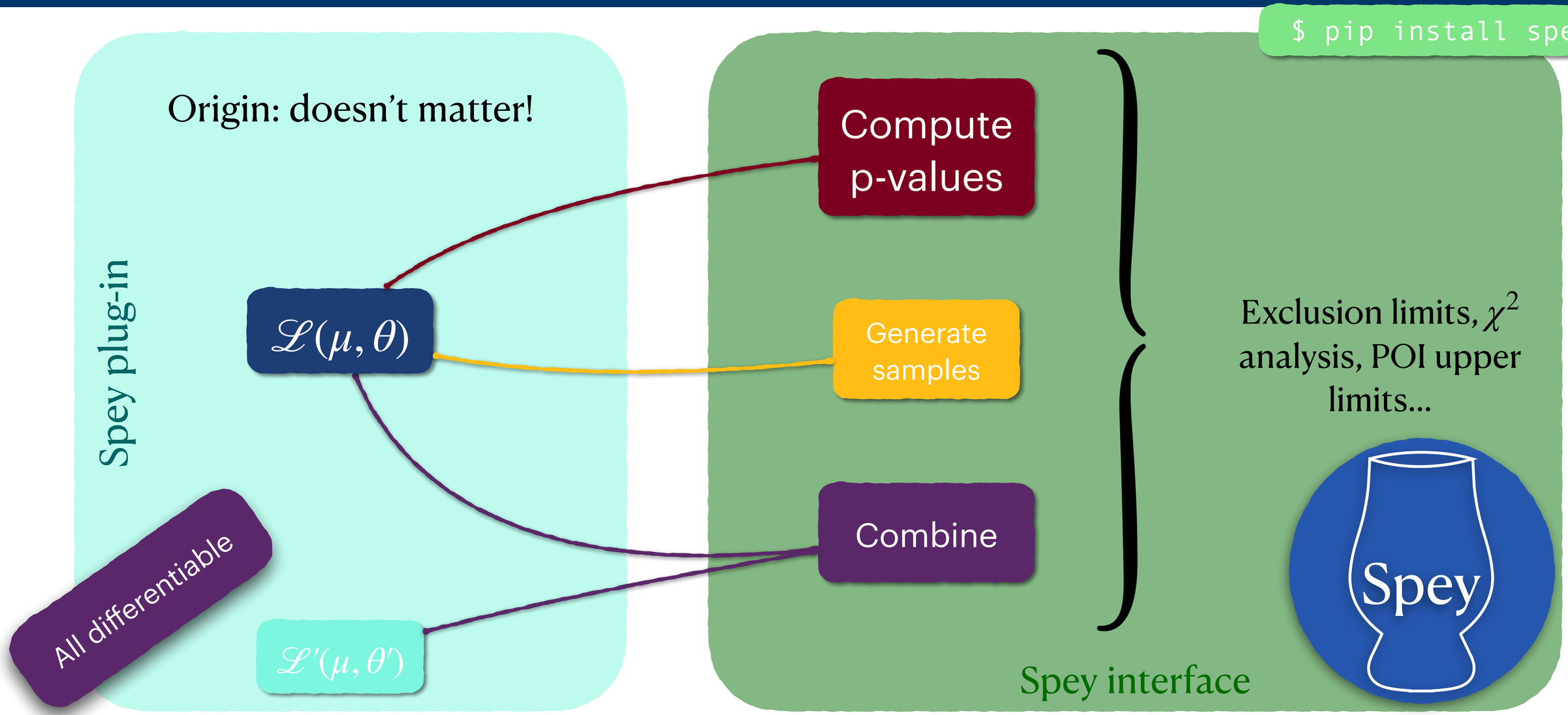
OLD DOMINION UNIVERSITY

What is the next step?

- Not all analyses come with “pyhf” likelihoods (I wish!)
- Full likelihoods can become computationally intensive; can we simplify them without sacrifice?
- There are many different software for hypothesis testing; how can we unite them?
- How can we combine statistical models?



Uniting statistical modelling under Spey



JYA; arXiv:2307.06996

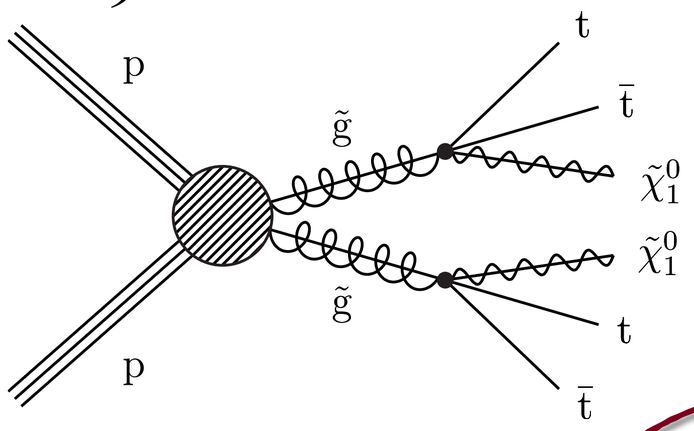
Simplified Likelihoods

Not all analyses come with “pyhf” likelihoods

Jack Y. Araz

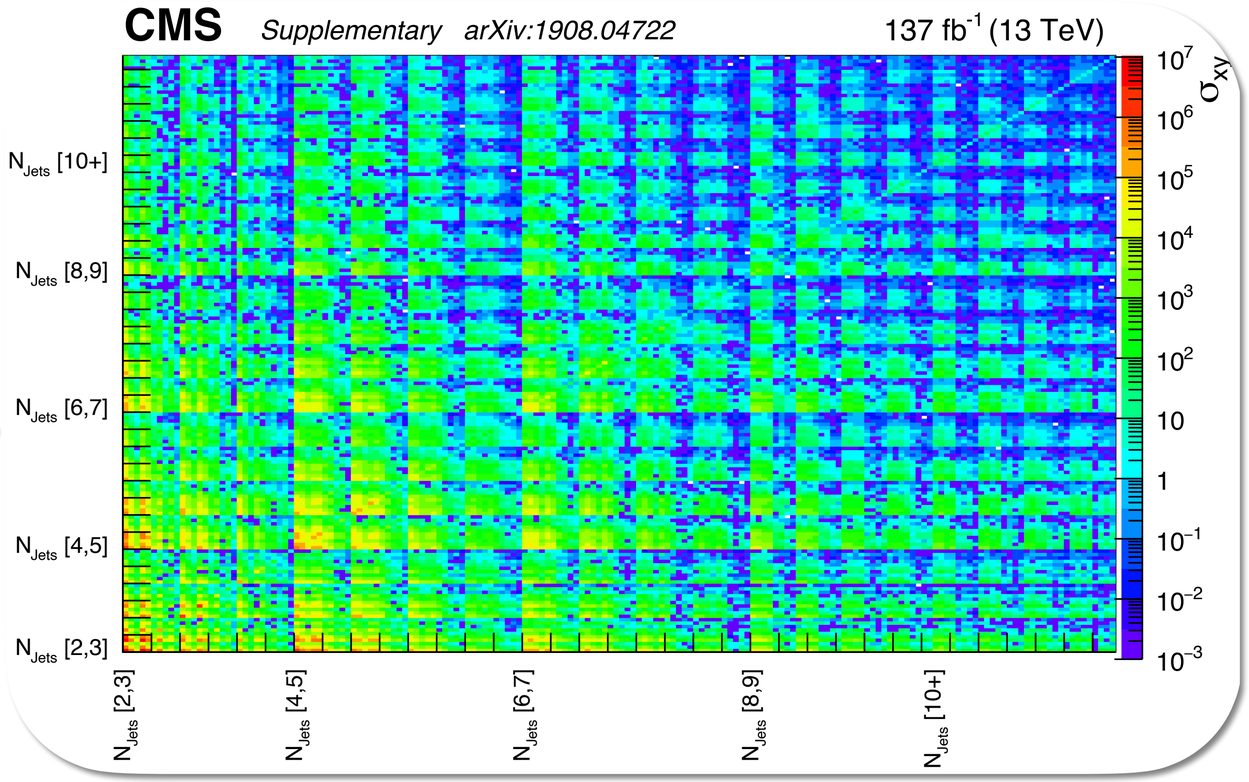
Simplified likelihoods

CMS-SUS-19-006



$$\mathscr{L}(\mu,\theta) = \left[\prod_{i\in\text{bins}} \text{Poiss}\left(n^i \,|\, \mu n_s^i + n_b^i + \theta^i \sigma_b^i\right)\right] \cdot \mathcal{N}\left(\theta \,|\, 0,\rho\right)$$

CMS-NOTE-2017-001



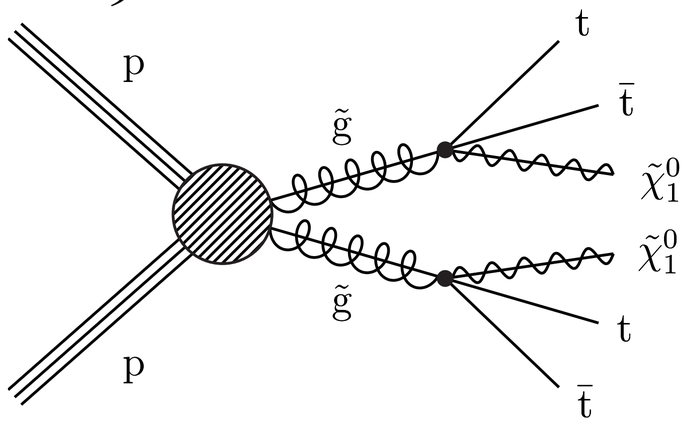
186 signal regions!

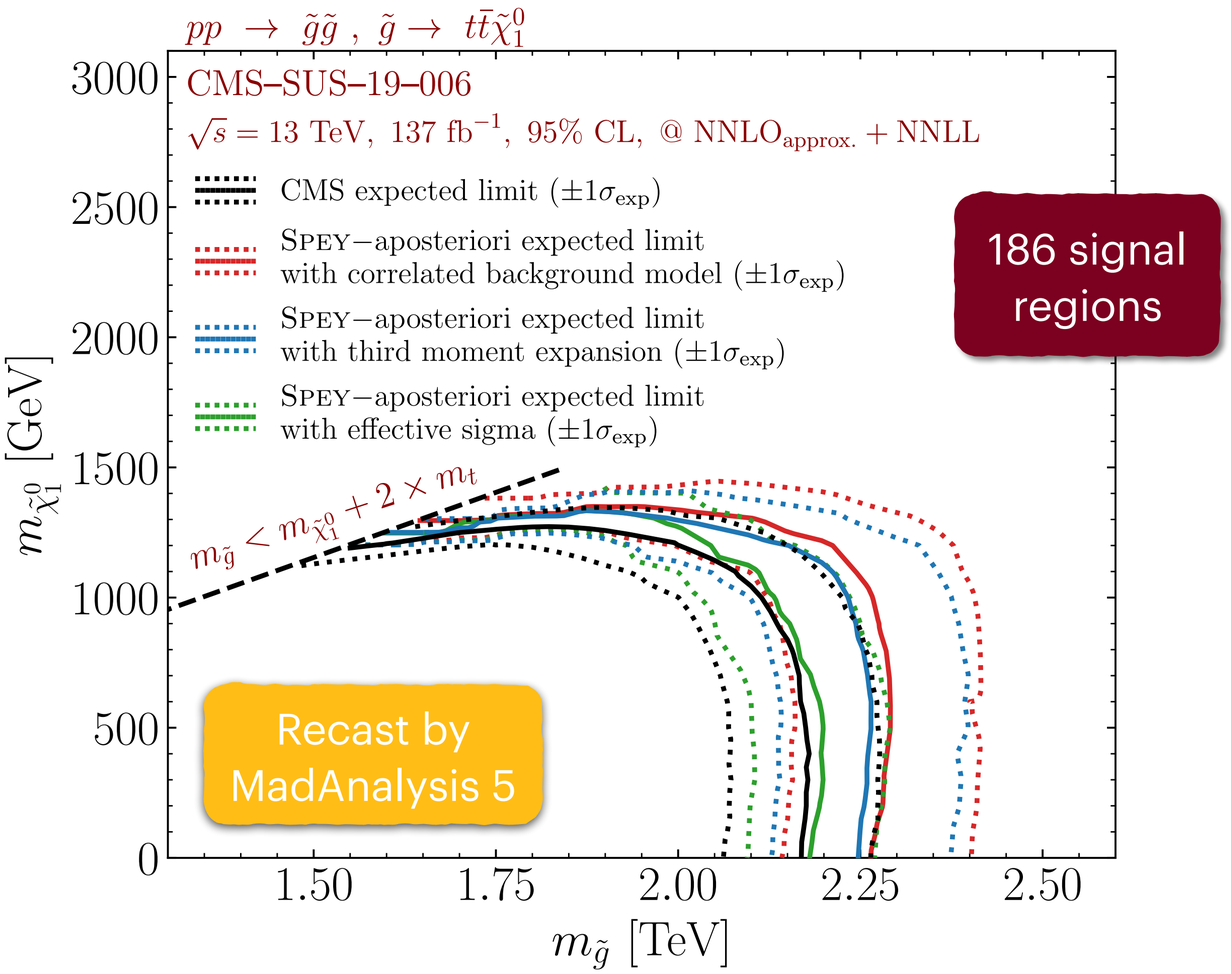
Can be expanded via third moments or using asymmetric uncertainties

Simplified likelihoods

JYA; arXiv:2307.06996

CMS-SUS-19-006





186 signal regions

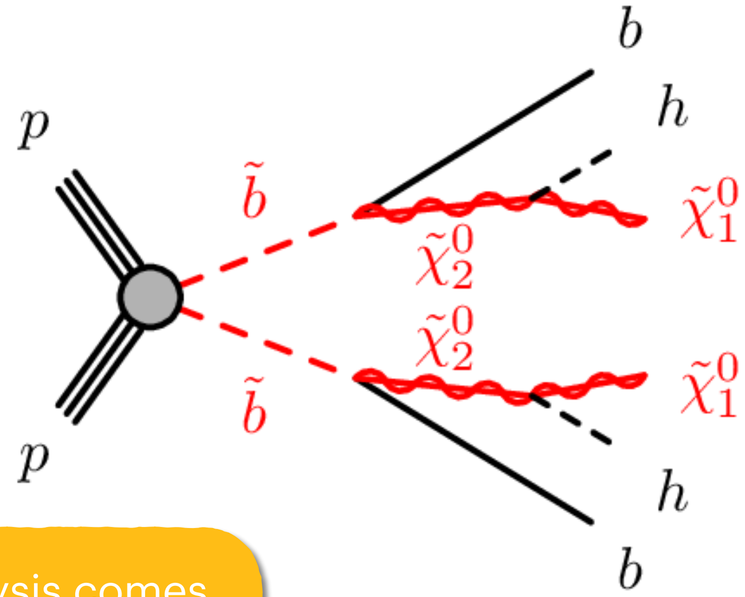
Recast by MadAnalysis 5

Combining Statistical Models

Combining Statistical Models

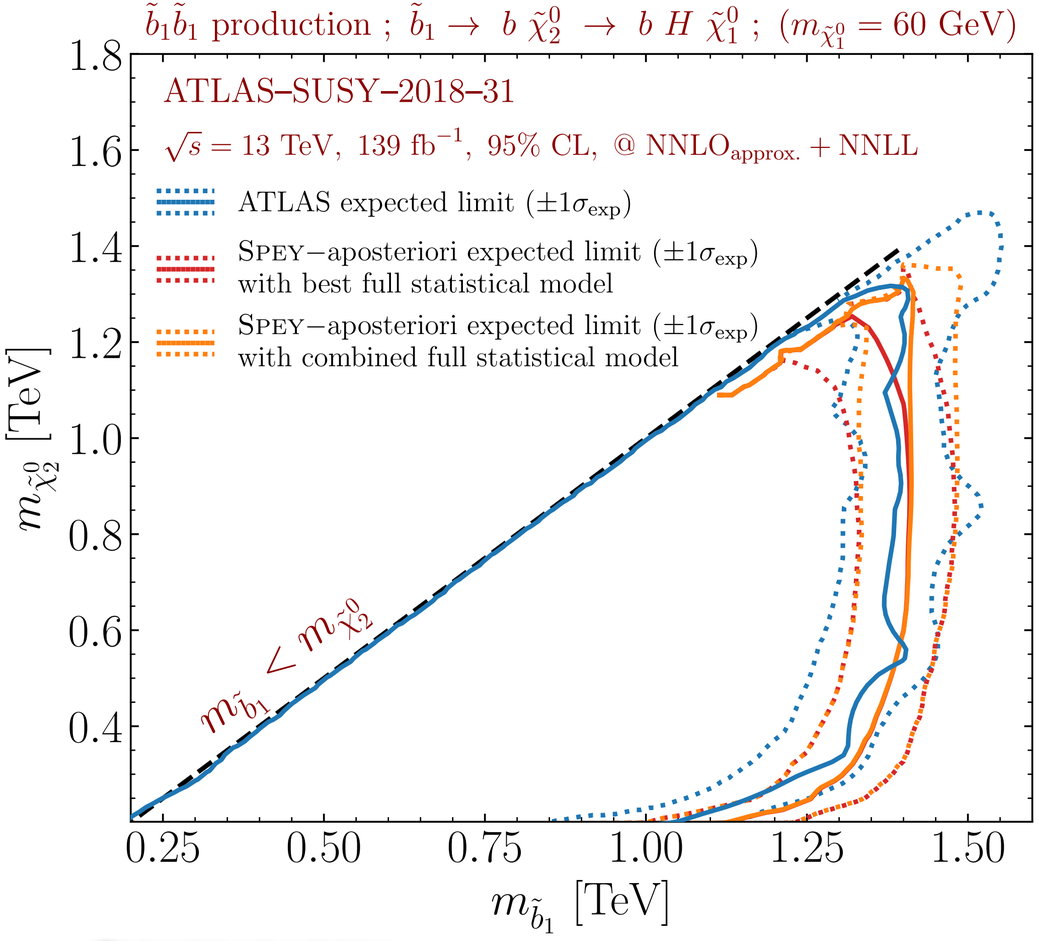
```
$ pip install spey-pyhf
```

JYA; arXiv:2307.06996



This analysis comes with three different super regions!

Full likelihoods include all the necessary information to mix and match nuisance parameters to combine them!



Recast by MadAnalysis 5

Combining Statistical Models

JYA; arXiv:2307.06996

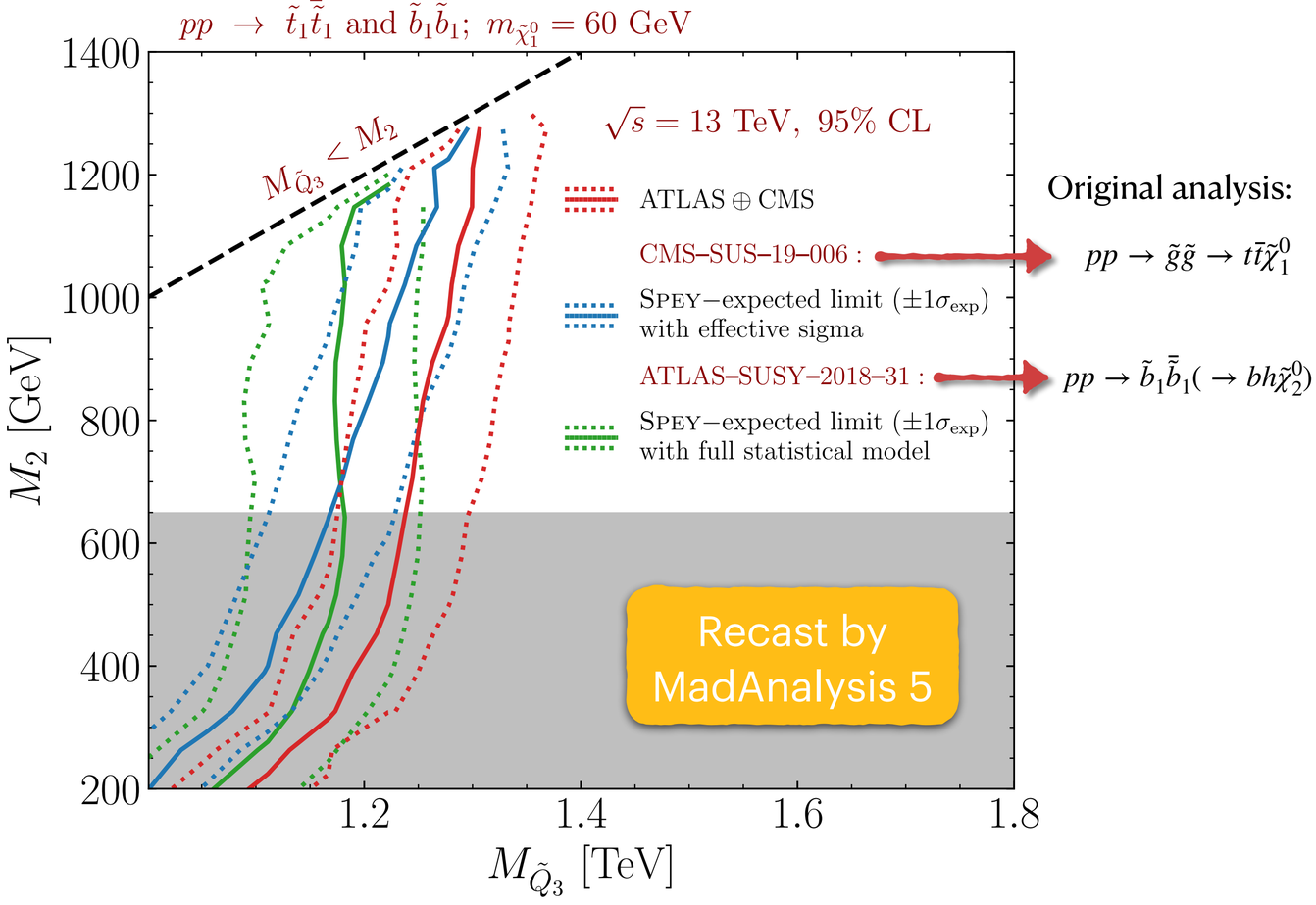
$$\mathscr{L}' = \mathscr{L}_{\text{ATLAS}} \oplus \mathscr{L}_{\text{CMS}}$$

Full likelihood

Simplified likelihood with effective sigma model

A combination of analyses, rather than regions, contains much more information!

! Assumes that likelihoods are not correlated !



Original analysis:

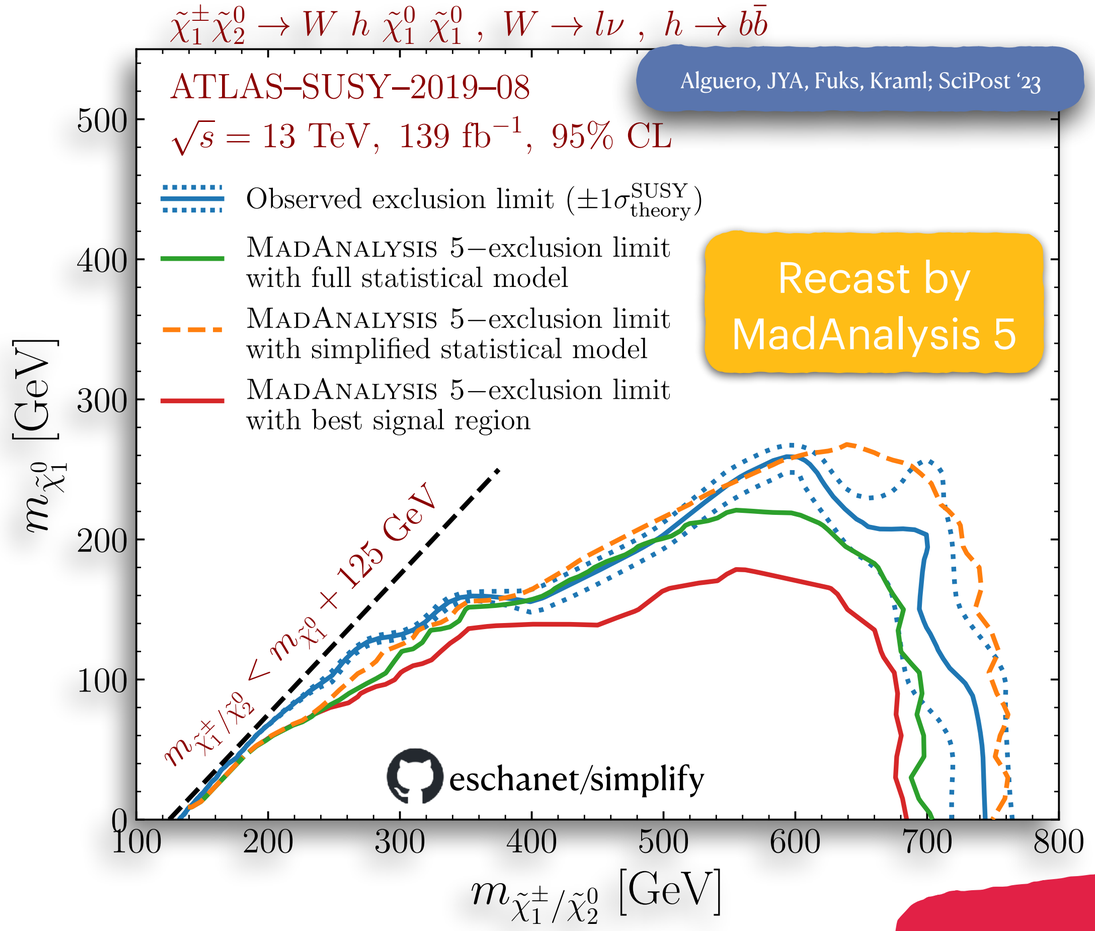
CMS–SUS–19–006 : $pp \to \tilde{g}\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$

ATLAS–SUSY–2018–31 : $pp \to \tilde{b}_1\bar{\tilde{b}}_1(\to bh\tilde{\chi}_2^0)$

How do we avoid the computational cost of full likelihoods?

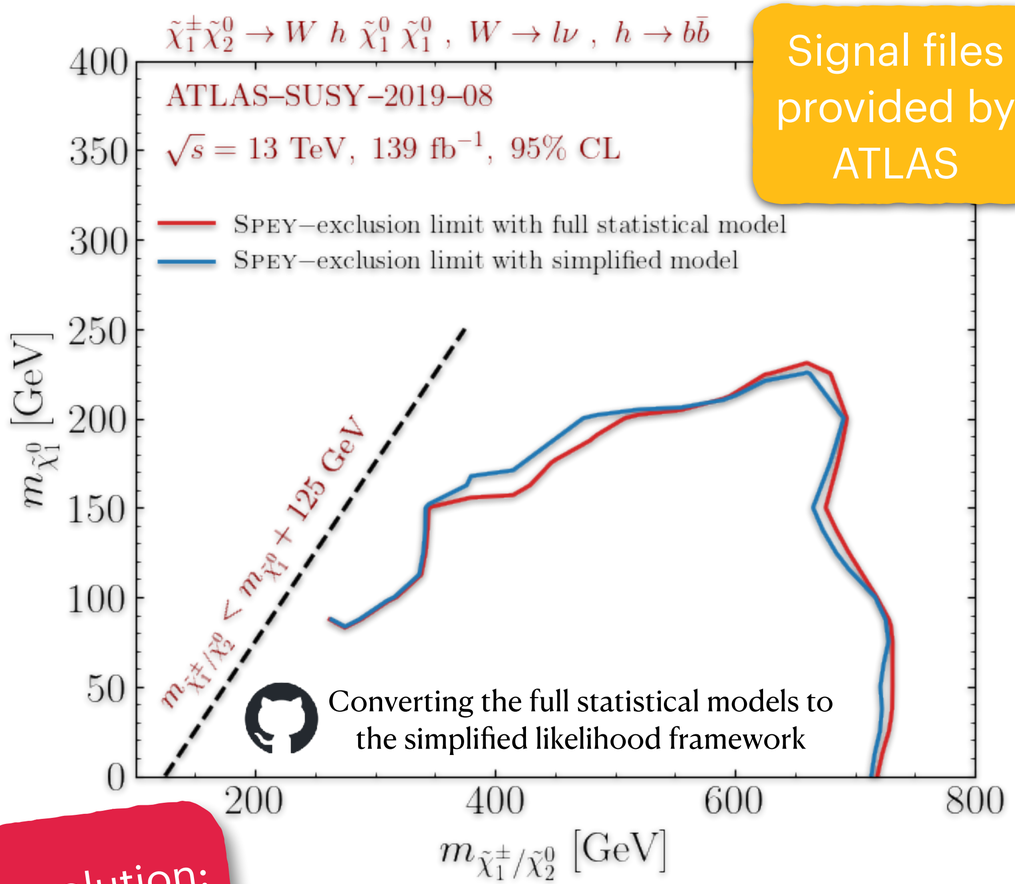
Faster inference!

JYA; arXiv:2307.06996



Alguero, JYA, Fuks, Kraml; SciPost '23

Recast by MadAnalysis 5



Signal files provided by ATLAS

The next step in the evolution: Machine-learned likelihoods with Humberto & Wolfgang

Conclusion
&
Future Directions

Conclusion

Spey online documentation

Spey arXiv:2307.06996

- CMS planning to release full likelihoods via the “combine” package
- Spey acts as a hub for various likelihood implementations for hypothesis testing.
- The ability to study likelihoods in a backend-agnostic way opens up various possibilities, such as combinations.
- There are other statistical model implementations! fastprof-hep/fastprof

Smarter simplified likelihood implementations

Machine Learned likelihoods

Dedicated custom likelihoods for other experiments

Thanks to pyhf-dev team, who pushed all this endeavour!