Institute for Theoretical Particle Physics and Cosmology
RWTH AACHEN UNIVERSITY

Collaborative Research Center TRR 257
Particle Physics Phenomenology after the Higgs Discovery

Open library of learned likelihoods of LHC experiments

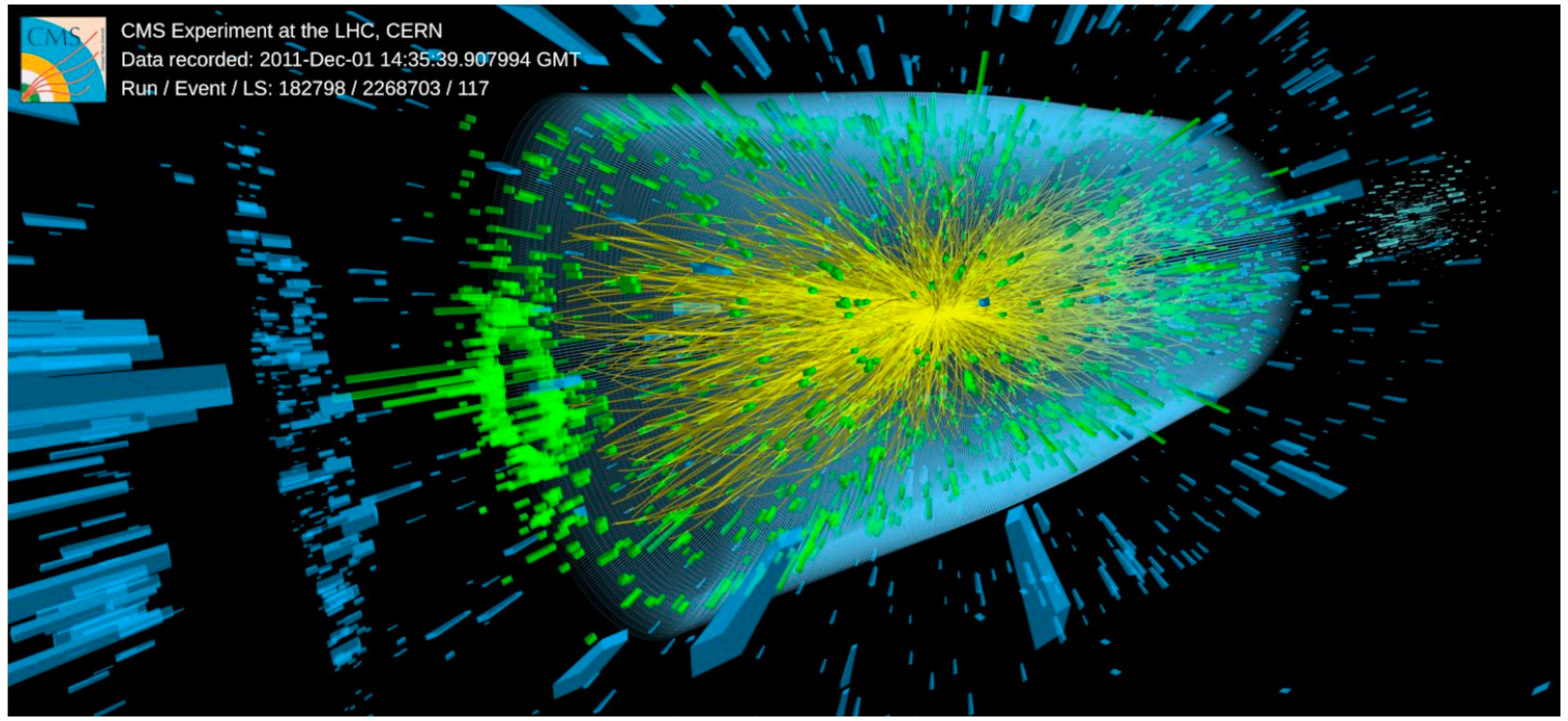
Humberto Reyes-González

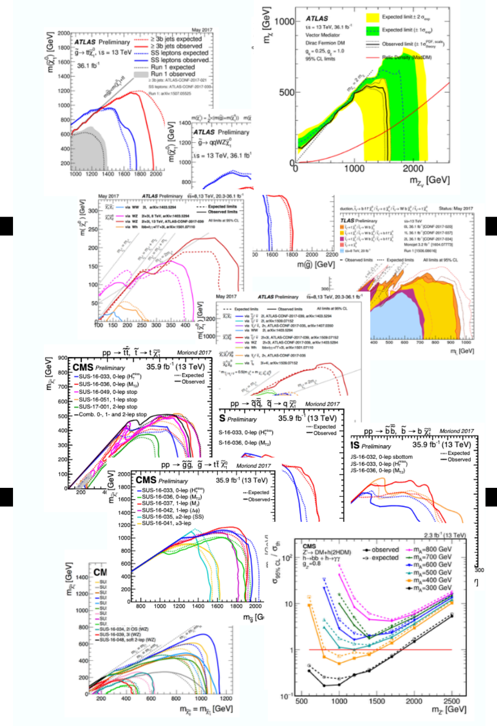
with Jack Y. Araz, Anja Butter, Joaquin Iturriza, Sabine Kraml, Rafal Maselek, Wolfgang Waltenberger

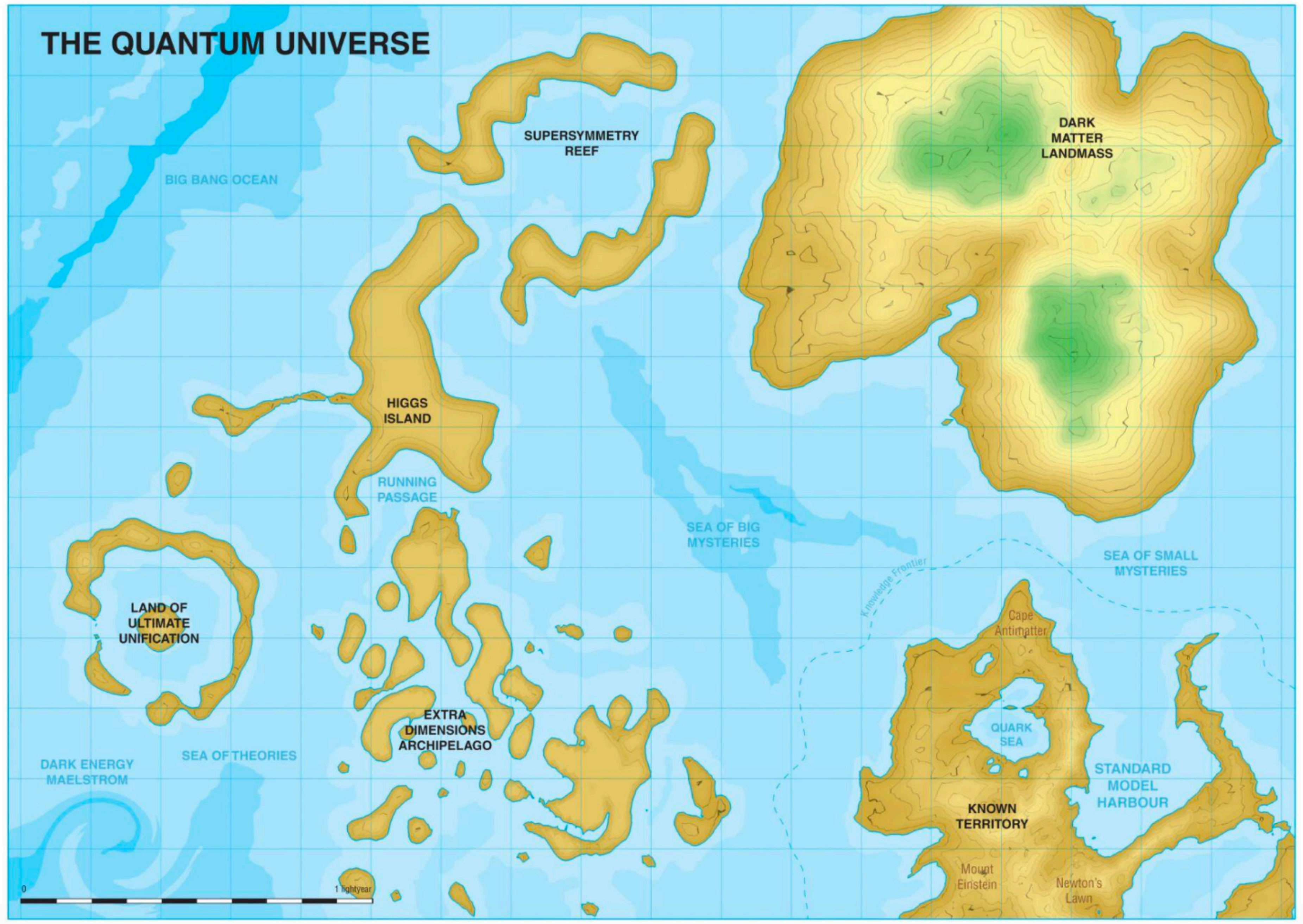
ML4JETS ▪ VIENNA ▪ 14–18 SEPTEMBER 2026

Introduction

Experiment — **(Re) interpretation** — **Theory**







- A key component is the statistical model.
- It encapsulates the full information of an analysis.
- ATLAS analyses are provided trough the PyHF framework.

LHC likelihoods in a nutshell

Bayes theorem:

$$P(\Theta, x) = P_x(x\,|\,\Theta)\pi_\Theta(\Theta) = P_\Theta(\Theta\,|\,x)\pi_x(x)$$

LHC Statistical model:

$$P(\mu, \theta; \mathrm{data}) = \Pi_{k=1}^{n_c} P[n_i; \mu\epsilon_{i,k}(\vec{\theta})N_{S,i,k}(\vec{\theta}) + B_{i.k}(\vec{\theta})]\Pi_{j=1}^{n_{syst}} G(\theta_j^{obs}; \theta_j; 1)$$

(Observed) data

Nuisance parameters (uncertainties)

(Auxiliary) data

Parameters of Interest (signal strength, observables, etc.)

With the SM we perform global fits, exclude BSM models, find upper limits, search for SM deviations, etc.

The profiled likelihood

$$P(\mu, \theta; \text{data}) = \Pi_{k=1}^{n_c} P[n_i; \mu\epsilon_{i,k}(\vec{\theta})N_{S,i,k}(\vec{\theta}) + B_{i.k}(\vec{\theta})]\Pi_{j=1}^{n_{syst}} G(\theta_j^{obs}; \theta_j; 1)$$

- In the Profiled Likelihood (PL) the nuisance parameters are fixed such that the likelihood is maximised given a signal strength μ.
- **The PL is a function of the signal yields.**

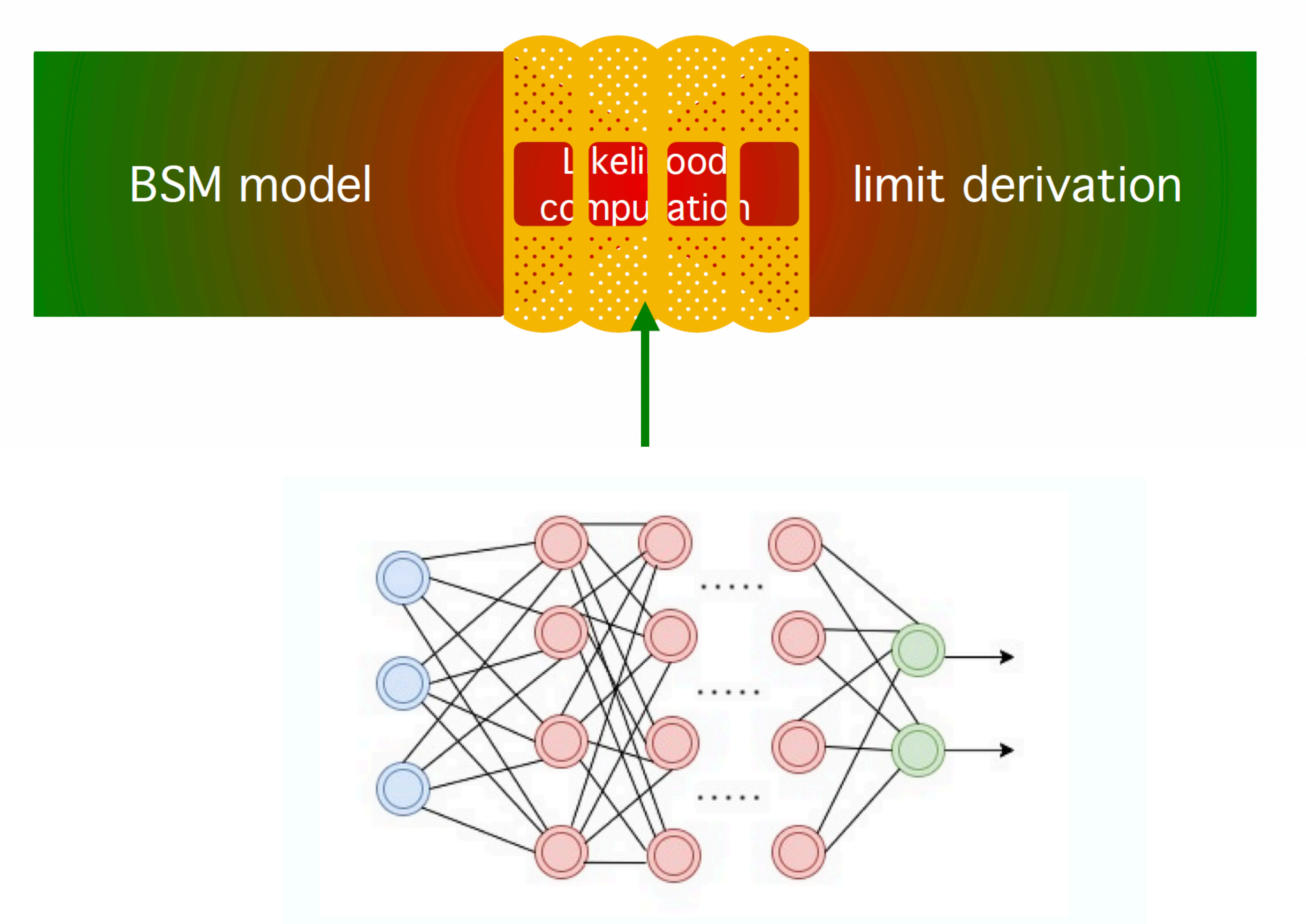
$$L(n_i \,|\, \mu; \hat{\theta}(\mu))$$

- With the PL we construct Log Likelihood Ratio (LLR) test

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

Costly Profiled Likelihood computation.

ML parametrisation fixes the problem



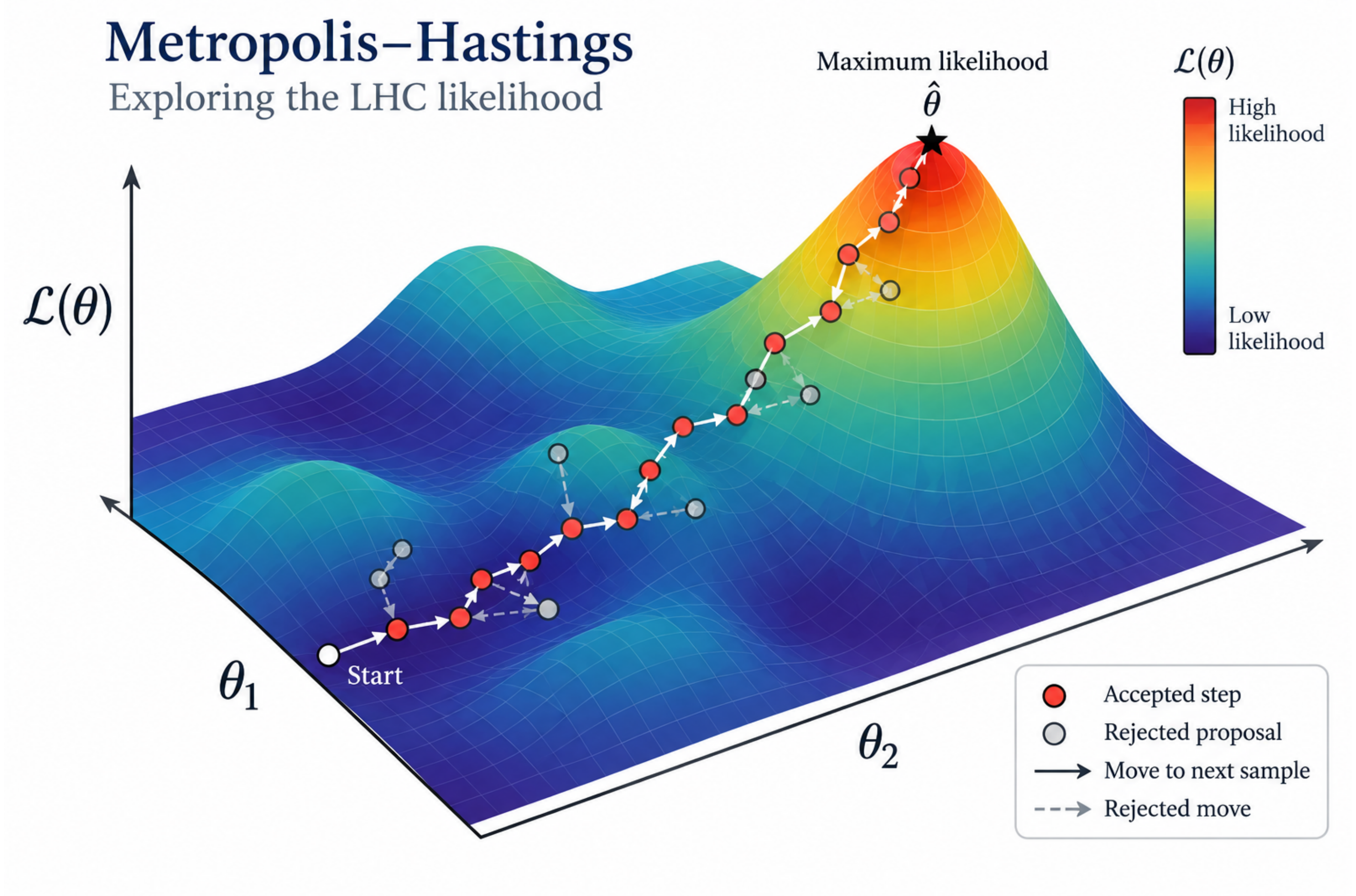
General pipeline

$$\mathrm{N}_i^{\max} = B_i + \max(0, \mathrm{N}_i^{\mathrm{obs}} - 5\Delta B_i),$$

$$\Delta N_i \equiv \min(-B_i + \Delta B_i, \mathrm{N}_i^{\mathrm{obs}} - B_i),$$

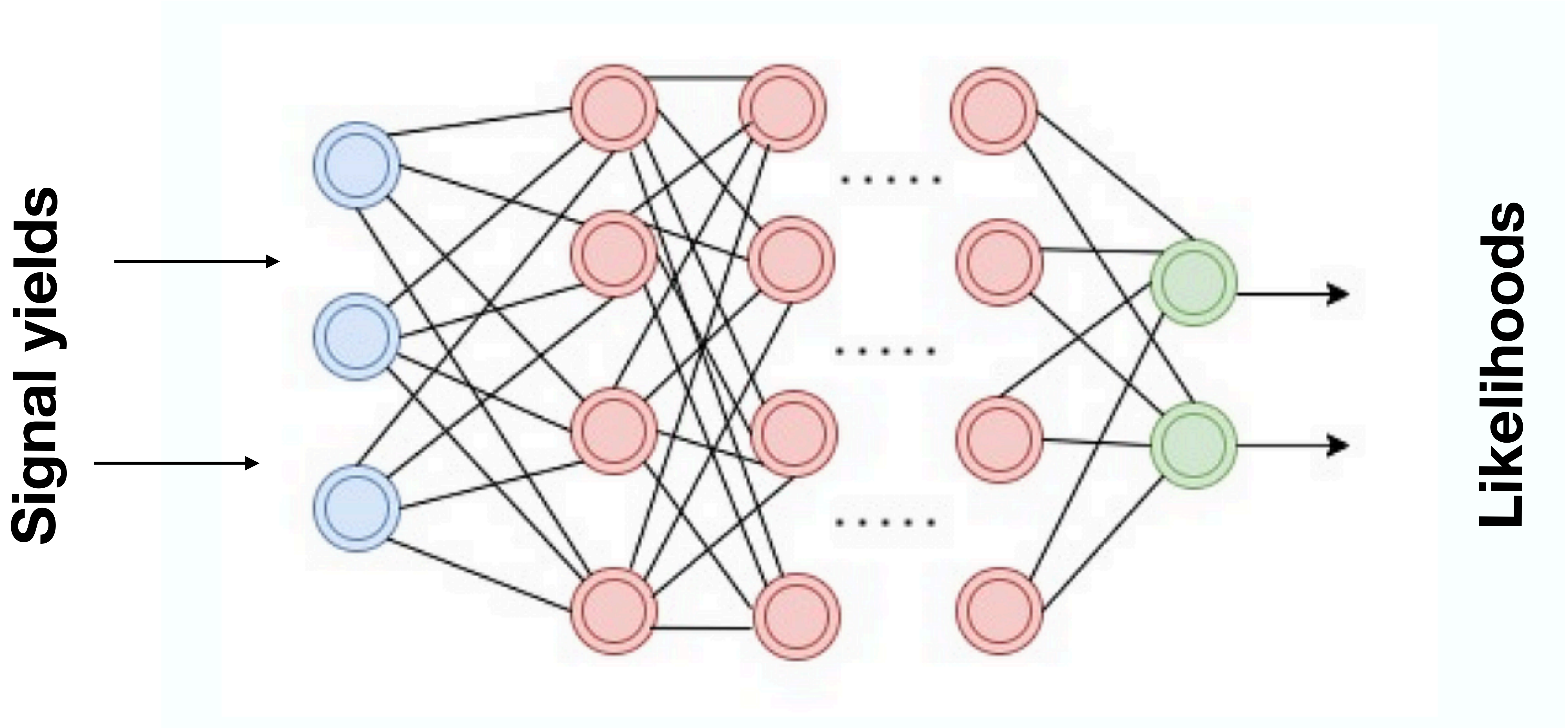
$$\mathrm{N}_i^{\min} = \begin{cases} B_i + \Delta N_i, & \text{if } \Delta N_i < 0 \\ 0, & \text{if } \Delta N_i \geq 0 \end{cases},$$

Defining the parameter space of the starting point



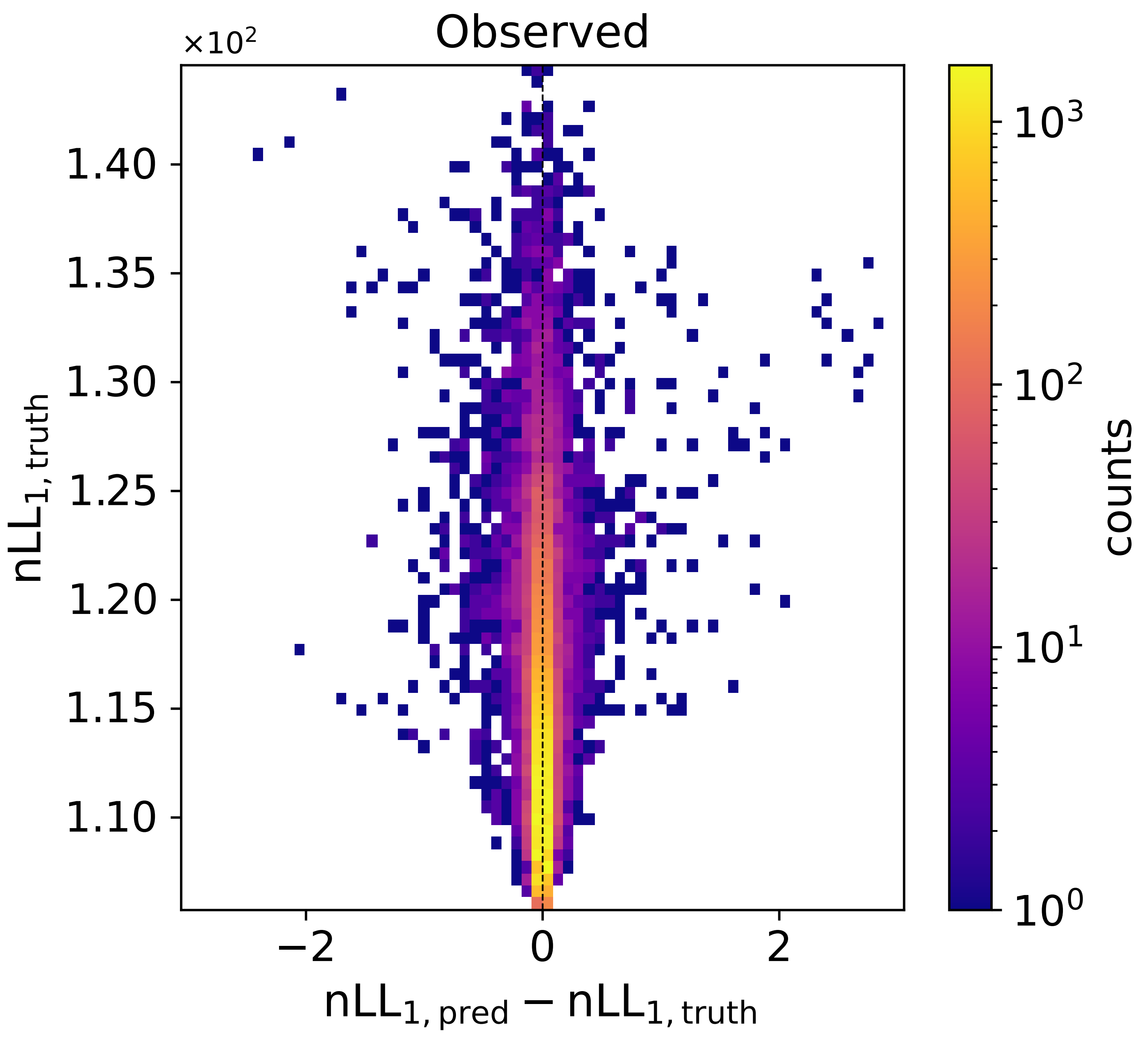
Metropolis-Hastings MCMC Sampling

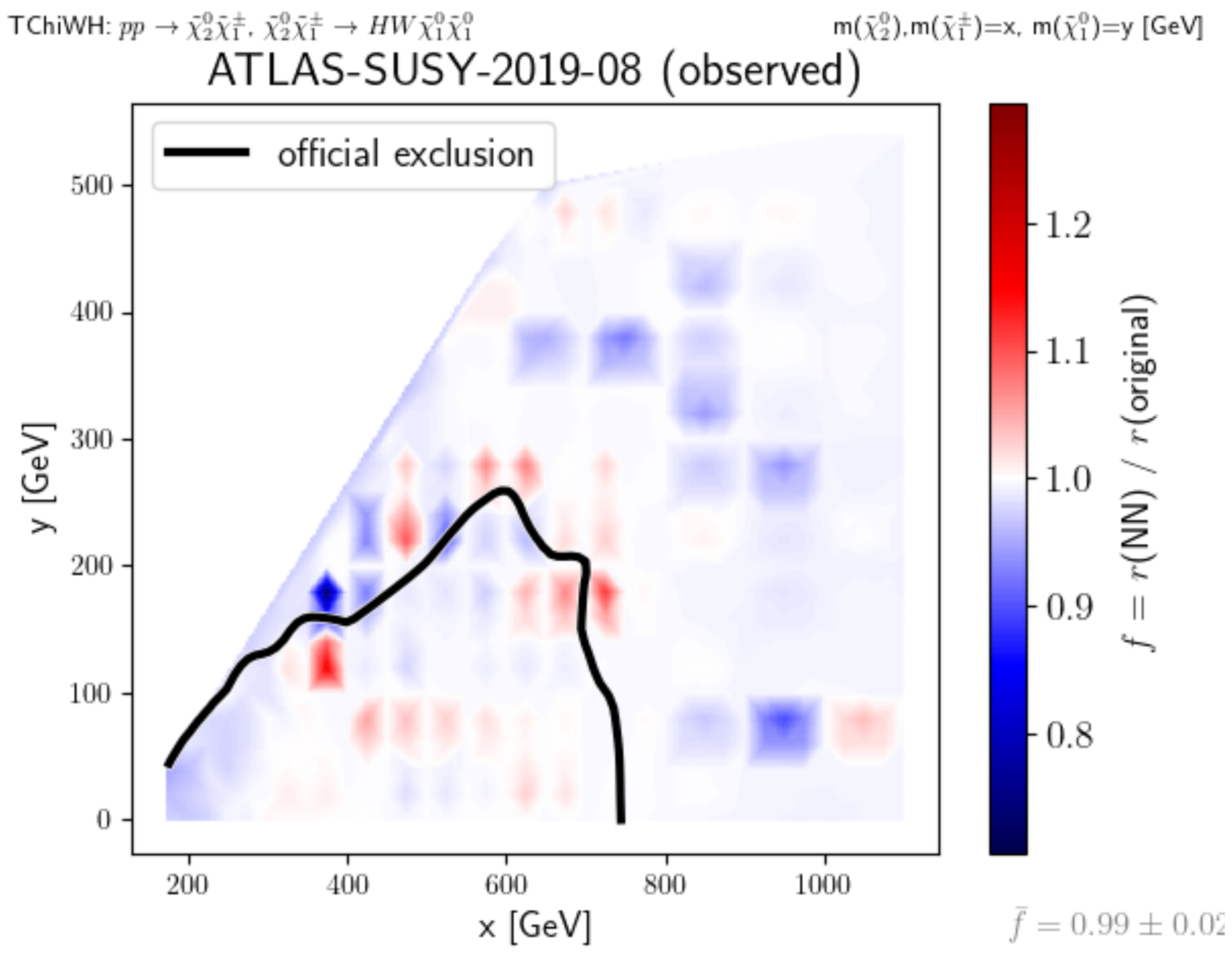
General pipeline

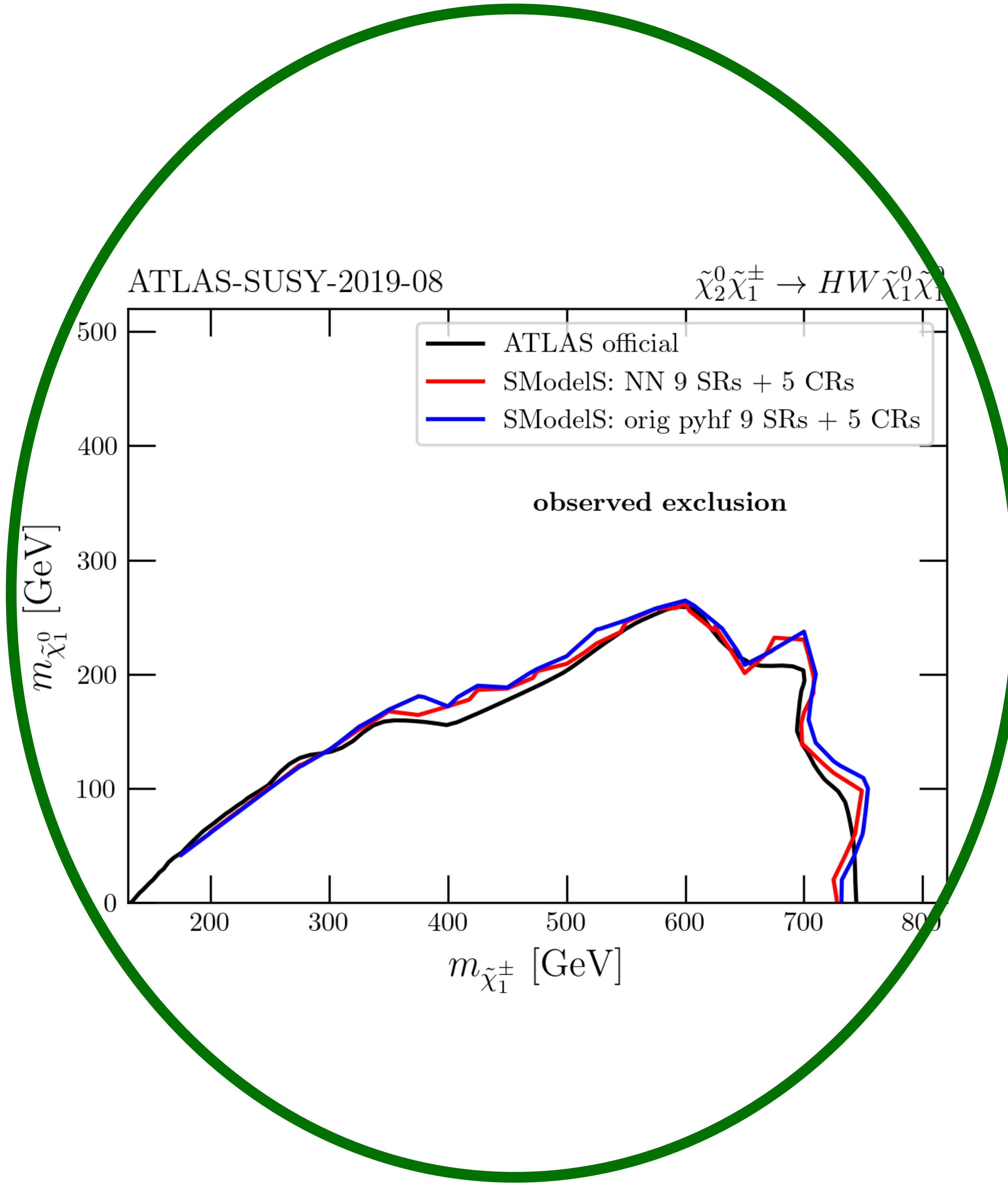


Training the (MLP) Neural Networks
with heteroscedastic loss
Uncertainty estimation included!

Three stages of validation







The official way

General pipeline

Deployment as ONNX model

Framework independent

Metadata included

The devil is in the details

Proof concept ? EASY!
Reliable deployment, CHALLENGING

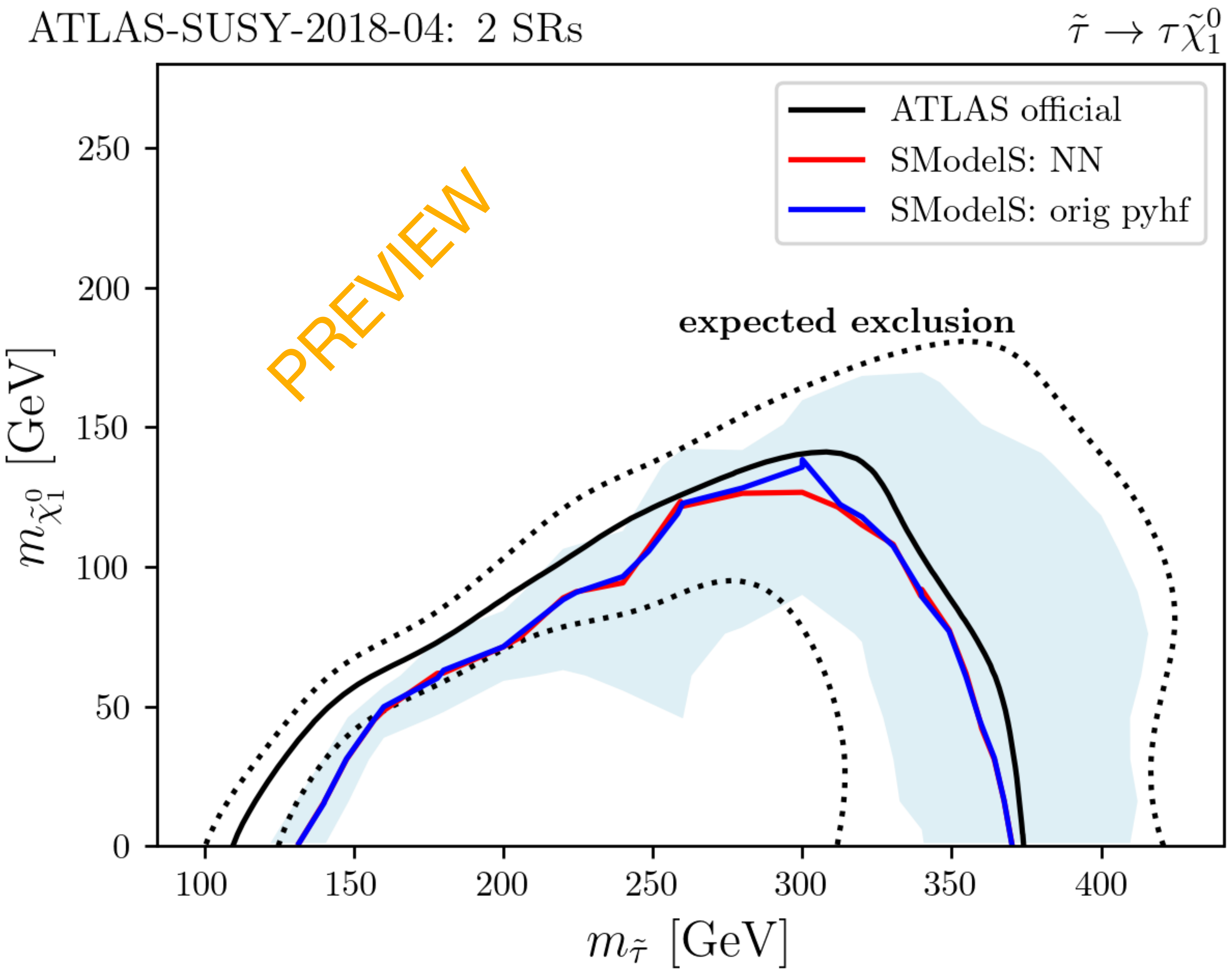
- Are the analyses divided in several statistical models?
- Do we add control regions to the likelihood construction?
- Do we take into account signal leakage?
- Do we construct a priori or a posteriori expected likelihoods?
- What about Asimov likelihoods?
- …

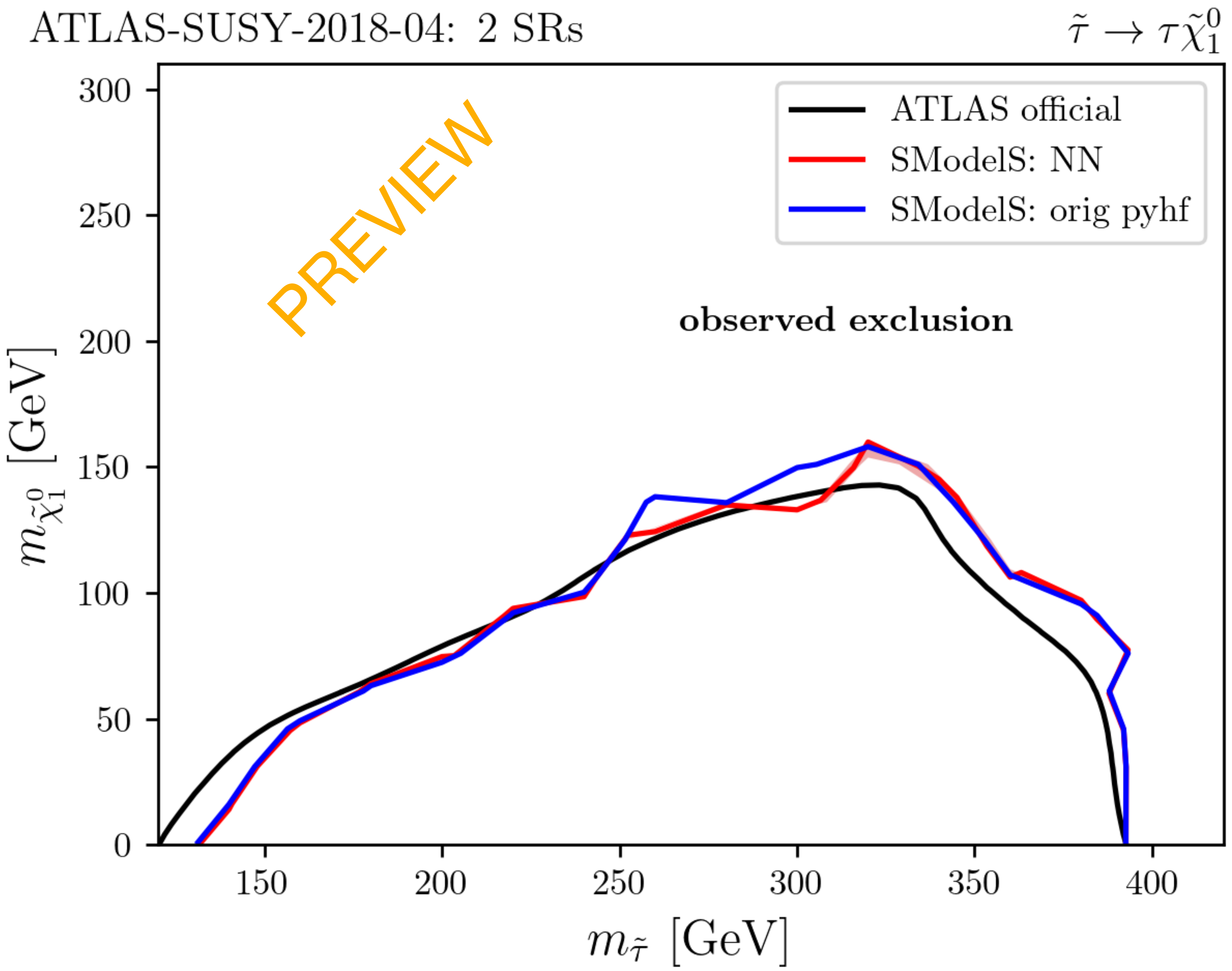
The devil is in the details

- Observed and expected likelihoods and the corresponding Asimov likelihoods are learned
- A probability of signal leakage is accepted.
- Addition of control regions on a case by case basis
- One ONNX for each statistical model

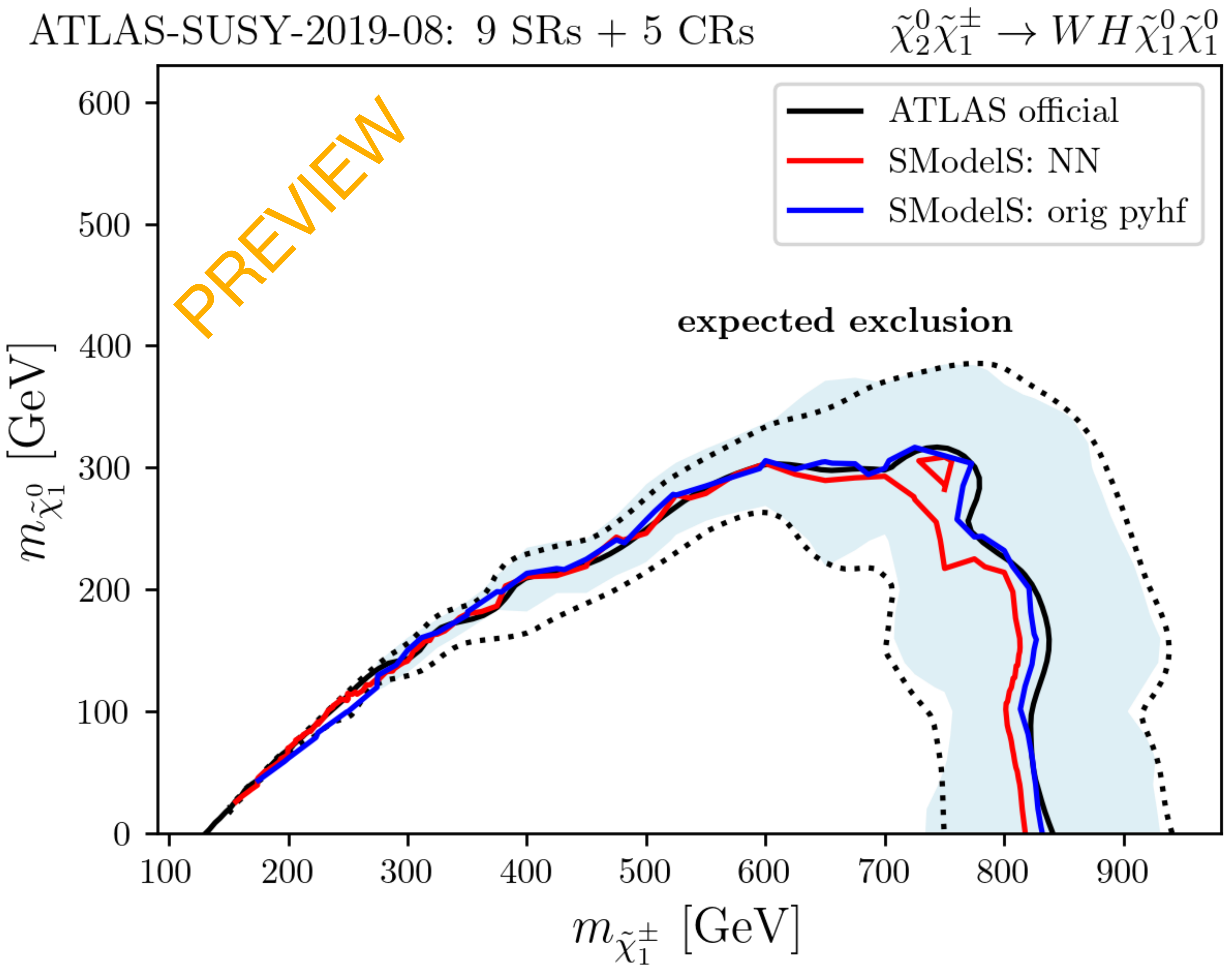
ATLAS ID	Short description	SRs	CRs	comment
SUSY-2018-04	Stau search, 2 taus	2	3	w/o CRs
SUSY-2018-16	2 soft lept., EW-ino	44	6	needs SRs & CRs
	2 soft lept., slepton	32	6	needs SRs & CRs
SUSY-2018-32	2 OS leptons	36	3	needs SRs & CRs
SUSY-2019-08	EW-inos, $Wh(\to b\bar{b})$	9	5	w/ or w/o CRs
SUSY-2019-09	EW-inos, WZ onshell	20	3	needs SRs & CRs
	EW-inos, WZ offshell	31	1	needs SRs & CRs

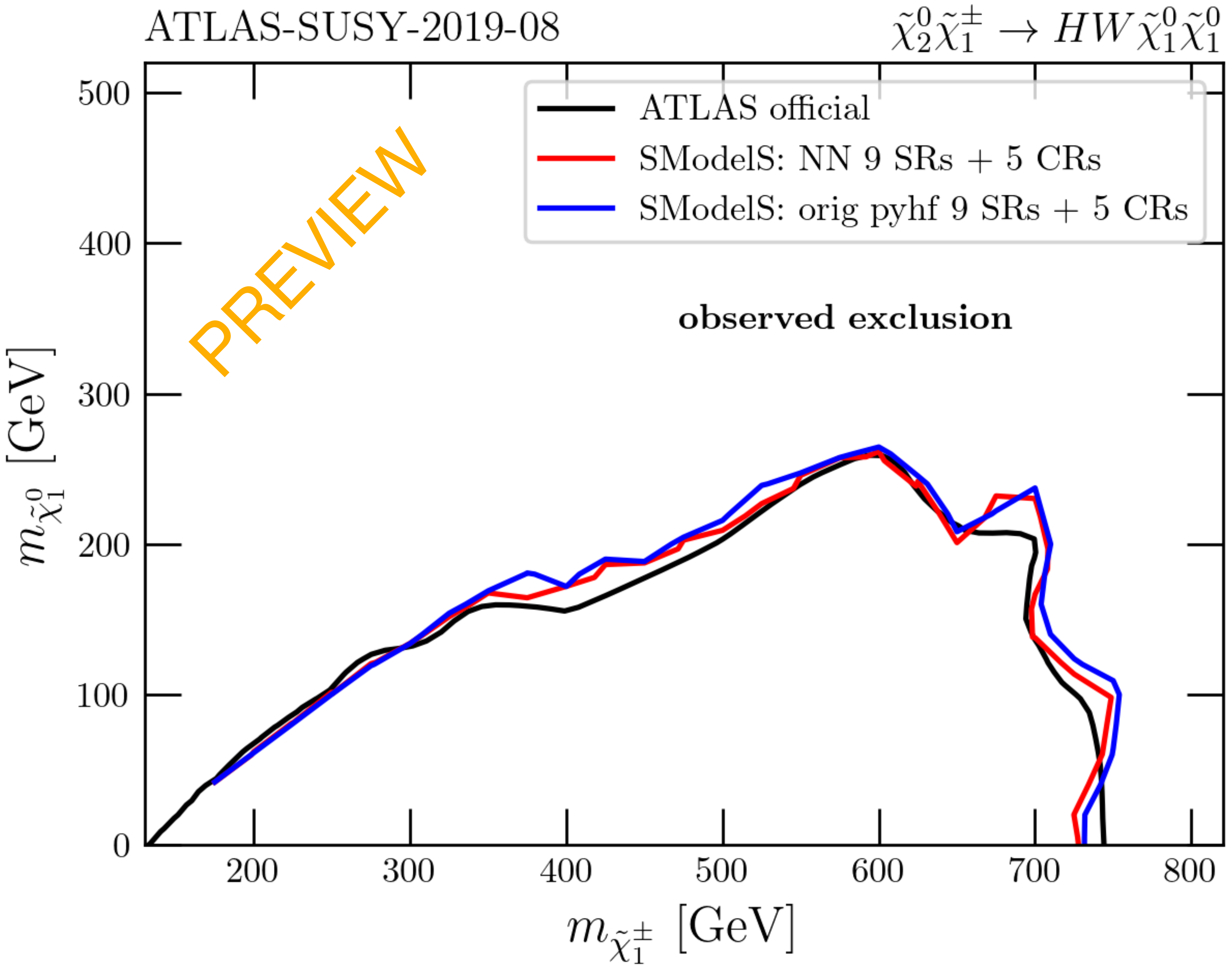
ATLAS-SUSY-2018-04



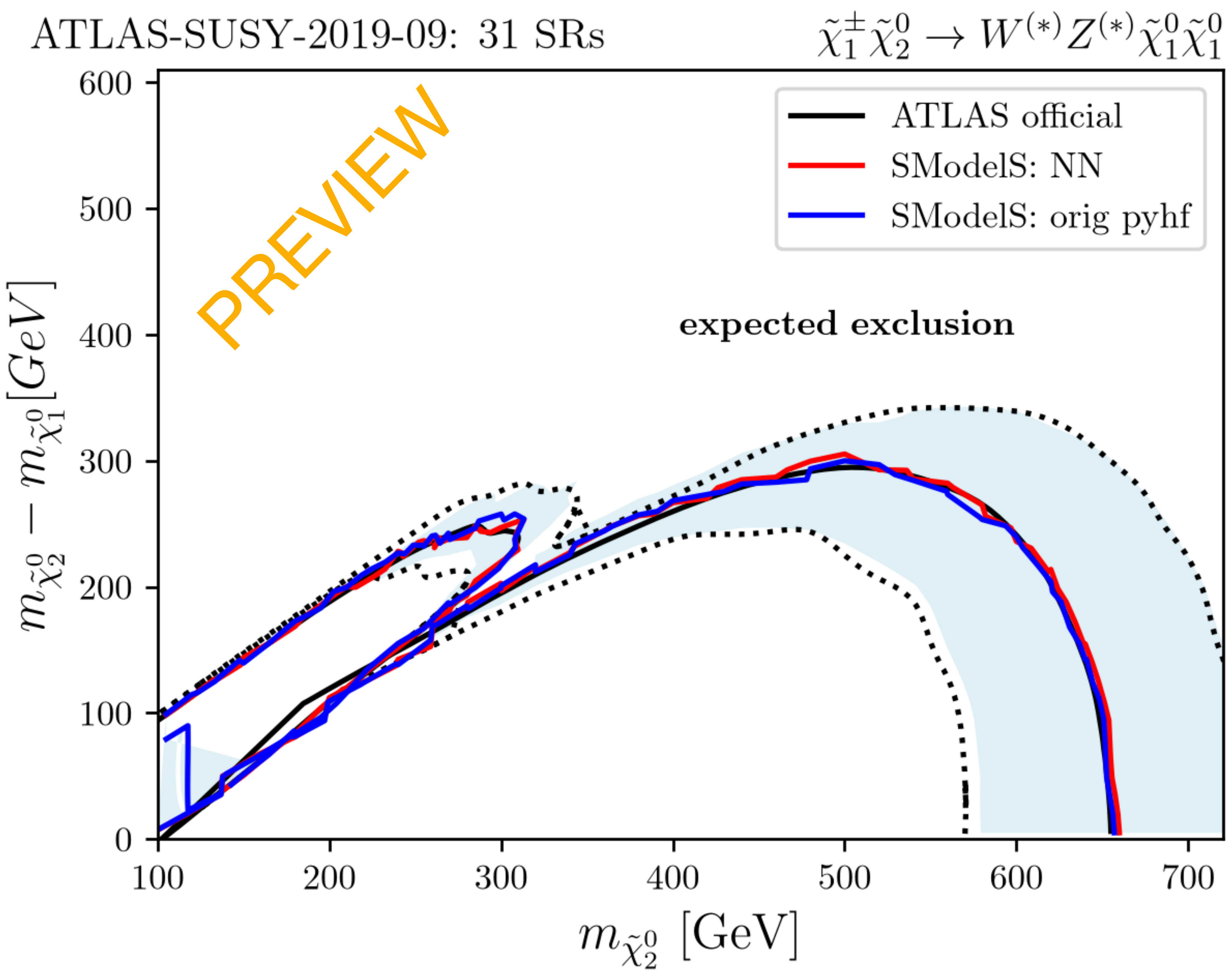


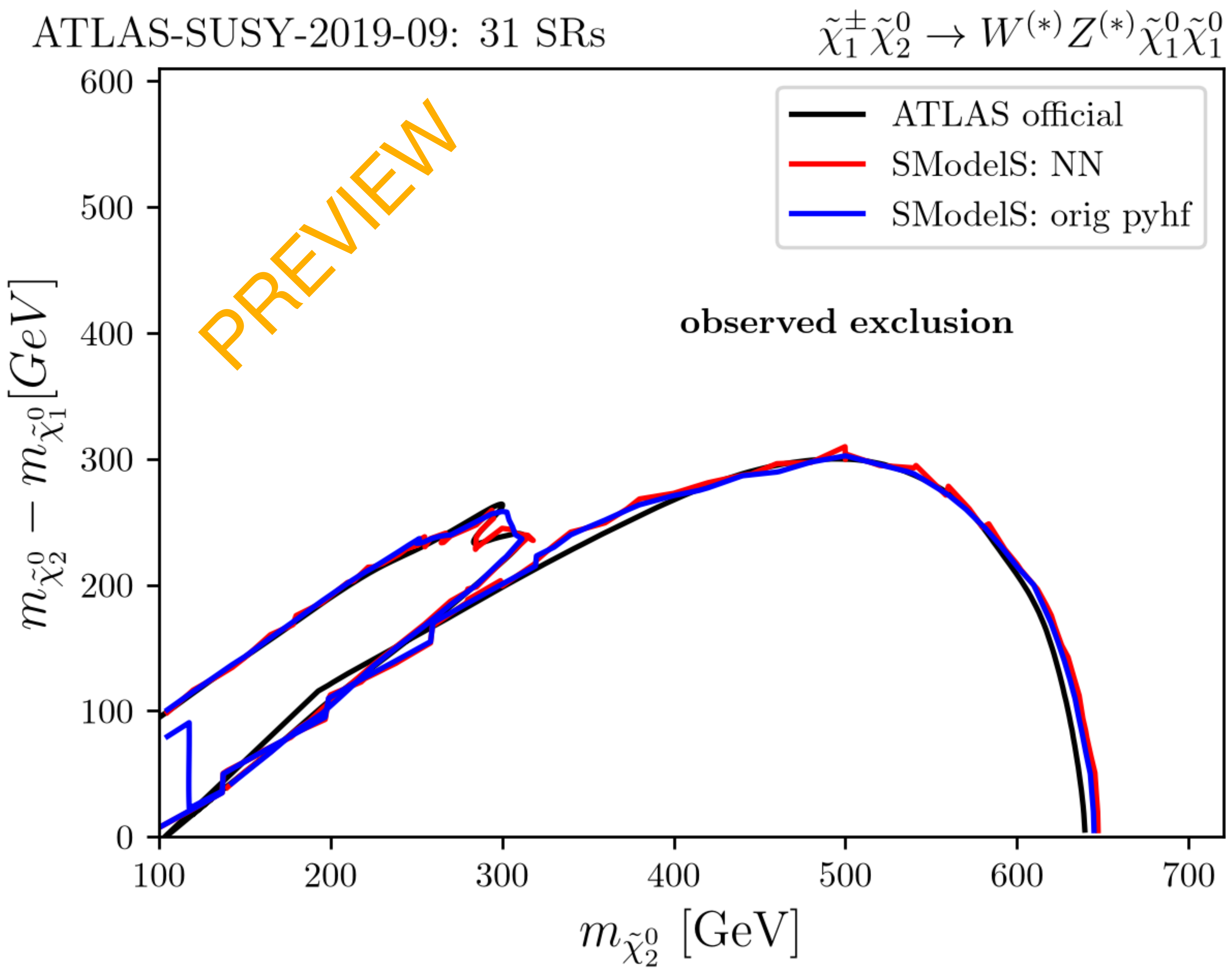
ATLAS-SUSY-2019-08



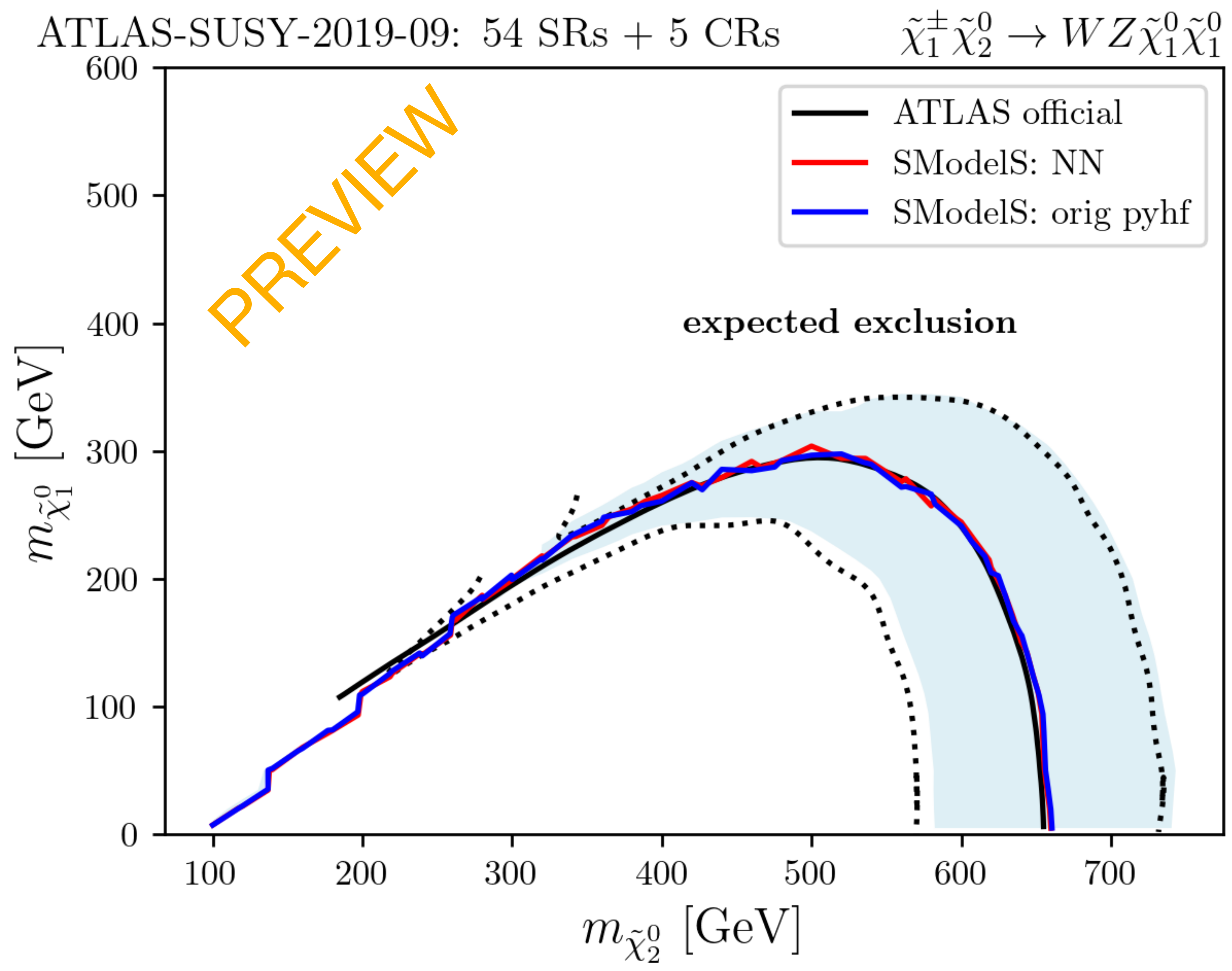


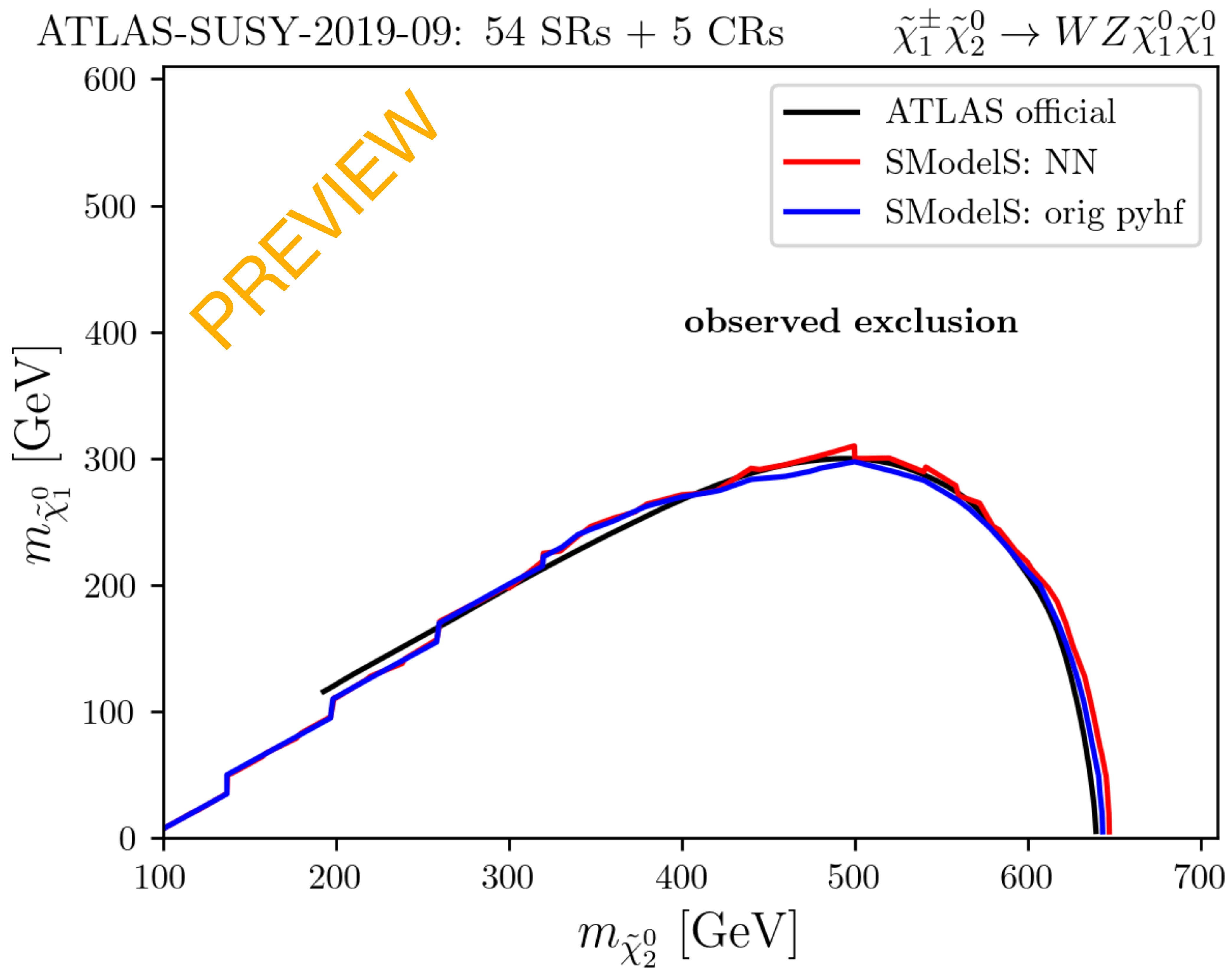
ATLAS-SUSY-2019-09 (Off shell)



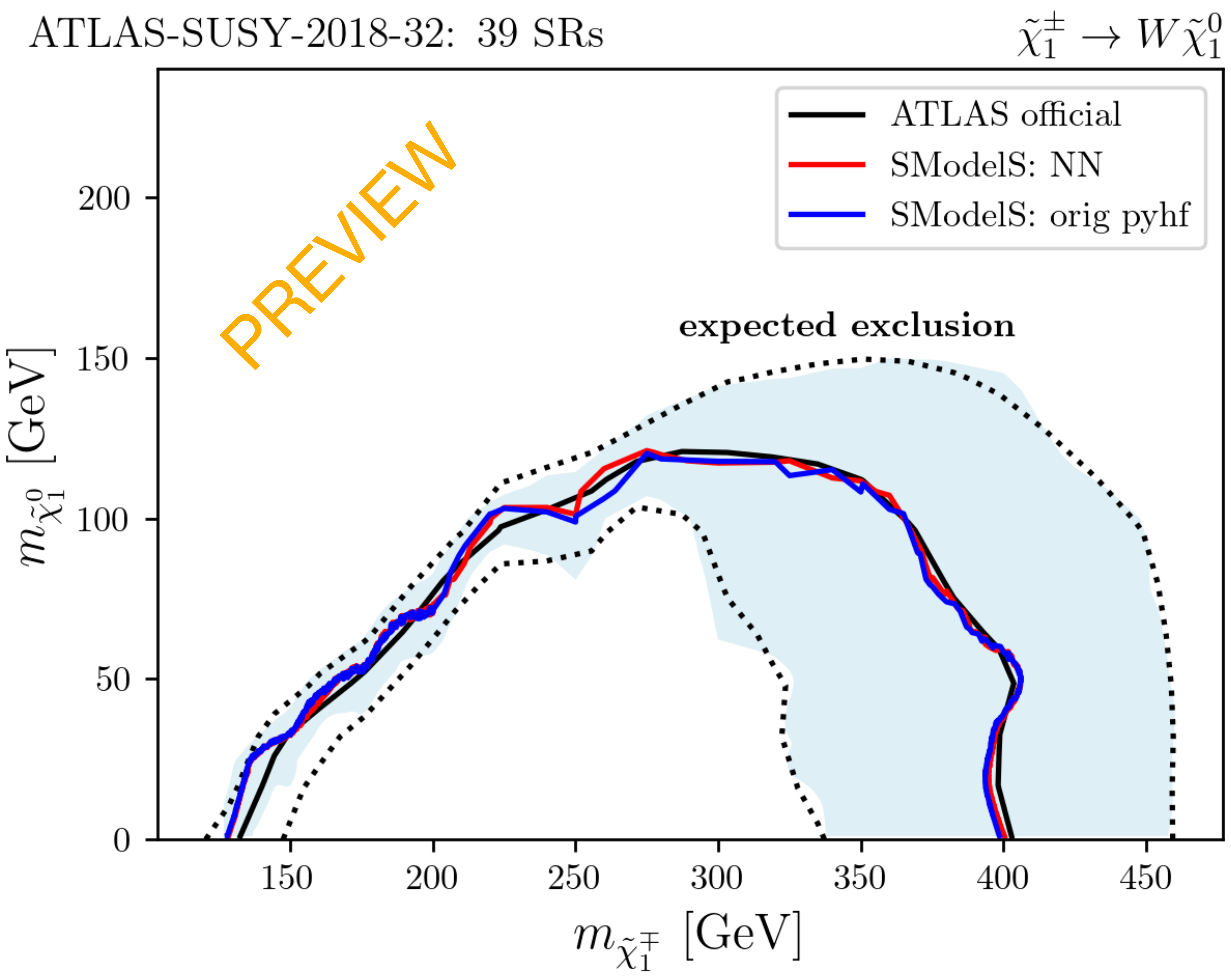


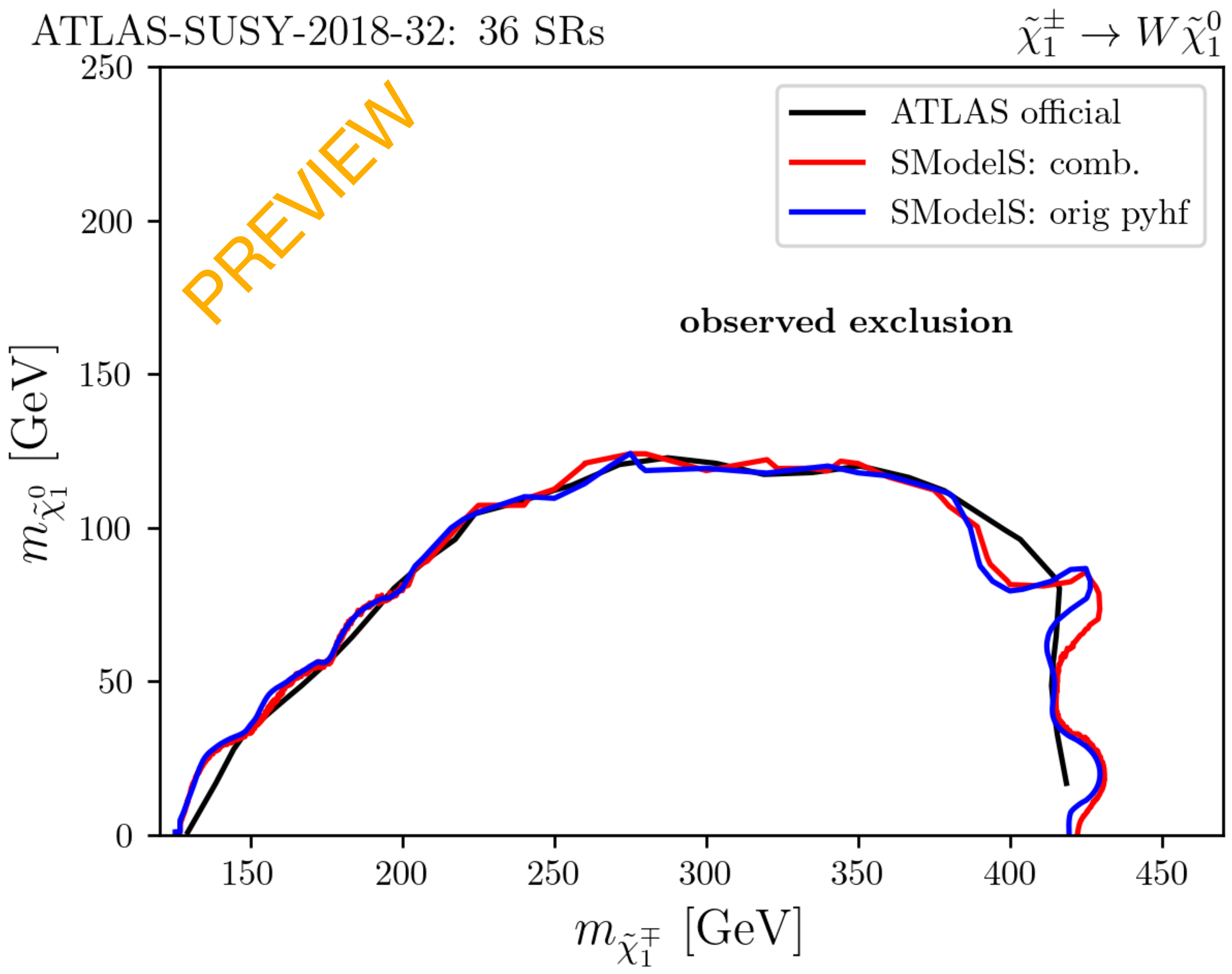
ATLAS-SUSY-2019-09 (On shell)





ATLAS-SUSY-2018-32





Available now!!

GitHub | pypi package 3.2.0 | Anaconda.org 3.1.1 | docs main

2 Sep 2026: SModelS version 3.2.0 available (what's new)

- First support for NN surrogate statistical models (ONNX format)

https://smodels.github.io/

Conclusions

- We are releasing **OLLL the open library of learned likelihoods of LHC experiments**.
- Fast and ~~furious~~ precise LHC reinterpretation possible.
- Already available on SModelS
- Models on Zenodo. Code on Github

What's next?

- Releasing the paper of course :)
- Interface to other reinterpretation tools.
- Protomodel vs LHC analyses exploration.
- Adding more analyses to the library. CMS too!
- AI assistance to optimise the pipeline.

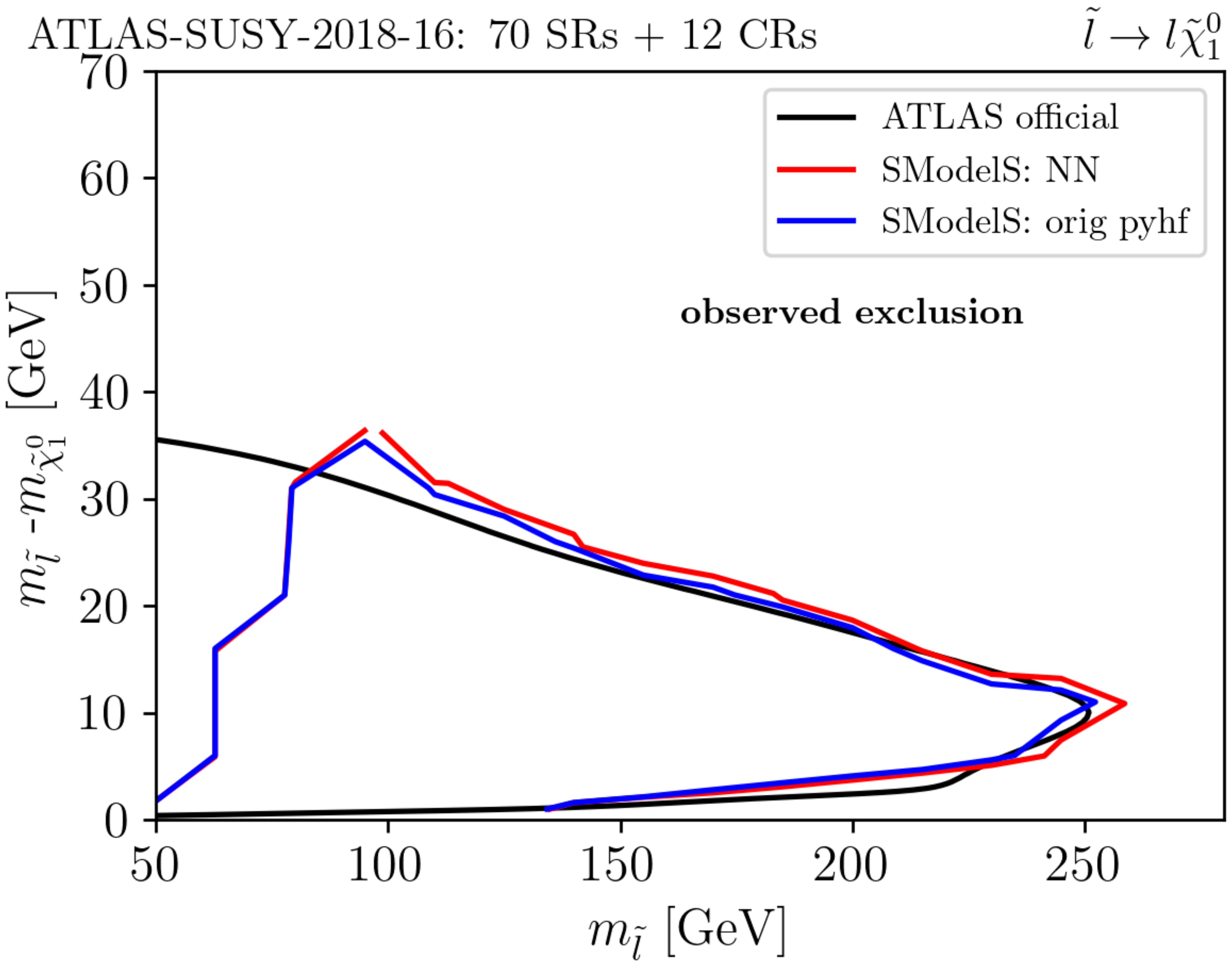
Final message:

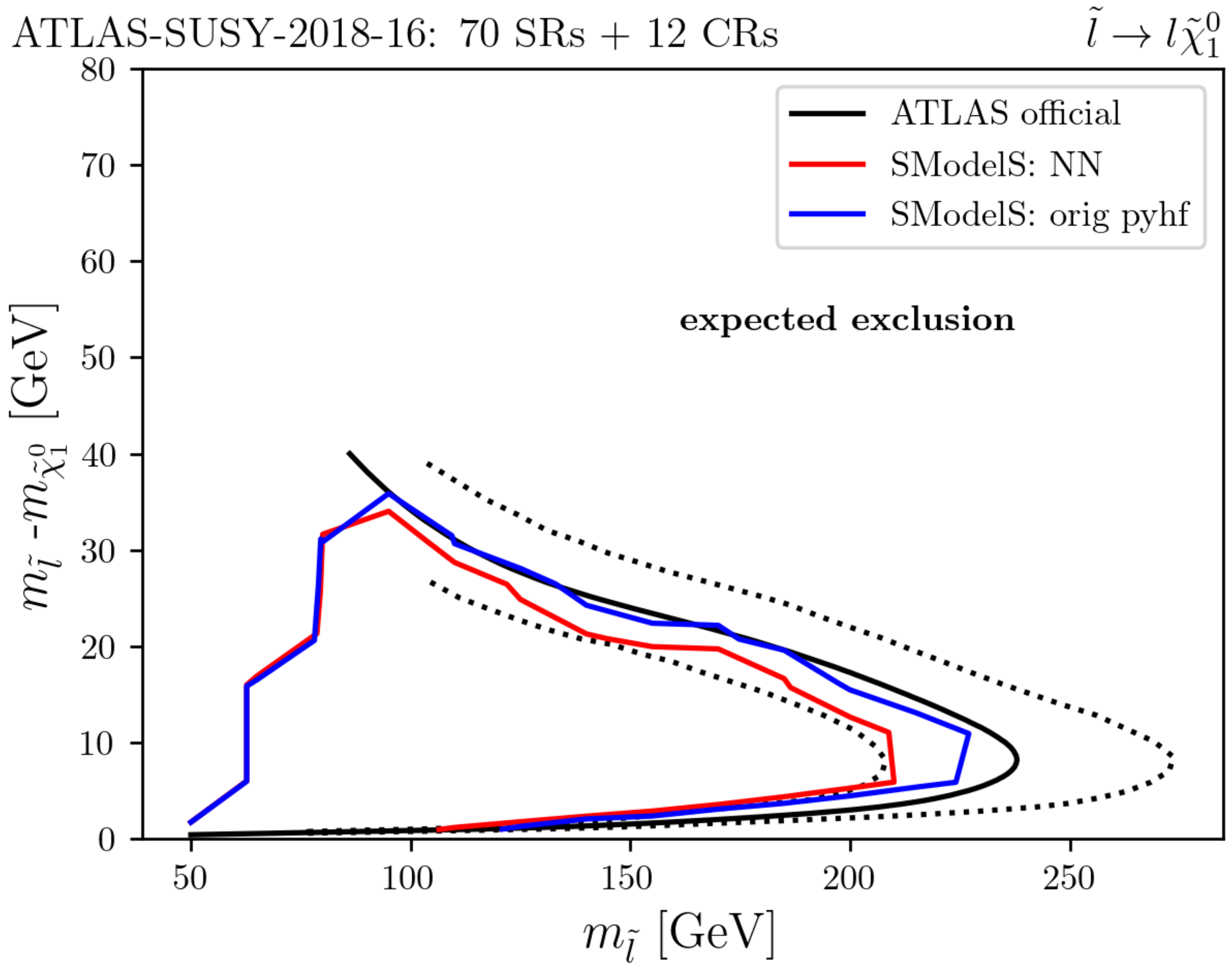
The path from concept to production is dark and full of terrors

THANK YOU

BACK-UP

ATLAS-SUSY-2018-06 (sleptons)





ATLAS-SUSY-2018-16 (electroweakinos)