# Machine Learning LHC Likelihoods


Humberto Reyes-Gonzalez
RWTH Aachen


3/12/2024
HEP group, University of Siegen

# Collaboration.

## ML profile likelihoods arXiv:2502.XXXX

**J. Araz, A. Butter, J. Iturriza, S. Kraml, R. Maselek, C. Krause, W. Waltenberger, H.R.G.**

## NF-likelihoods arXiv:2309.09743

**H.R.G., Riccardo Torre.**

# Motivation

- **Likelihood functions (full statistical models)** parametrise the full information of an LHC analysis; wether it is New Physics (NP) search or an SM measurement.

- Their **preservation** is a key part of the **LHC legacy.**

- **Usage:** Resampling, Reinterpretation in the context of different NP models and/or with with different statistical approaches,….

- They are particularly relevant in the context of **reinterpretation of LHC results**.

- The challenge: Full statistical models, are often high dimensional functions that are both hard to parametrize and evaluate.

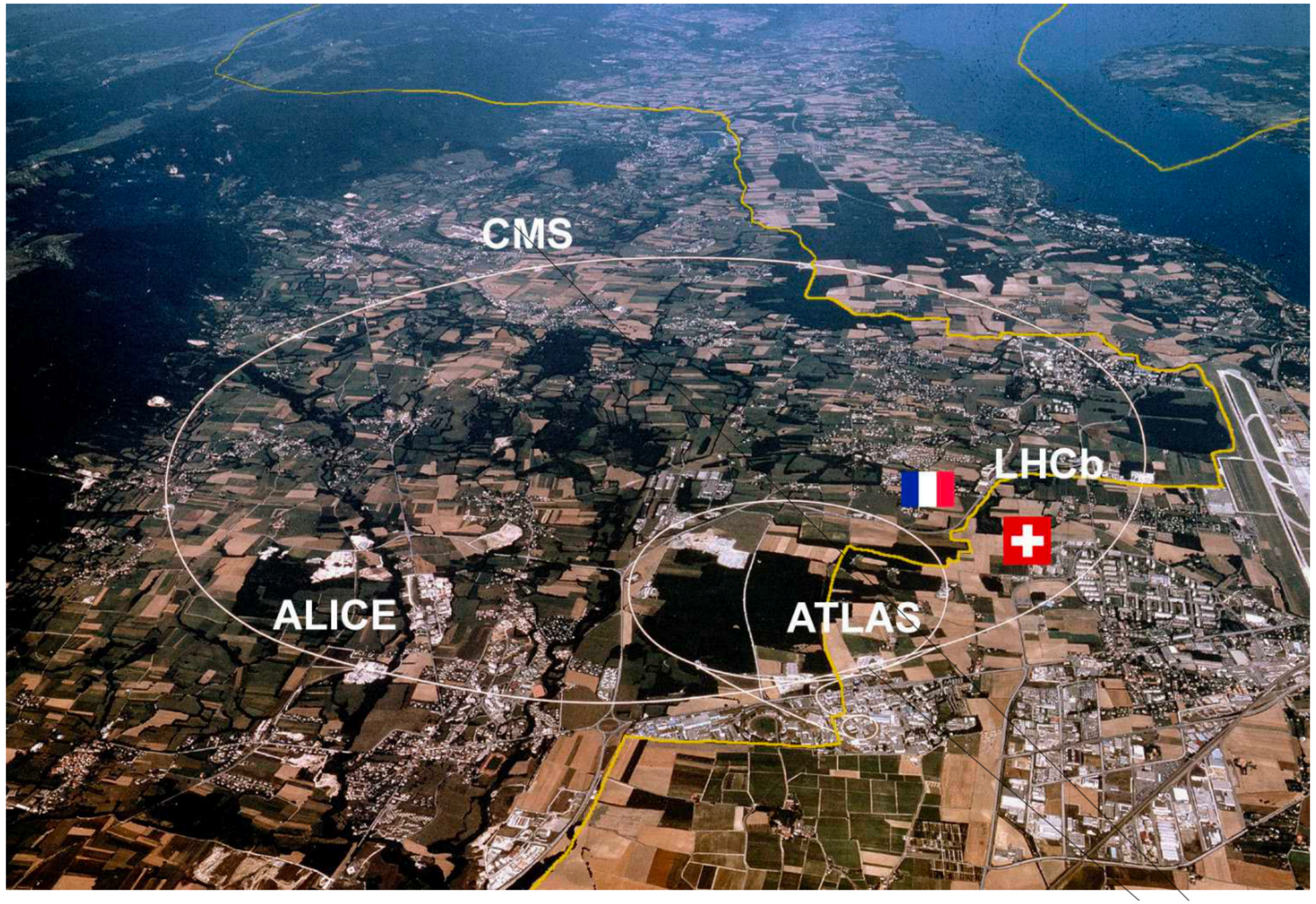
- The proposed solution: **We learn the  likelihood functions as Machine Learning models.**

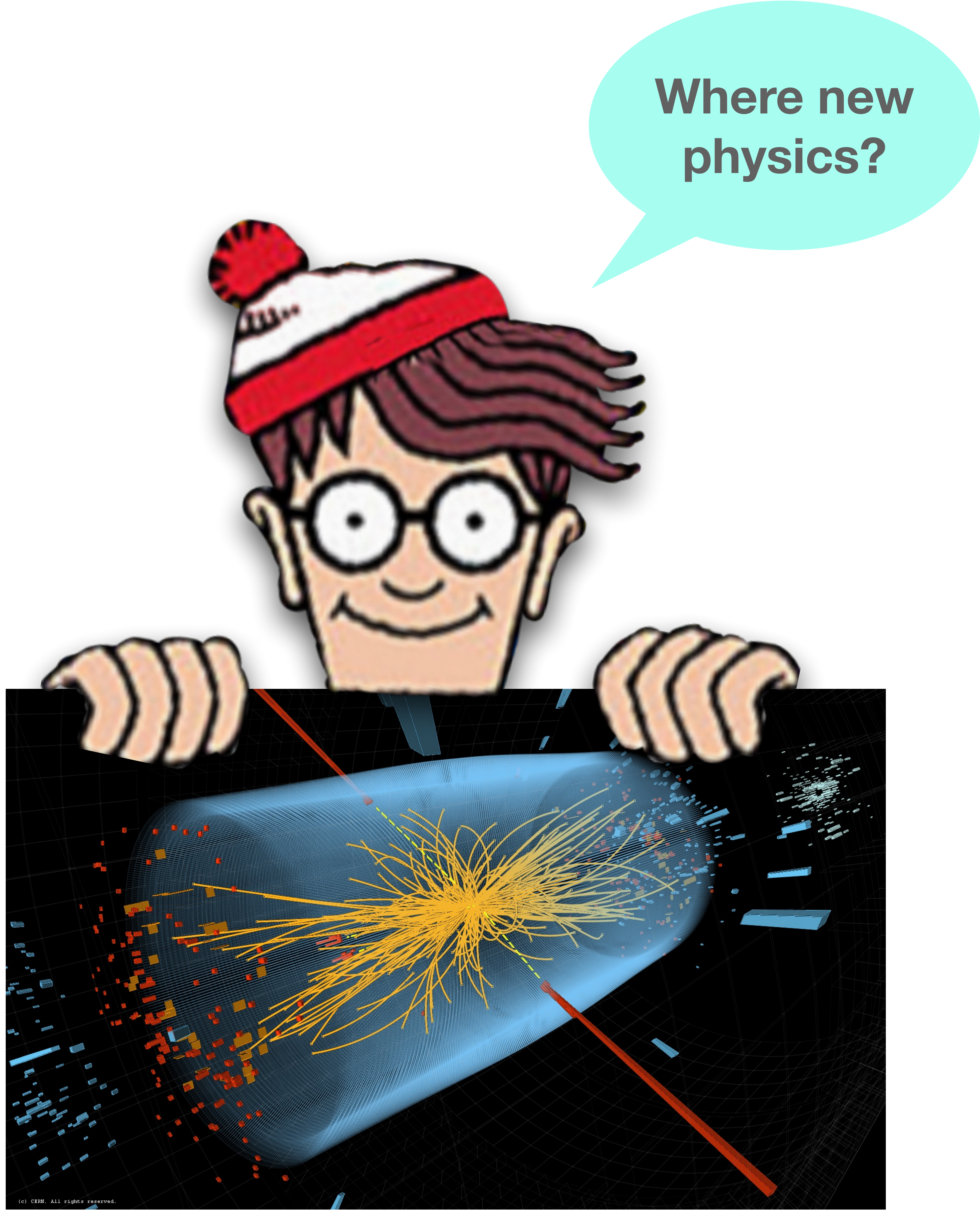
# Overview

- Reinterpretation of LHC results
- LHC likelihoods
- Machine Learning the profile likelihoods of LHC results.
- NFLikelihood: Unsupervised learning LHC likelihoods.
- Conclusions

# Reinterpretation of LHC results

# The LHC



# New physics searches at the LHC

RESONANCES

DARK MATTER

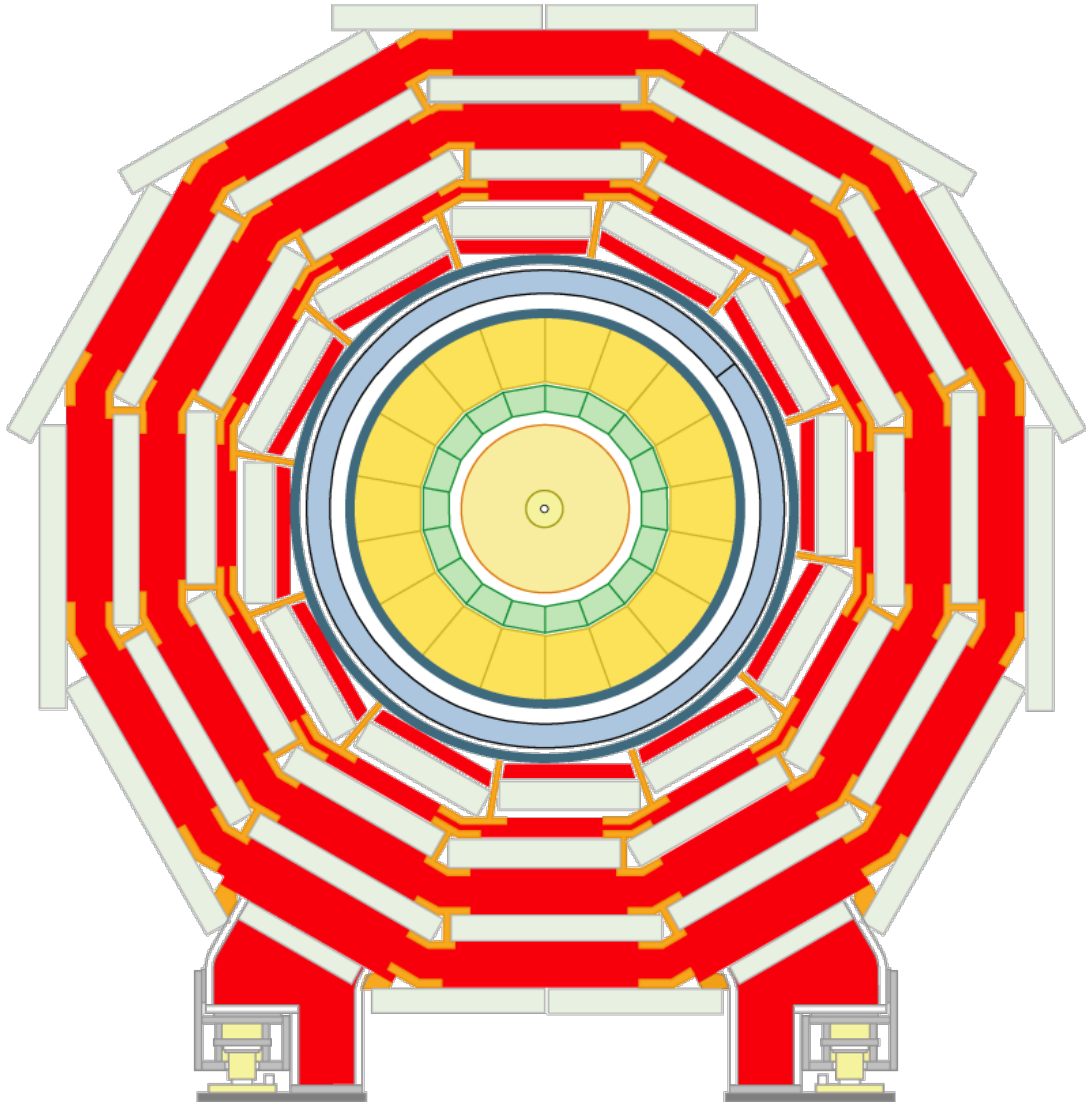
EXOTICS

LEPTOQUARK SUPERSYMMETRY

LONG LIVED PARTICLES

MULTIPLE HIGGSES

?

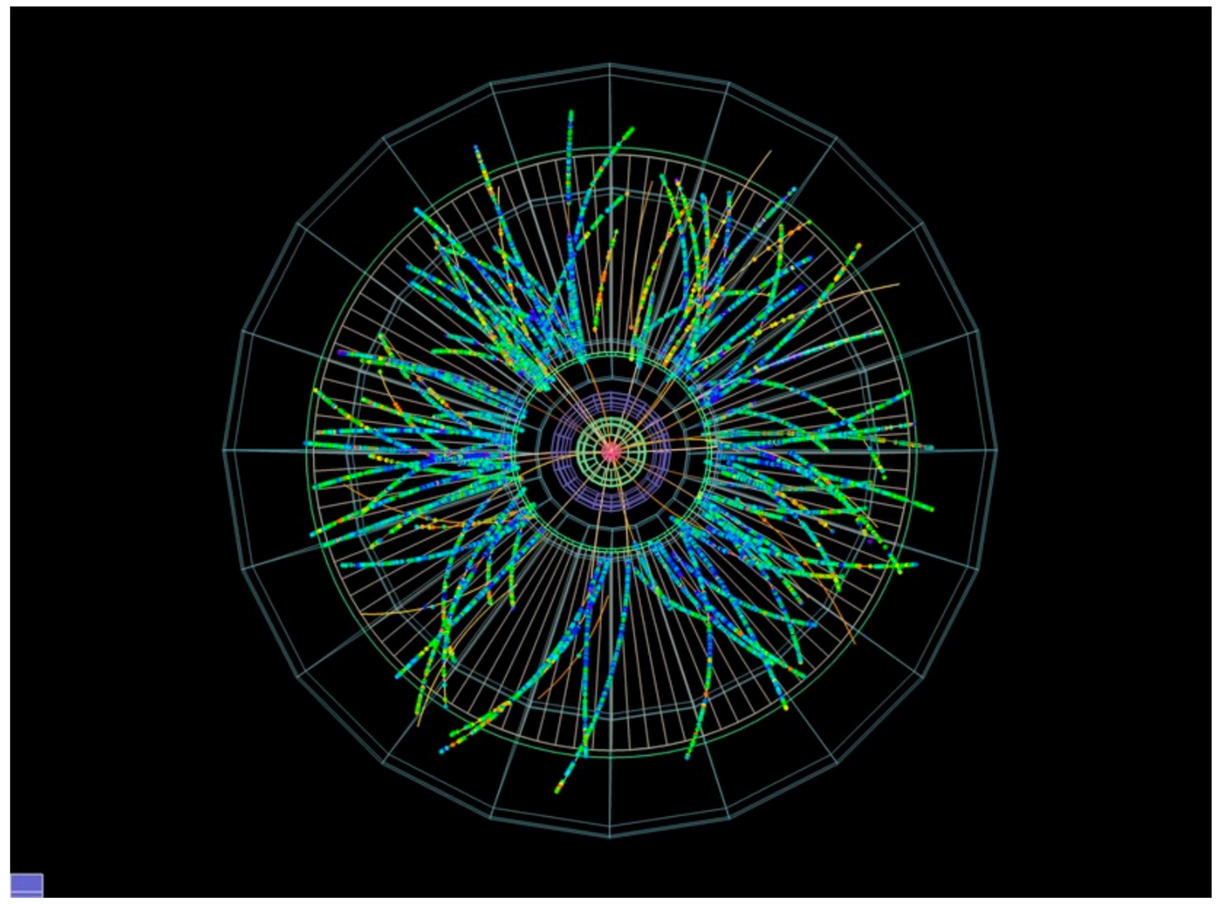


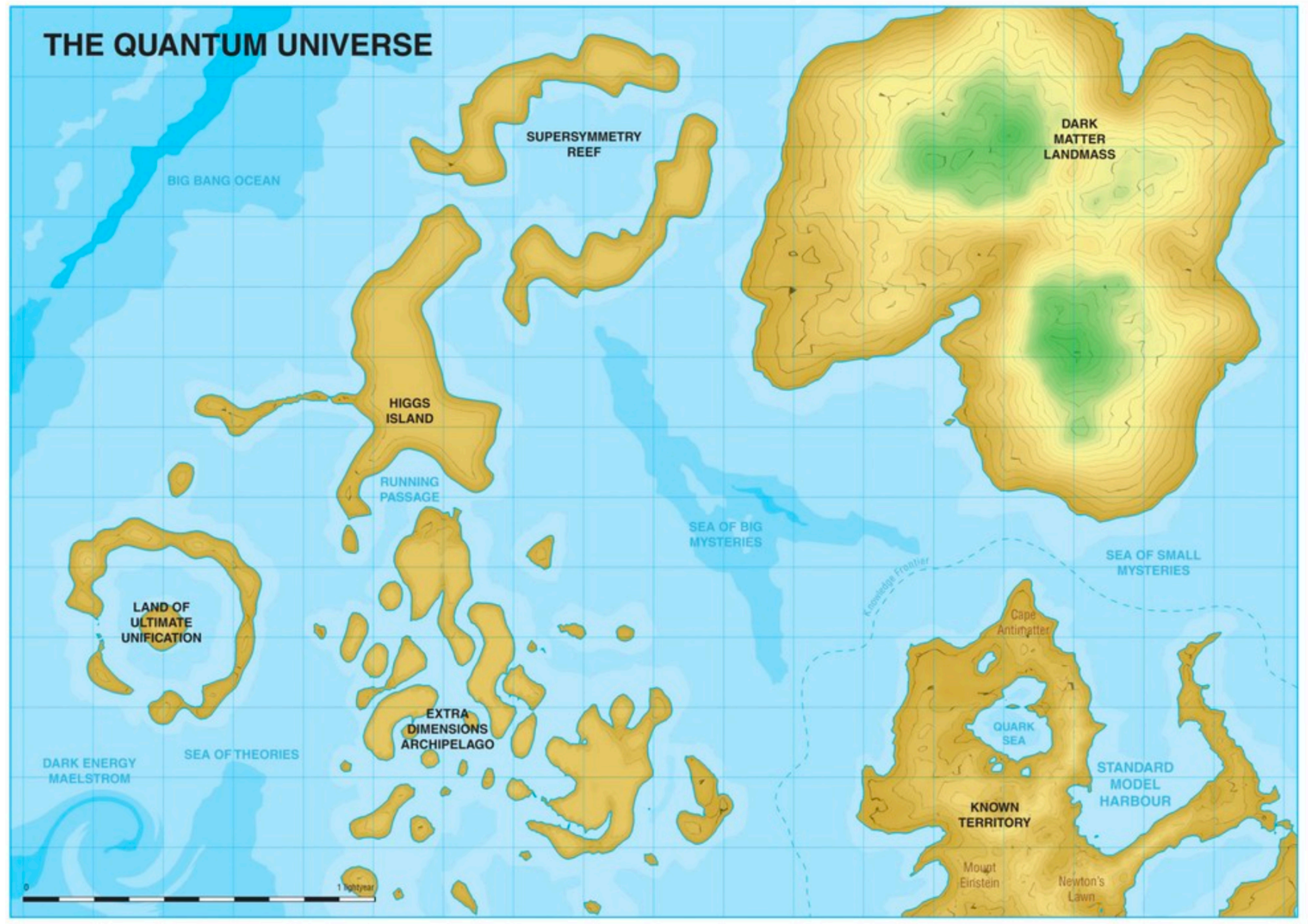


Courtesy of R. Maselek

# Reinterpretation of LHC results

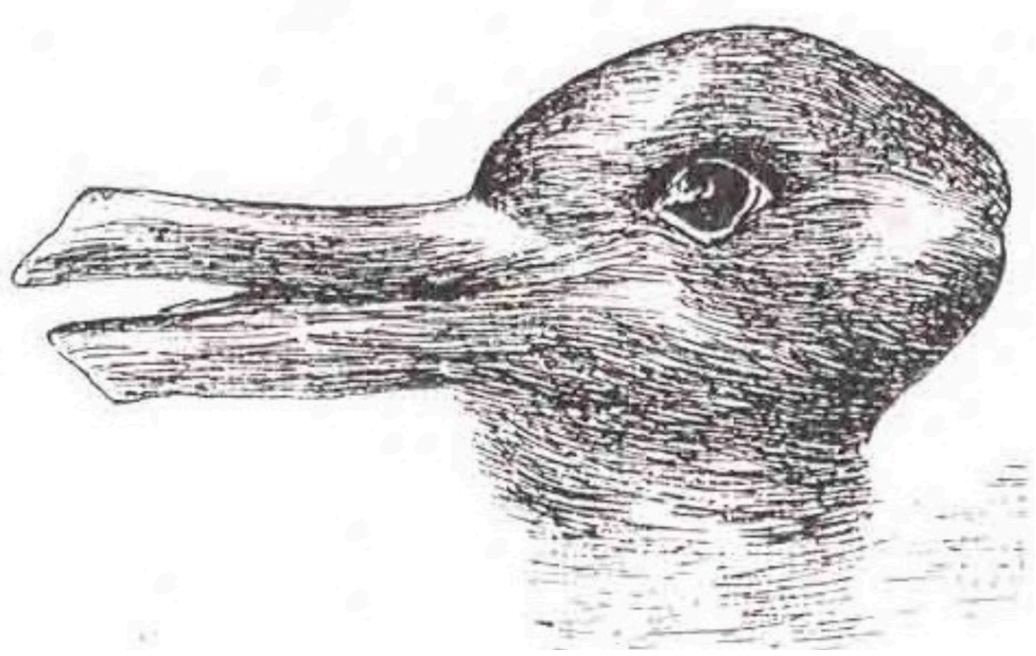
- ATLAS and CMS has an extensive program of searches for new physics.
- Experimental analyses are often optimized and interpreted in the context of simplified models, EFTs, or specific BSM scenarios.






- However, there is a sea of proposed theories/scenarios for new physics.
- Many are non-minimal, less-known, not-thought-of- yet... theories that are not directly interpreted by LHC searches.

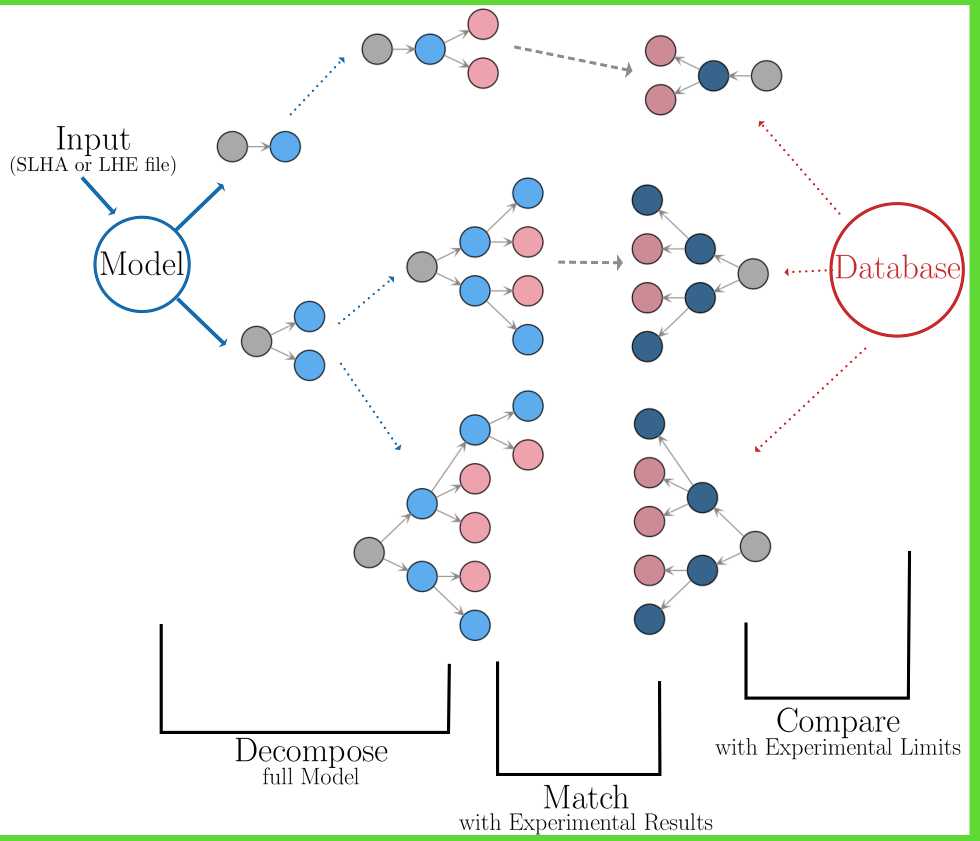
- The aim of the **LHC reinterpretation framework** is to be able to test any BSM theory against LHC results.
- A very active field with strong communication between theorists and experimenters.



https://twiki.cern.ch/twiki/bin/view/LHCPhysics/InterpretingLHCresults

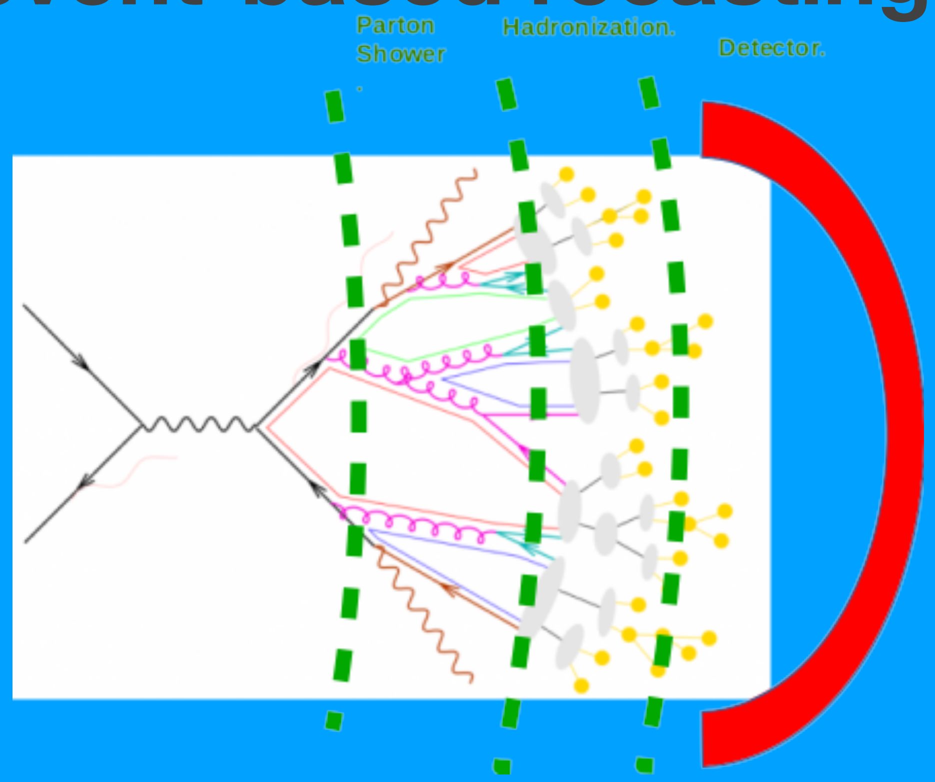
# Reinterpretation of LHC results

## Simplified Model Spectra.




SModelS

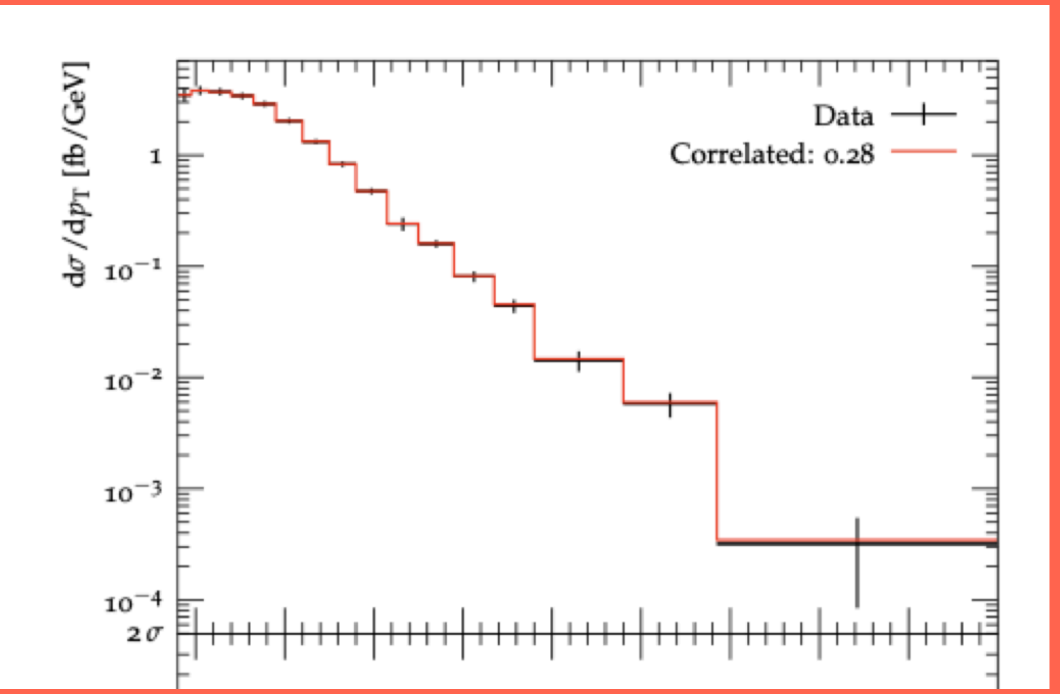
## MC event-based recasting.




CHECKMATE

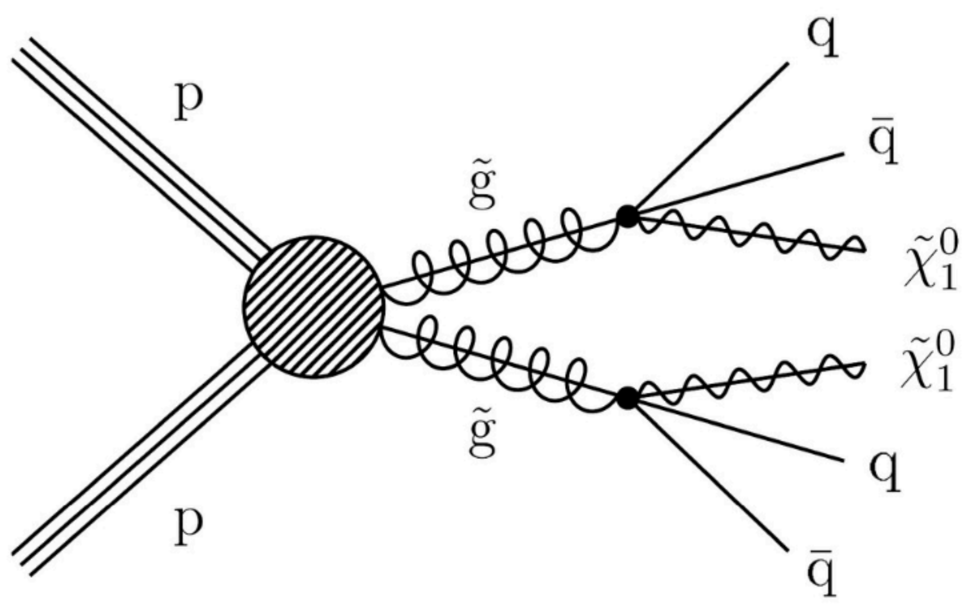
MADAnalysis 5

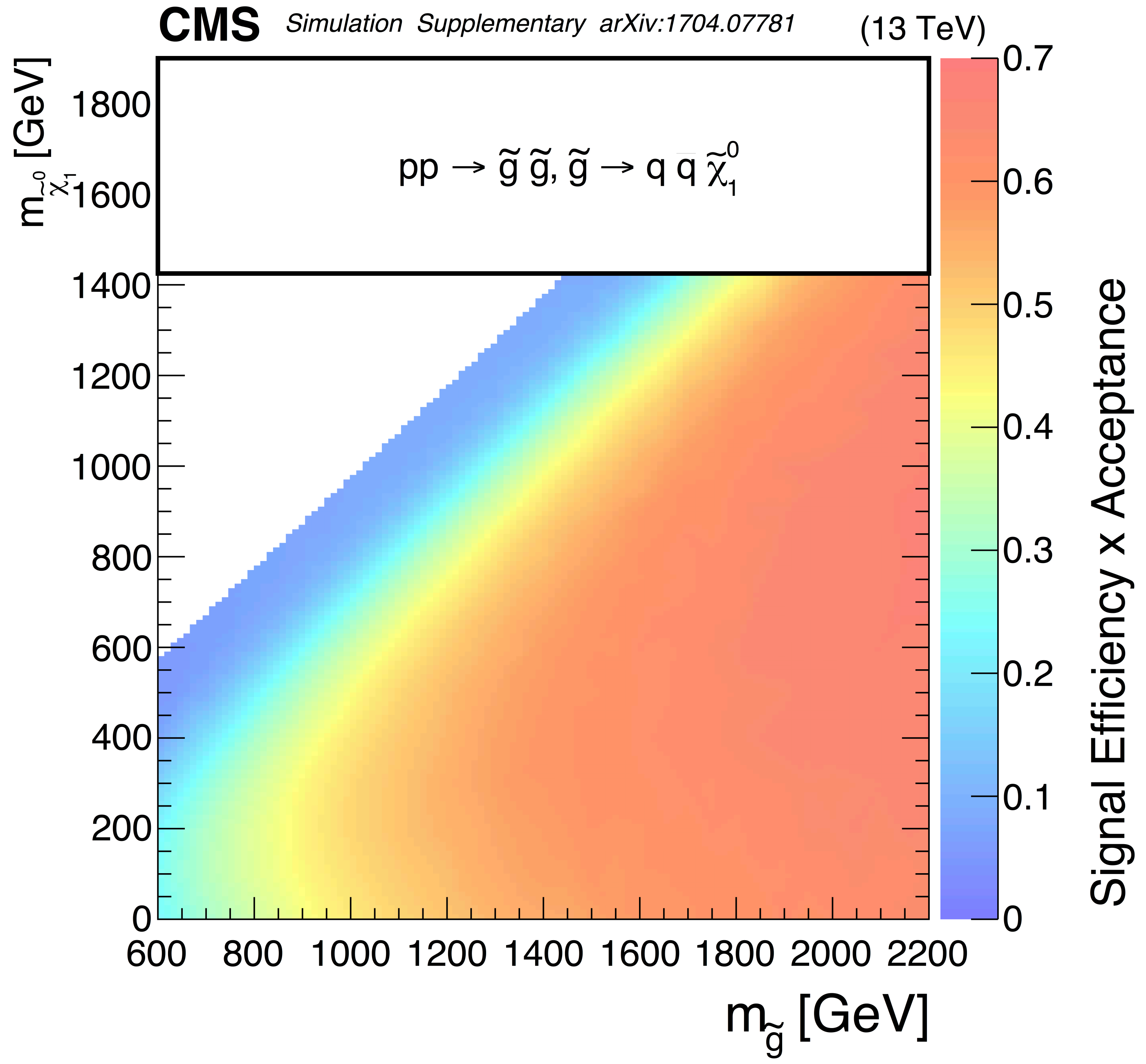
## Recasting measurements




CONTUR

# Simplified Model Spectra re-interpretation




- A lot of LHC new physics searches present their results as upper limit and efficiency maps in the context of simplified model spectra (SMS).
- SMS results are relatively straightforward to reinterpret.
- Direct comparison of $\sigma \times BR$ of a BSM theory with $(\sigma \times BR)_{UL}$
- **Very fast!** No MC required.

.

# SmodelS: 10 years of LHC reinterpretation

- SModelS is based on a general procedure to decompose Beyond the Standard Model (BSM) collider signatures into Simplified Model Spectrum (SMS) topologies.
- Previously, it focused only on $Z_2$ signatures . As of version 3, this has been extended arXiv:2409.12942.

**Working principle.**

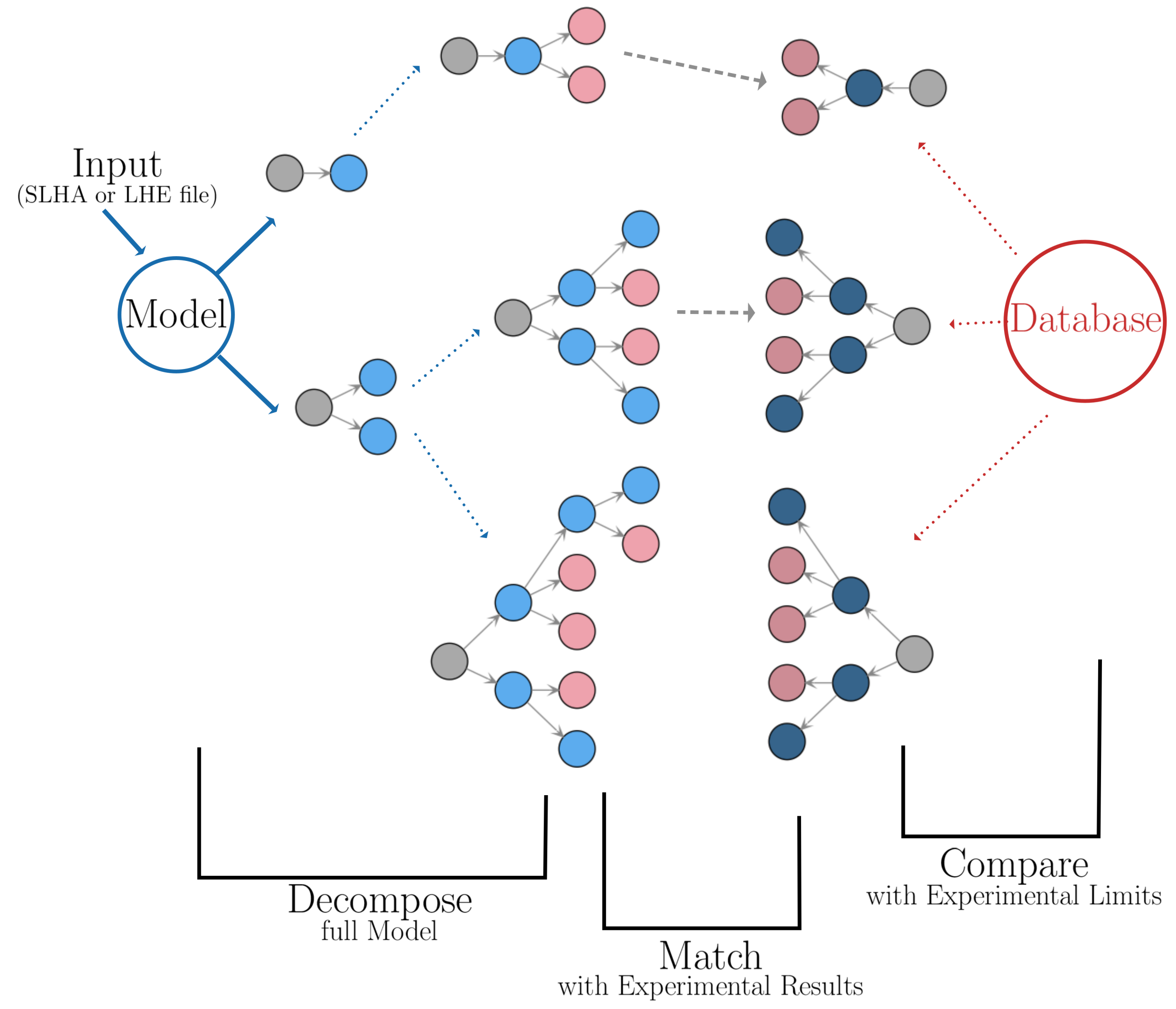
- the decomposition of the BSM spectrum into SMS topologies
- a database of **more than a hundred experimental SMS results.**
- the interface between decomposition and results database to compute limits.
- Including finding the best combination of results for each decomposition. A la TACO: arXiv:2209.00025

**A tool for interpreting simplified-model results from the LHC**

Mohammad Mahdi Altakach, Sabine Kraml, Andre Lessa, Sahana Narasimha, Timothée Pascal, Camila Ramos, Humberto Reyes-González, Théo Reymermier, Yoxara Villamizar, Wolfgang Waltenberger

Previously involved in SModelS: Gaël Alguero, Federico Ambrogi, Jan Heisig, Charanjit K. Khosa, Juhi Dutta, Suchita Kulkarni, Ursula Laa, Veronika Magerl, Wolfgang Magerl, Philipp Neuhuber, Doris Proschofsky, Jory Sonneveld, Michael Traub, Matthias Wolf, Alicia Wongel

https://smodels.github.io/



# Likelihoods of LHC new physics searches.

# Milestones on full statistical model publication

- ATLAS started publishing full statistical models of NP searches (2019) ATL-PHYS-PUB-2019-029.
- First ML learned likelihoods. The (supervised) DNNLikelihood (2019) arXiv:1911.03305
- Release of the pyhf package to construct statistical models (2020) 10.21105/joss.02823
- Interface with reinterpretation tools: SmodelS (2020) arXiv:2009.01809, MadAnalysis (2022) arXiv:2206.14870,
- Spey: Generalised framework for likelihood handling (2023) arXiv:2307.06996.
- Unsupervised Likelihood learning. The NFLikelihood (2023) arXiv:2309.09743
- COMBINE: The CMS statistical analysis and combination tool (2024) arXiv:2404.06614

**Let's keep it going!**

# Likelihoods of LHC results

***Bayes theorem:***

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

Likelihood
Prior
Posterior
Evidence

***LHC Statistical model:***

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

signal/control yields
(Observed) data
Nuisance parameters (uncertainties)
(Auxiliary) data
Parameters of Interest (signal strength, observables, etc.)

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

# The profile likelihood

- For new physics searches reinterpretation, we are often interested in **the Profile Likelihood (PL)**.

$$P(x\,|\,\mu;\hat{\theta}(\mu))$$

- The PL is defined as a function where the nuisance parameters are fixed such that the likelihood is maximised given a signal strength $\boldsymbol{\mu}$.
- In the case of positive signal $\boldsymbol{\mu}$=1, otherwise if data is Standard Model like $\boldsymbol{\mu}$=0.
- The PL is a function of the signal yields (data), $\boldsymbol{N_s}$.
- Or  of the signal yields $\boldsymbol{n_s}$ AND the control yields $\boldsymbol{N_c}$ *. Depending on how is fitted.



*Relevant in case of e.g. signal leakage

# Likelihood ratio test statistic

- With the PL we construct **Log Likelihood Ratio (LLR) tests, e.g.**

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

- Depending on how the PL is fitted and defined, we can derive upper limits, exclusion confidence levels and discoveries, by computing a corresponding p-value.

$$p_{\mu,\mathrm{obs}} = \int_{\tilde{q}_{\mu},\mathrm{obs}}^{\infty} f\left(\tilde{q}_{\mu} \,|\, \mu'\right) d\tilde{q}_{\mu}$$

# Machine Learning the profile likelihoods

# Motivation.

- In LHC-reinterpretation, to exclude a BSM model, we are mostly interested in the profiled likelihood given a signal strength.

- Optimally, **we can compute the profiled likelihood from pyhf's full statistical models**.

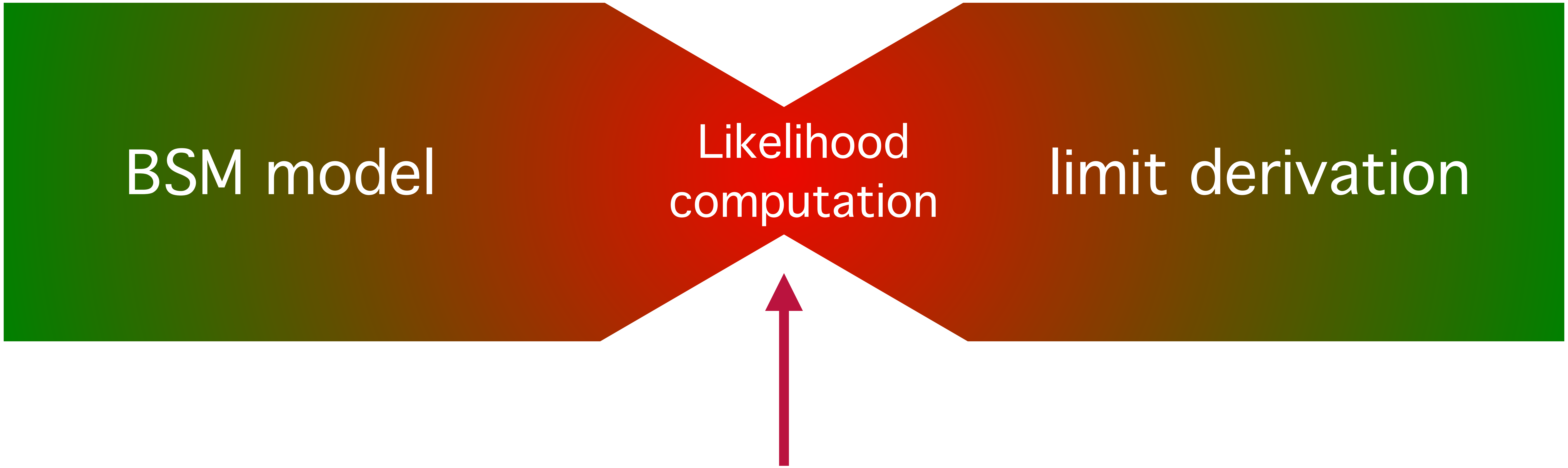
- However, this computation can take several minutes per parameter point. A pheno study often requires to survey thousands of points.

- This considerably scales-up the time consumption. Specially for fast reinterpretation approaches.

- **Using Neural Networks  provides a fast and compact way using profiled likelihoods in our day-to-day pheno studies.**

- **We will super useful for anomaly surveys, proto-models style (arXiv:2105.09020).**

# Challenge: Computational bottleneck

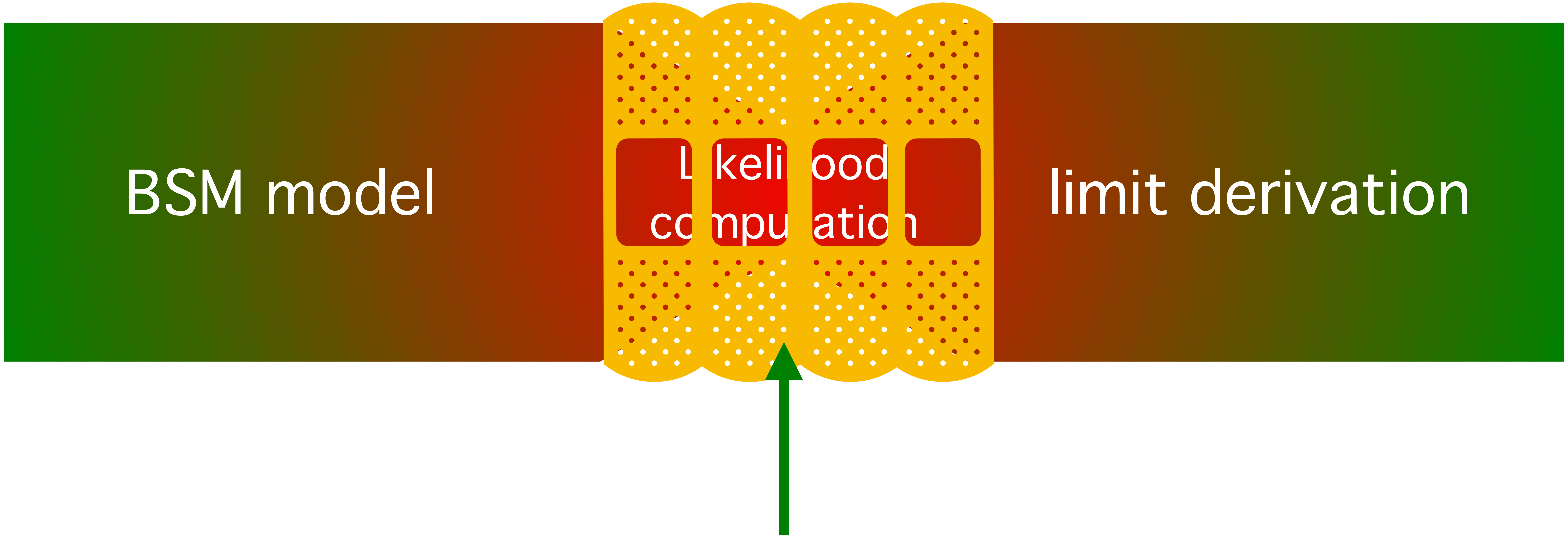



Full statistical model calculations enter here



Courtesy of R. Maselek

# Solution: Machine Learning




Machine Learning enters here



Courtesy of R. Maselek

# The usual suspects to learn

- Likelihood Observed.
- Likelihood a priori Expected.
- Likelilhood a posteriori Expected.
- Asimov Likelihood.

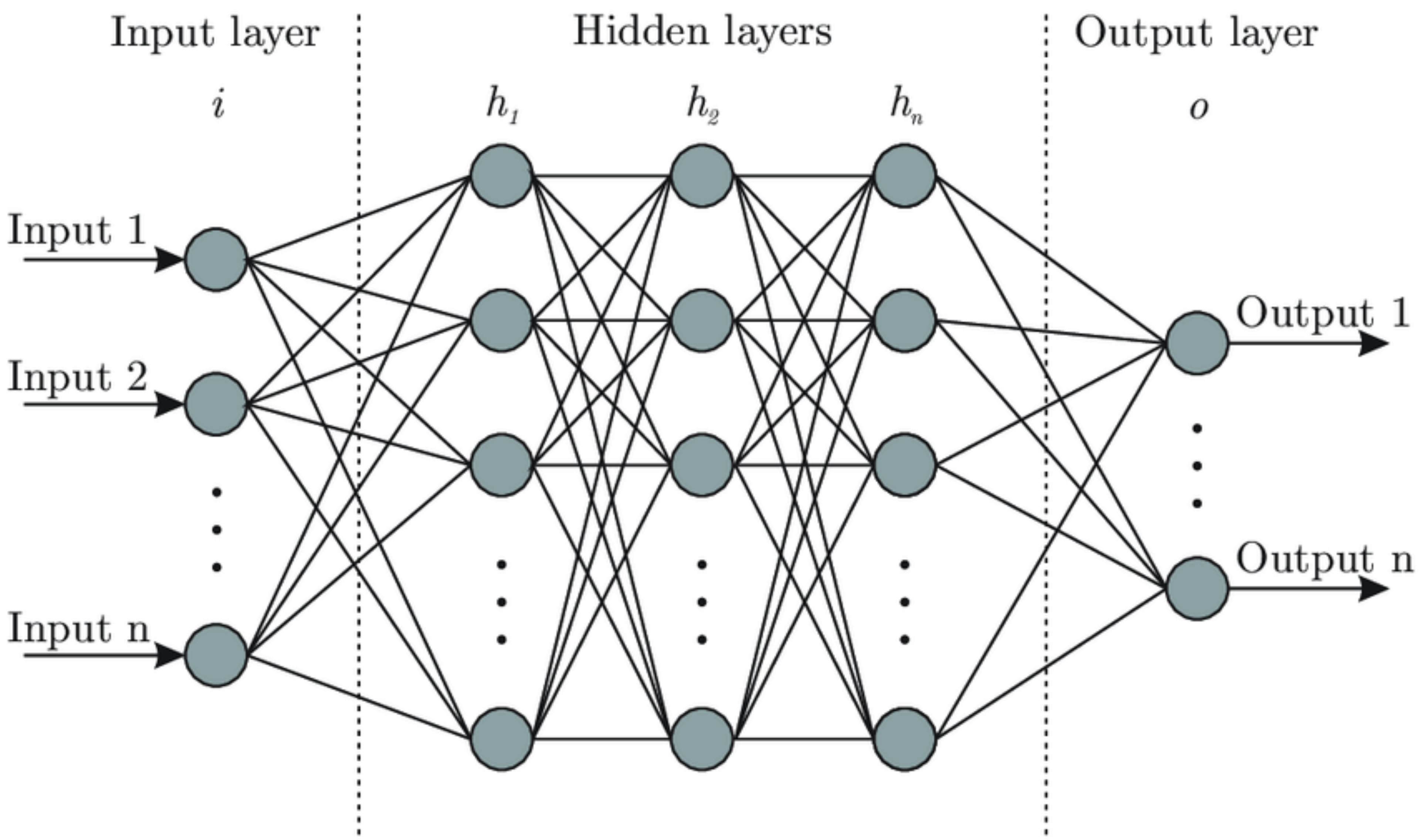
# The usual suspects to learn

## For now we learn:

- Likelihood Observed.
- Likelihood a priori Expected.
- Likelihood a posteriori Expected.
- Asimov Likelihood.

# Training strategy

- SAMPLING: MCMC Metropolis-Hastings.
- INPUTS: event yields in all bins and channels (including CRs)
- ARCHITECTURE: 5X256
- OUTPUTS: negative log likelihoods (for μ=0 and μ=1), for expected and observed data
- LOSS FUNCTION: MSE but others tested
- OPTIMIZER: ADAM
- SCHEDULER: Cosine Decay with warmup



# Deployment strategy

**SAVING:** After training, the best models for each analysis are ensemble together and saved as **ONNX** files and stored on Github or Zenodo.
This ensures framework-independent usability. In the spirit of arXiv:2312.14575.

**Les Houches guide to reusable ML models in LHC analyses**

*Jack Y. Araz[1], Andy Buckley[2], Gregor Kasieczka[3], Jan Kieseler[4], Sabine Kraml[5], Anders Kvellestad[6], Andre Lessa[7], Tomasz Procter[2], Are Raklev[6], Humberto Reyes-Gonzalez[8,9,10], Krzysztof Rolbiecki[11], Sezen Sekmen[12], Gokhan Unel[13]*

ONNX

**USAGE:** The NN likelihoods will be available for statistical studies via an Spey (arXiv:2307.06996) backend.

SPEY: smooth inference for reinterpretation studies

Jack Y. Araz

```python
from spey import BackendBase, ExpectationType
from spey.base.model_config import ModelConfig

import onnx, onnxruntime
import numpy as np


class MLlikelihoods(BackendBase):
    """
    Spey plug-in to evaluate machine learned likelihoods

    Args:
        signal_yields (``List[float]``): Signal yields
        network_path (``str``): Path to the network onnx file.
    """

```

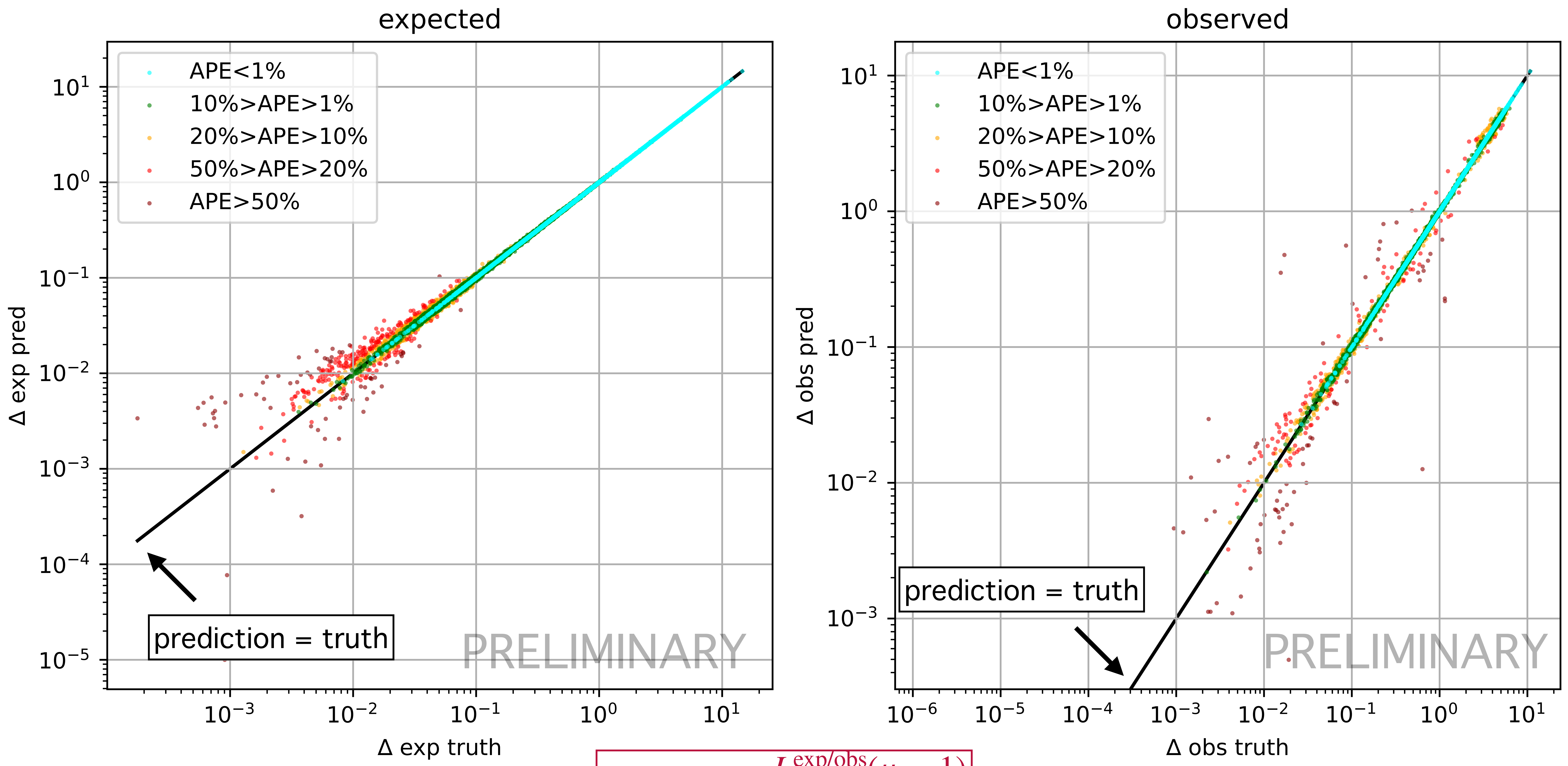
# Example Likelihoods

- **ATLAS-SUSY-2018-04**
- Search for direct stau production in events with two -leptons
- 2 signal bins, 3 control bins
- DOI: 10.1103/PhysRevD.101.032009

- **ATLAS-SUSY-2019-08**
- Search for direct production of e-winos in final states with 1 lepton, MET and a Higgs boson decaying into 2 -jets
- 9 signal bins, 5 control bins
- DOI: https://doi.org/10.17182/hepdata.90607.v4

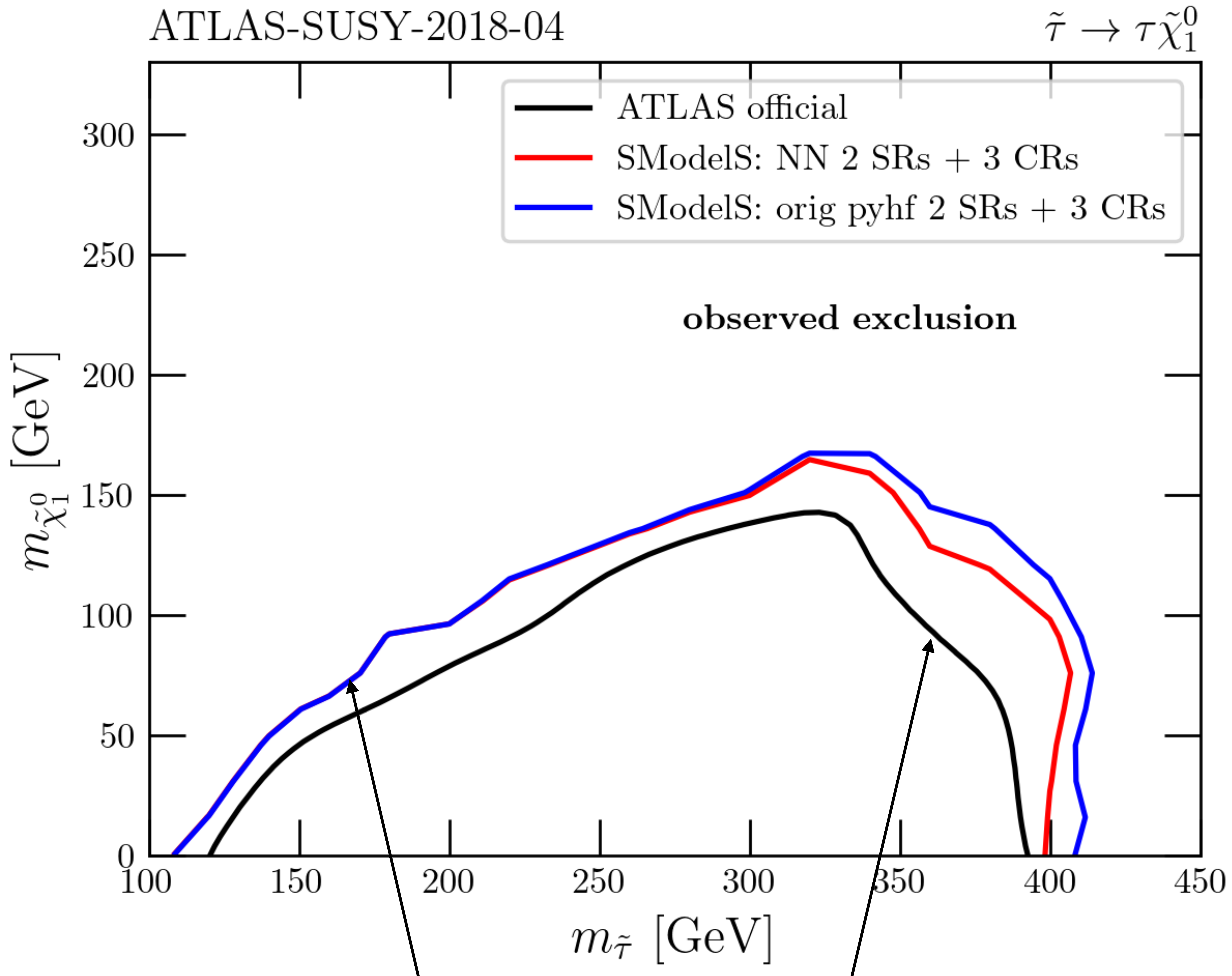
- **1911.12606**
- Searches for electroweak production of supersymmetric particles with compressed mass spectra in $\sqrt{s} = 13$ TeV in collisions with the ATLAS detector.
- Four separate statistical models.
- EWKinos (43,5), Selectrons(15,6) Sleptons(32,5), Smuons(15,5)
- DOI: 10.1103/PhysRevD.101.052005

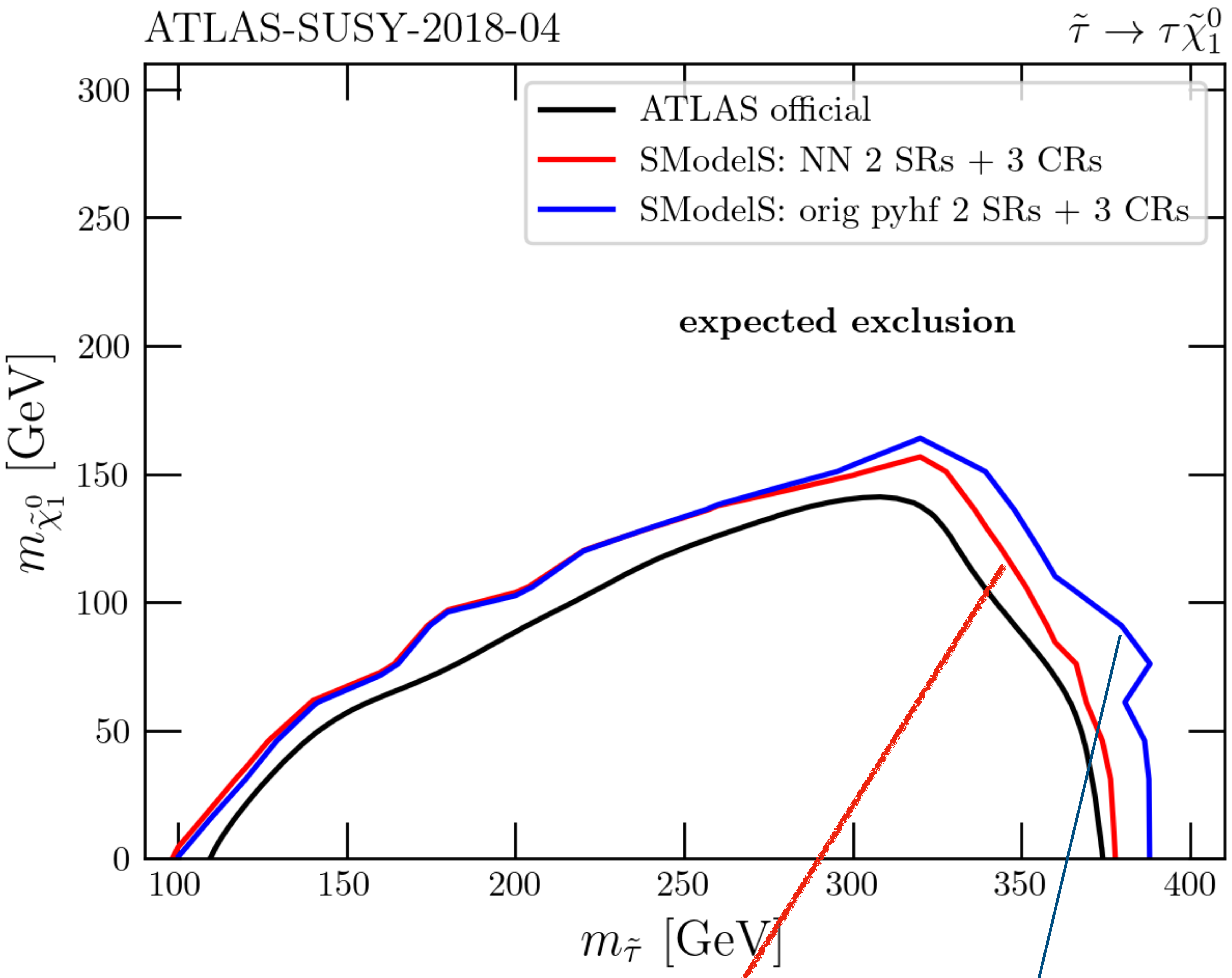
# Results ATLAS-SUSY-2018-04




$$\Delta^{\text{exp/obs}} = \ln \frac{L^{\text{exp/obs}}(\mu = 1)}{L^{\text{exp/obs}}(\mu = 0)}$$

# Results ATLAS-SUSY-2018-04



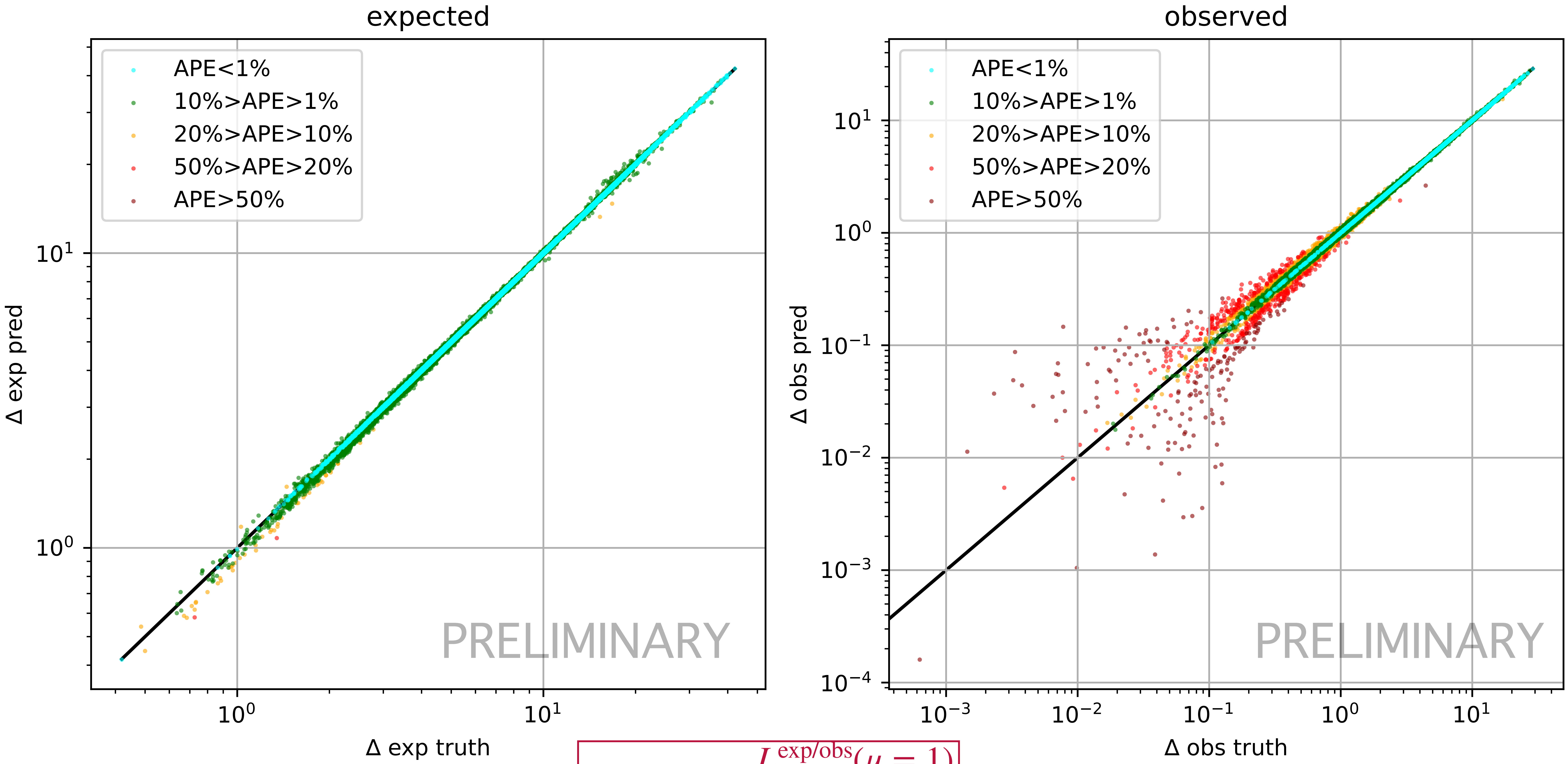

L fitted with Control Regions

L NOT fitted with Control Regions

A priori prescription

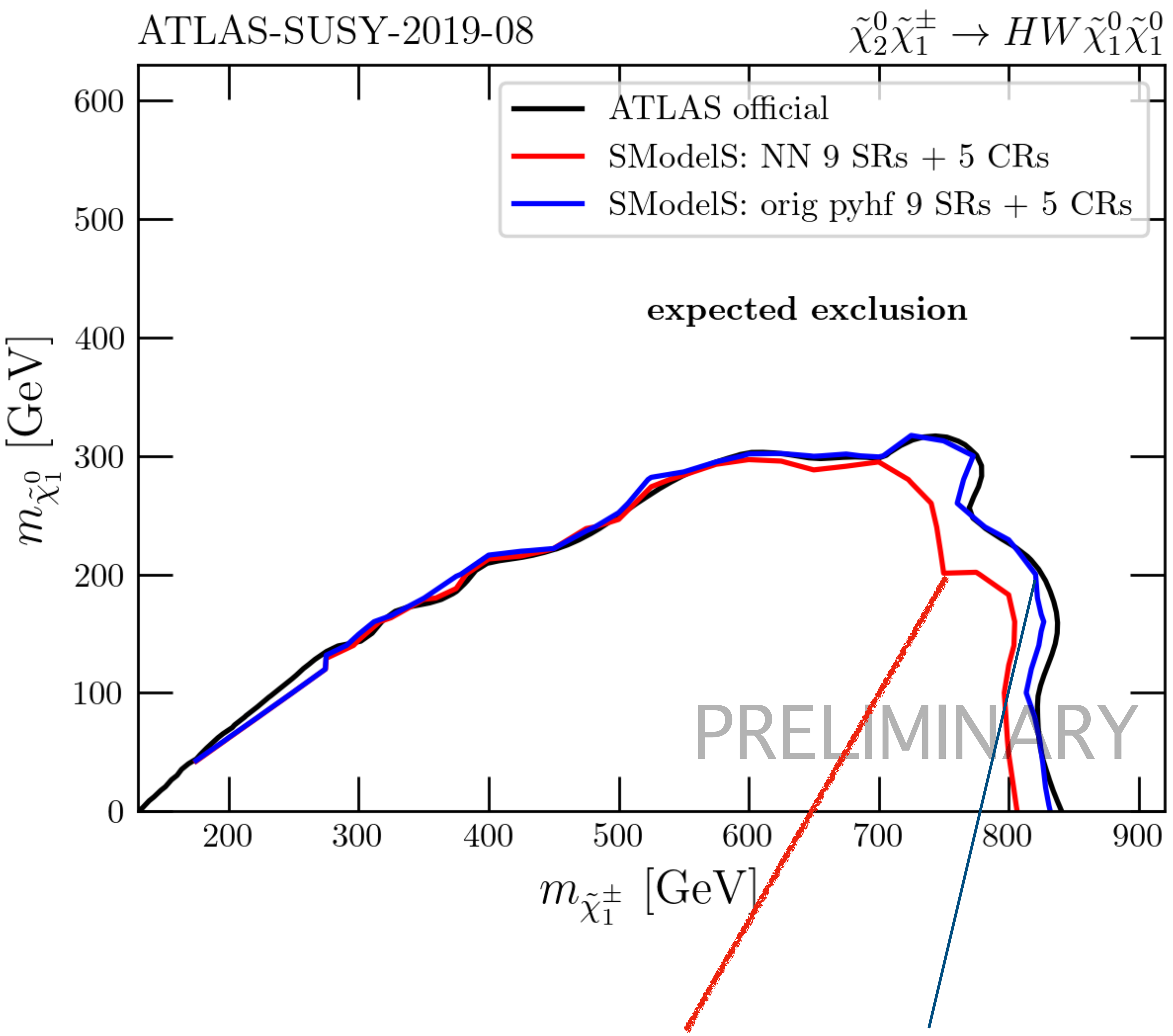
Asimov prescription

# Results ATLAS-SUSY-2019-08



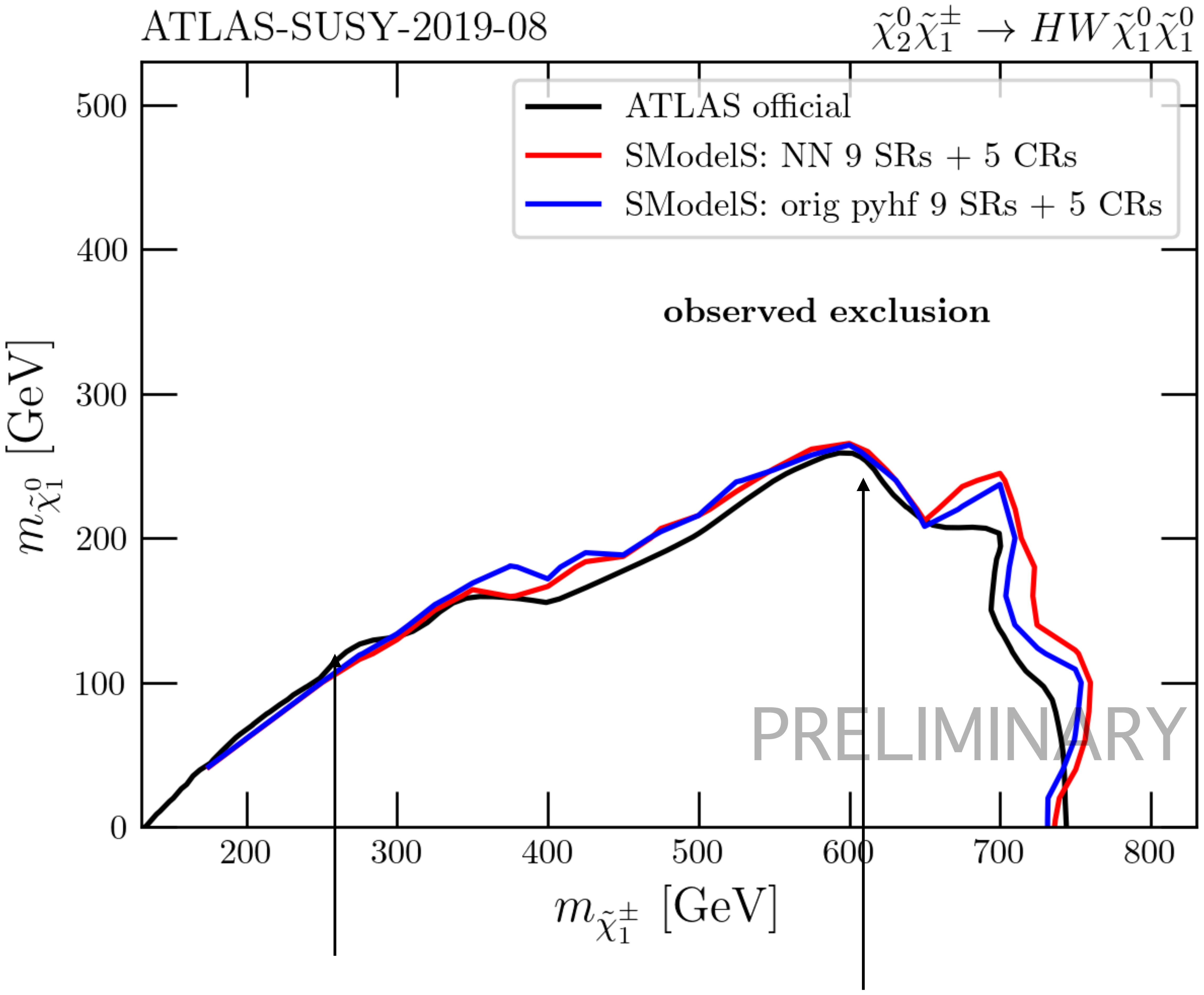

$$\Delta^{\mathrm{exp/obs}} = \ln \frac{L^{\mathrm{exp/obs}}(\mu = 1)}{L^{\mathrm{exp/obs}}(\mu = 0)}$$

# Results ATLAS-SUSY-2019-08




A priori prescription

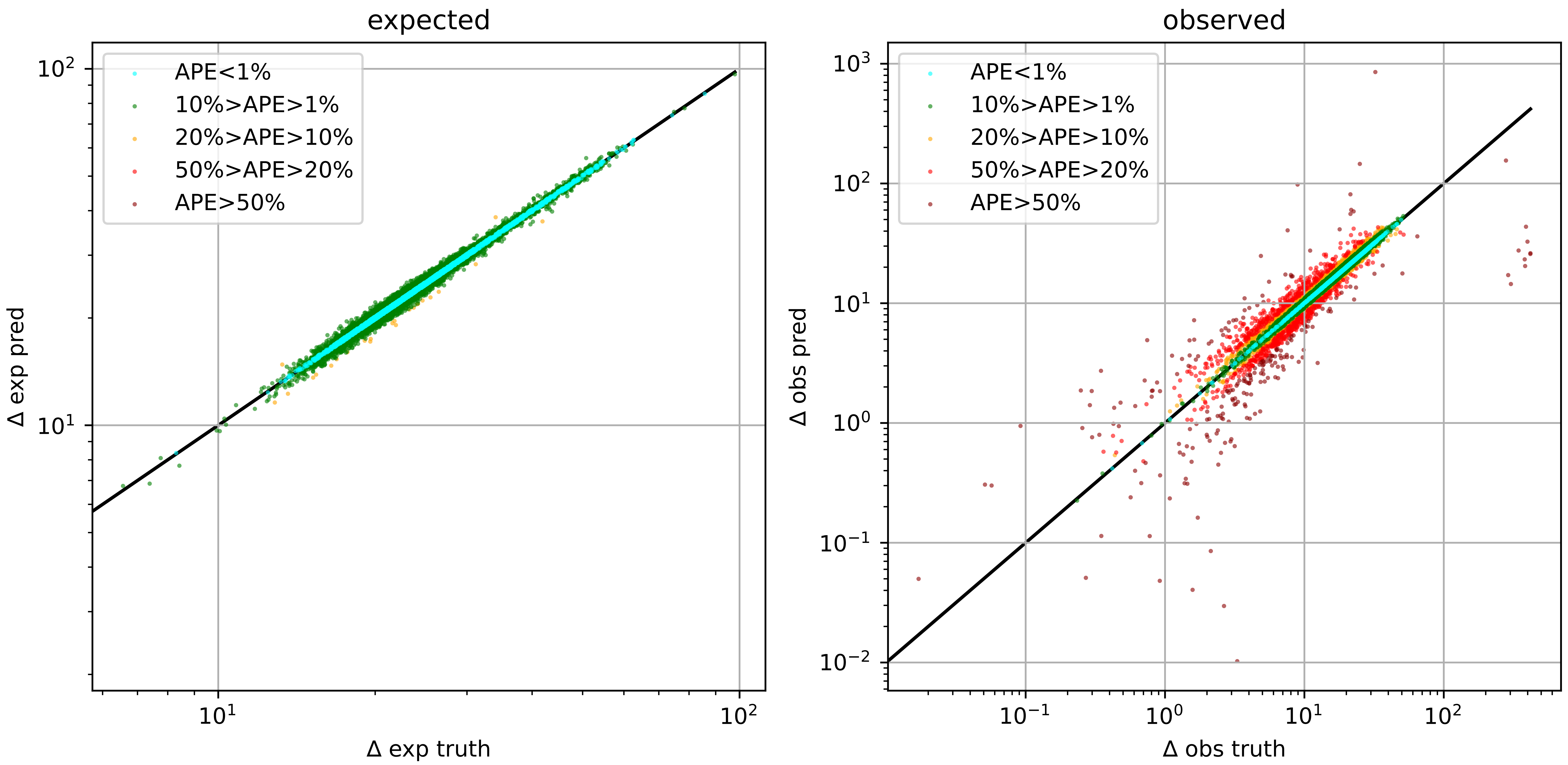
Asimov prescription




All include control regions

VERY GOOD MATCH

# Results 1911.12606-EWKino-leakage




$$\Delta^{\mathrm{exp/obs}} = \ln \frac{L^{\mathrm{exp/obs}}(\mu = 1)}{L^{\mathrm{exp/obs}}(\mu = 0)}$$

# NF-likelihoods

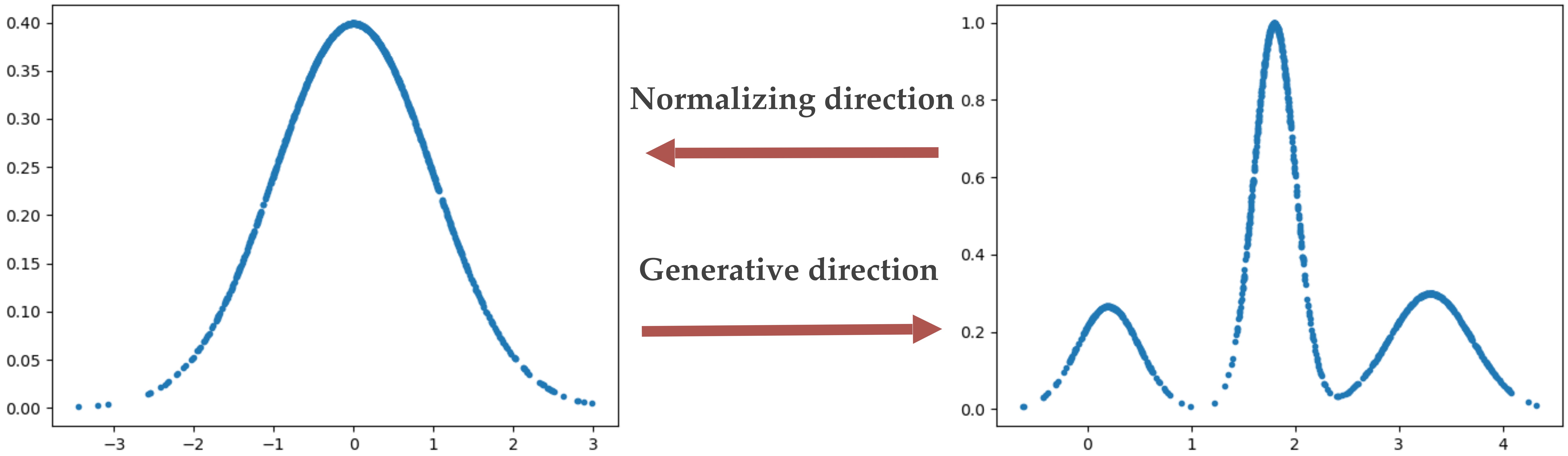
# Motivation

- Now we look at the other side of the statistical model.
- We focus on the likelihood function of the parameters, given an observation.
- This is relevant for measurements and global fits of parameters to data.
- The challenge: To parametrize all the parameters of interest (POIs) and nuisance parameters.
- Proposed solution: To model them as Normalizing Flows.

$$P_{\Theta}(\Theta \,|\, x = \mathrm{obs})$$

# Basic principle

Following the change of variables formula, perform a series of **bijective, continuous, invertible** transformations on a *simple* probability density function (pdf) to obtain a *complex* one.

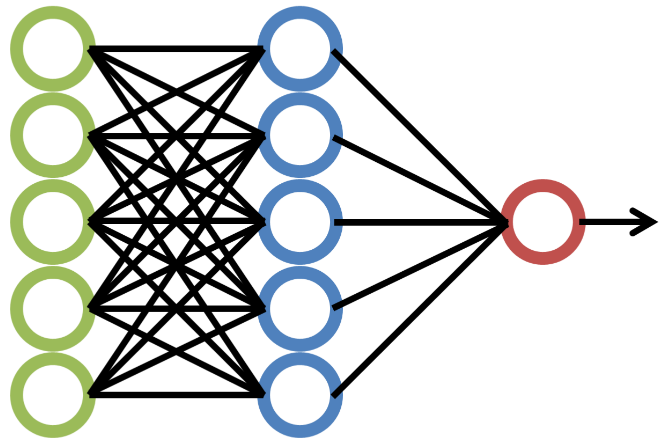



See review: Ivan K. et. al.
arXiv:1908.09257

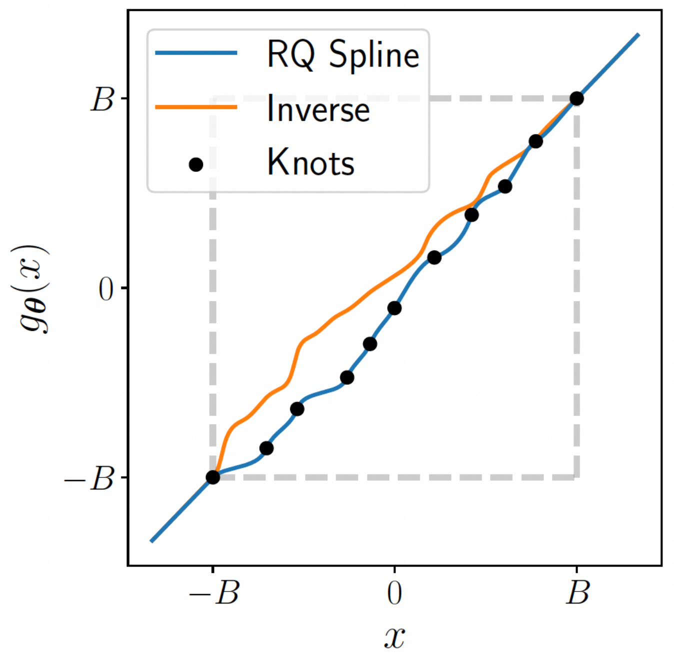
# Choosing the transformations

**THE OBJECTIVE:**
To perform the right transformations to accurately estimate the complex underlying distribution of some observed data.

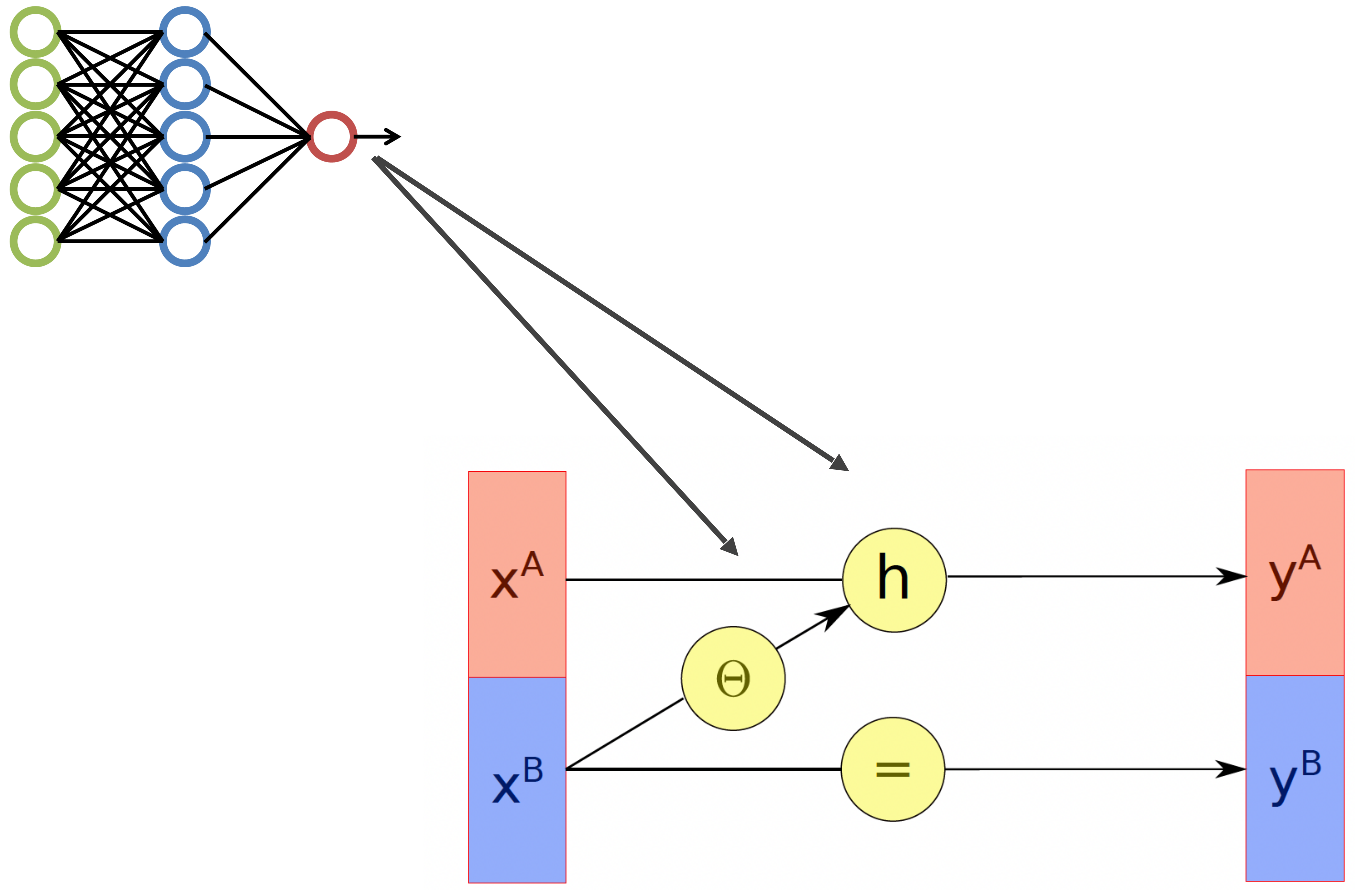

**THE RULES OF THE GAME:**
- The transformations (bijectors) must be invertible
- They should be sufficiently expressive
- And computationally efficient (including Jacobian)



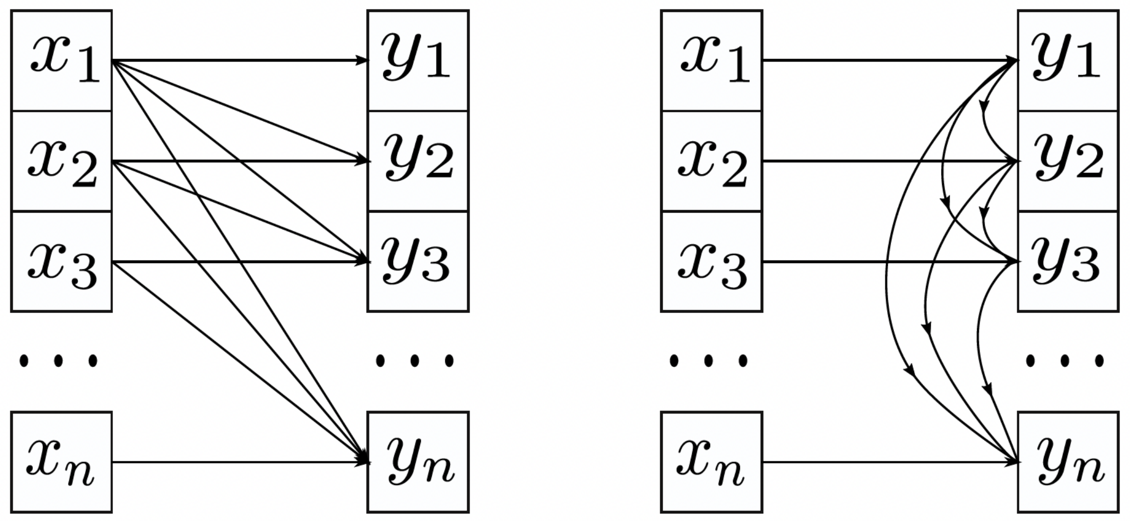
**THE STRATEGY:**
Let ***Neural Networks*** learn the parameters of ***Autoregressive* Normalizing Flows.***



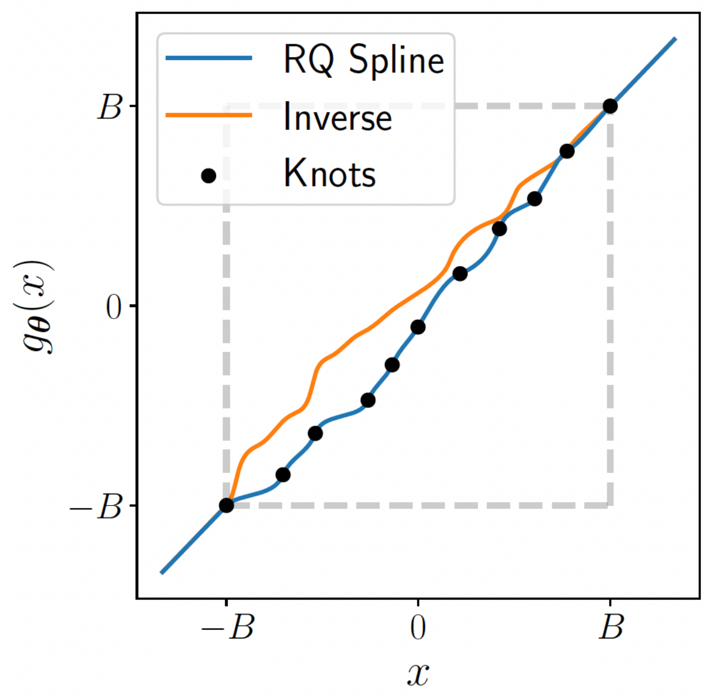
# Choosing the transformations




| Bijector / NF type | Affine | Rational Quadratic Spline |
|---|---|---|
| Coupling | **RealNVP** | **C-RQS** |
| Autoregressive | **MAF** | **A-RQS** |

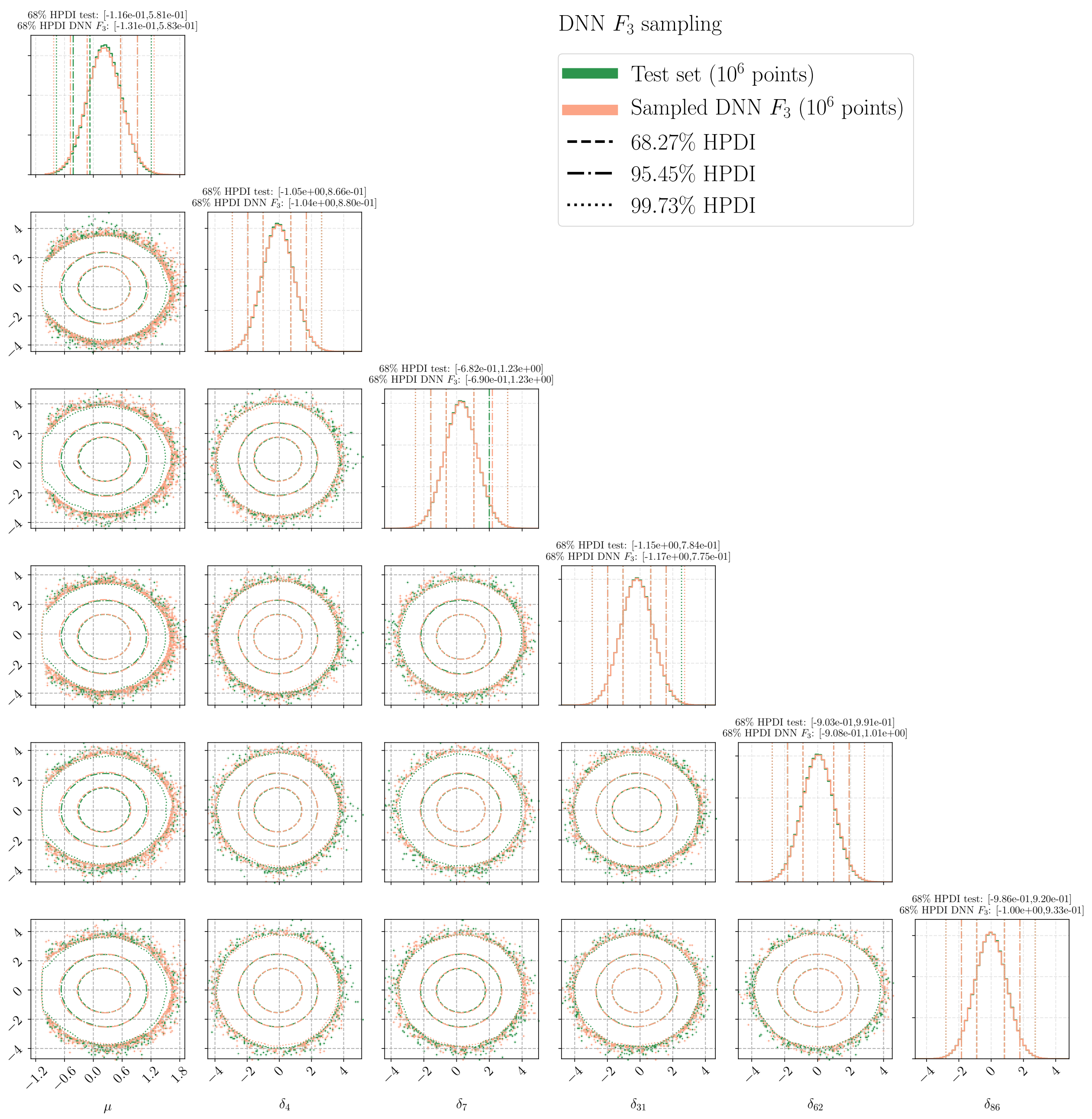



$$y = \Theta x + h$$



# Previous results.

- **DNNLikelihood**
- Supervised Learning with Deep Neural Networks.
- 95-dim LHC-like toy likelihood.
- arXiv:1911.03305 (A. Coccaro, M. Pierini, L. Silvestrini, R. Torre )



# Example Likelihoods

$$P_\Theta(\Theta \,|\, x = \mathrm{obs})$$

**LHC-like toy likelihood.**

- Simplified likelihood (Multivariate-Gaussian)
- 1 parameter of interest (signal strength)
- 89 nuisance parameters.
- Ref. arXiv:1809.05548

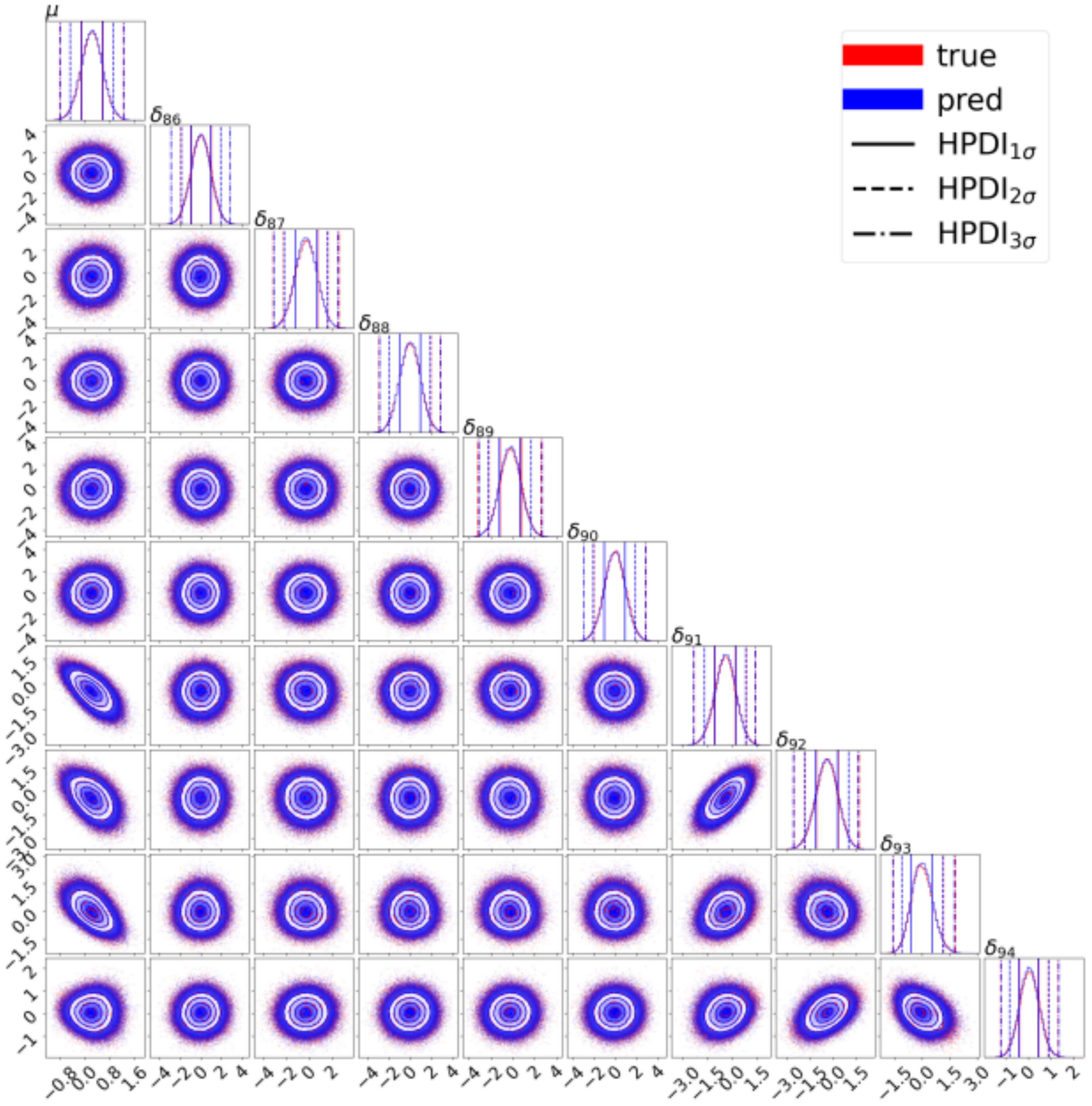
**Flavor fit likelihood**

- Flavor observables related to
- 12 parameters of interest (Wilson coefficients)
- 77 nuisance parameters.
- Ref. arXiv:1809.05548

**ElectroWeak fit Likelihood**

- EW observables.
- Including recent measurements of top mass (CMS) and W mass (CDF).
- 8 parameters of interest (Wilson coefficients of SMEFT operators)
- 32 nuisance parameters.
- Ref. arXiv:2204.04204

# LHC-like toy likelihood.




**Hyperparameters for Toy Likelihood**

| # of samples | hidden layers | algorithm | # of bijec. | spline knots | range | L1 factor | patience | max # of epochs |
|---|---|---|---|---|---|---|---|---|
| $\mathbf{2 \cdot 10^5}$ | $3 \times 64$ | MAF | 2 | - | - | 0 | 20 | 200 |

Table 1: Hyperparameters leading to the best determination of the Toy Likelihood.

**Results for Toy Likelihood**

| # of samples | Mean KS-test | Mean SWD | $\mathrm{HPDIe}_{1\sigma}$ | $\mathrm{HPDIe}_{2\sigma}$ | $\mathrm{HPDIe}_{3\sigma}$ | time (s) |
|---|---|---|---|---|---|---|
| $\mathbf{2 \cdot 10^5}$ | $0.4893 \pm .0292$ | $0.03947 \pm .0019$ | 0.02073 | 0.01207 | 0.01623 | 133 |

Table 2: Best results obtained for the Toy Likelihood.

**Results for Toy Likelihood POI**

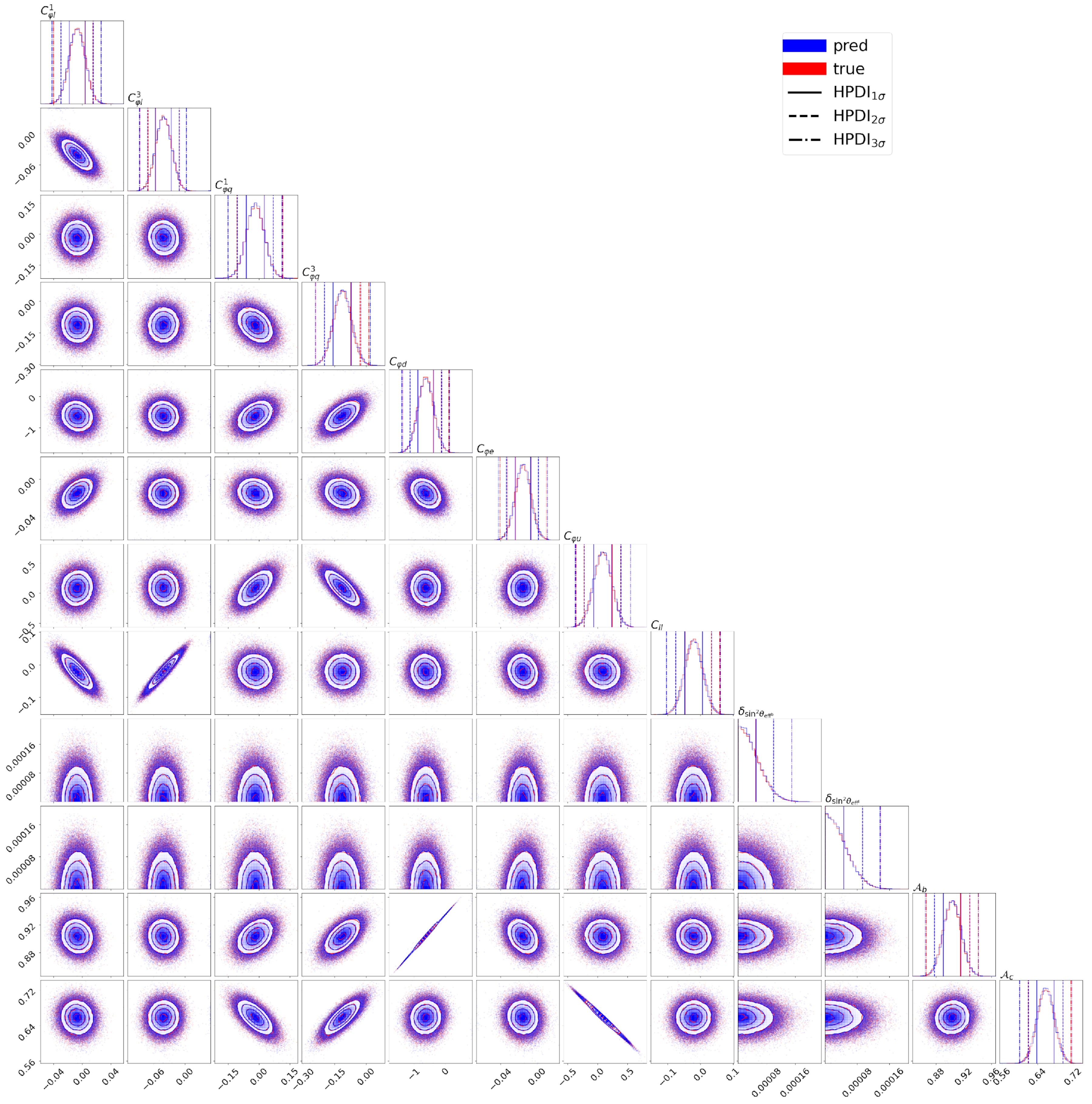
| POI | KS-test | $\mathrm{HPDIe}_{1\sigma}$ | $\mathrm{HPDIe}_{2\sigma}$ | $\mathrm{HPDIe}_{3\sigma}$ |
|---|---|---|---|---|
| $\mu$ | 0.54 | 0.02742 | 0.01359 | 0.01786 |

Table 3: Results for the POI in the Toy Likelihood.



R. Torre, HRG. arxiv:2309.09743

# ElectroWeak fit Likelihood




| Hyperparameters for the EW Likelihood | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| # of samples | hidden layers | # of bijec. | algorithm | spline knots | range | L1 factor | patience | # of epochs |
| $\mathbf{2 \cdot 10^5}$ | 2 | $3 \times 128$ | A-RQS | 4 | -6 | 0 | 20 | 800 |

Table 4: Hyperparameters leading to the best determination of the EW Likelihood.

| Results for the EW Likelihood | | | | | | |
|---|---|---|---|---|---|---|
| # of samples | Mean KS-test | Mean SWD | HPDIe$_{1\sigma}$ | HPDIe$_{2\sigma}$ | HPDIe$_{3\sigma}$ | time (s) |
| $\mathbf{2 \cdot 10^5}$ | $0.4307 \pm 0.06848$ | $0.003131 \pm 0.00053$ | 0.000339 | 0.0008664 | 0.006973 | 7255 |

Table 5: Best results obtained on the EW Likelihood.

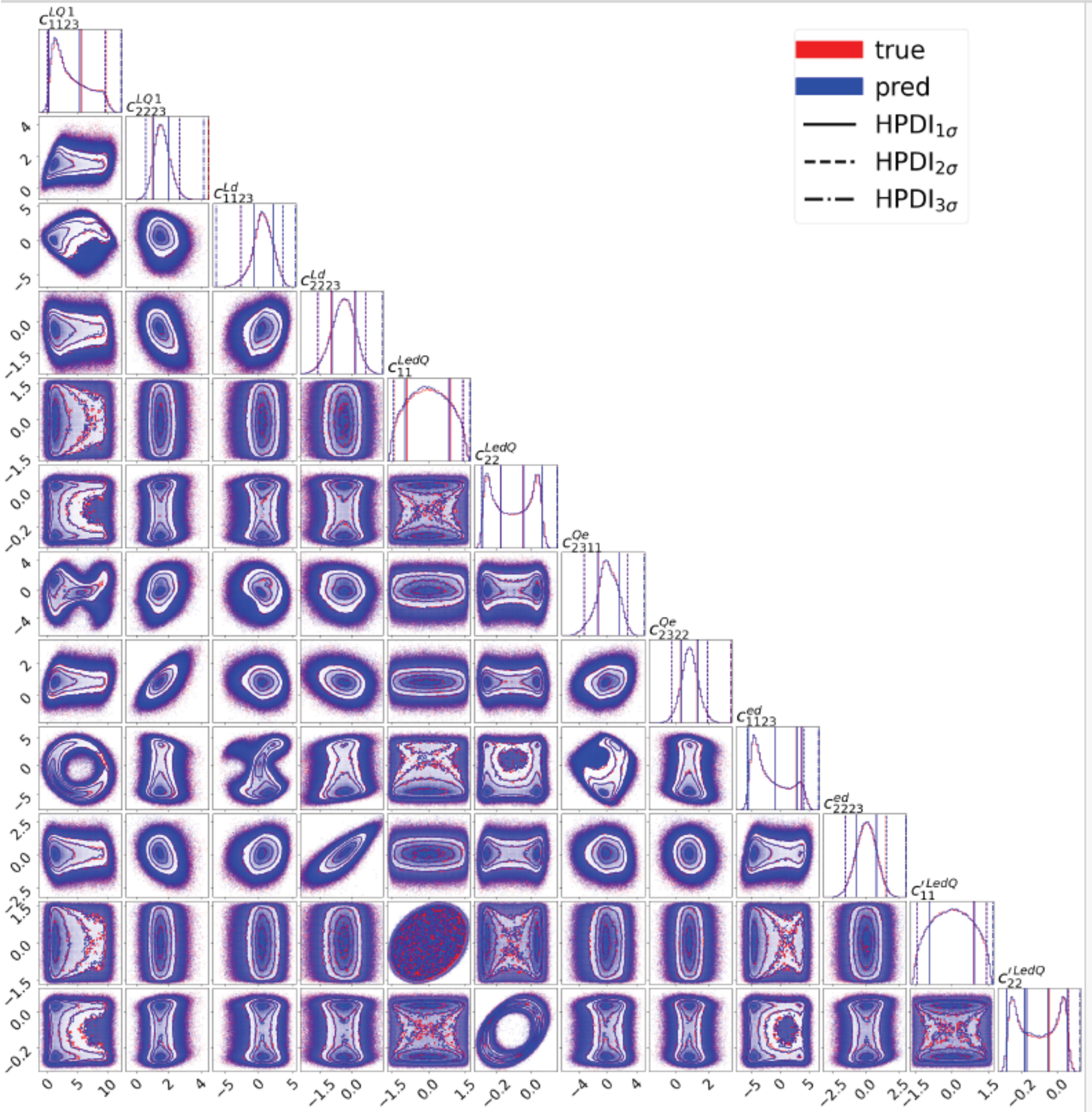
| Results for EW Likelihood | | | | |
|---|---|---|---|---|
| POI | KS-test | HPDIe$_{1\sigma}$ | HPDIe$_{2\sigma}$ | HPDIe$_{3\sigma}$ |
| $c^1_{\varphi l}$ | 0.1901 | 0.08384 | 0.09787 | 0.437 |
| $c^3_{\varphi l}$ | 0.2078 | 0.0346 | 0.1039 | 0.4967 |
| $c^1_{\varphi q}$ | 0.4581 | 0.02279 | 0.01131 | 0.04866 |
| $c^3_{\varphi q}$ | 0.4989 | 0.01219 | 0.01439 | 4.1017 |
| $c_{\varphi d}$ | 0.5221 | 0.01713 | 0.03808 | 0.09952 |
| $c_{\varphi e}$ | 0.4885 | 0.01453 | 0.2146 | 0.1401 |
| $c_{\varphi u}$ | 0.5259 | 0.005409 | 0.005082 | 0.341 |
| $c_{ll}$ | 0.2193 | 0.1667 | 0.08047 | 0.0713 |

Table 6: Results for the Wilson coefficients in the EW Likelihood.



R. Torre, HRG. arxiv:2309.09743

# Flavor fit likelihood




| Hyperparameters for the Flavor Likelihood | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| # of samples | hidden layers | # of bijec. | algorithm | spline knots | range | L1 factor | patience | max # of epochs |
| **$10^6$** | $3 \times 1024$ | 2 | A-RQS | 8 | -5 | 1e-4 | 50 | 12000 |

Table 7: Hyperparameters leading to the best determination of the Flavor Likelihood.

| Results for the Flavor Likelihood | | | | | | |
|---|---|---|---|---|---|---|
| # of samples | Mean KS-test | Mean SWD | HPDIe$_{1\sigma}$ | HPDIe$_{2\sigma}$ | HPDIe$_{3\sigma}$ | time (s) |
| **$5 \cdot 10^5$** | $0.4237 \pm 0.03405$ | $0.02717 \pm 0.002374$ | 0.00867 | 0.007346 | 1.419e-07 | 9550 |

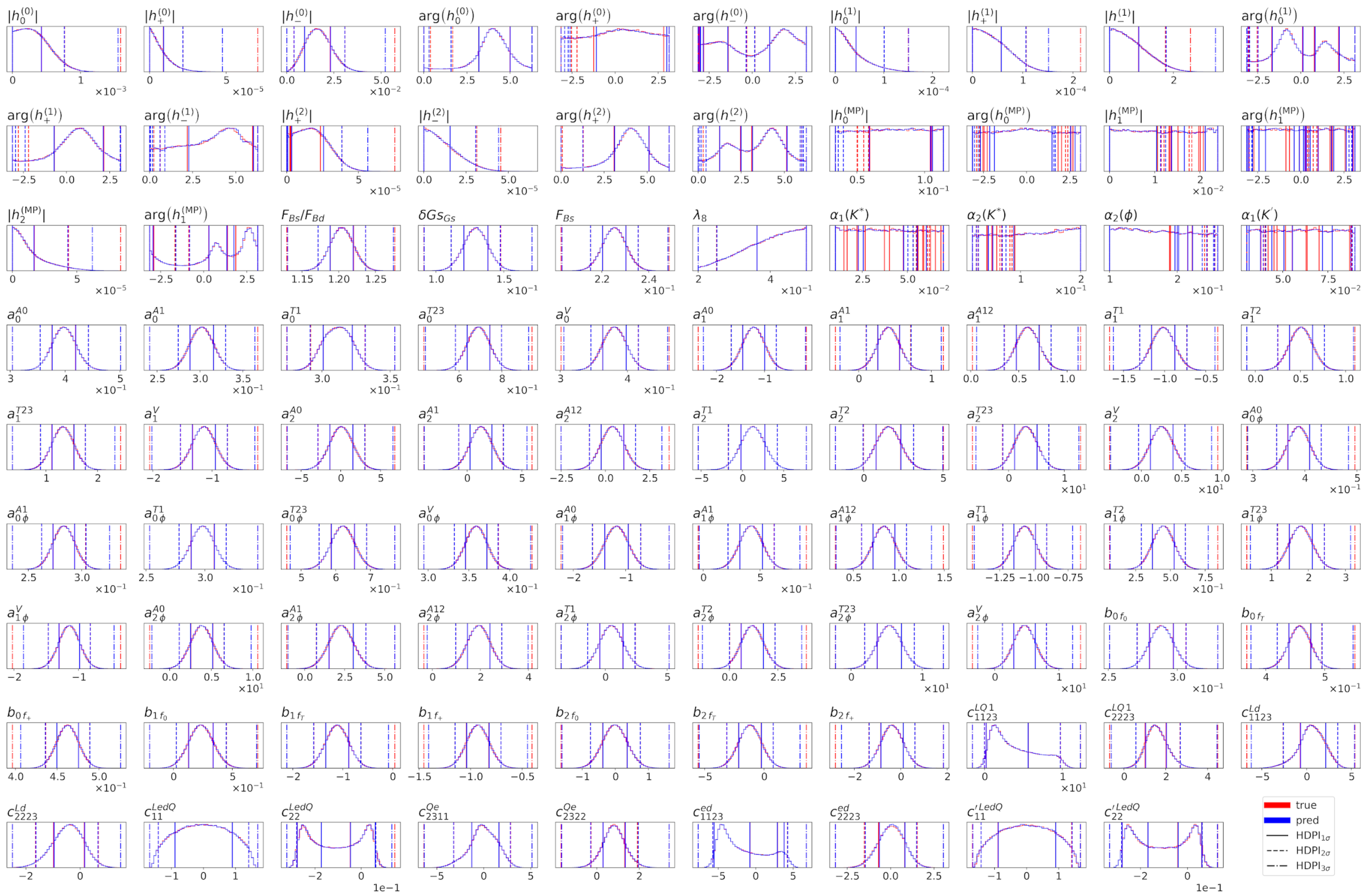
Table 8: Best results obtained for the Flavor Likelihood.

| Results for Flavor Likelihood POIs | | | | |
|---|---|---|---|---|
| POI | KS-test | HPDIe$_{1\sigma}$ | HPDIe$_{2\sigma}$ | HPDIe$_{3\sigma}$ |
| $c_{1123}^{LQ\,1}$ | 0.4346 | 0.007251 | 1.83e-05 | 4.731e-08 |
| $c_{2223}^{LQ\,1}$ | 0.4736 | 0.01249 | 0.00162 | 0.03575 |
| $c_{1123}^{Ld}$ | 0.486 | 0.01466 | 0.006628 | 0.002338 |
| $c_{2223}^{Ld}$ | 0.4138 | 0.0513 | 0.02446 | 2.398e-08 |
| $c_{11}^{LedQ}$ | 0.5362 | 0.00738 | 0.004683 | 5.387e-08 |
| $c_{22}^{LedQ}$ | 0.5161 | 0.02799 | 0.001639 | 2.155e-09 |
| $c_{2311}^{Qe}$ | 0.4476 | 0.01389 | 0.007458 | 1.419e-07 |
| $c_{2322}^{Qe}$ | 0.382 | 0.02132 | 0.02496 | 0.0004609 |
| $c_{1123}^{ed}$ | 0.4789 | 0.04076 | 0.00333 | 5.602e-08 |
| $c_{2223}^{ed}$ | 0.4436 | 0.008685 | 0.016 | 1.502e-08 |
| $c_{11}^{\prime\,LedQ}$ | 0.3203 | 0.09194 | 0.007041 | 8.011e-08 |
| $c_{22}^{\prime\,LedQ}$ | 0.4157 | 0.03001 | 0.008749 | 4.374e-08 |

Table 9: Results for the Wilson coefficients in the Flavor Likelihood.

R. Torre, HRG. arxiv:2309.09743

# Flavor fit likelihood



# Outlook

- Learn more profile likelihoods from ATLAS statistical models. Possibly including a posteriori and Asimov.
- Investigate how to learn profile likelihoods from CMS statistical models. Other technical developments required.
- Saving NFlikelihoods in ONNX format.
- Aim for learning full statistical models with (conditional) normalizing flows.

# Conclusions

- Multivariate phenomenological studies require efficient handling of likelihoods.
- NNs provide an orders of magnitude faster alternative for LHC likelihood publication. From dozens of minutes to less than a second per point!
- Profile likelihoods are easily learnable by NNs.
- They can easily be integrated into modern reinterpretation frameworks.
- Complete infrastructure for ML Likelihoods under construction.
- Normalizing Flows show great capacity of learning complex high dimensional functions. Specially, the A-RQS.
- NFs can accurately and efficiently  model LHC and EFT-fit likelihoods.
- Further developments under way. Stay tuned!

# Thank you!

# ATLAS full statistical models

- Full statistical models by ATLAS are available on HEPData
- They are provided as JSON files
- There are background files and signal patches
- Each patch corresponds to some signal point and contains modifiers to the background files
- There can be hundreds of modifiers
- Spey/PyHF can load and process these files


```
"patch": [
    {
        "op": "add",
        "path": "/channels/0/samples/0",
        "value": {
            "data": [
                2.3051342964172363
            ],
            "modifiers": [
                {
                    "data": null,
                    "name": "lumi",
                    "type": "lumi"
                },
                {
                    "data": [
                        0.6571804118166927
                    ],
                    "name": "staterror_QCR1cut_cuts",
                    "type": "staterror"
                },
                {
                    "data": {
                        "hi": 1.06675,
                        "lo": 0.911403
                    },
                    "name": "PRW_DATASF",
                    "type": "normsys"
                },
```




Courtesy of R. Maselek