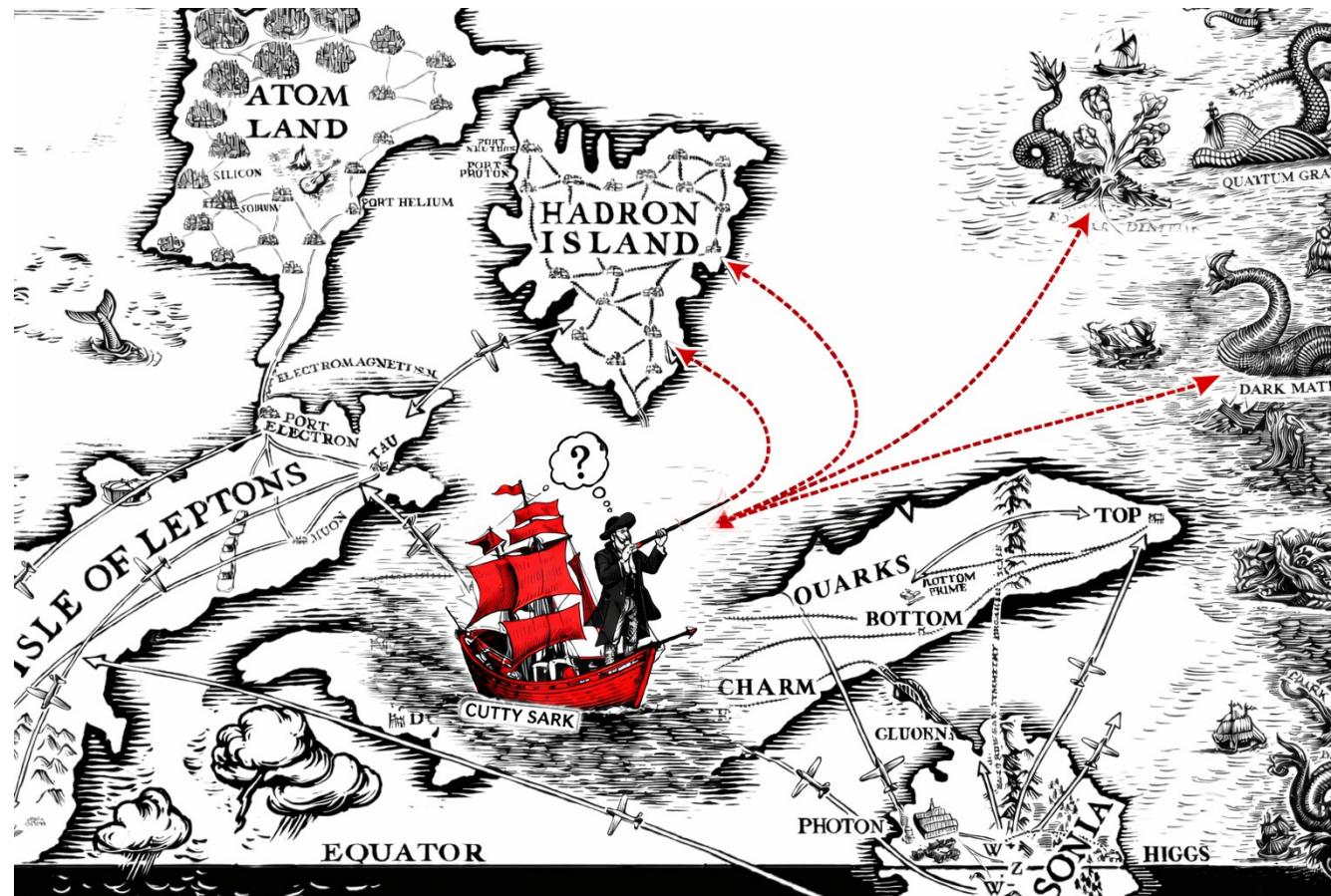


# LHC Searches and the Next Standard Model: A Meta-Statistical View

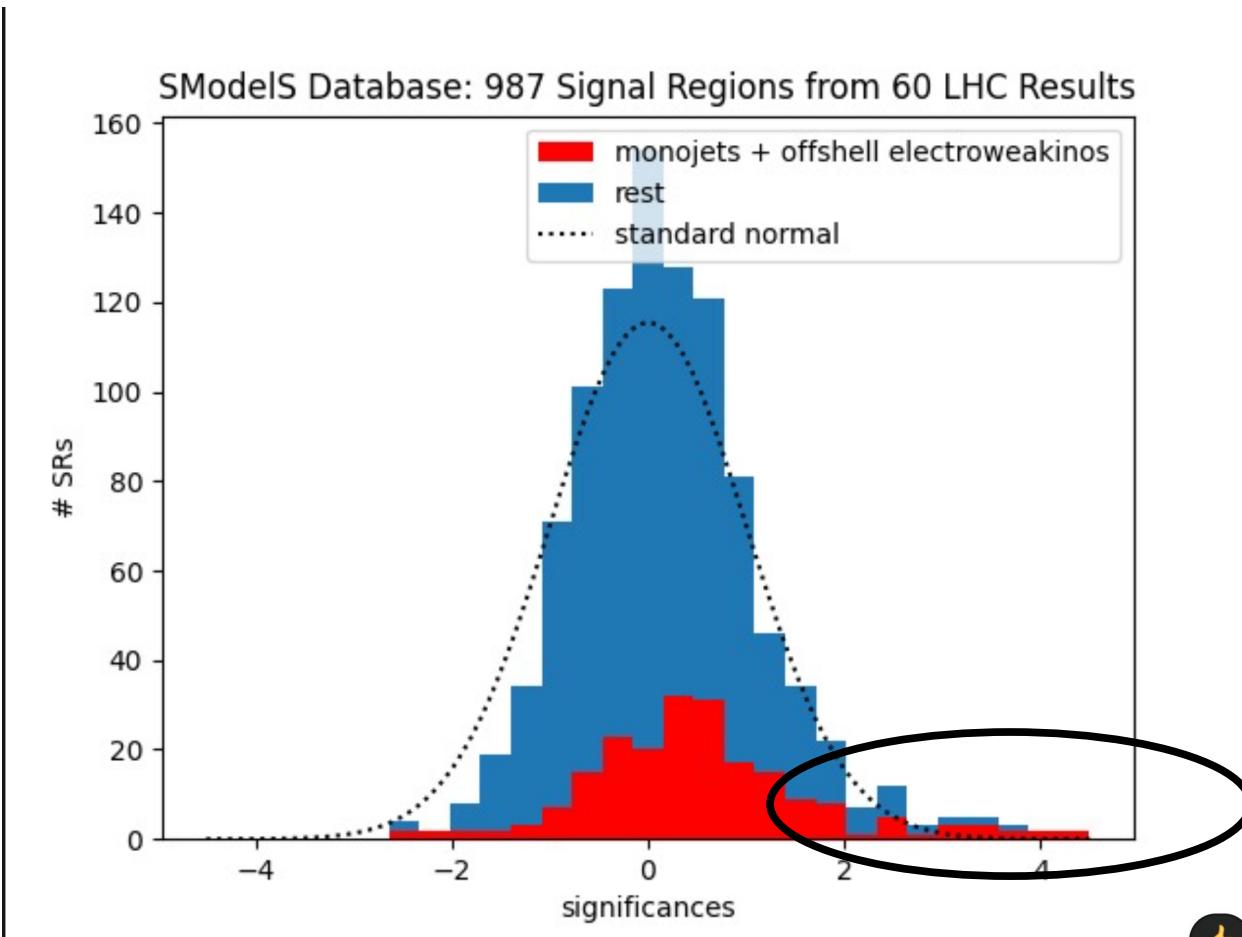


Wolfgang Waltenberger (MBI and Uni Wien)

Imperial College London, Fundamental Physics  
Seminar, January 2026



We, the SModelS collaboration have been collecting the published search results of CMS and ATLAS since the dawn of the LHC. Here is v3.1.0 of our database:



Distribution of **p-values** of all the signal regions we have in our database, Standard Model hypothesis. If no new physics is in the data, it should follow a standard normal (dashed black). Excesses are at + infinity

We have **125 analyses** in our database, but mostly because of lacking reinterpretation material, **only 60 made it into the plot above**. Will get to this later.



(a similar plot posted  
For medical research.)

No offense meant,  
medical researchers, we  
are very happy you are  
doing all the hard  
research )

Associate Professor of Public Policy, Politics, Education @UVAbatten I share cool social science.

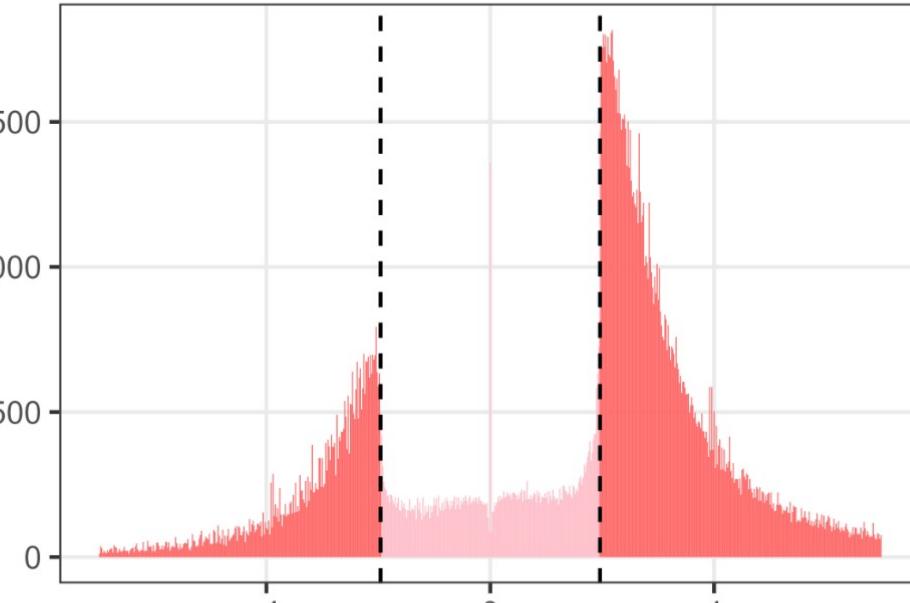
[View all](#)

 John B. Holbein ✅ @JohnHolbein1 · Nov 4

Look at the distribution of z-values from medical research!

∅ ...

### Distribution of z-values



A histogram showing the distribution of z-values. The x-axis is labeled "z-value" and ranges from -4 to 4 with major grid lines every 4 units. The y-axis is labeled "Count" and ranges from 0 to 7500 with major grid lines every 2500 units. The distribution is highly right-skewed, with the highest frequency occurring between 2 and 3 standard deviations above the mean (around 2.5 to 3). The distribution is centered near zero. Three vertical dashed lines are drawn at approximately z = -1.5, z = 0, and z = 1.5, highlighting the tails of the distribution.

109

377

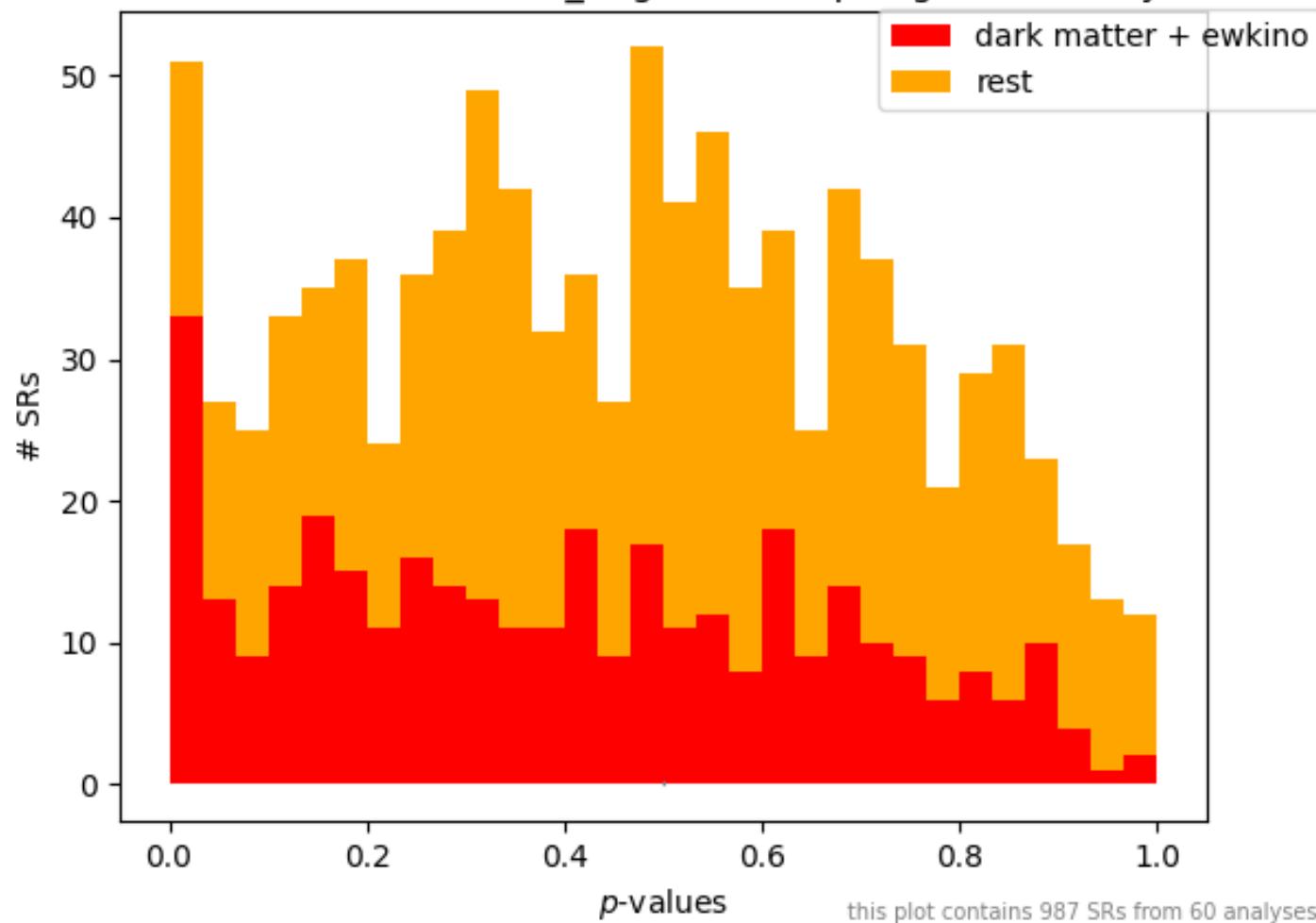
5.3K

662K

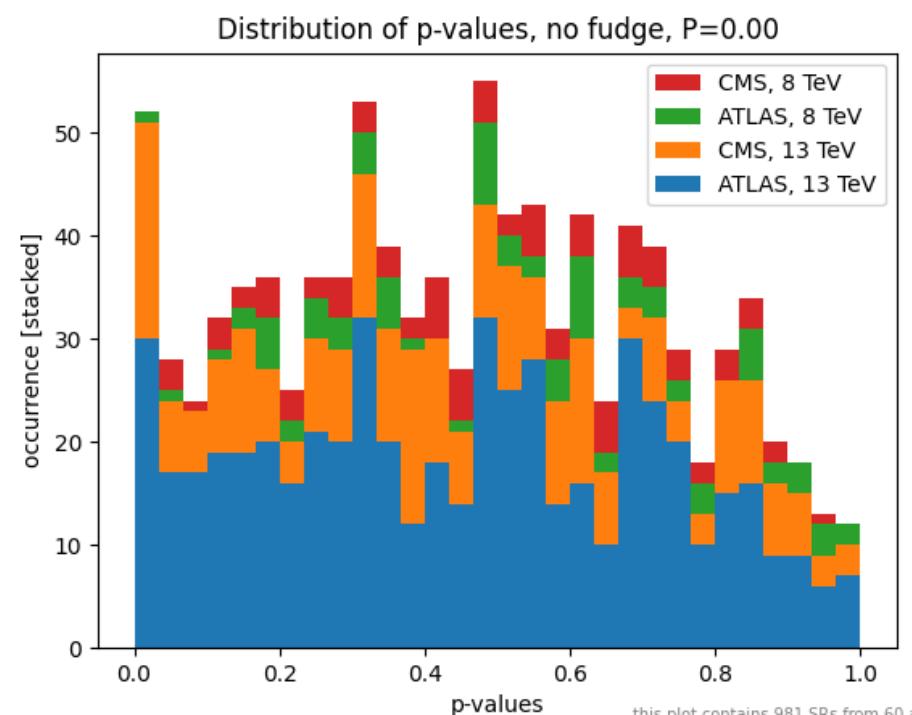
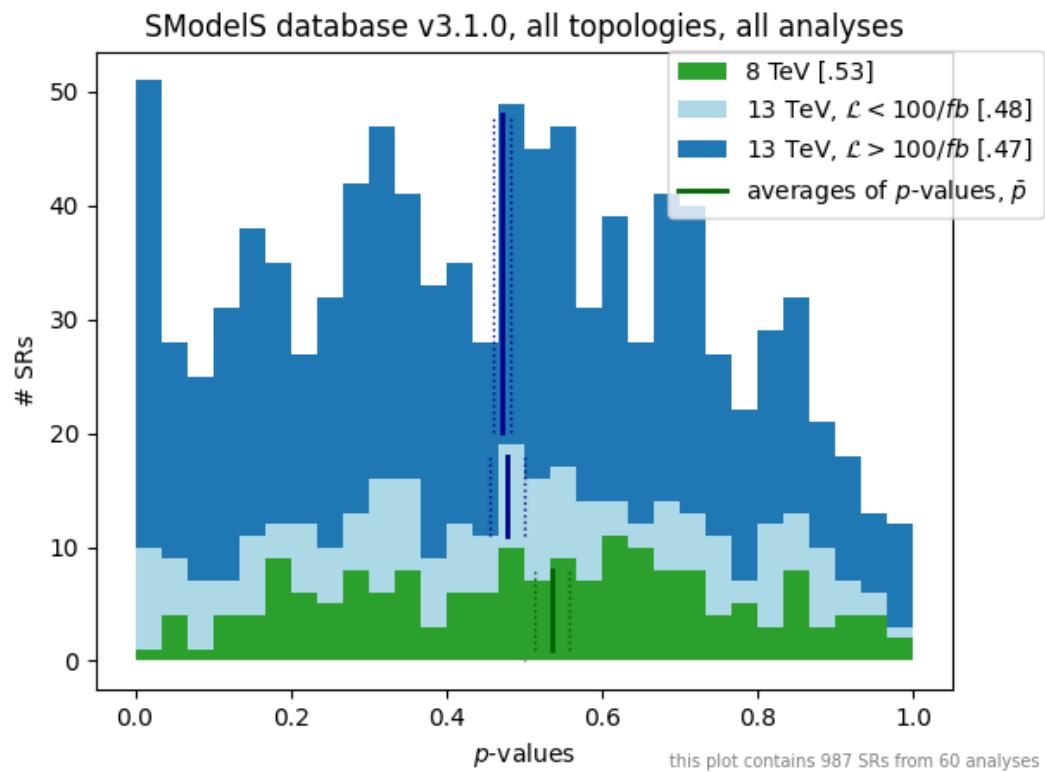
Bookmark Up



SModelS database v3.1.0\_original, all topologies, all analyses



Similar information but p-values instead of significances. New physics would manifest at  $p \sim 0$ , but only in some classes of results.  
Red: electroweakino and dark matter searches only.



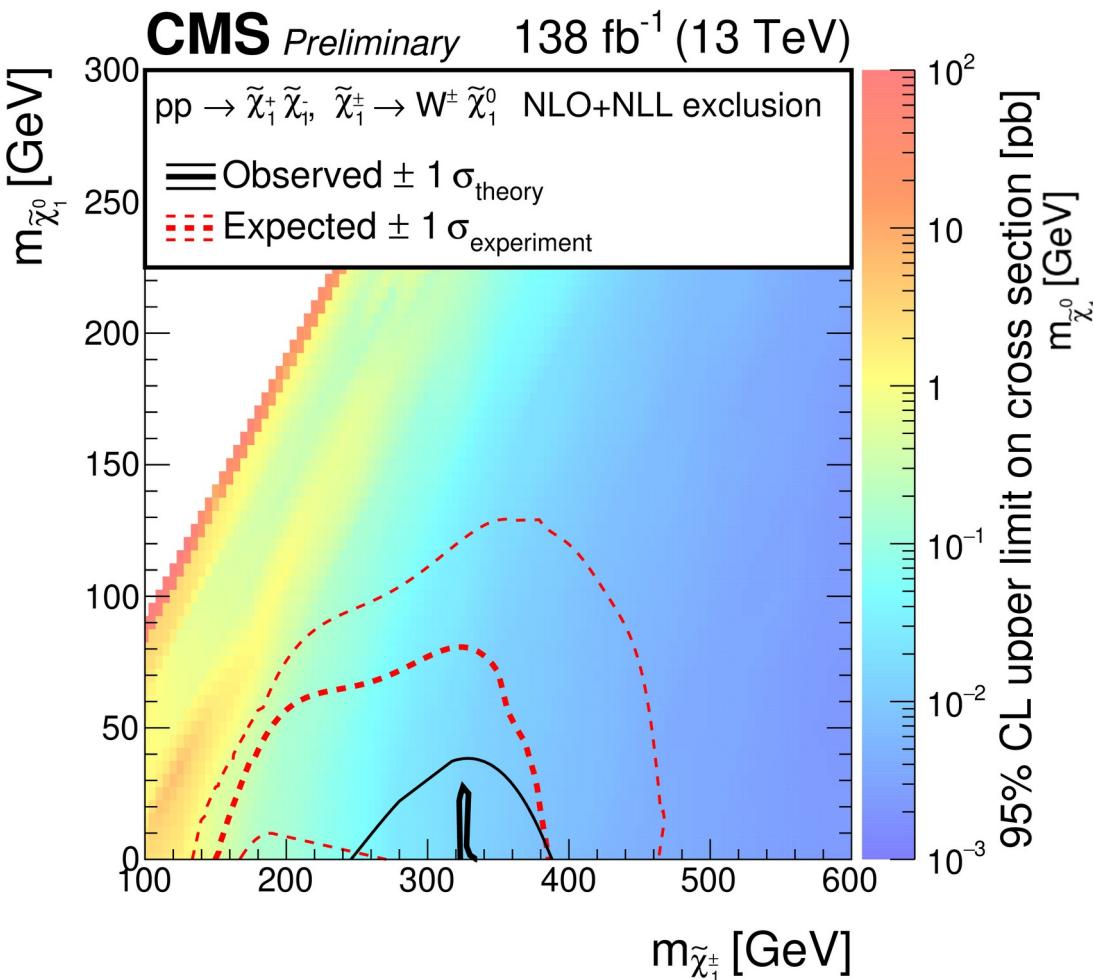


So the meta-statistics of almost all classes of results seem to look as expected (congratulations, LHC community!) except for electroweakino, dark matter searches.

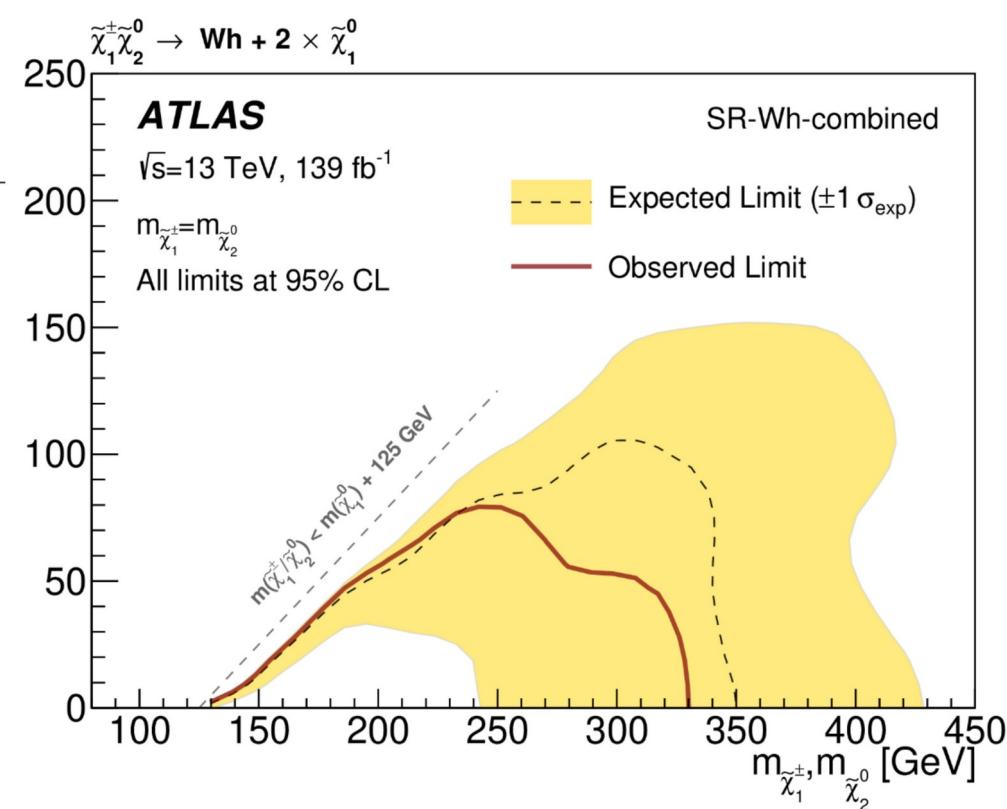
Results of about 50% of electroweakino/dark matter searches are not in the database (either not yet, or technical obstacles).

So what about missing, or more recent mono-X and electroweakino searches?

# RESULTS NOT (YET) IN SMODELS

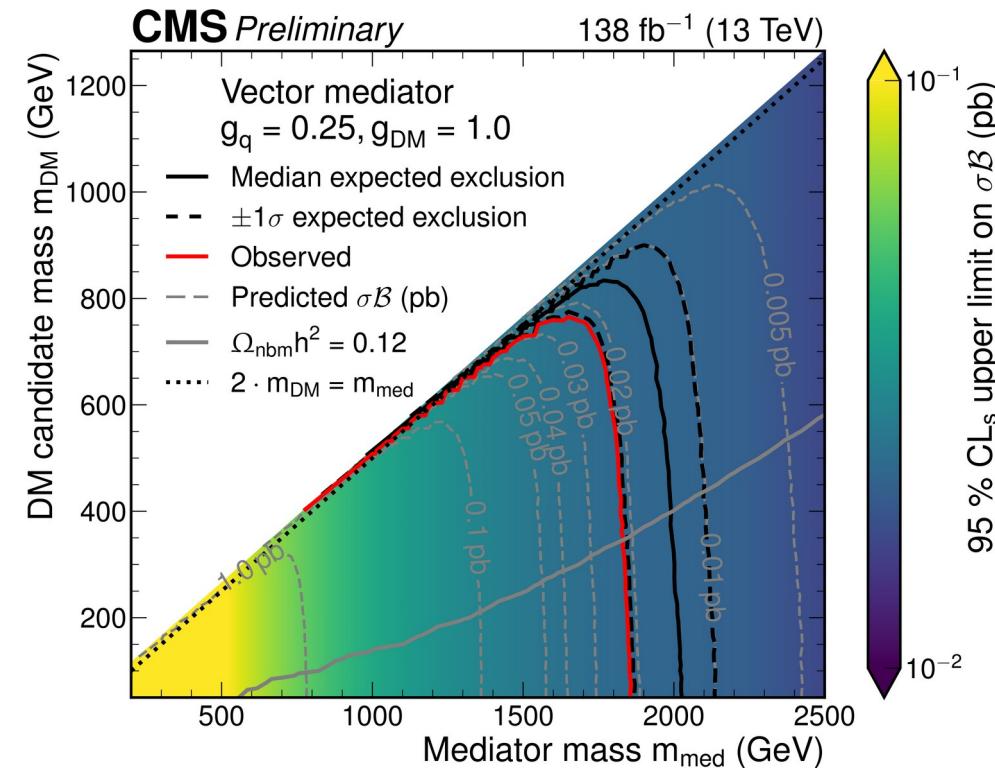
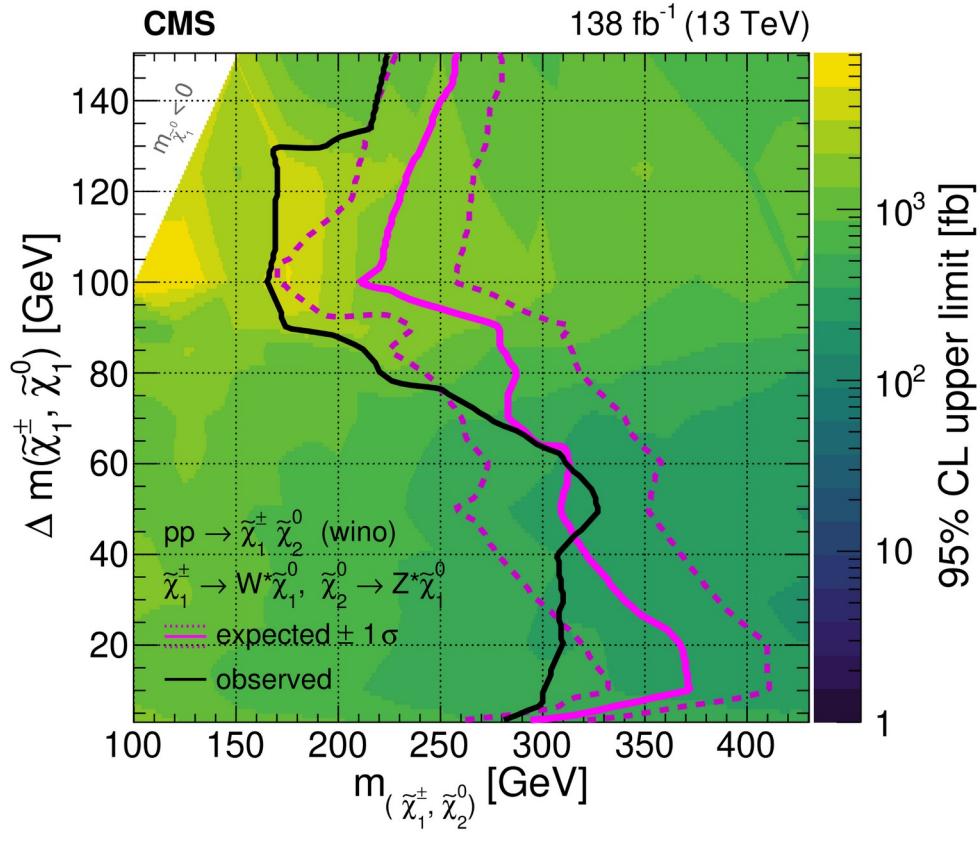


CMS-SUS-23-002, chargino pair production



ATLAS-SUSY-2023-03,  
 Charginos + neutralinos

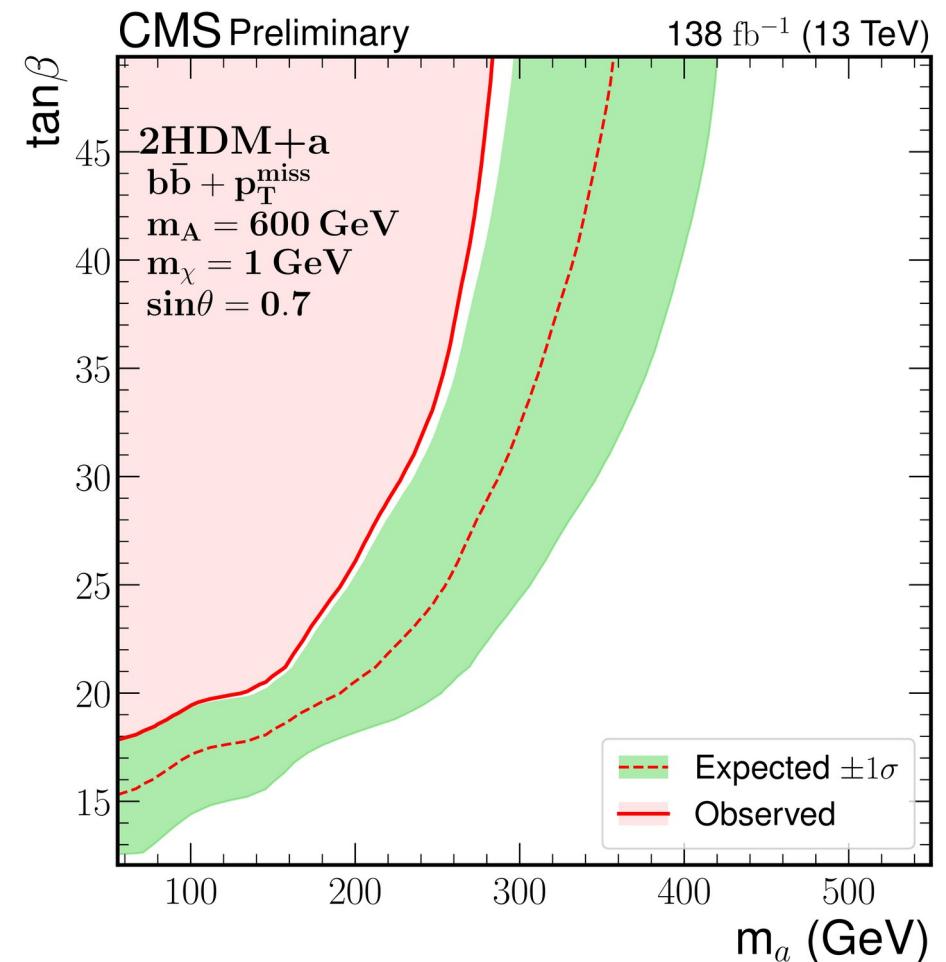
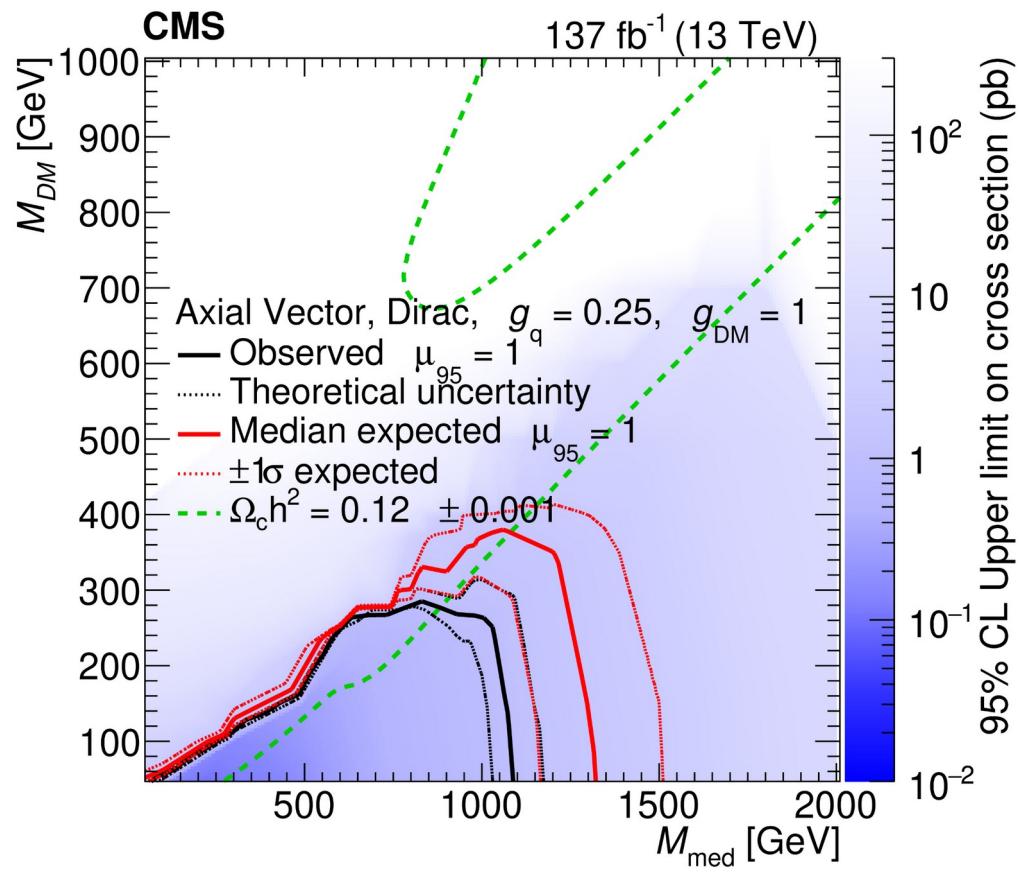
# RESULTS NOT (YET) IN SMODELS



CMS-SUS-23-003, chargino neutralino production

CMS-SUS-23-004, mono-top

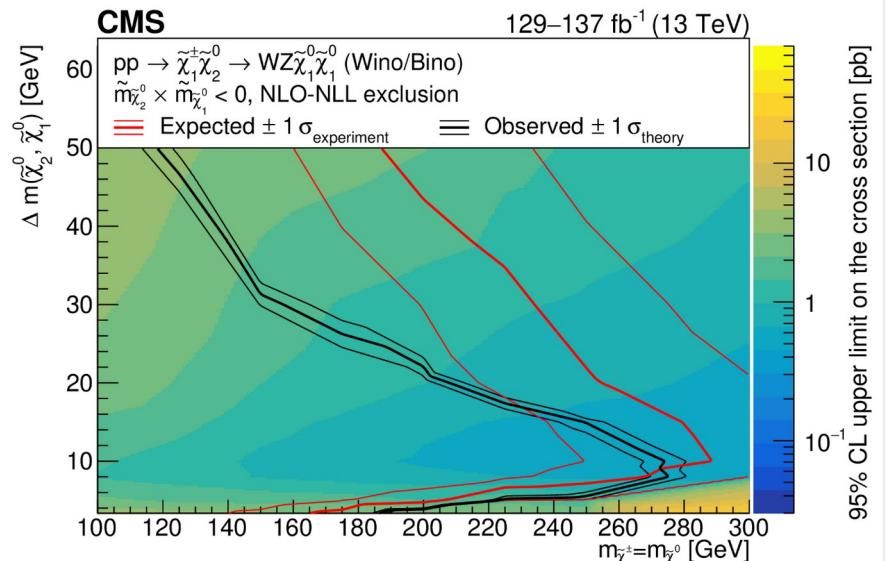
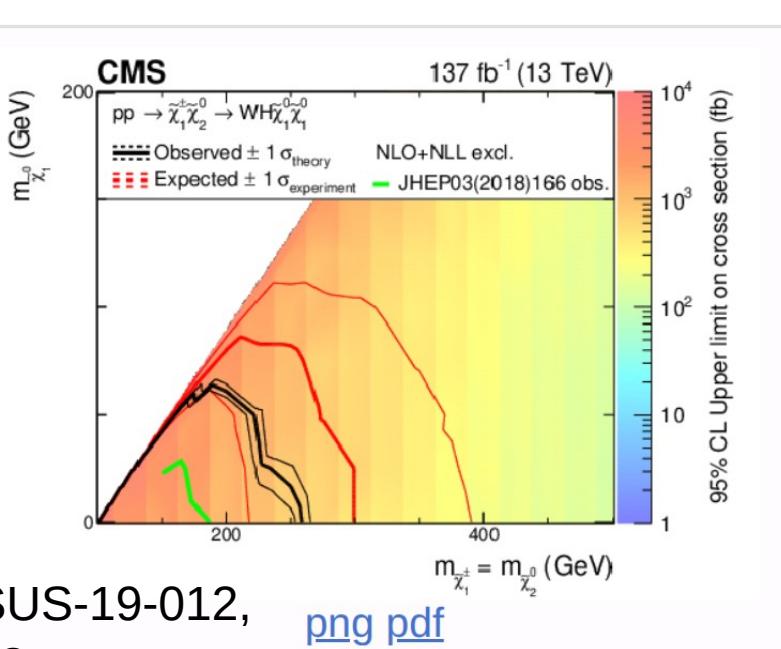
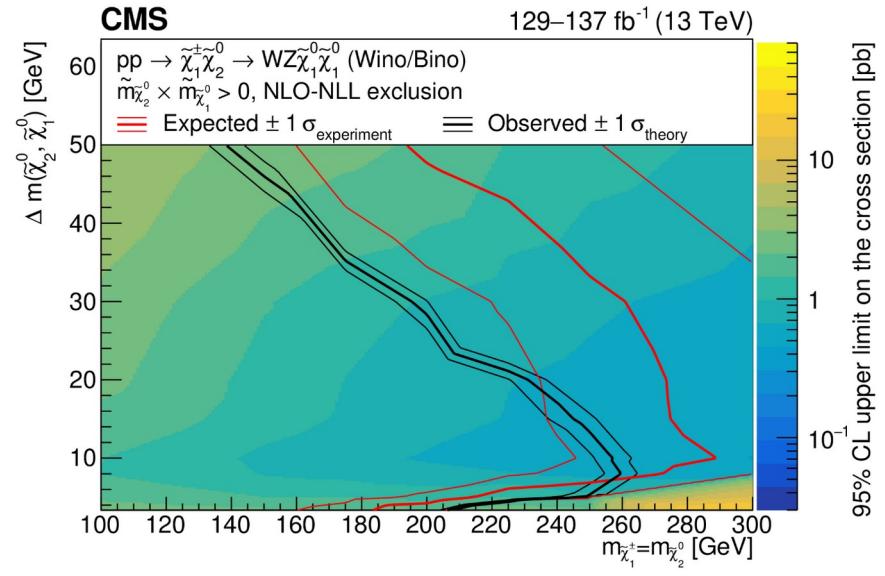
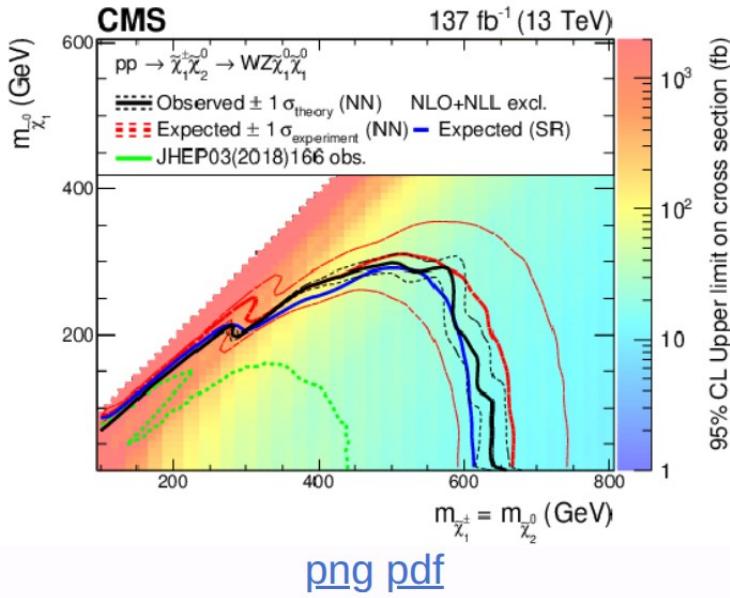
# RESULTS NOT (YET) IN SMODELS



CMS-SUS-23-016, MET + gamma

CMS-SUS-23-008, MET+bb

# RESULTS NOT (YET) IN SMODELS

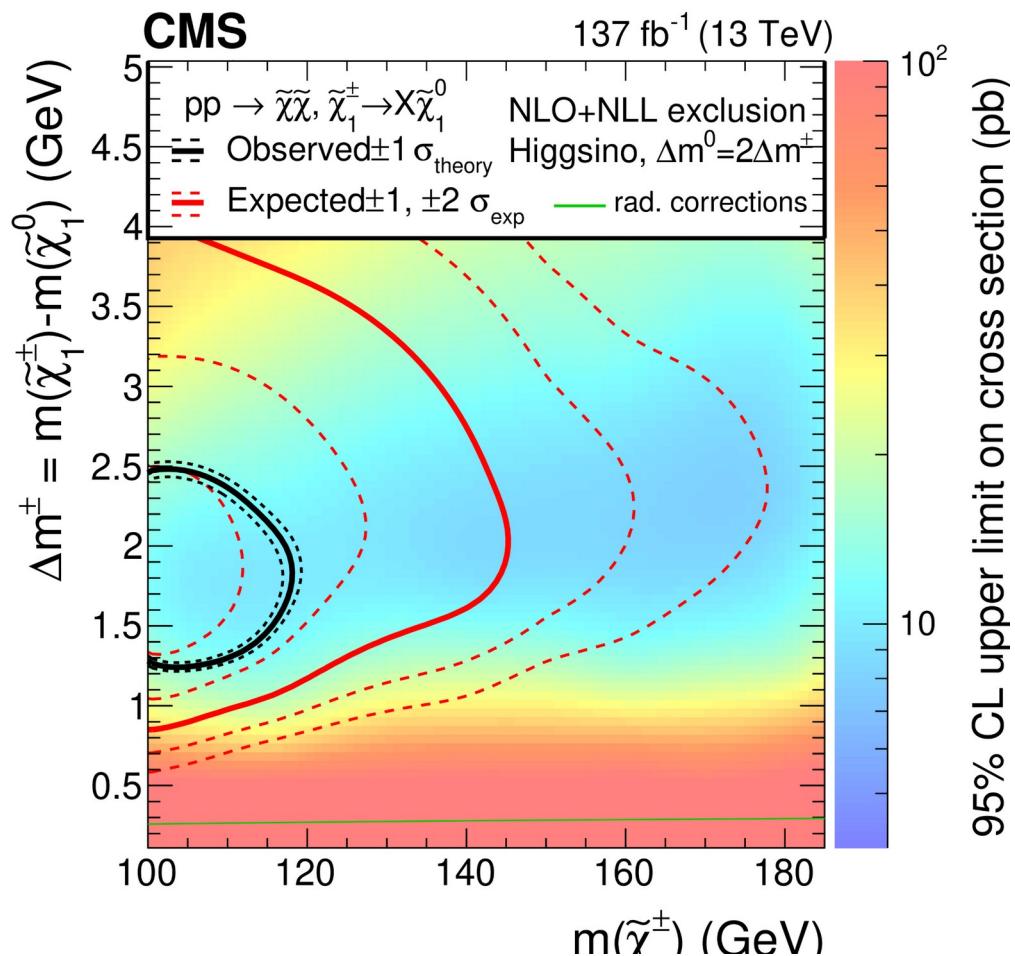


CMS-SUS-19-012,  
ewkinos

[png](#) [pdf](#)

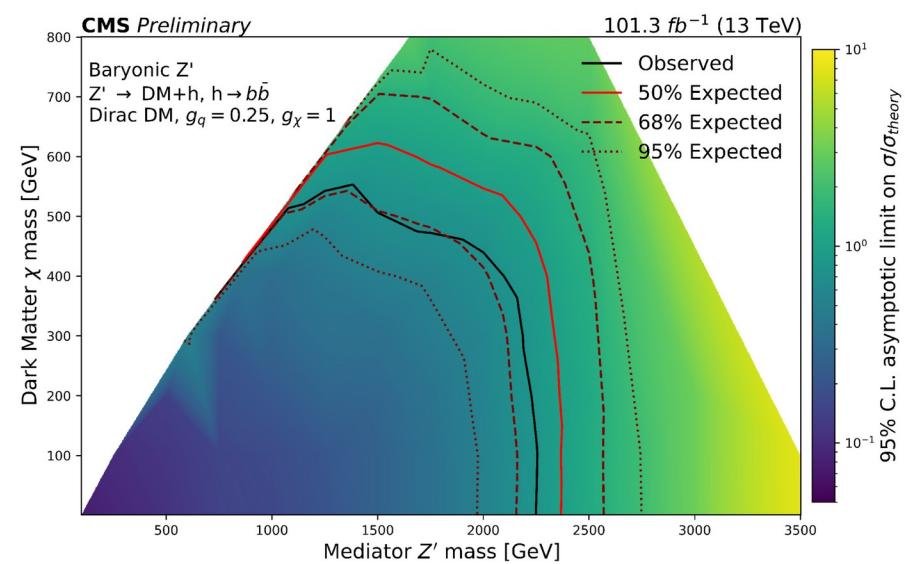
CMS-SUS-18-004, ewkinos

# RESULTS NOT (YET) IN SMODELS



CMS-SUS-24-003

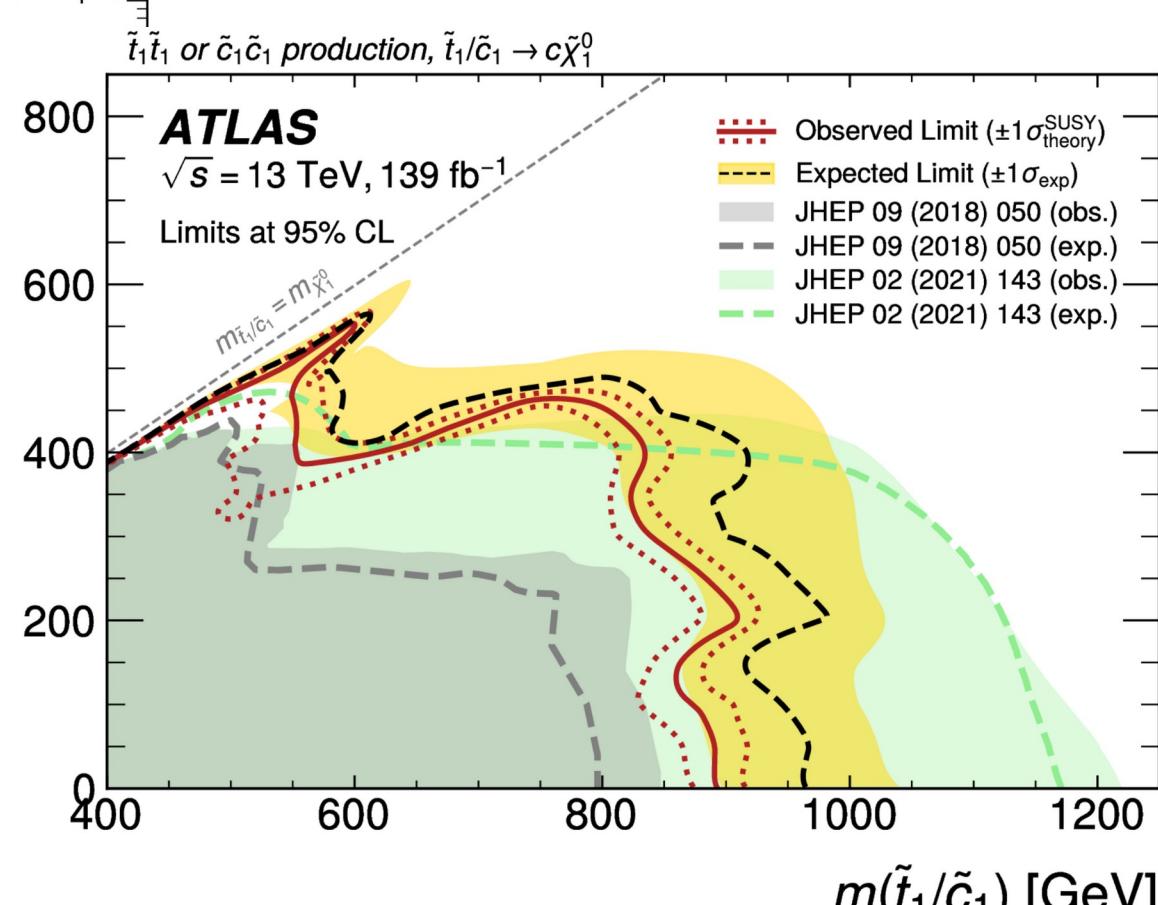
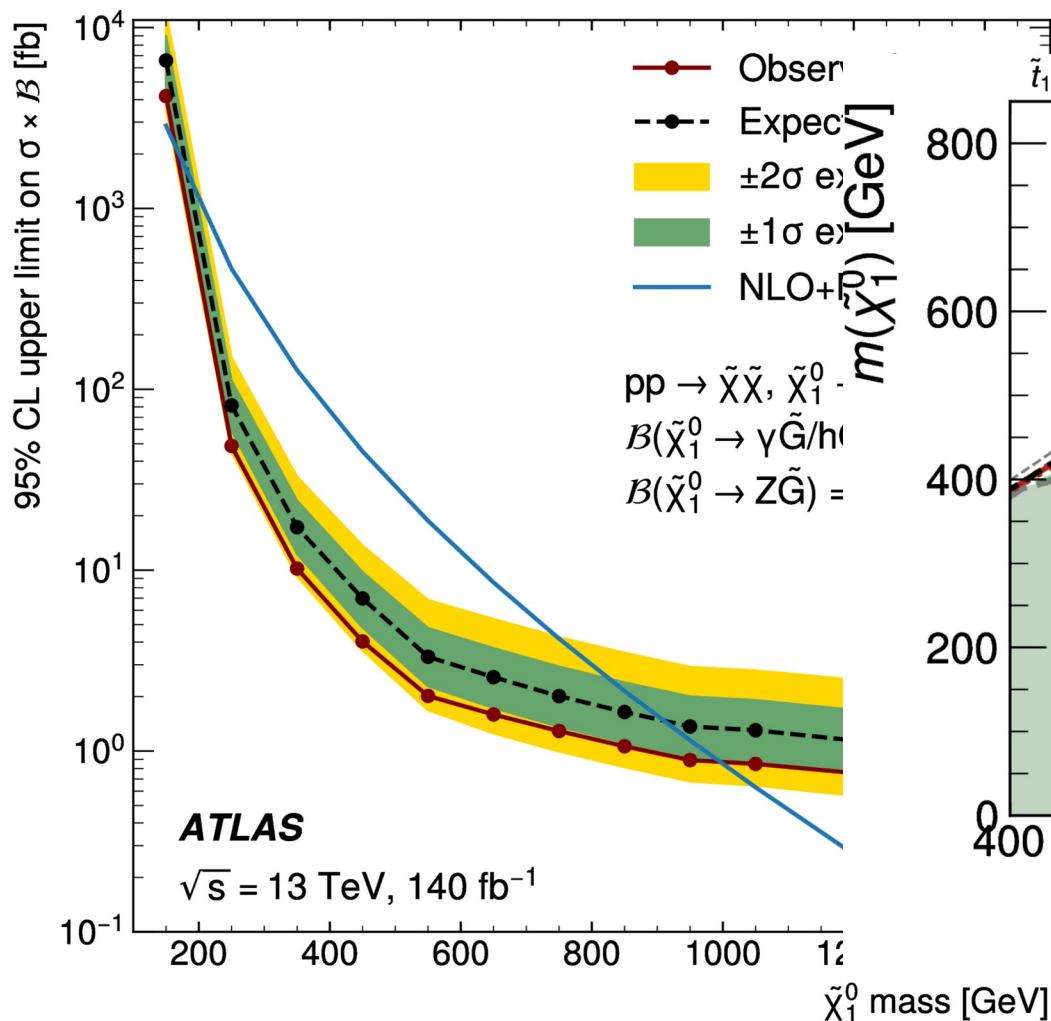
Ewkinos, small mass splitting



CMS-SUS-24-007

MET + higgs

# RESULTS NOT (YET) IN SMODELS



ATLAS-SUSY-2021-07  
Ewkinos (GMBS)

ATLAS-SUSY-2018-25  
Cc + MET

So overall:

- \*) too many small excesses for these classes of searches
- \*) not enough underfluctuations

Where do we go from here?

We find it imperative for experimentalists to be radically conservative:

- \*) how can we argue this away?
- \*) can we really draw a consistent picture or are there inconsistencies between all these excesses?
- \*) and in case we cannot “argue this away”, let’s make sure all tools, all infrastructure that is needed for evidence/discovery of new physics, is in fact there

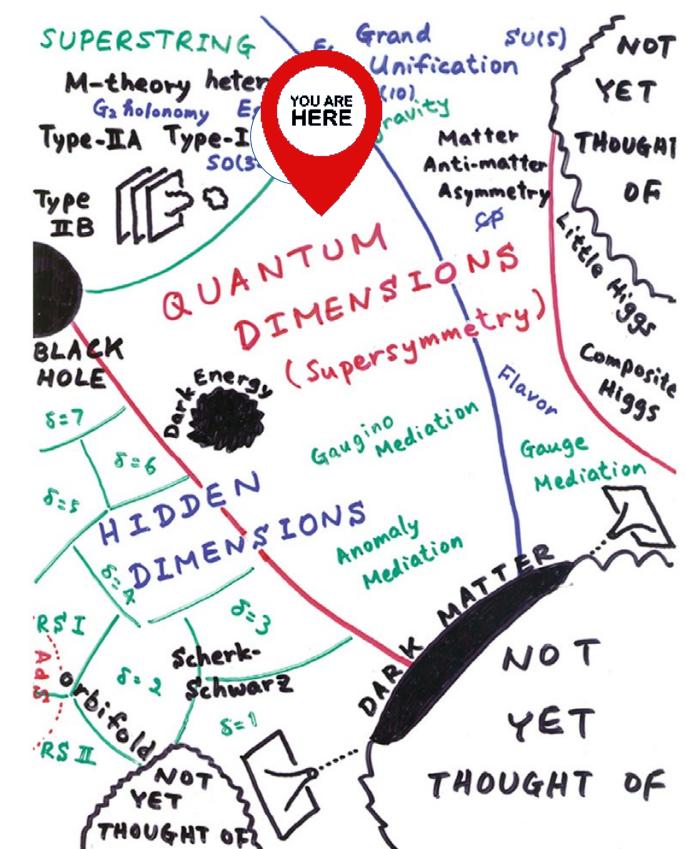
It is crucial for theorists to be conservatively radical:

\*) can we really draw a consistent picture or are there inconsistencies between all these excesses?

\*) if it really is due to new physics,  
what's the global picture,  
what do we infer globally?

\*) how would we construct a  
Next Standard Model from  
such experimental evidence?

A situation very much unlike that  
of the Higgs discovery!

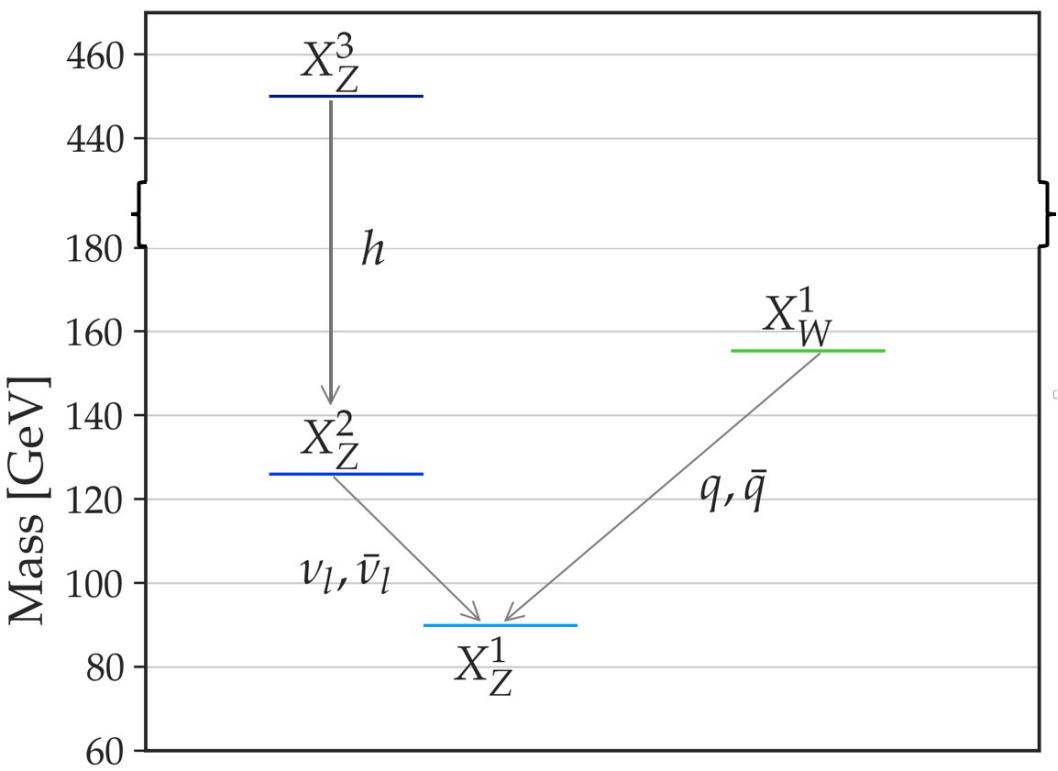


Let's try and find a simple prototype  
BSM theory that explains the  
excessive number of excesses!

Apart from it being interesting in its own right,  
such an attempt may help us also answering the  
radically conservatives' questions.

# Here is such a prototype model, a “protomodel”

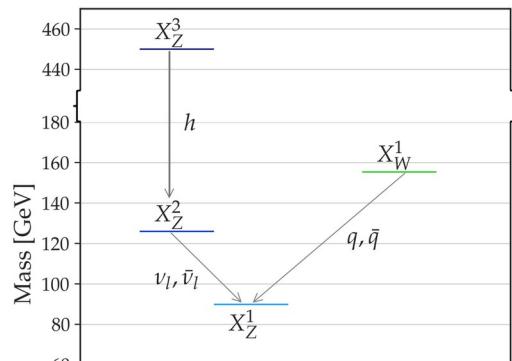
(We introduced this notion of a “protomodel”. It’s an incomplete BSM model, driven by data, not by theory considerations. LHC **production cross sections are free parameters**, BSM particle spins are left unspecified)



$X_z^i$  = “neutralino-like”

$X_w^i$  = “chargino-like”

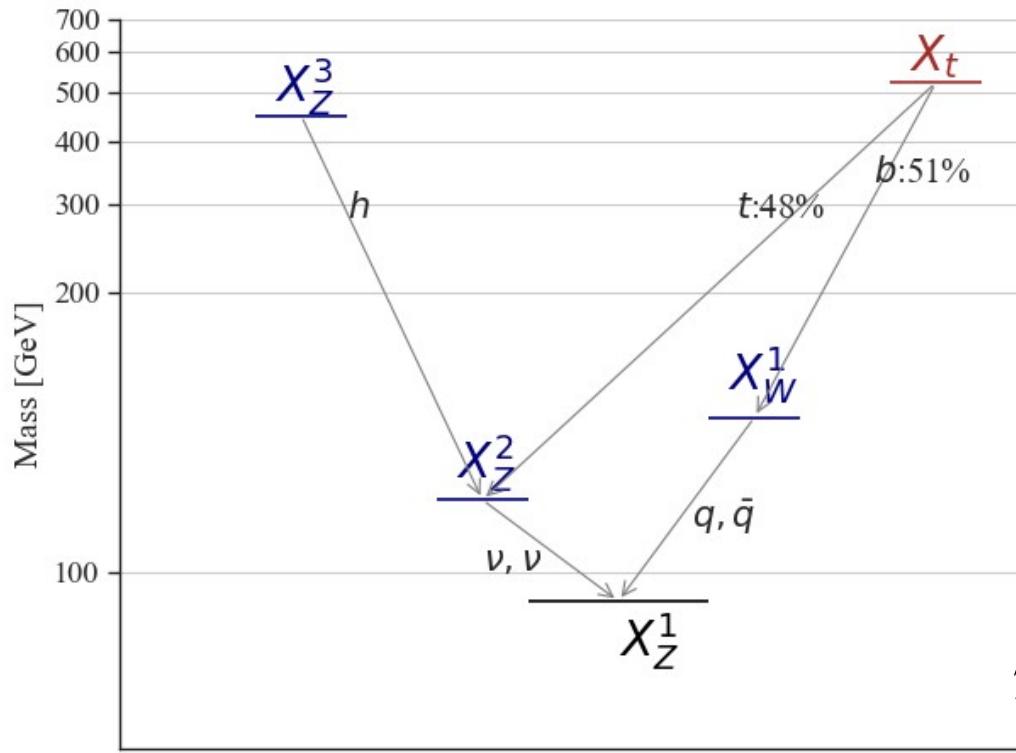
Why do we consider this to be an interesting protomodel?



Analysis Name	Type	Dataset	Topos	Observed	Expected	Approx $\sigma$	Particles
<a href="#">CMS-EXO-20-004</a>	comb	-	TChiISR, TChiZISRqq	2e+02 fb	1e+02 fb	4.6 $\sigma$	$X^1_W, X^1_Z,$ $X^2_Z$
<a href="#">ATLAS-EXOT-2018-06</a>	em	EM10	TChiISR, TChiZISRqq	413	359.00 +/- 10.00	2.5 $\sigma$	$X^1_W, X^1_Z,$ $X^2_Z$
<a href="#">CMS-SUS-20-004</a>	comb	-	TChiHH	0.3 fb	0.2 fb	2.1 $\sigma$	$X^3_Z, X^2_Z$

- \*) It contextualizes three of the biggest excesses we are seeing in the databases, that are considered mutually independent.
- \*) we can thus approximately combine the likelihoods of these three analyses – by multiplying them.
- \*) Another result exists, but is correlated, so we cannot add it in that fashion (but it would also exhibit an excess)

$$T_L := 2 \ln \left( \frac{L_{\text{BSM}}^c(\hat{\mu}) \cdot \pi_{\text{BSM}}}{L_{\text{SM}}^c \cdot \pi_{\text{SM}}} \right) = 33.69 = 5.8^2 \quad \text{"local 5.8 sigma" (if ndf=1) – with a GIGANTIC look elsewhere effect!}$$

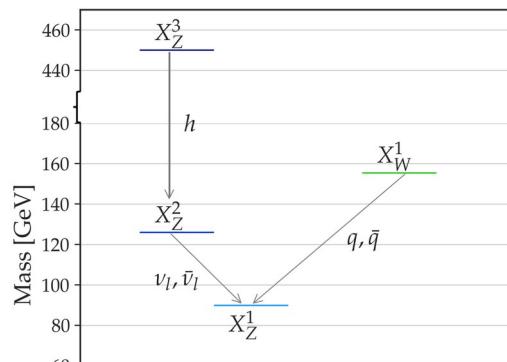


# Here is another such protomodel

$$T_L := 2 \ln \left( \frac{L_{\text{BSM}}^c(\hat{\mu}) \cdot \pi_{\text{BSM}}}{L_{\text{SM}}^c \cdot \pi_{\text{SM}}} \right) = 39.06 = 6.25^2$$

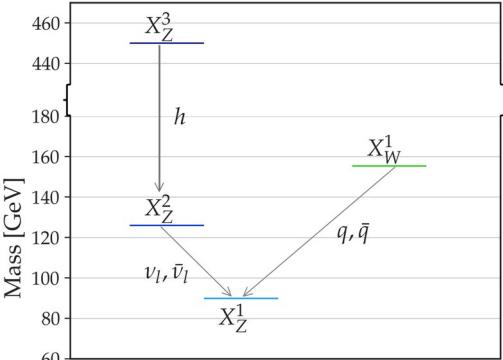
Analysis Name	Type	Dataset	Topos	Observed	Expected	Approx σ	Particles
<a href="#">CMS-EXO-20-004</a>	comb -		TChiISR, TChiZISRqq	2e+02 fb	1e+02 fb	4.6 σ	X <sup>1</sup> <sub>W</sub> , X <sup>1</sup> <sub>Z</sub> , X <sup>2</sup> <sub>Z</sub>
<a href="#">ATLAS-EXOT-2018-06</a>	em	EM10	TChiISR, TChiZISRqq	413	359.00 +/- 10.00	2.5 σ	X <sup>1</sup> <sub>W</sub> , X <sup>1</sup> <sub>Z</sub> , X <sup>2</sup> <sub>Z</sub>
<a href="#">ATLAS-SUSY-2018-12</a>	em	SRBTT	T2tt	67	46.00 +/- 7.00	2.1 σ	X <sub>t</sub> , X <sup>2</sup> <sub>Z</sub>
<a href="#">CMS-SUS-20-004</a>	comb -		TChiHH	0.3 fb	0.2 fb	2.0 σ	X <sup>3</sup> <sub>Z</sub> , X <sup>2</sup> <sub>Z</sub>
<a href="#">ATLAS-SUSY-2018-08</a>	em	SRSF140	T2tt	9	5.10 +/- 0.90	1.5 σ	X <sub>t</sub> , X <sup>2</sup> <sub>Z</sub>
<a href="#">ATLAS-SUSY-2016-16</a>	em	tN_med	T2tt	50	36.30 +/- 6.60	1.5 σ	X <sub>t</sub> , X <sup>2</sup> <sub>Z</sub>
<a href="#">ATLAS-SUSY-2013-16</a>	em	SRA4	T2tt	4	2.40 +/- 0.70	0.9 σ	X <sub>t</sub> , X <sup>2</sup> <sub>Z</sub>
<a href="#">ATLAS-SUSY-2013-15</a>	em	tNboost	T2tt	5	3.30 +/- 0.70	0.9 σ	X <sub>t</sub> , X <sup>2</sup> <sub>Z</sub>
<a href="#">CMS-PAS-SUS-13-015</a>	em	pTmiss350_Nb2	T2tt	15	8.60 +/- 7.10	0.8 σ	X <sub>t</sub> , X <sup>2</sup> <sub>Z</sub>
<a href="#">CMS-SUS-16-050-agg</a>	comb -		T2tt	2 fb	2 fb	-0.2 σ	X <sub>t</sub> , X <sup>2</sup> <sub>Z</sub>
<a href="#">CMS-SUS-13-011</a>	em	LM150	T2tt	227	251.00 +/- 50.00	-0.5 σ	X <sub>t</sub> , X <sup>2</sup> <sub>Z</sub>

# But let's focus on the simpler protomodel



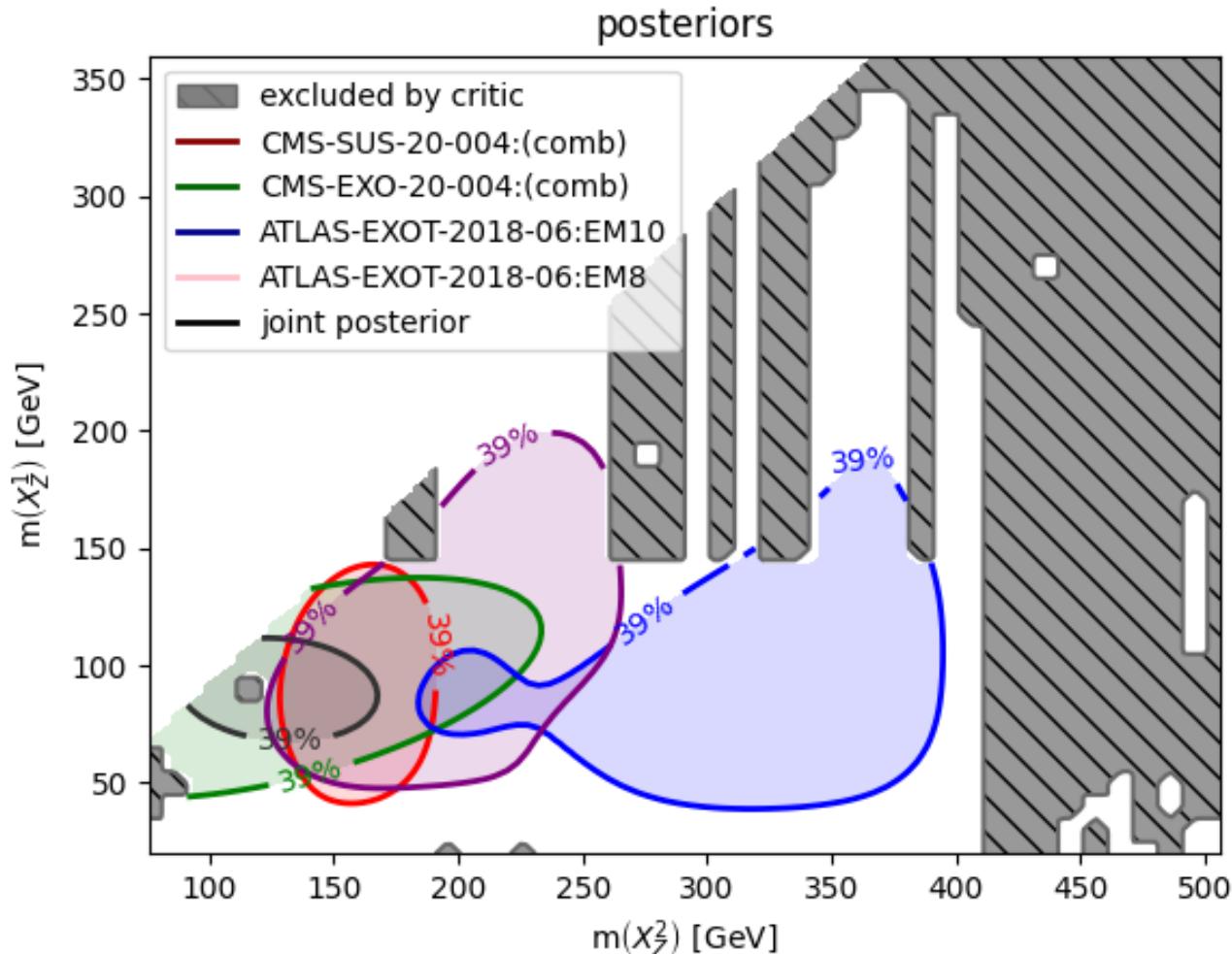
Analysis Name	Type	Dataset	Topos	Observed	Expected	Approx $\sigma$	Particles
<a href="#">CMS-EXO-20-004</a>	comb -		TChiISR, TChiZISRqq	2e+02 fb	1e+02 fb	4.6 $\sigma$	$X^1_W, X^1_Z,$ $X^2_Z$
<a href="#">ATLAS-EXOT-2018-06</a>	em	EM10	TChiISR, TChiZISRqq	413	359.00 +/- 10.00	2.5 $\sigma$	$X^1_W, X^1_Z,$ $X^2_Z$
<a href="#">CMS-SUS-20-004</a>	comb -		TChiHH	0.3 fb	0.2 fb	2.1 $\sigma$	$X^3_Z, X^2_Z$

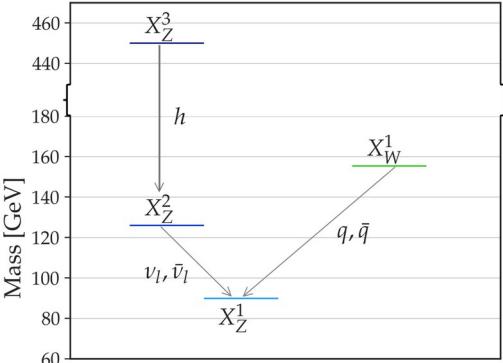
Do we consider these excesses to be mutually compatible?



Analysis Name	Type	Dataset	Topos	Observed	Expected	Approx $\sigma$	Particles
<a href="#"><u>CMS-EXO-20-004</u></a>	comb	-	TChiISR, TChiZISRqq	2e+02 fb	1e+02 fb	4.6 $\sigma$	$X^1_W, X^1_Z,$ $X^2_Z$
<a href="#"><u>ATLAS-EXOT-2018-06</u></a>	em	EM10	TChiISR, TChiZISRqq	413	359.00 +/- 10.00	2.5 $\sigma$	$X^1_W, X^1_Z,$ $X^2_Z$
<a href="#"><u>CMS-SUS-20-004</u></a>	comb	-	TChiHH	0.3 fb	0.2 fb	2.1 $\sigma$	$X^3_Z, X^2_Z$

# Combining Likelihoods

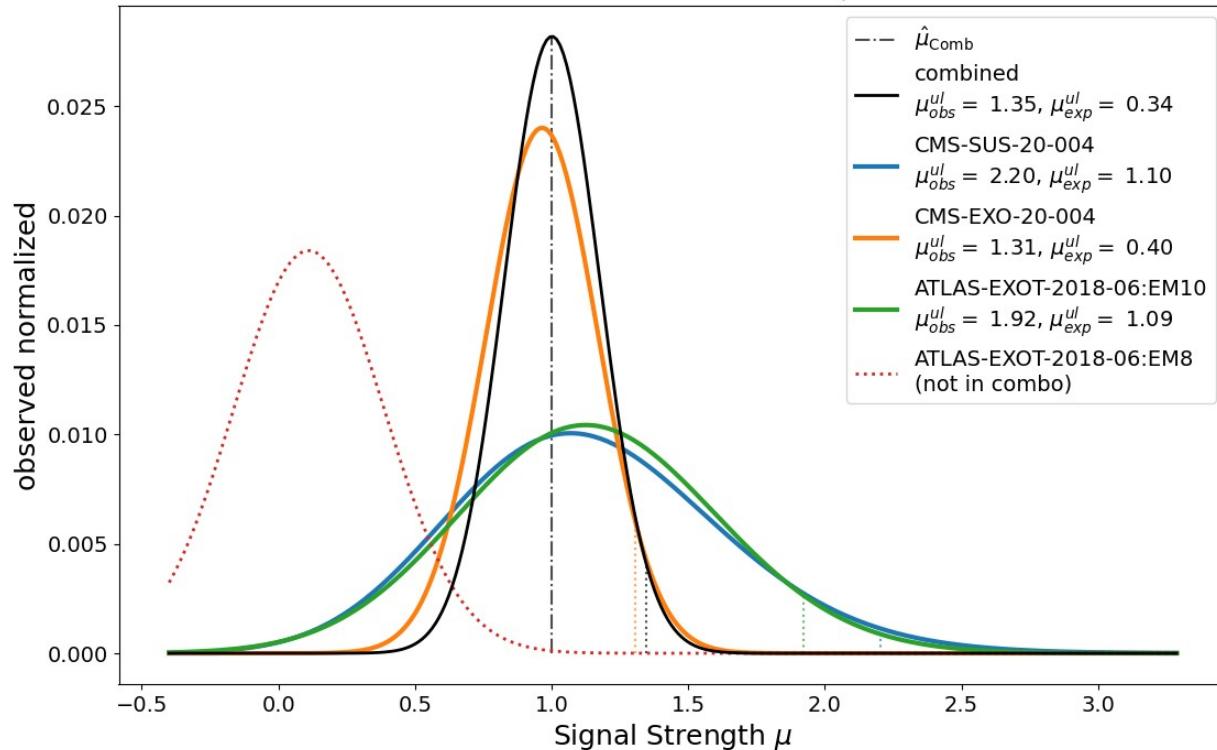




Analysis Name	Type	Dataset	Topos	Observed	Expected	Approx $\sigma$	Particles
<a href="#">CMS-EXO-20-004</a>	comb -	TChiISR, TChiZISRqq		2e+02 fb	1e+02 fb	4.6 $\sigma$	$X_W^1, X_Z^1,$ $X_Z^2$
<a href="#">ATLAS-EXOT-2018-06</a>	em	EM10	TChiISR, TChiZISRqq	413	359.00 +/- 10.00	2.5 $\sigma$	$X_W^1, X_Z^1,$ $X_Z^2$
<a href="#">CMS-SUS-20-004</a>	comb -	TChiHH		0.3 fb	0.2 fb	2.1 $\sigma$	$X_Z^3, X_Z^2$

# Combining Likelihoods

contributions of individual likelihoods, hiscore model



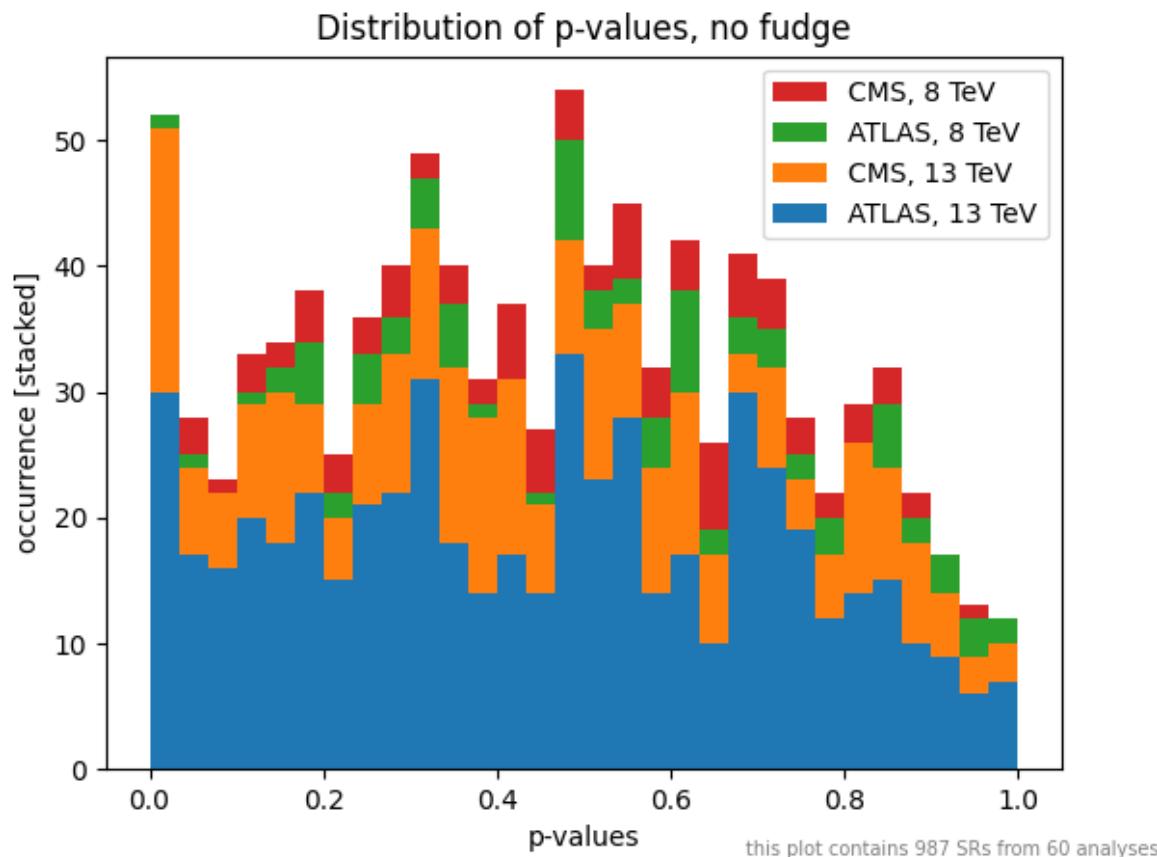
Likelihoods as functions of an overall signal strength multiplier

Taking this model with mu as the only free parameter, we could formulate a likelihood ratio test:

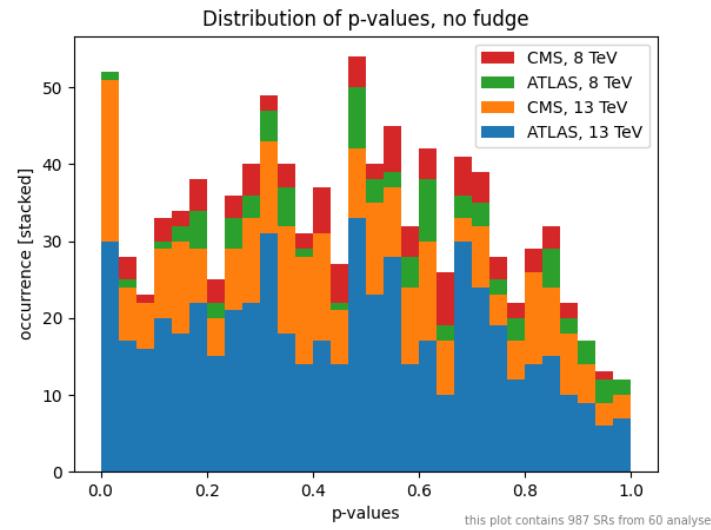
$$T_L := 2 \ln \left( \frac{L_{\text{BSM}}^c(\hat{\mu}) \cdot \pi_{\text{BSM}}}{L_{\text{SM}}^c \cdot \pi_{\text{SM}}} \right) = 33.69 = 5.8^2$$

Now that we have a protomodel, what about the meta-statistics plots that I showed you earlier? We have a new hypothesis, so we can compute the p-values for that!

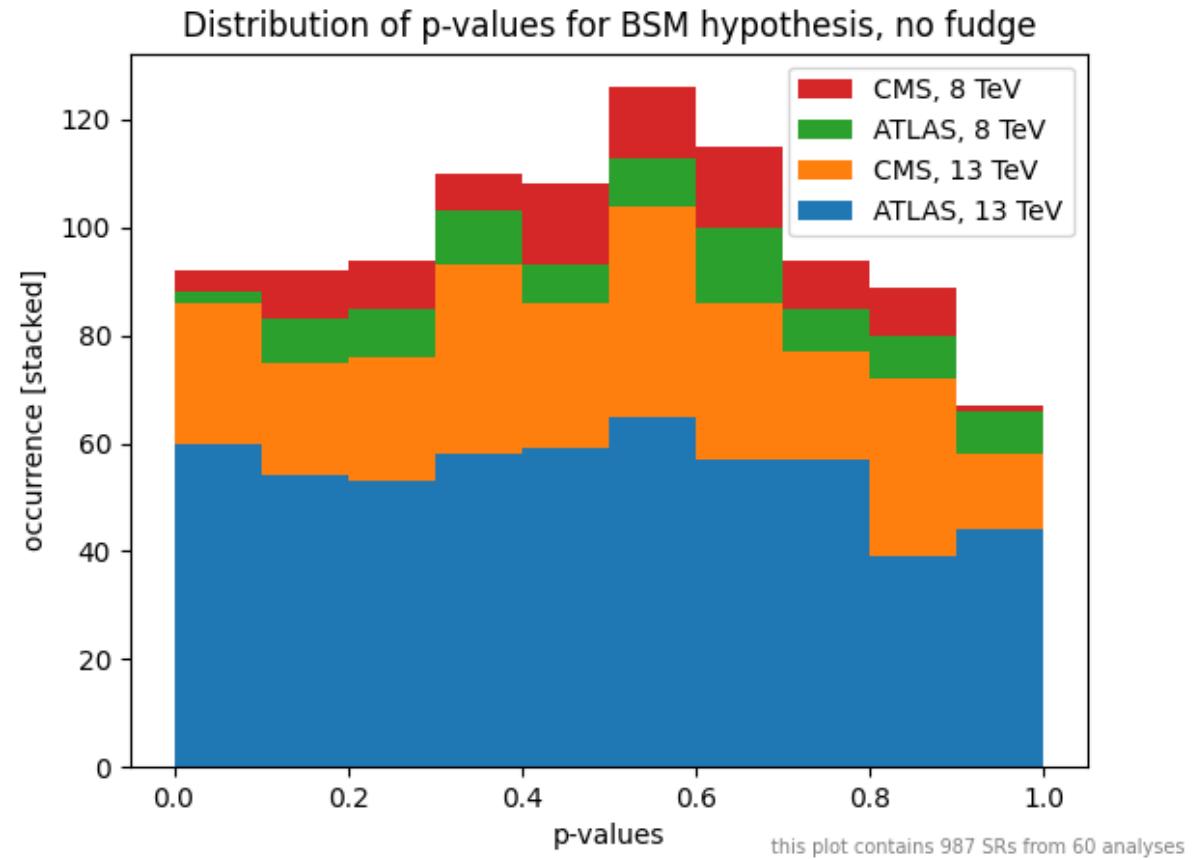
For the SM hypothesis:



# p-values for The SM hypothesis

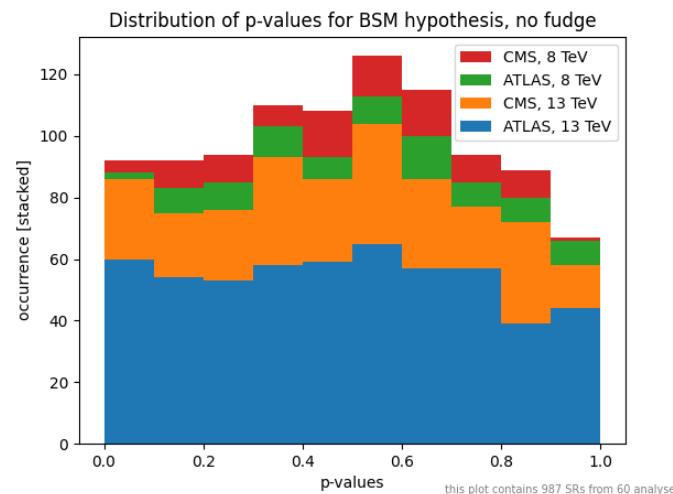


Instead of the SM, let's compute the p-values for the BSM hypothesis:

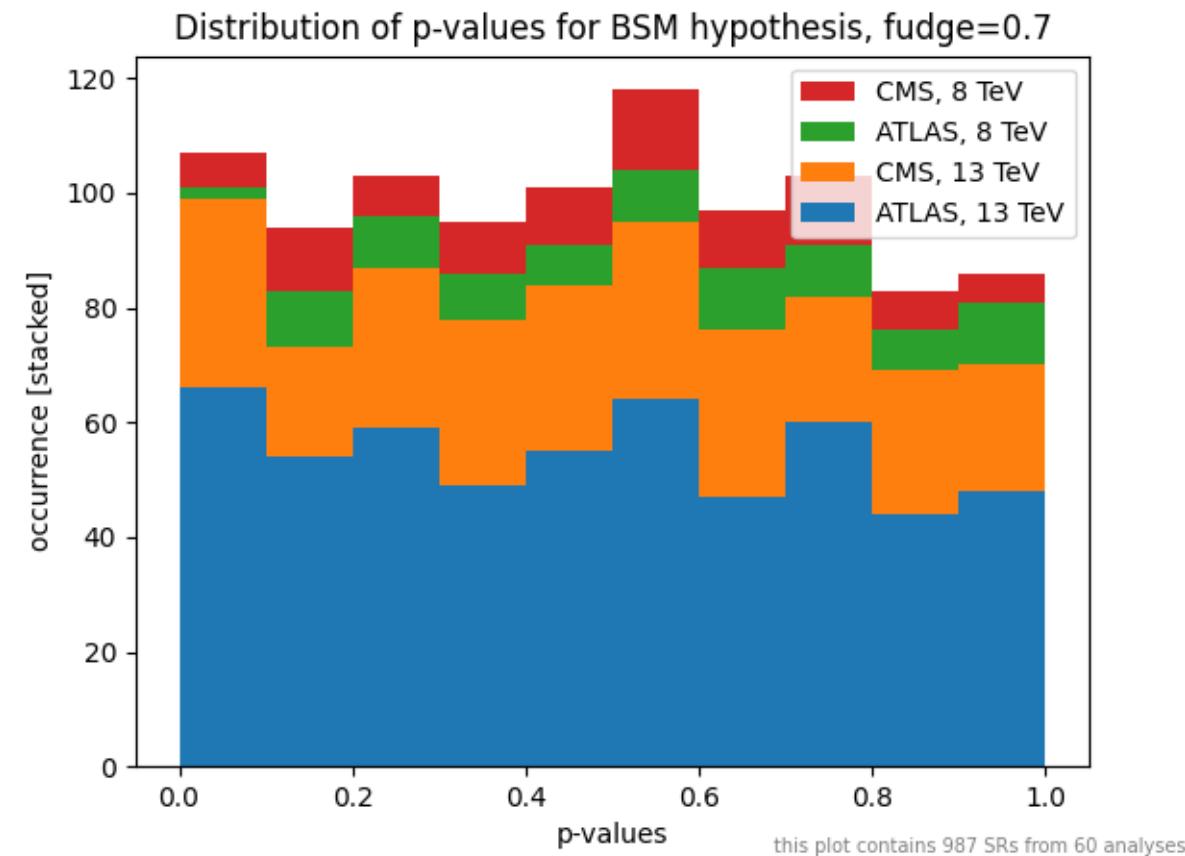


Instead of the SM, let's compute the p-values for the BSM hypothesis:

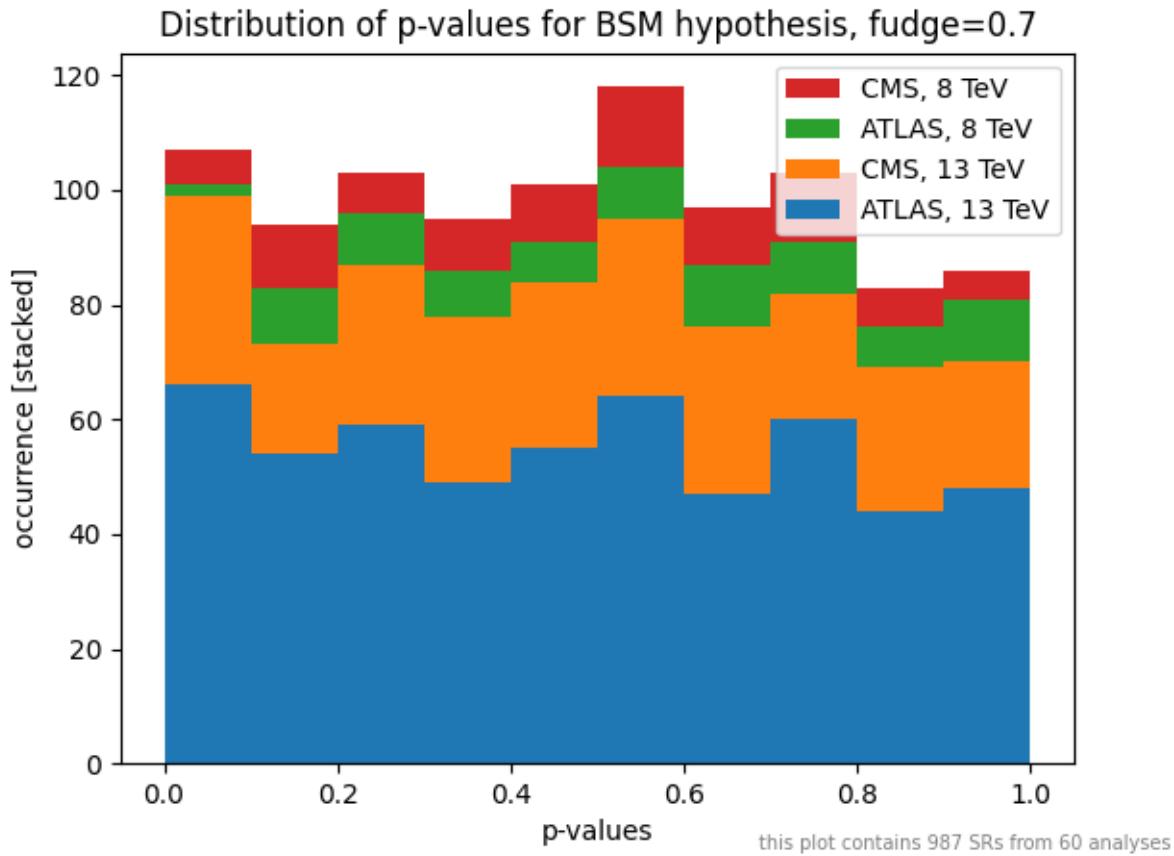
The convex shape makes sense if we assume the systematic errors to be conservatively estimated (as they should).



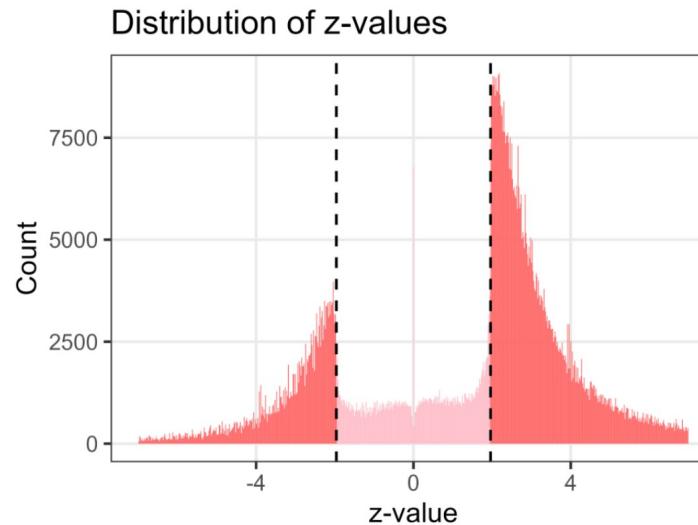
Here is the distribution of the P-values for the protomodel hypothesis, assuming all **systematic** errors to be  $\sim 40\%$  overestimated



Here is the distribution of the p-values for the protomodel hypothesis, assuming all systematic errors to be  $\sim 40\%$  overestimated

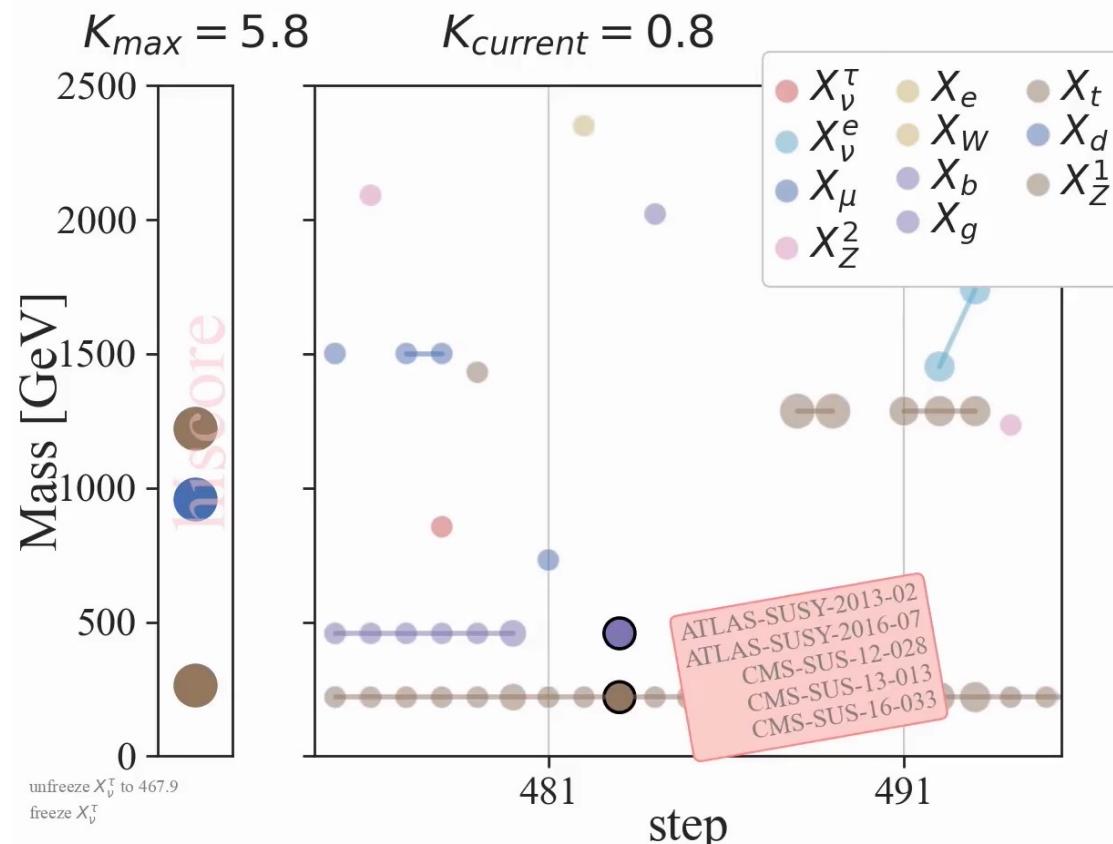
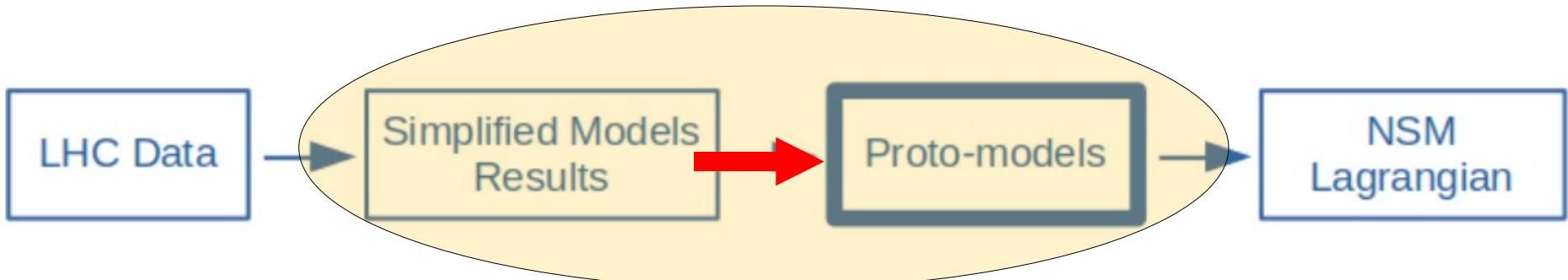


(This feels like nothing short of a miracle.  
Remember the medical research plot?)



But how did we obtain that  
protomodel?

# PROTOMODELS

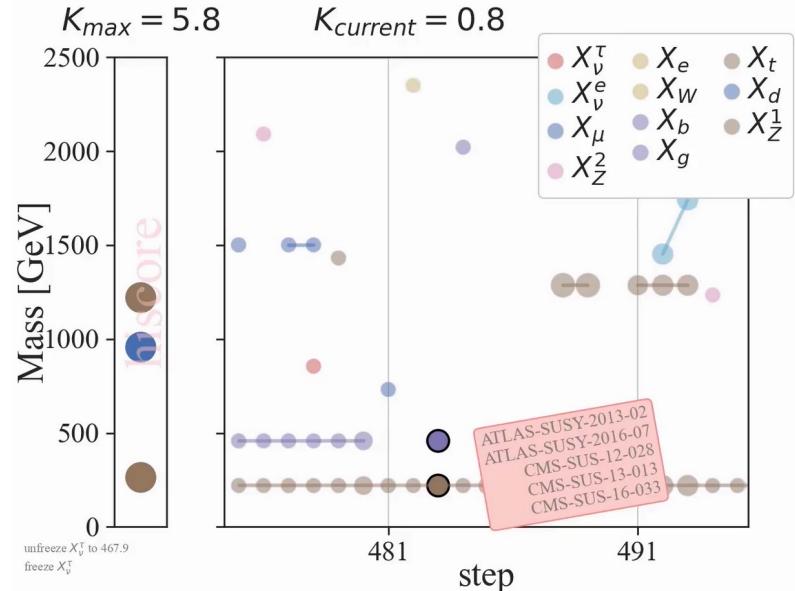


“Likelihood-based automated model building” (in a wider sense of the word ‘model building’)

# PROTOMODELS

It's like an MCMC walk as used for e.g. Bayesian inference, except:

- \*) we allow the machine to add / remove particles, decay channels, production mechanisms (variable dimensionality)
- \*) the machine combines analysis results depending on a notion of “combinability” and “completeness”
- \*) we look at likelihood ratios –  $L(\text{BSM})/L(\text{SM})$  – not likelihoods
- \*) we thus look for that legal combination that maximally violates The SM hypothesis
- \*) we added the notion of “critics” that ensure we remain consistent with all negative results in our database
- \*) we think of production cross sections as free parameters
- \*) we published a prototype v1 in 2020:



Artificial Proto-Modelling: Building Precursors of a Next Standard Model from Simplified Model Results

Wolfgang Waltenberger,<sup>a,b,1</sup> André Lessa,<sup>c</sup> Sabine Kraml<sup>d</sup>

<sup>a</sup>Institut für Hochenergiephysik, Österreichische Akademie der Wissenschaften, Nikolsdorfer Gasse 18, A-1050 Wien, Austria

<sup>b</sup>University of Vienna, Faculty of Physics, Boltzmanngasse 5, A-1090 Wien, Austria

<sup>c</sup>Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, Santo André, SP - Brasil

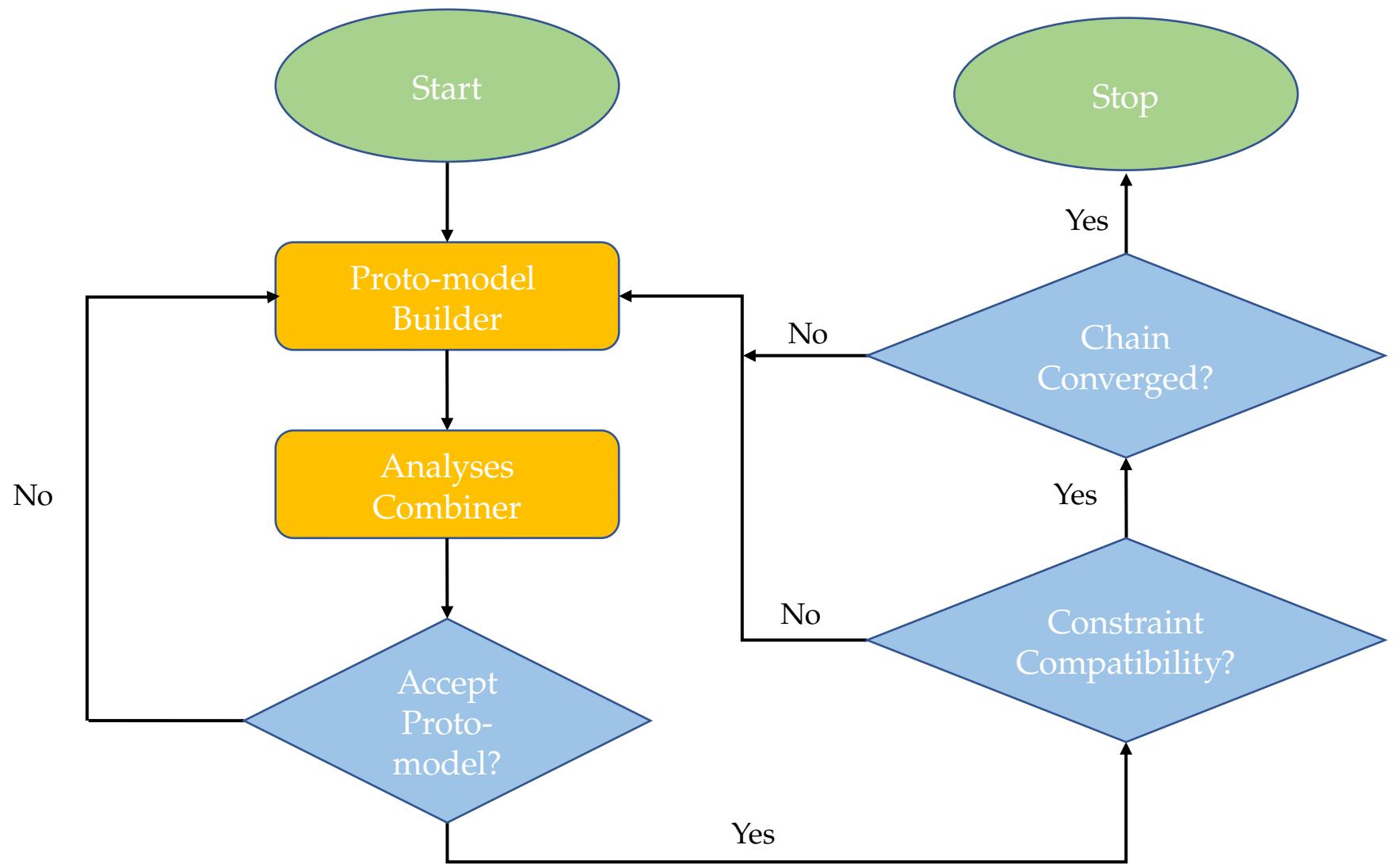
<sup>d</sup>Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3,

53 Avenue des Martyrs, F-38026 Grenoble, France

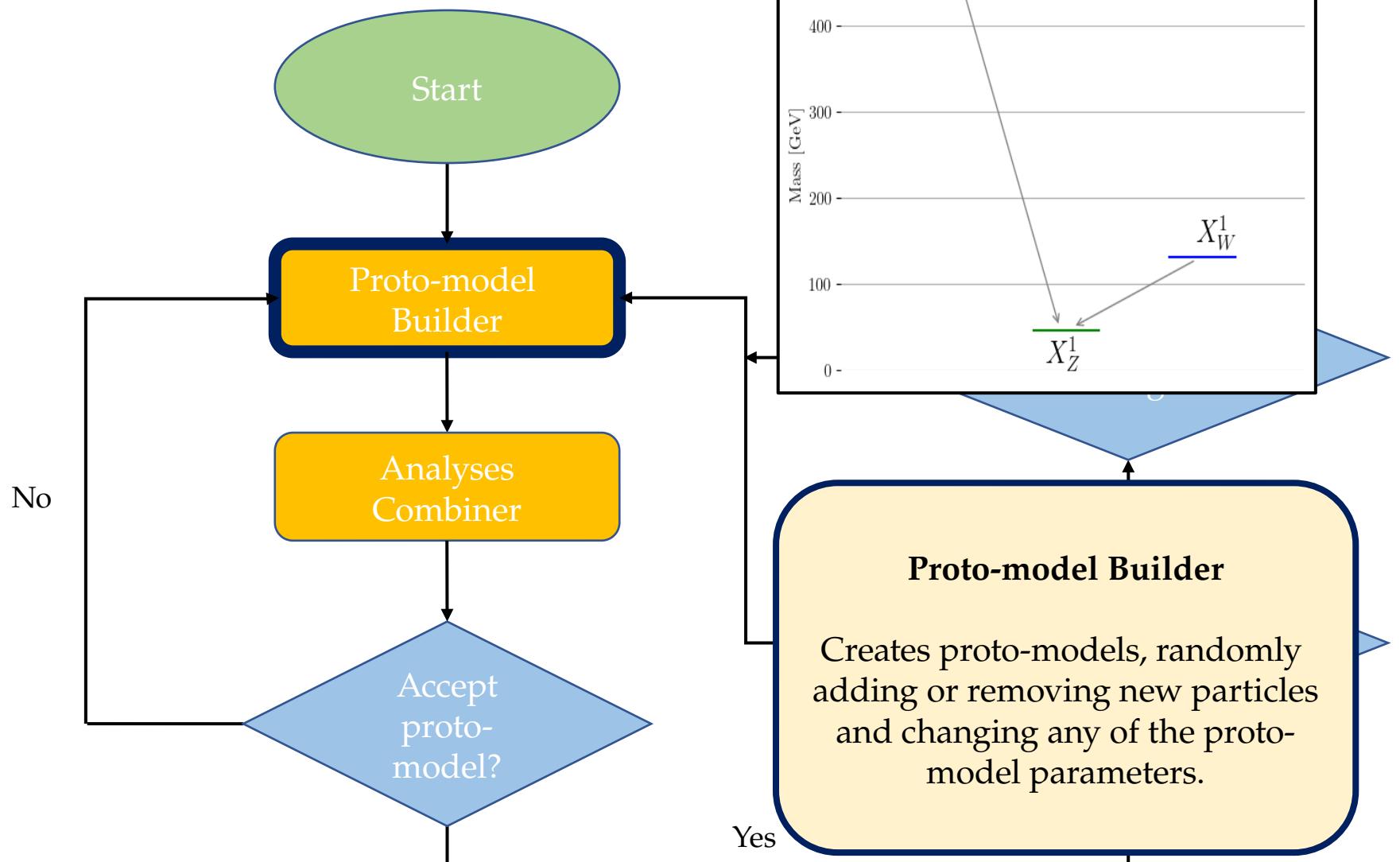
E-mail: [wolfgang.waltenberger@oeaw.ac.at](mailto:wolfgang.waltenberger@oeaw.ac.at), [andre.lessa@ufabc.edu.br](mailto:andre.lessa@ufabc.edu.br),

[sabine.kraml@lpsc.in2p3.fr](mailto:sabine.kraml@lpsc.in2p3.fr)

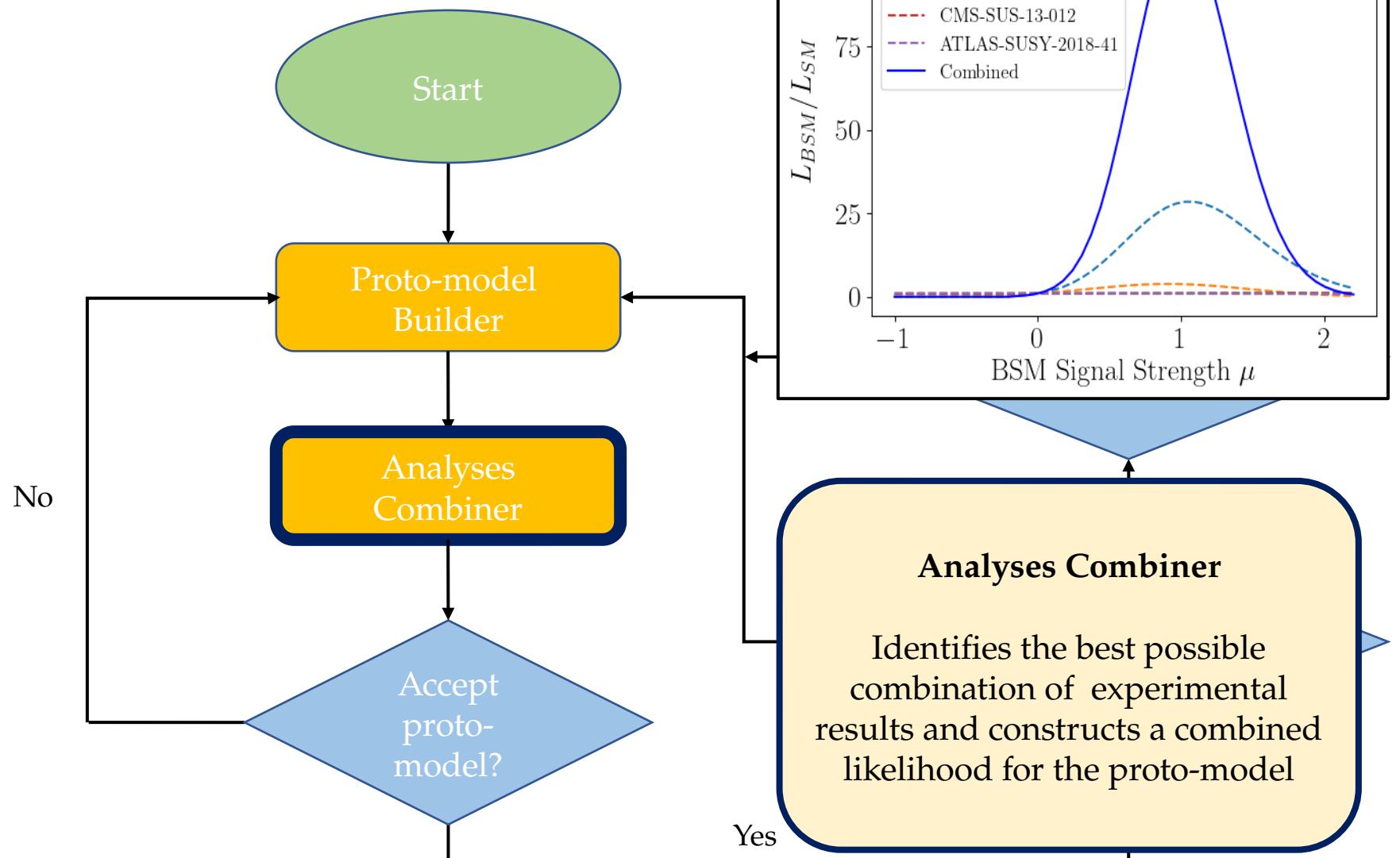
# Algorithm Flowchart



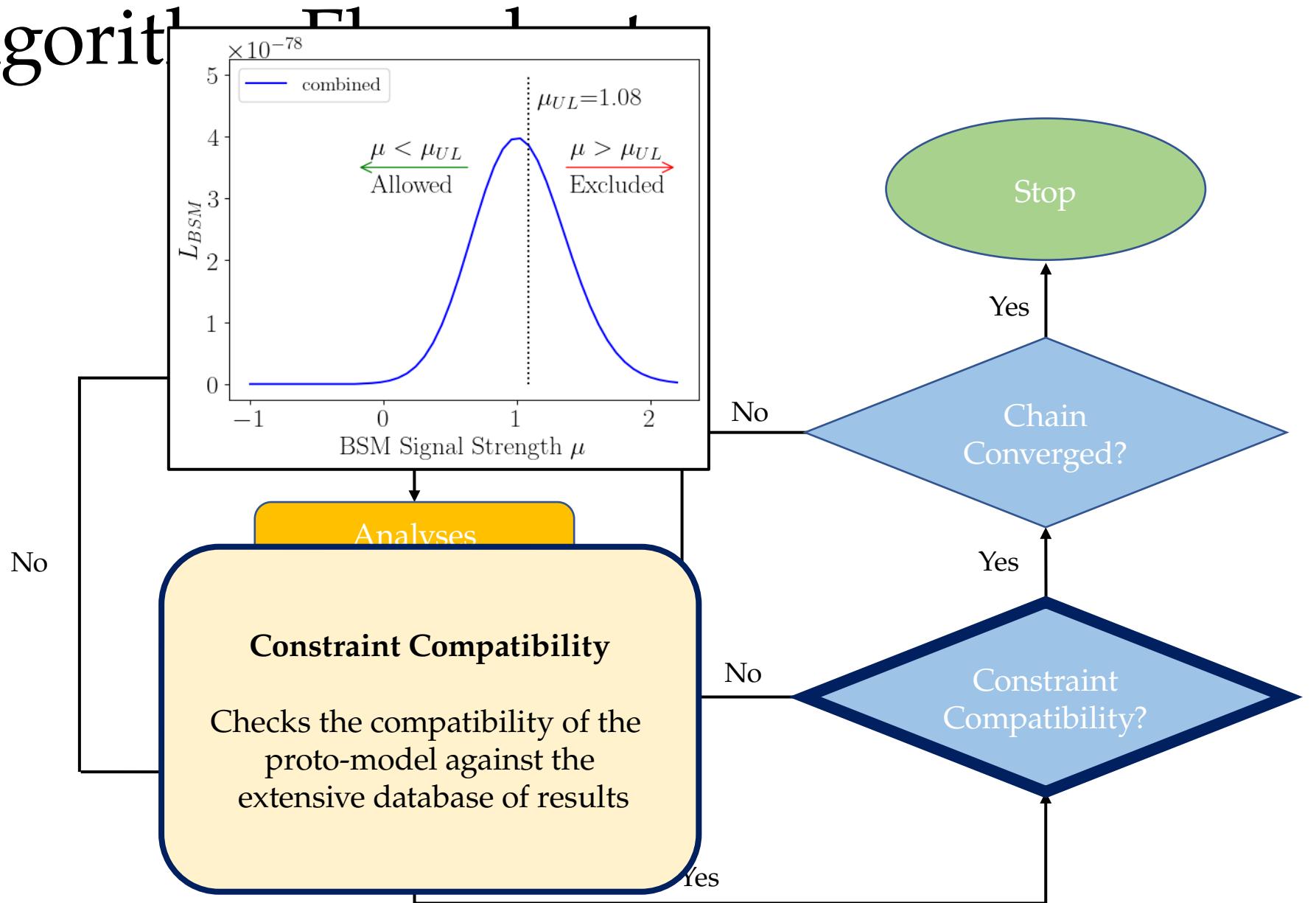
# Algorithm Flowchart



# Algorithm Flowchart



# Algorithm



# THE TEST STATISTIC

The test statistic  $K^c$  is a likelihood-ratio test that quantifies how much better the proto-model describes the data than the Standard-Model (plus a penalty for model complexity).

The diagram shows the formula for the test statistic  $K^c$  as a blue circle:

$$K^c := -2 \ln \frac{L_{\text{SM}}^c \cdot \pi(\text{SM})}{L_{\text{BSM}}^c(\hat{\mu}) \cdot \pi(\text{BSM})}$$

Annotations explain the components:

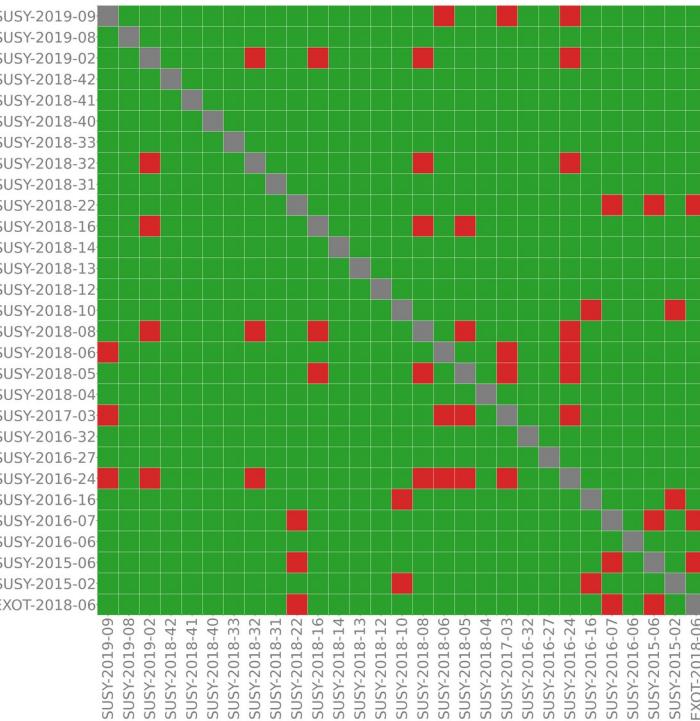
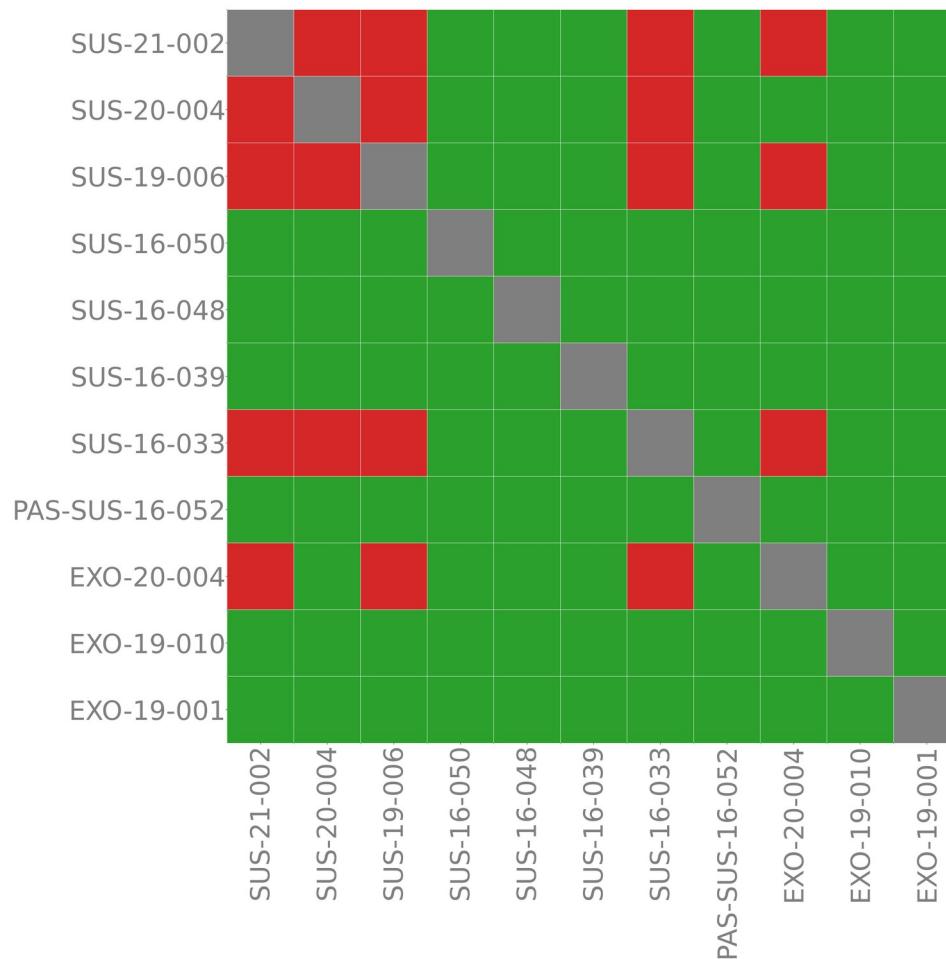
- A green arrow points to the term  $L_{\text{SM}}^c \cdot \pi(\text{SM})$  with the text "Quantifies violation of Standard Model hypothesis".
- A red arrow points to the term  $\pi(\text{BSM})$  with the text "Penalizes for model complexity (among other things)".
- A text box below the equation states: "Main penalty is similar to Akaike Information Criterion:"

$$\text{AIC} = 2k - 2 \ln(\hat{L})$$

We search for proto-models and combinations of results / likelihoods that maximize  $K^c$   
*while remaining compatible with all negative results in our database.*

# COMBINABILITIES

CMS, 13 TeV



Green: can combine

Red: cannot combine

\*) Only 11 (ATLAS: 29) Run-2 CMS analyses published with all material needed to be usable for protomodels, though situation has been greatly improving recently

\*) Longer term vision: full stats model for all analyses, naming conventions for nuisances. Correct combinations, taking into account correlations.

# PROTOMODELS V2

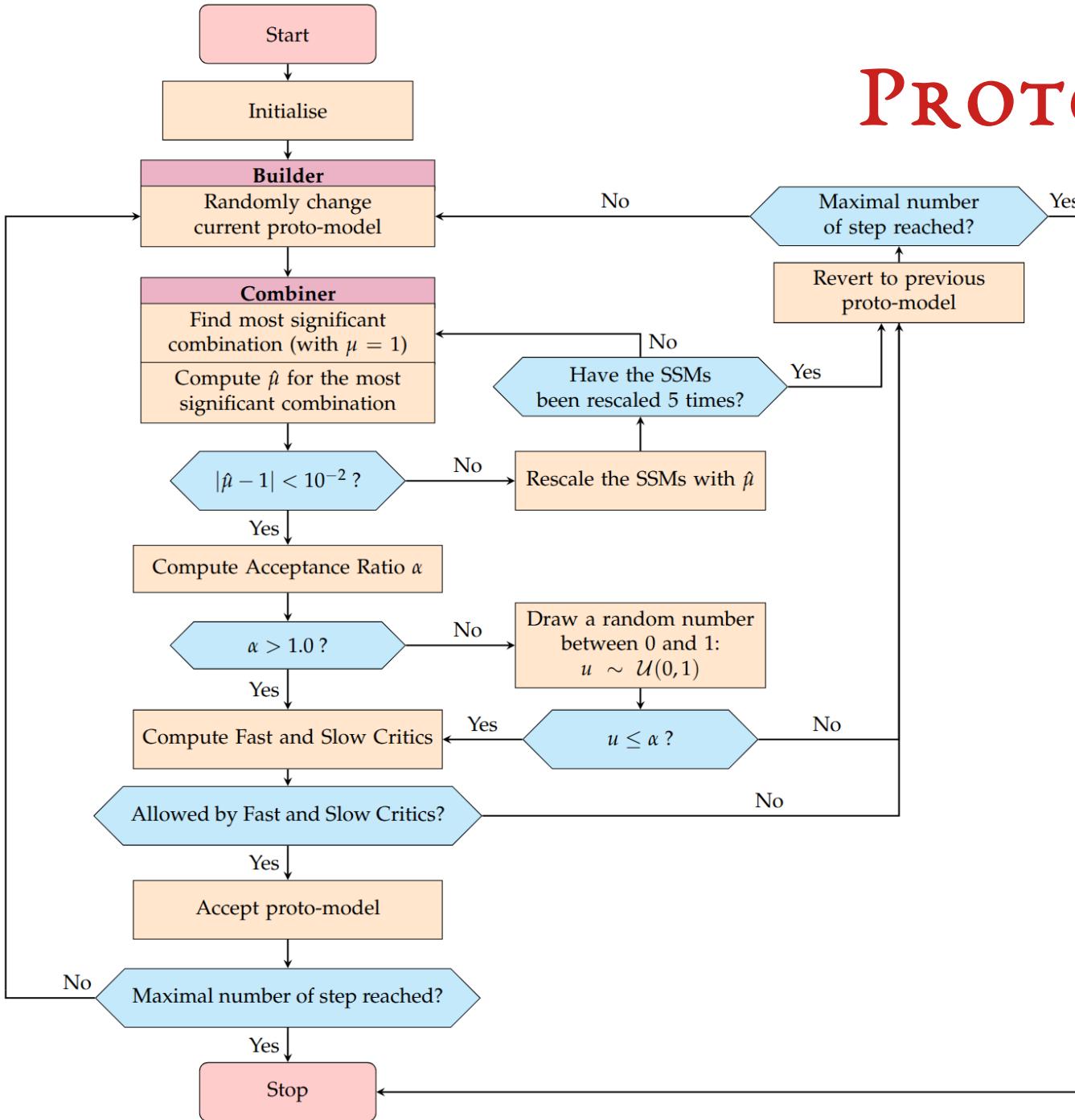
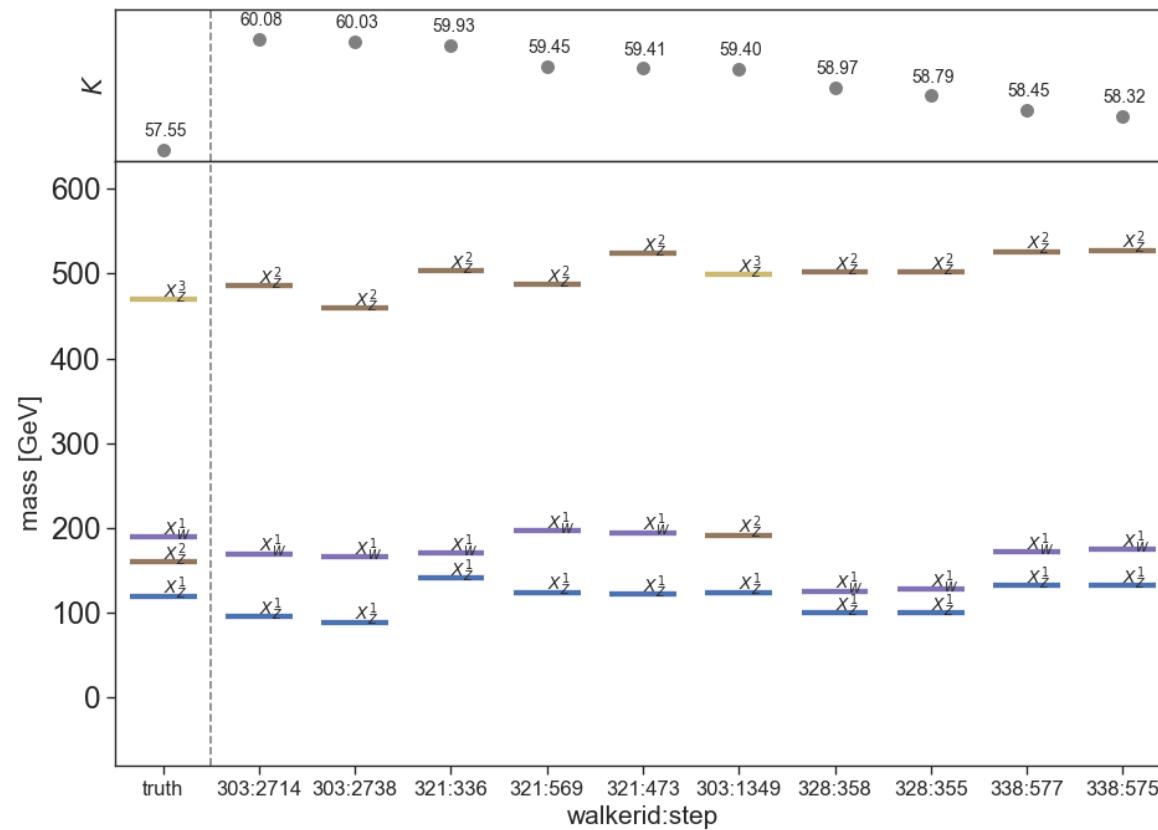


FIGURE 6.2: Simplified flowchart of the new code architecture.

Here is a flow chart of our much refined v2 of the machinery

# Closure test

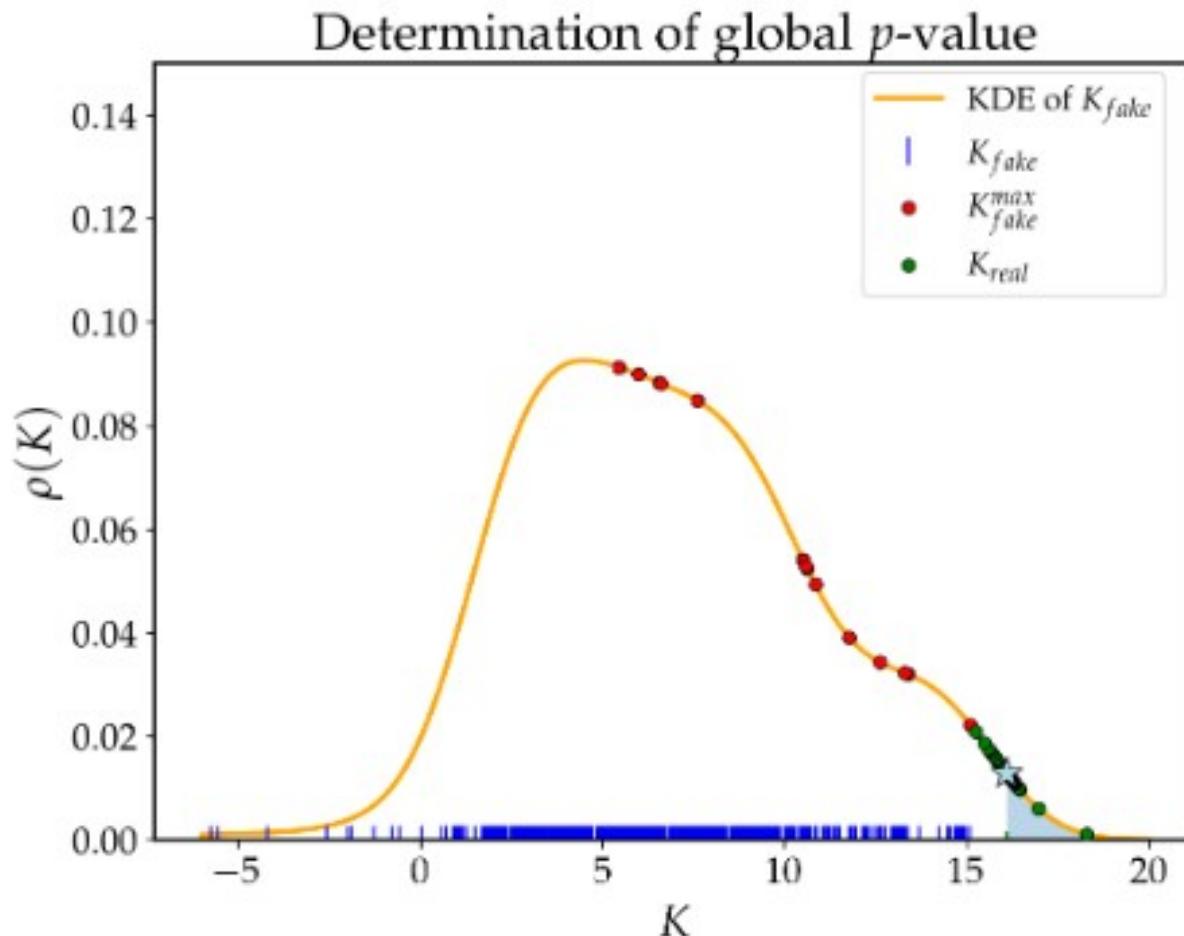
If we create fake data with an injected signal, then run our algorithm over it, do we get the truth back out?



For protomodels similar to our hiscore model, we do.

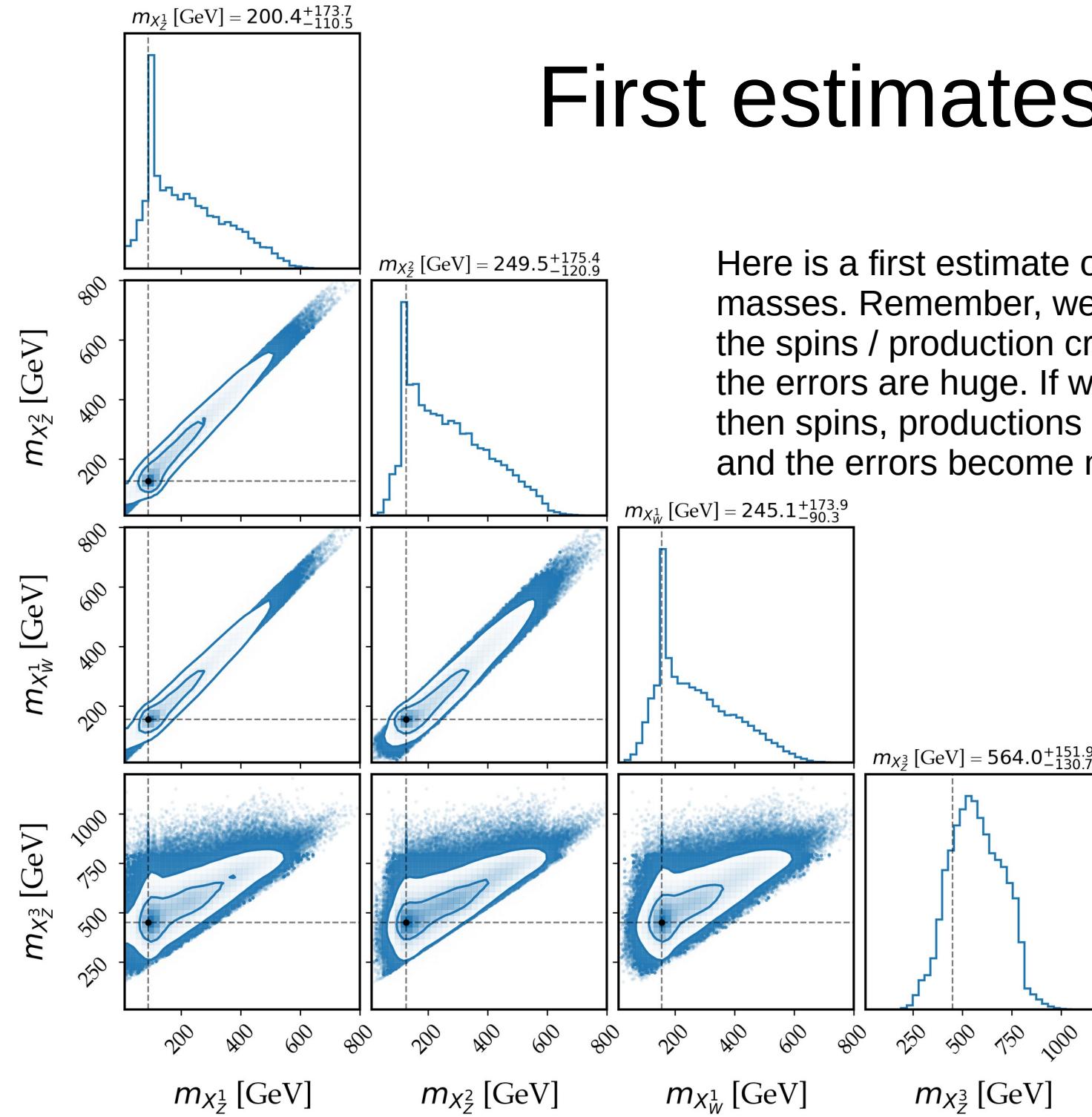
# global p-value

If we synthesize SM-only versions of our database, and run our algorithm on these synthetic data, we can compute a p-values for the SM-hypothesis



Extremely costly computation (1 – 100 millions of CPU hours), so only very rough estimate of global p-value (though the estimate is likely conservative)

# First estimates of errors



Here is a first estimate of e.g. errors on particle masses. Remember, we do not fixate the spins / production cross sections of our model, so the errors are huge. If we assume e.g. the MSSM, then spins, productions cross sections etc are fixed and the errors become much smaller

# Other pheno groups

Have other pheno groups noticed / noted on these similar compatibilities between excesses?

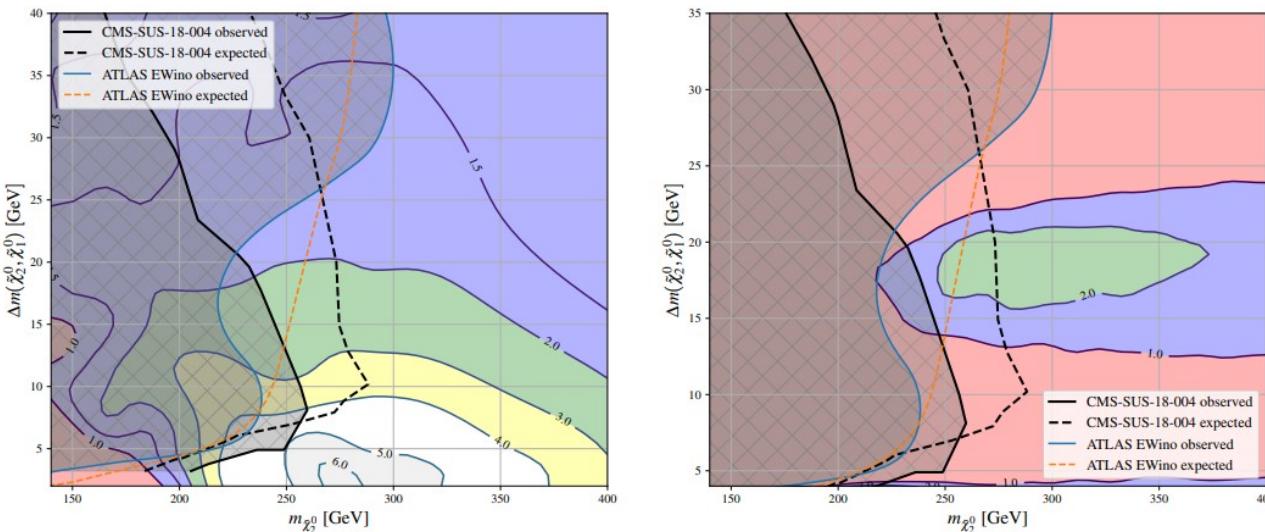
## A joint explanation for the soft lepton and monojet LHC excesses in the wino-bino model

Diyar Agin<sup>\*1</sup>, Benjamin Fuks<sup>†1</sup>, Mark D. Goodsell<sup>‡1</sup>, and Taylor Murphy<sup>§2</sup>

### GUT-based NMSSM with Singlino Dark Matter: Describing the ATLAS/CMS Excesses

E. Bagnaschi,<sup>a</sup> M. Chakraborti,<sup>b</sup> S. Heinemeyer<sup>c,\*</sup> and I. Saha<sup>d</sup>

partners still allowed with masses as low as a few hundred GeV. Over the last years, searches for the “golden channel”,  $pp \rightarrow \tilde{\chi}_2^0 \tilde{\chi}_1^\pm \rightarrow \tilde{\chi}_1^0 Z^{(*)} \tilde{\chi}_1^0 W^{\pm(*)}$  show consistent excesses between CMS and ATLAS in the 2 soft-lepton and 3 soft-lepton plus  $\cancel{E}_T$  searches, favoring mass scales of  $m_{\tilde{\chi}_2^0} \approx m_{\tilde{\chi}_1^\pm} \gtrsim 200$  GeV and  $\Delta m_{21} := m_{\tilde{\chi}_2^0} - m_{\tilde{\chi}_1^0} \approx 20$  GeV. We analyze these excesses in the framework of the Next-to-Minimal Supersymmetric Standard Model (NMSSM). We assume

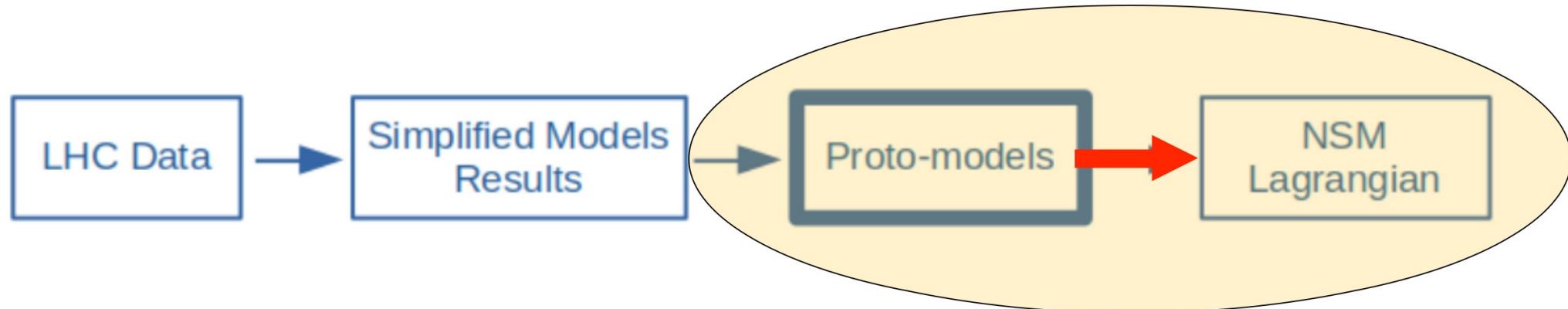


**Figure 4:** Results for CMS monojet searches (left) and ATLAS soft-lepton+ $E_T^{\text{miss}}$  (right) in the wino-bino scenario, overlaid with soft-lepton excluded regions reported by ATLAS and CMS. Top: Expected and observed limits. Bottom: Contours of Bayes factor  $B_{10}$ .

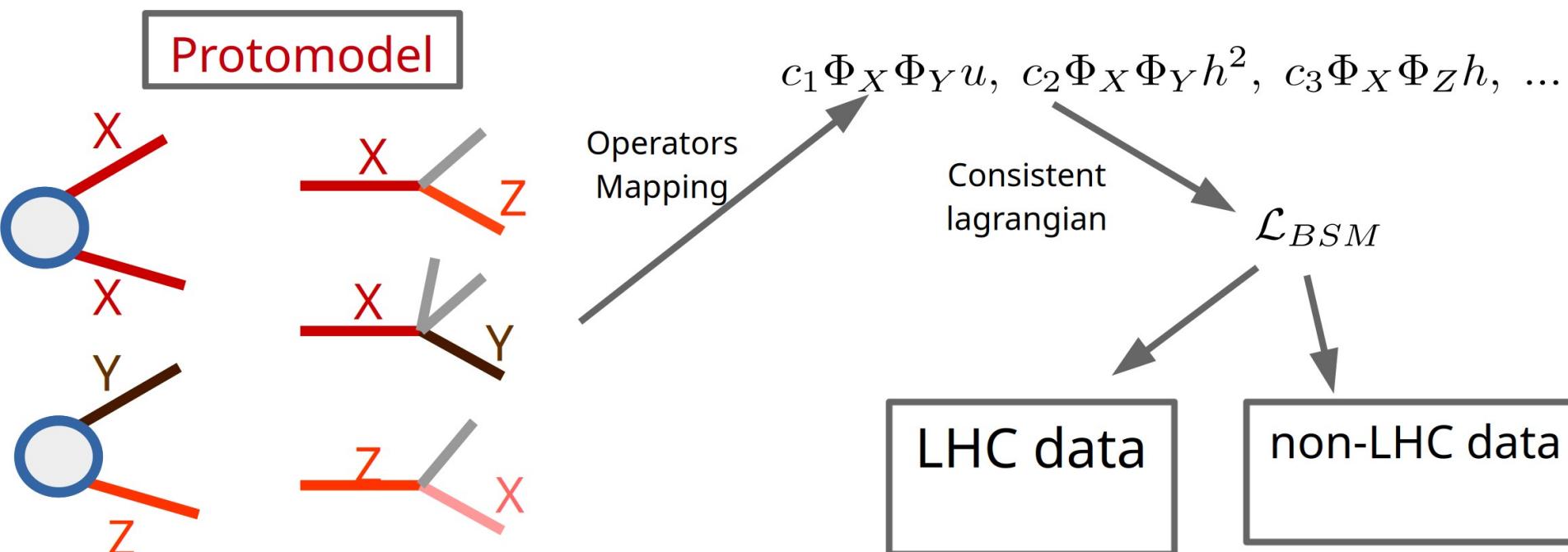
# Protomodels – what will we do with them?

(apart from them serving as a high-level summary of what we are seeing, possibly informing the search programmes about where to look next)

# NSM Lagrangian from the proto-models, work in progress



Work from protomodels to UV complete theories has begun  
(John Gargalionis) – lots of combinatorics!



# Wanna help?

v1 was a feasibility study

v2 will (soon) be the first version producing protomodels that are of physics relevance

v3 should give a comprehensive summary of much of the LHC programme (especially SUSY)

We have started to plan a v3 of the project

There is more than plenty to do, from curating the inputs to speeding things up with ML and gradients to connecting the output protomodels with UV complete models.

Full scale combinations within the collaborations, etc etc.

# Wanna help?

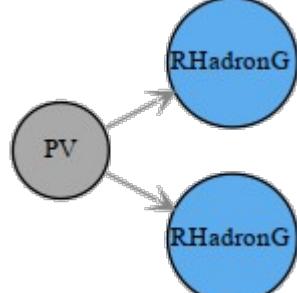
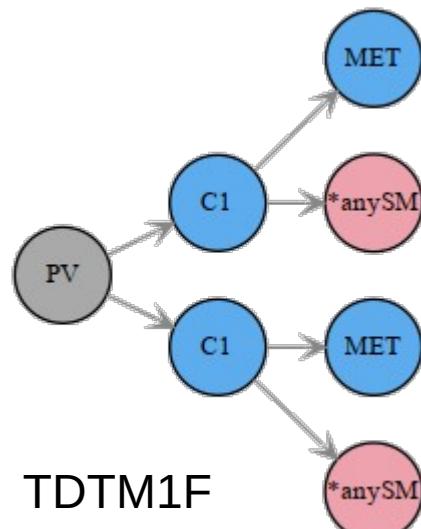
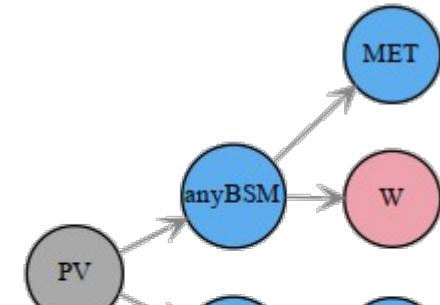
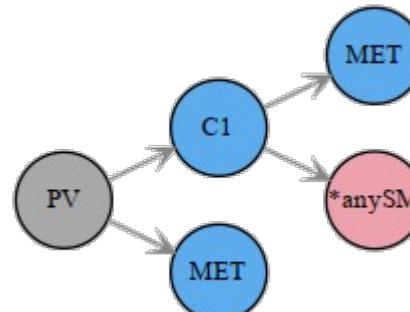
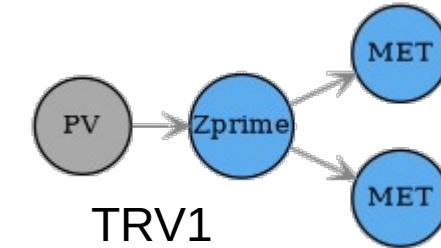
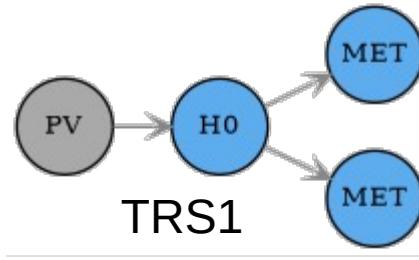
## TODO

- Implement more analyses in our database
- machine learn surrogate statistical models to speed up evaluation
- work on hs3/combine support in our software
- work on code that can simplify statistical models, to speed things up
- combinations within the MSSM, NMSSM, etc
- introduce gradients in our inference chain
- extend the space of theories we can cover
- Recast older analyses, recast for new simplified models
- Compare with e.g. dark matter relic density, XENONnT
- Connect with measurements, EFTS
- Protomodels but full-scale combinations (possibly within the experimental collaborations?)
- Run interesting models with full recasts
- Identification of compatible UV complete theories

Thank you!

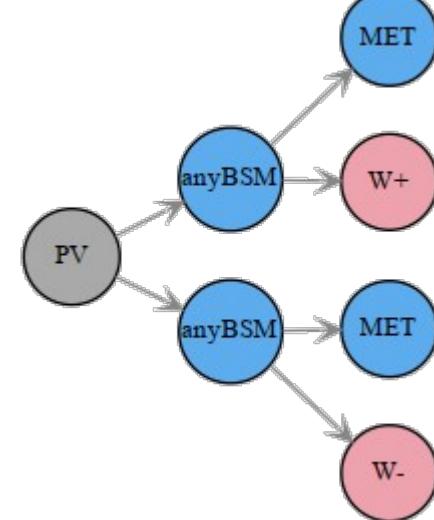
Naively: (without assuming conservatism) we would expect  
 $\approx 960 * \Phi(2.40) \approx 8$  SRs.  
 We have 28 SRs with  $Z > 2.40$ , 4 SRs with  $Z < -2.40$

# LIST OF SRs WITH $Z > 2.40$

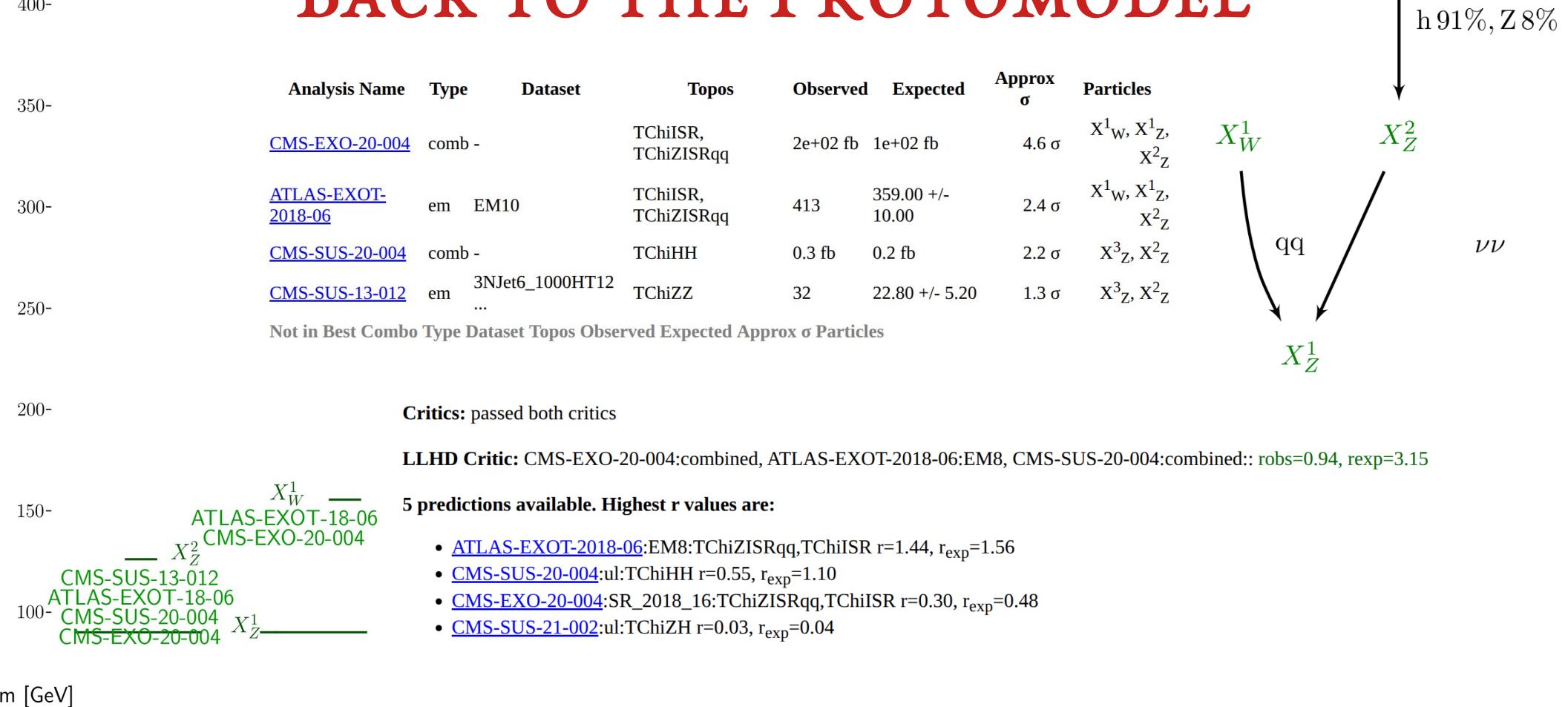


Z	analysis	SR	topo	TDTM2F	expected
4.47	CMS-EXO-20-004	SR_2018_8	TChiISR ...	8429	$730.98 \pm 152.03$
4.31	CMS-EXO-20-004	SR_2018_4	TChiISR ...	35833	$33037.13 \pm 618.12$
4.11	CMS-EXO-20-004	SR_2018_6	TChiISR ...	14266	$13209.01 \pm 250.28$
3.85	CMS-EXO-20-004	SR_2018_9	TChiISR ...	5406	$4914.64 \pm 101.04$
3.80	ATLAS-SUSY-2018-42	SR1400_High_long	T1DispT ...	7	$0.70 \pm 0.40$
3.78	CMS-EXO-20-004	SR_2018_1	TChiISR ...	191829	$178470.12 \pm 3615.00$
3.72	CMS-EXO-20-004	SR_2018_2	TChiISR ...	104522	$97428.42 \pm 1896.81$
3.63	CMS-EXO-20-004	SR_2018_3	TChiISR ...	59398	$55592.19 \pm 1059.45$
3.46	ATLAS-SUSY-2018-42	SR1500_High_long	TDTM1FT ...	6	$0.60 \pm 0.40$
3.35	CMS-EXO-20-004	SR_2018_11	TChiISR ...	2671	$2428.39 \pm 55.08$
3.30	ATLAS-SUSY-2018-16-hino	SR_ewk_2l_med	TChiWWo ...	170	$112.40 \pm 12.60$
3.21	ATLAS-SUSY-2018-42	SR1300_High_long	TDTM1FT ...	6	$0.80 \pm 0.40$
3.21	ATLAS-SUSY-2018-16	SR_ewk_2l_ee_high_d	TChiWZoff	16	$4.80 \pm 1.60$
3.17	CMS-EXO-20-004	SR_2018_12	TChiISR ...	1613	$1441.83 \pm 35.64$
3.17	ATLAS-SUSY-2018-42	SR1600_High_long	T1DispT ...	5	$0.54 \pm 0.35$
3.01	ATLAS-SUSY-2018-42	SR1200_High_long	T1DispT ...	6	$0.90 \pm 0.50$
3.01	CMS-EXO-20-004	SR_2018_7	TChiISR ...	11730	$10991.22 \pm 213.74$
2.93	ATLAS-SUSY-2017-03	SR31_ISR	TChiWZ	12	$3.90 \pm 1.00$
2.83	ATLAS-SUSY-2018-42	SR1800_High_long	T1DispTRHadGM1	4	$0.44 \pm 0.32$
2.75	CMS-EXO-20-004	SR_2018_5	TChiISR ...	21854	$20714.72 \pm 391.42$
2.56	ATLAS-SUSY-2019-09	SRlow_0Jd	TChiWZoff	77	$54.00 \pm 4.00$
2.51	CMS-SUS-20-004	SR11	T5HHTChiHH	4	$0.06 \pm 0.11$
2.51	CMS-EXO-20-004	SR_2018_10	TChiISR ...	3285	$3065.69 \pm 66.79$
2.49	ATLAS-SUSY-2018-42	SR2400_High_long	T1DispTRHadGM1	3	$0.29 \pm 0.28$
2.47	ATLAS-EXOT-2018-06	EM10	TChiISR ...	413	$359.00 \pm 10.00$
2.45	ATLAS-SUSY-2018-42	SR2600_High_long	T1DispTRHadGM1	3	$0.25 \pm 0.31$
2.44	ATLAS-SUSY-2018-42	SR2000_High_long	T1DispTRHadGM1	3	$0.36 \pm 0.25$
2.42	ATLAS-SUSY-2018-42	SR2200_High_long	T1DispTRHadGM1	3	$0.33 \pm 0.28$

- \*) CMS-EXO-20-004 is yes overrepresented  
 (simplified likelihood fit possible culprit?)
- \*) "small mass gap + MET" topologies are dominant



# BACK TO THE PROTOMODEL

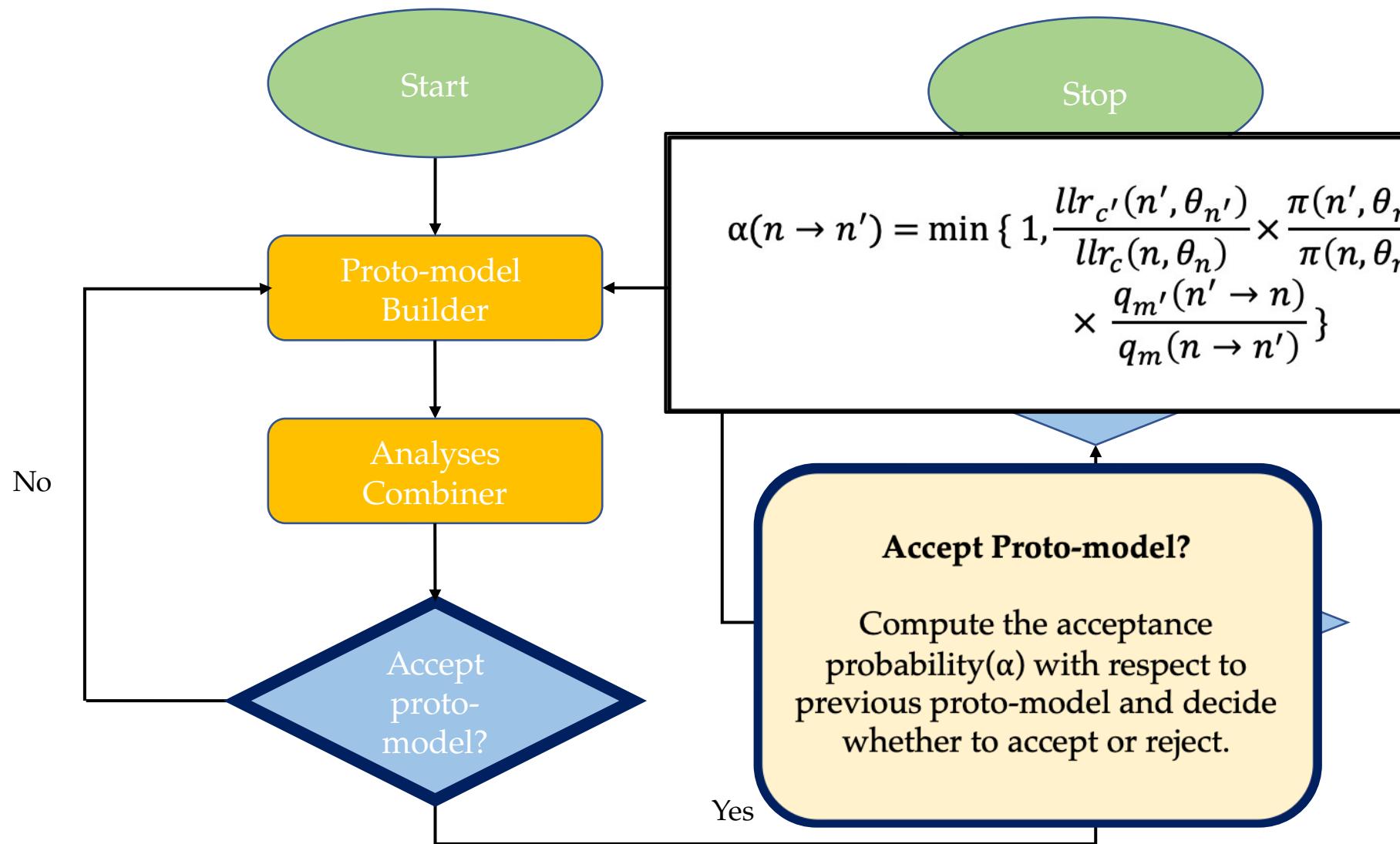


The builder was able to build a fairly simple protomodel that can contextualize a combination of 4 approximately independent results with a joint significance of 5.85 sigmas.

The critics could not kill the protomodel – to the contrary, they also see excesses. LHC results that apply that are not in our database (because lack of reinterpretation material) exist. A very rough inspection by naked eye suggests that there are more, compatible excesses.



# Algorithm Flowchart



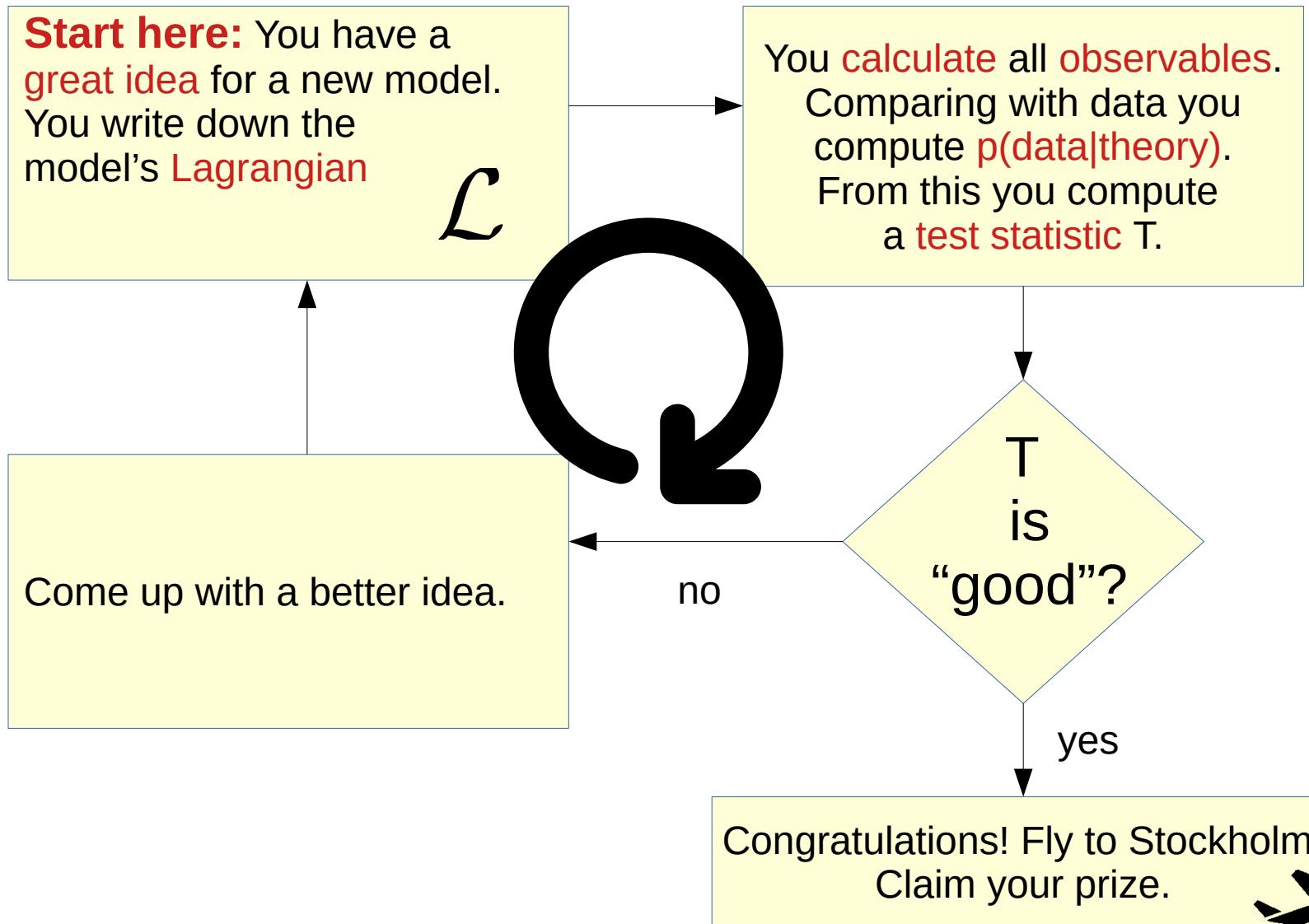
# OUR INVERSE PROBLEM

To connect our observations with contender theories for the Next Standard Model,

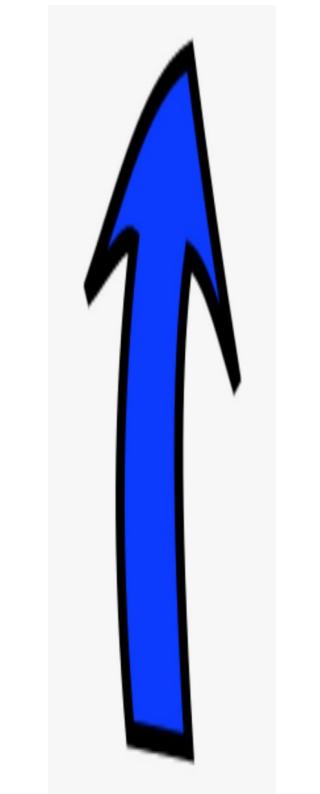
- In addition to lots of other information, we need statistical models (in one way or another)
- we can either start by:  
*postulating a theory, then fit its parameters*  
("top-down")
- or:  
*try to build our model directly from data*  
("bottom-up")

# Top-down versus bottom-up

**Top-  
Down:**



# Top-down versus bottom-up



**Bottom-  
Up:**

**Start here:** You describe your experimental findings in a language amenable to theoretical physics, e.g. simplified models for on-shell effects (“searches”), effective field theories for off-shell effects (“measurements”).

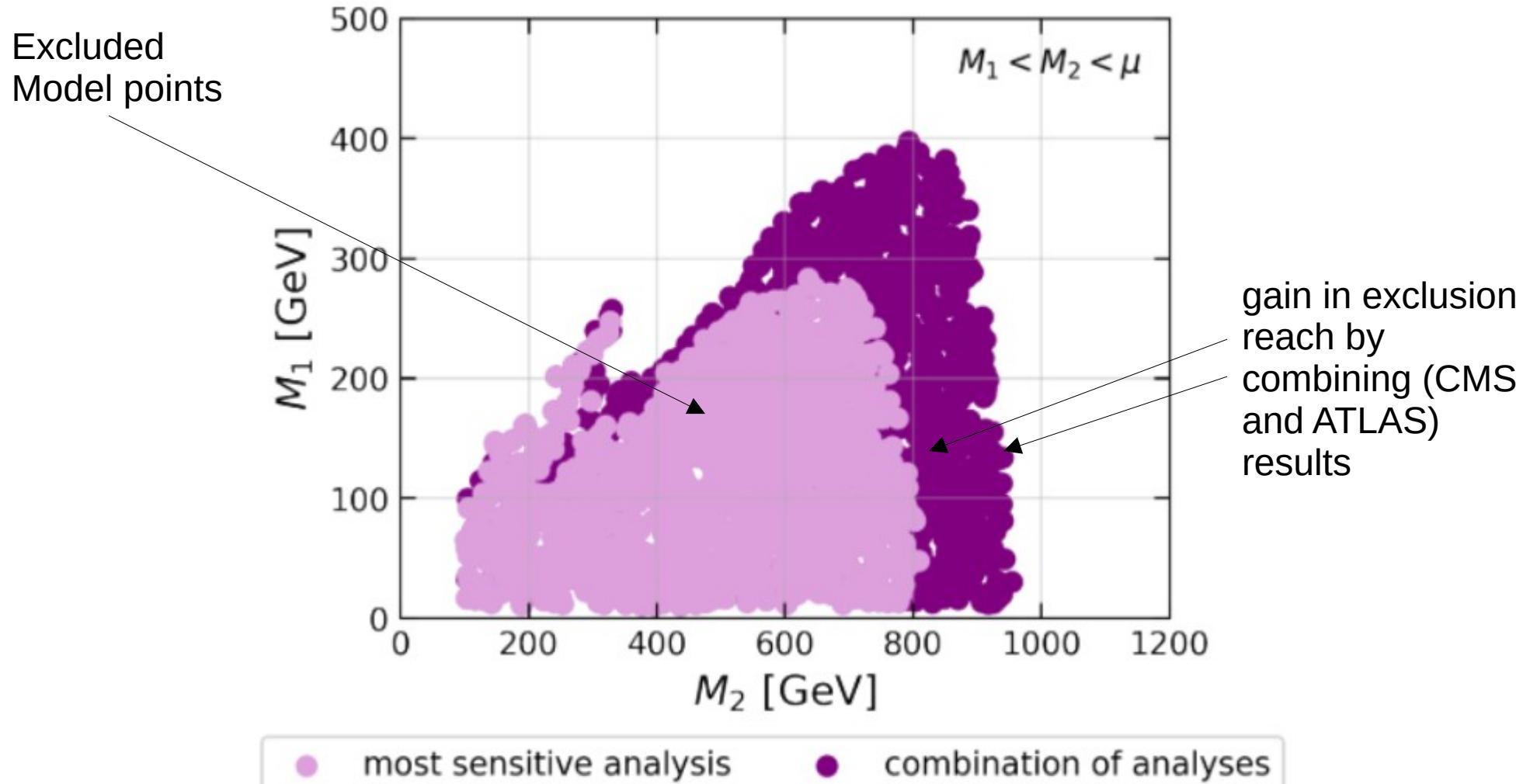
From the descriptions you try and construct precursor theories to the NSM that describe everything you really know about TeV-scale (and below) physics

Only now do you think about symmetries, gauge groups, etc that may underlie all observations. Construct your Lagrangian.

$\mathcal{L}$

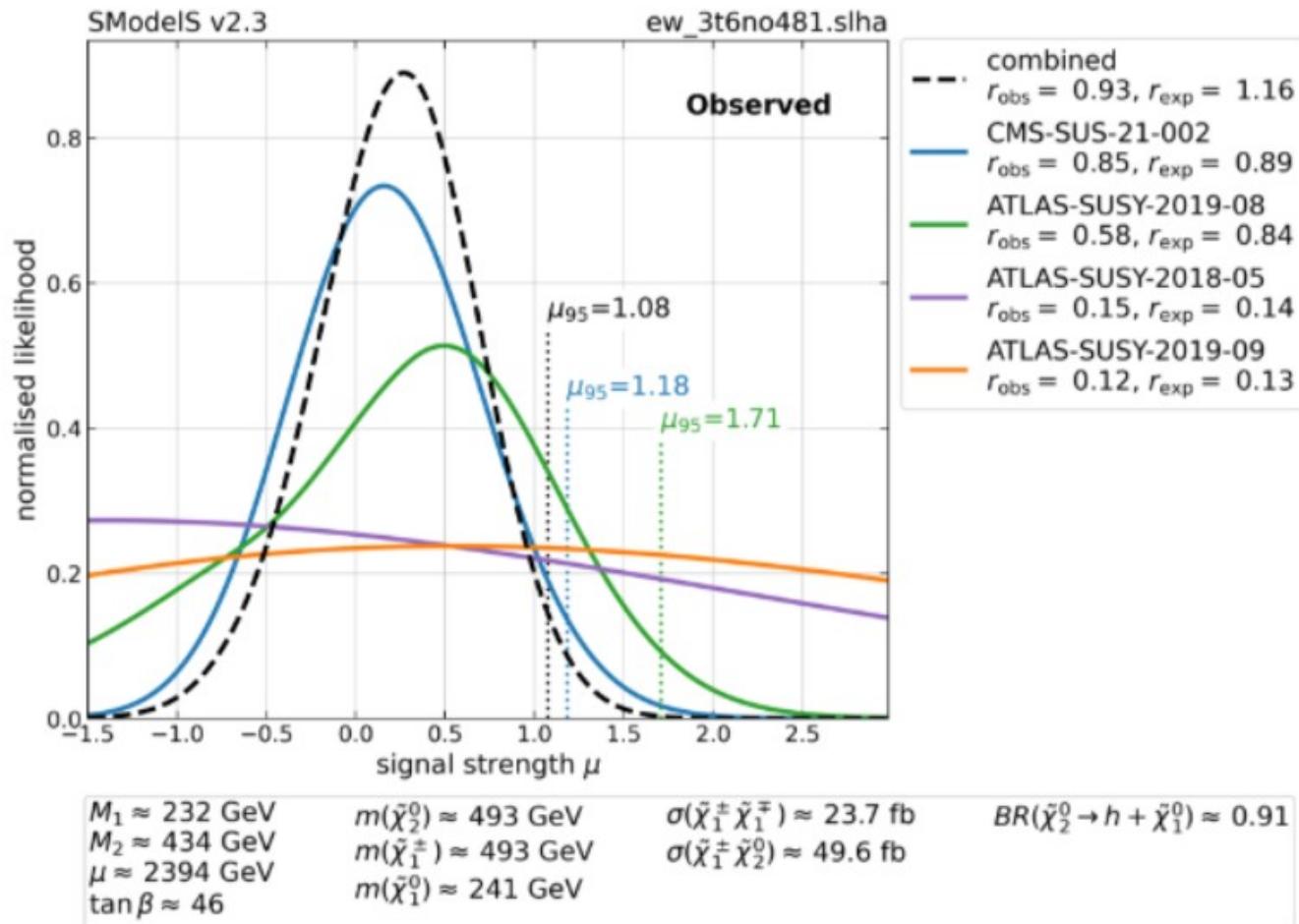
# Top-down: example

“Global LHC constraints on electroweakinos with SModelS v2.3”  
Model: MSSM, but looking at its charginos and neutralinos only



# Top-down: example

P3



Combination at work, for one model point. Likelihoods here are functions of a single parameter (the overall “signal strength”). Combining several likelihoods (colored lines) results in a narrower combined likelihood (dashed line).

# STATISTICAL MODELS IN THE SMODELS DATABASE

ATLAS, Run-2, Full Lumi only:

ID	Short Description	$\mathcal{L} [\text{fb}^{-1}]$	$\text{UL}_{\text{obs}}$	$\text{UL}_{\text{exp}}$	EM	comb.
ATLAS-EXOT-2018-48	di-top resonance	139.0	✓	✓		
ATLAS-EXOT-2019-03	dijet resonance	139.0	✓	✓		
ATLAS-SUSY-2018-04	2 hadronic $\tau$	139.0	✓		✓	PYHF
ATLAS-SUSY-2018-05	$2\ell + \text{jets}$ , EWK	139.0	✓		✓	PYHF
ATLAS-SUSY-2018-06	$3\ell$ , EWK	139.0	✓	✓	✓	
ATLAS-SUSY-2018-08	2 OS $\ell$	139.0	✓		✓	
ATLAS-SUSY-2018-09	2 SS $\ell$	139.0	✓			
ATLAS-SUSY-2018-10	$1\ell + \text{jets}$	139.0	✓		✓	
ATLAS-SUSY-2018-12	$0\ell + \text{jets}$	139.0	✓	✓	✓	
ATLAS-SUSY-2018-13	displaced jets	139.0			✓	SLv1
ATLAS-SUSY-2018-14	displaced vertices	139.0			✓	PYHF
ATLAS-SUSY-2018-16	$2 \text{ soft } \ell + \text{jets}$ , EWK	139.0	✓	✓	✓	PYHF
ATLAS-SUSY-2018-22	multi-jets	139.0	✓		✓	
ATLAS-SUSY-2018-23	$Wh(\gamma\gamma)$ , EWK	139.0	✓	✓		
ATLAS-SUSY-2018-31	$2 b\text{-jets} + 2 h$	139.0	✓		✓	PYHF
ATLAS-SUSY-2018-32	2 OS $\ell$	139.0	✓		✓	PYHF
ATLAS-SUSY-2018-40	$2 b\text{-jets} + 2 h$	139.0	✓	✓	✓	
ATLAS-SUSY-2018-41	hadr. EWK	139.0	✓	✓	✓	SLv1
ATLAS-SUSY-2018-42	charged LLPs, $dE/dx$	139.0	✓	✓	✓	
ATLAS-SUSY-2019-02	$2 \text{ soft } \ell$ , EWK	139.0	✓		✓	SLv1
ATLAS-SUSY-2019-08	$1\ell + h(bb)$ , EWK	139.0	✓		✓	PYHF
ATLAS-SUSY-2019-09	$3\ell$ , EWK	139.0	✓	✓	✓	PYHF

Type of statistical model

# STATISTICAL MODELS

- Only exclusions for particular models

If only “exclusion lines” are given, without upper limits, we can do nothing

- Observed 95% CL upper limits on production cross sections only:

cannot construct likelihood, binary decision “excluded” / “not-excluded” only (“critic”)

- Expected and observed 95% CL upper limits

can construct an approximate likelihood with truncated Gaussian,  
cannot combine topologies, very crude approximation

- Signal Efficiency “maps” for “simplified models”

can construct a likelihood as Gaussian (for the nuisances) \* Poissonian  
(for yields), can work per SR, and combine topologies in each SR [\*]

- Efficiency maps + correlation matrices

can combine signal regions via multivariate Gaussian \* Poissonians

- Efficiency maps + full likelihoods

full realism, correct statistical model

- Efficiency maps + full likelihoods + naming convention

+ knowledge about “data overlaps”

can combine publications at full realism

Likelihoods  
Combos



BETTER

[\*] if efficiency maps are not supplied, we can try to produce them with recasting frameworks

# HEPDATA



Search HEPData

Search

About

Submission Help

File Formats

Sign in

Browse all

Tumasyan, Armen et al.

Last updated on 2022-01-11 08:41 | Accessed 1621 times

Cite

JSON

Hide Publication Information

Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at  $\sqrt{s} = 13$  TeV

The CMS collaboration

Tumasyan, Armen , Adam, Wolfgang , Andrejkovic, Janik  
Walter , Bergauer, Thomas , Chatterjee, Suman , Dragicevic,  
Marko , Escalante Del Valle, Alberto , Fruehwirth, Rudolf ,  
Jeitler, Manfred , Krammer, Natascha

JHEP 11 (2021) 153, 2021.

<https://doi.org/10.17182/hepdata.106115.v2>

[Journal](#) [INSPIRE](#) [Resources](#)

## Abstract

A search is presented for new particles produced at the LHC in proton-proton collisions at  $\sqrt{s} = 13$  TeV, using events with energetic jets and large missing transverse momentum. The analysis is based on a data sample corresponding to an integrated luminosity of  $101 \text{ fb}^{-1}$ , collected in 2017–2018 with the CMS detector. Machine learning techniques are used to define separate categories for events with narrow jets from initial-state radiation and events with large-radius jets consistent with a hadronic decay of a W or Z boson. A statistical

[Download All](#) ▾  
Version 2 ▾

[Filter 55 data tables](#)

Background prediction and observed data yields in the signal region bins. The background yields are obtained from the background-only fit...

Simplified likelihood:  
covariance matrix  
(Monojet)

Supplementary material  
[10.17182/hepdata.106115.v2/t14](https://doi.org/10.17182/hepdata.106115.v2/t14)  
Matrix of covariance coefficients between signal region bins. The coefficients are obtained from the background-only fit to the control regions,...

[Simplified likelihood: Yields \(Monojet\)](#) ▾

Supplementary material  
[10.17182/hepdata.106115.v2/t15](https://doi.org/10.17182/hepdata.106115.v2/t15)  
Background prediction and observed data yields in the signal region bins. The background yields are obtained from the background-only fit...

cmenergies

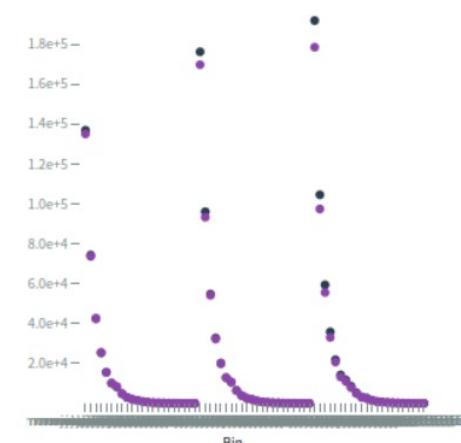
13000.0

Showing 50 of 66 values

Show All 66 values

Bin	Background yield	Data yield
monojet_2016_bin0	135004.8633	136865
monojet_2016_bin1	73681.17188	74340
monojet_2016_bin2	42448.17627	42540
monojet_2016_bin3	25541.23718	25316
monojet_2016_bin4	15454.10522	15653
monojet_2016_bin5	10162.67578	10092
monojet_2016_bin6	8473.070068	8298
monojet_2016_bin7	4853.264771	4906
monojet_2016_bin8	2960.103455	2987
monojet_2016_bin9	1906.687012	2032
monojet_2016_bin10	1498.361874	1514

## Visualize



Sum errors  Log Scale (Y)

Deselect variables or hide different error bars by clicking on them.

(CMS monojet search)

# SERIALIZED STATISTICAL MODELS

pyhf serializes the models into json files:

```
{  
    "channel": {  
        "type": "object",  
        "properties": {  
            "name": { "type": "string" },  
            "samples": { "type": "array", "items": {"$ref": "#/definitions/sample"}, "minItems": 1 }  
        },  
        "required": ["name", "samples"],  
        "additionalProperties": false  
    },  
}
```

Posterior distributions on nuisance parameters (such as jet energy scales, pdf uncertainties, reconstruction inefficiencies, etc etc) are described via “auxiliary measurements” → keeping things frequentist (if we wish)!

Statistics Standardization committee formed: **HS3**

<https://github.com/hep-statistics-serialization-standard/hep-statistics-serialization-standard>

These “closed world” approaches are general enough for users!

# SIMPLIFICATIONS AND INTEROPERABILITY

An ecosystem of tools around serialized statistical models will be crucial:

Currently we have:

**Simplify:** simplifies pyhf likelihoods, collapses nuisances.

<https://pypi.org/project/simplify/>

Seems to be working with only minor quirks, we are making use of it.

**Combine2pyhf:** converts combine datacards to pyhf and back

<https://github.com/kskovpen/combine2pyhf>

Orphaned, does not handle all cases.

**Pyhf-to-combine converter:**

<https://github.com/peterridolfi/Pyhf-to-Combine-converter>

Student project. Promising but incomplete.

# IN “USER SPACE”: MACHINE LEARNING SURROGATE MODELS

We are in the process of using surrogate models in SModelS (scheduled for SModelS v3.2.0)

If only likelihood is needed: simple, fully connected neural networks (Bayesian neural networks for adding errors on the estimates), e.g.:

CERN-TH-2019-187

## The DNNLikelihood: enhancing likelihood distribution with Deep Learning

Andrea Coccato<sup>a</sup>, Maurizio Pierini<sup>b</sup>, Luca Silvestrini<sup>b,c</sup>, and Riccardo Torre<sup>a,b</sup>

<sup>a</sup> INFN, Sezione di Genova, Via Dodecaneso 33, I-16146 Genova, Italy

<sup>b</sup> CERN, 1211 Geneva 23, Switzerland

<sup>c</sup> INFN, Sezione di Roma, P.le A. Moro, 2, I-00185 Roma, Italy

If full statistical model including toys is needed as well: generative network such as normalizing flows. e.g.:

SciPost Physics Submission

## The NFLikelihood: an unsupervised DNNLikelihood from Normalizing Flows

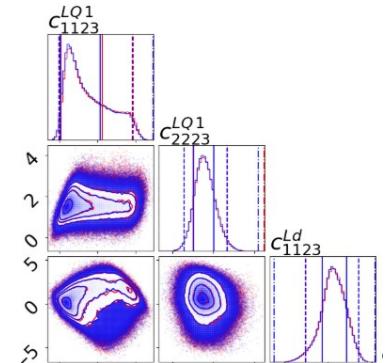
Humberto Reyes-González<sup>1,2,\*</sup> and Riccardo Torre<sup>2,†</sup>

**1** Department of Physics, University of Genova, Via Dodecaneso 33, 16146 Genova, Italy

**2** INFN, Sezione di Genova, Via Dodecaneso 33, I-16146 Genova, Italy

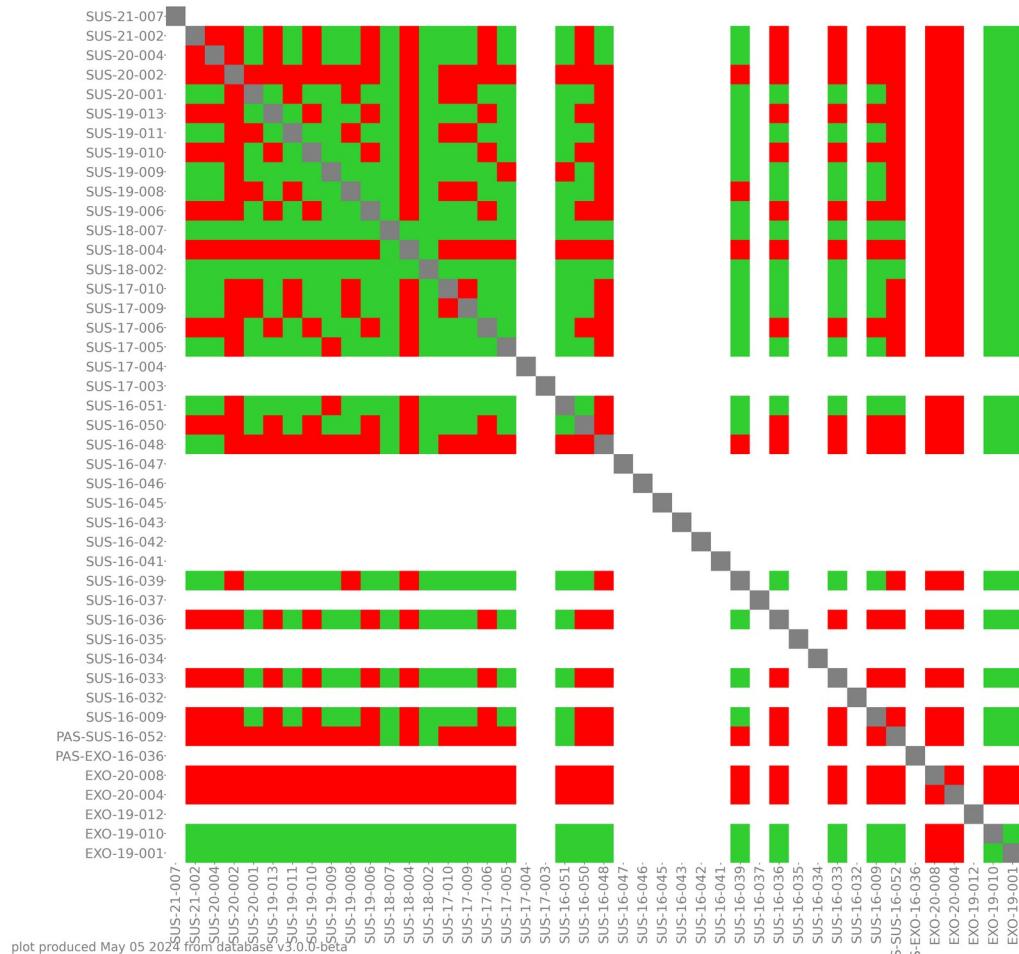
\* [humberto.reyes@rwth-aachen.de](mailto:humberto.reyes@rwth-aachen.de)

† [riccardo.torre@ge.infn.it](mailto:riccardo.torre@ge.infn.it)



# COMBINABILITY MATRIX, AND “RECASTS”

CMS, 13 TeV



- We keep track of which pairs of analyses have no overlaps in what LHC collision events enter the signal regions. In the “pheno community”, we combine only if no overlap.
- Currently most reliable source for the combinations matrix: **analysis recast frameworks**: generating events, track which signal regions are populated by which events
- Requires detailed documentation about the analysis
- Experimental collaborations could in principle produce this information as well (and better)

Long term vision: a naming convention is established, making it possible to correctly combine stats models!

Strength in numbers: Optimal and scalable combination of LHC new-physics searches

Jack Y. Araz, Andy Buckley, Benjamin Fuks, Humberto Reyes-Gonzalez, Wolfgang Waltenberger, Sophie L. Williamson, Jamie Yellen  
SciPost Phys. 14, 077 (2023) · published 20 April 2023

doi: 10.21468/SciPostPhys.14.4.077

pdf

BiBTeX

RIS

Submissions/Reports

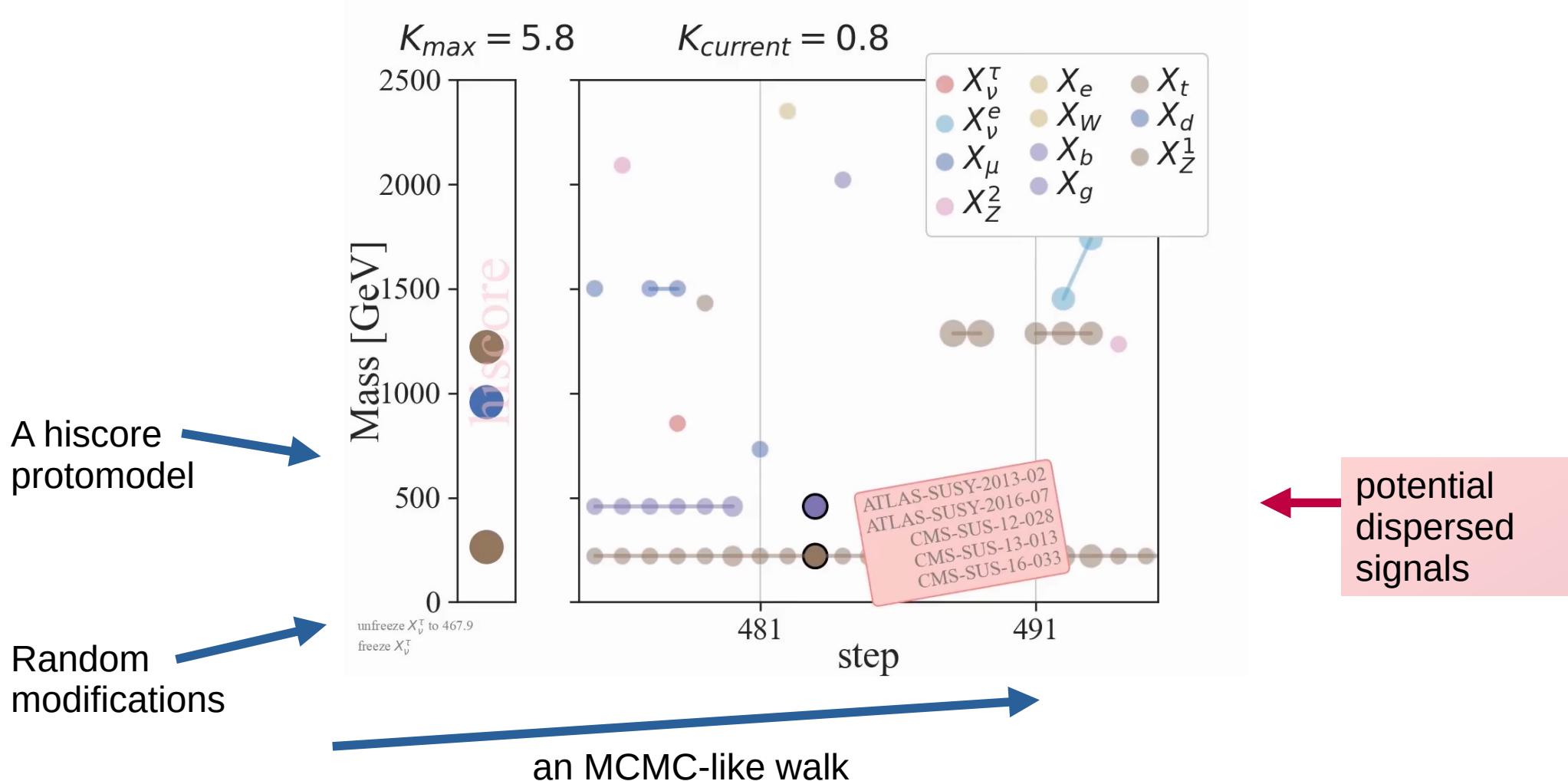
Check for updates

# REPRODUCING AN ANALYSIS

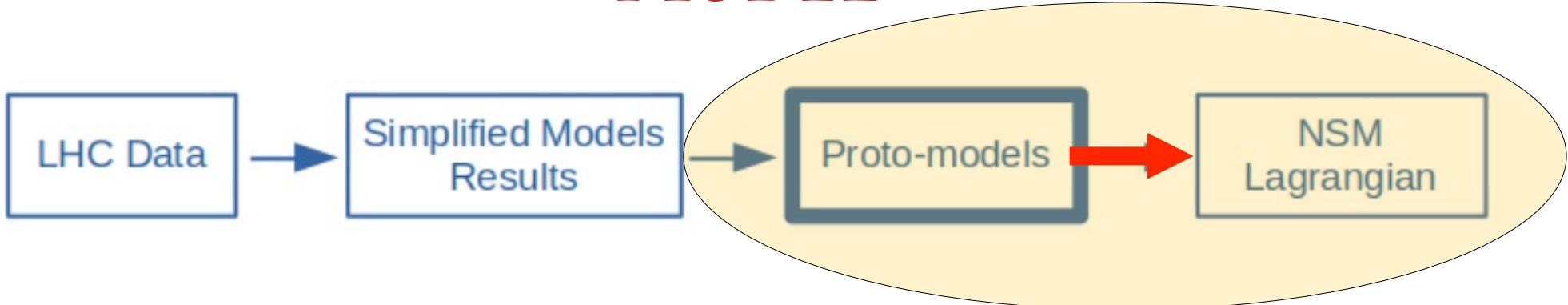
In order to facilitate reproduction of an entire analysis – to e.g. apply it to a new model, or to study analysis overlap (previous slide), the analysis needs to be documented in great detail:

- **Cutflow tables** of the analysis cuts, **reconstruction object efficiencies**, machine-learning models (e.g. as onnx files) need to be given in digitized form, for e.g. recasts, studies of analysis overlaps.
- Implementation in a simple framework / language: e.g. **SimpleAnalysis** or Analysis Description Language (**ADL**)
- For Monte Carlo samples used in publication: **generator datacards**, etc
- Upper limit plots, **signal efficiency maps** (of simplified models) **in digitized form**
- **Statistical models in reusable digitized form**

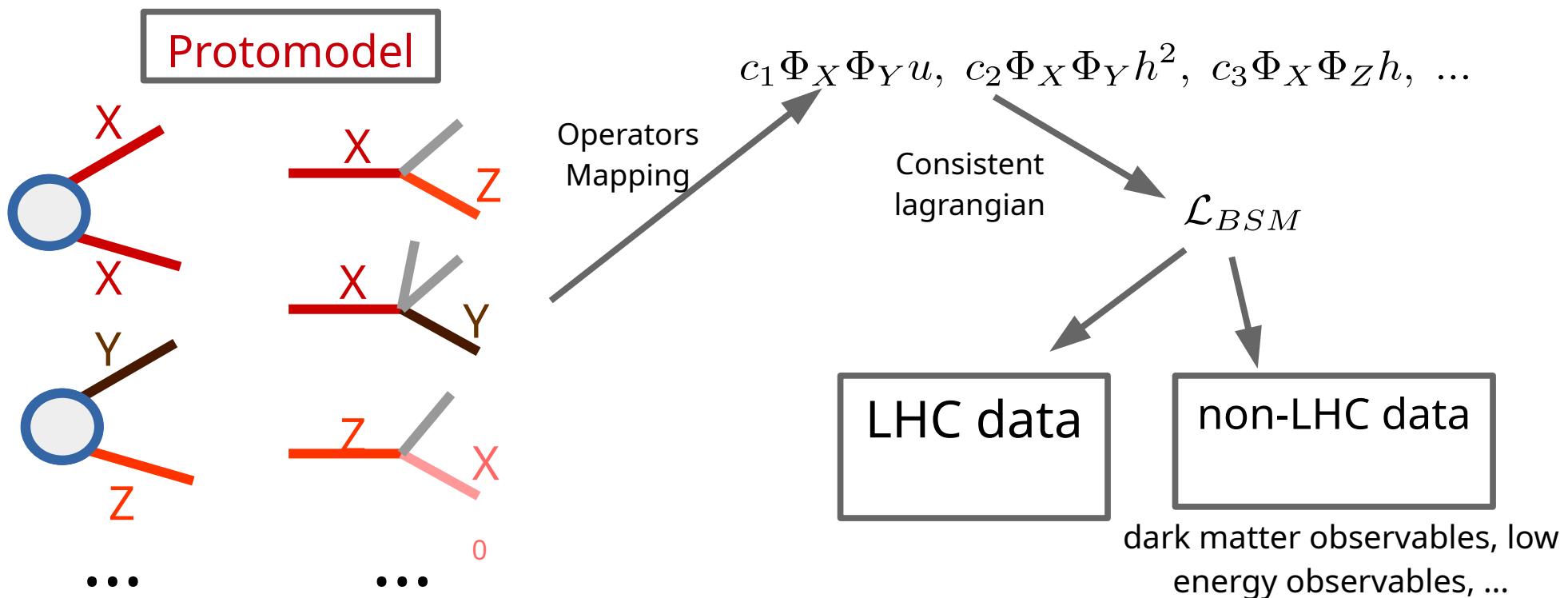
# EXPERIMENTAL RESULTS → PROTOMODELS



# PROTOMODELS → NEXT STANDARD MODEL



Work from protomodels to UV complete theories has begun  
(John Gargalionis, PostDoc in Valencia) – lot's of combinatorics!



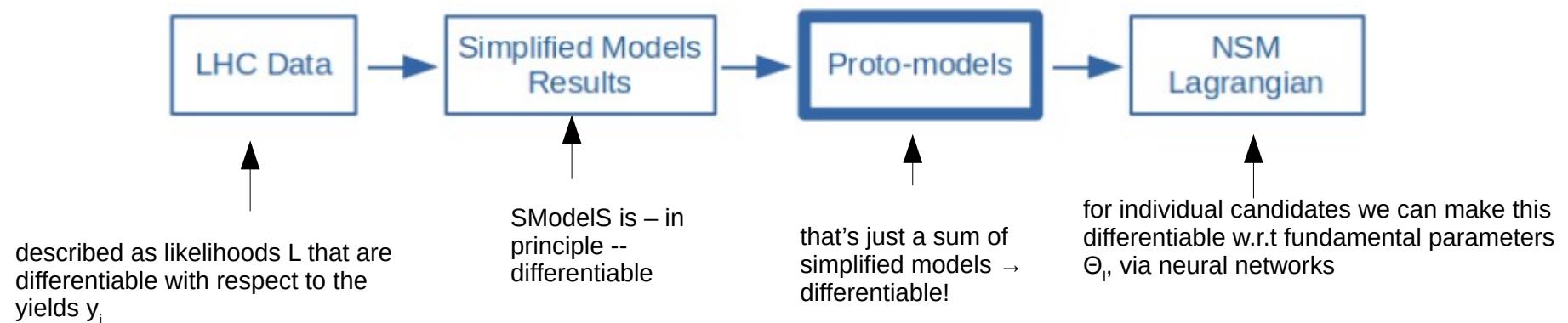
# SUMMARY

- Ideally, **all important aspects of a physics result get published**: cutflow tables, reconstruction object efficiencies, machine learned models, generator datacards, for SModelS: efficiency maps and: **serialized statistical models!**
- **HEPdata** is the collider physics results platform of choice
- **DOIs** for all digital objects
- **Global fits** are still an indispensable tool to obtain a “big picture” (see also gambit’s talk)
- The SModelS collaboration recently started to **automate model building based on the statistical models**

# DIFFERENTIABLY IF POSSIBLE



If we had gradients we could perform gradient descent to find the best model, and we could use e.g. the Fisher information to infer the error on its parameters (or, alternatively we can then MCMC-sample).



$$\frac{\partial L}{\partial \theta_l} = \frac{\partial L}{\partial y_i} \cdot \frac{\partial y_i}{\partial p_j} \cdot \frac{\partial p_j}{\partial(m_k, \Gamma_k, \sigma_k)} \cdot \frac{\partial(m_k, \Gamma_k, \sigma_k)}{\partial \theta_l}$$

**Not yet required** (theory space as well as space of measurements are still low-dimensional enough). Action item for now: keep choosing, supporting, developing technologies that support autograd