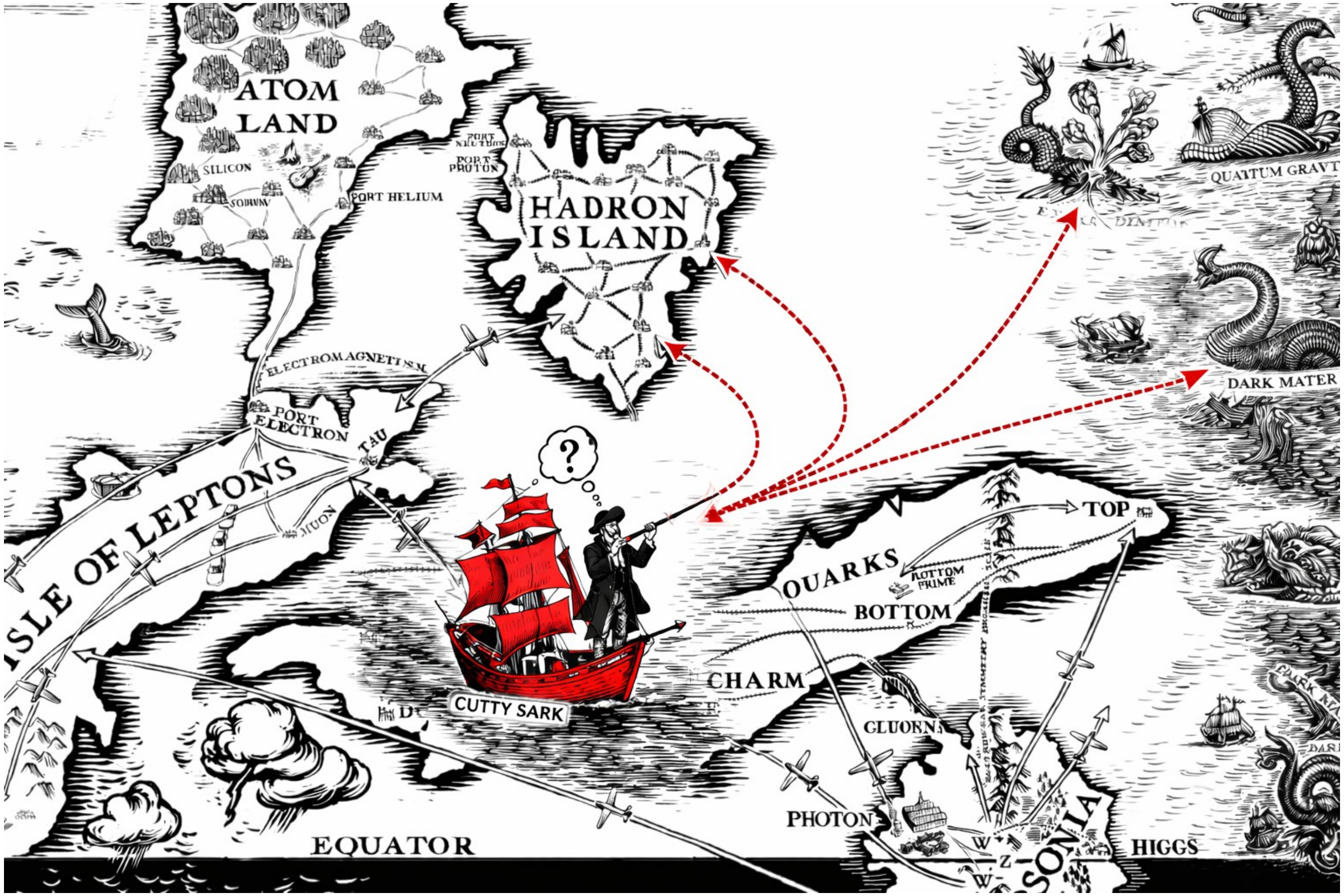
LHC Searches and the Next Standard Model: A Meta-Statistical View

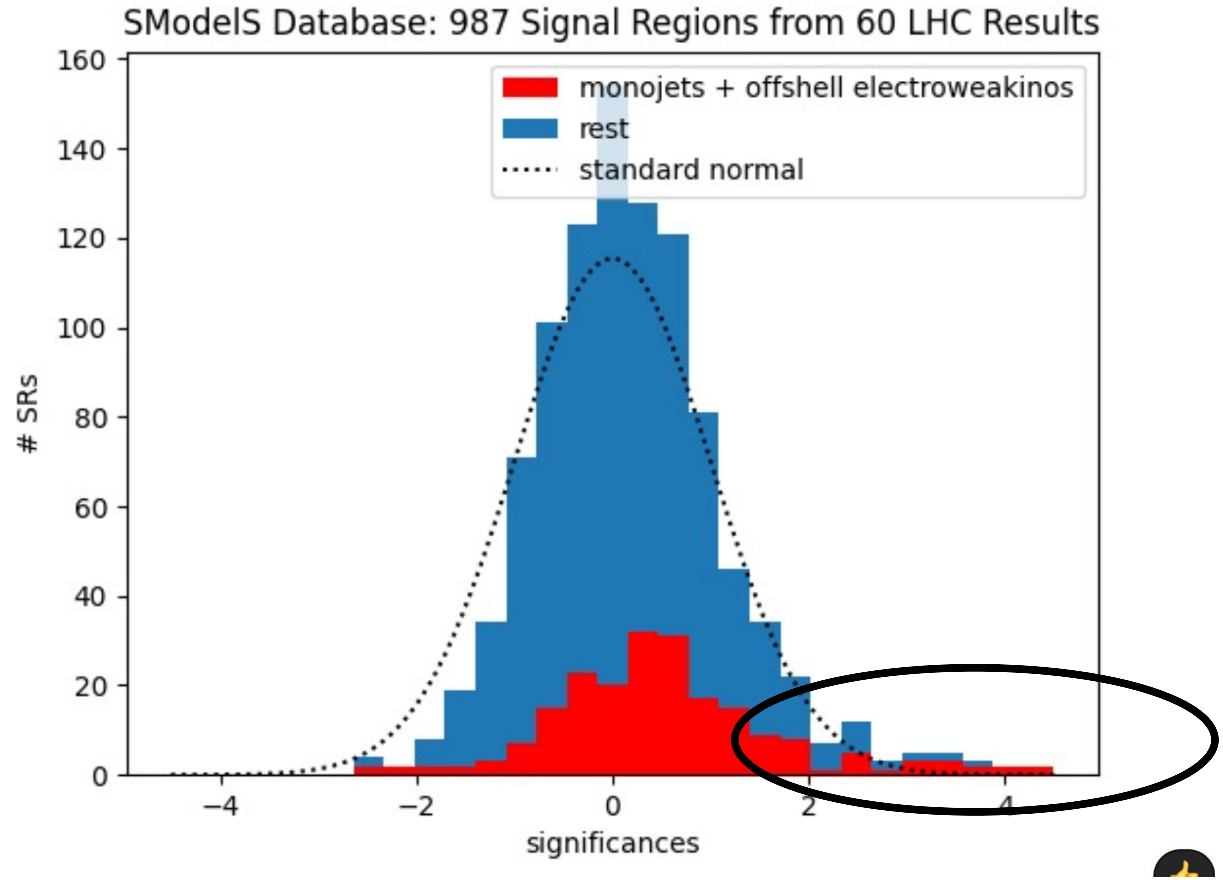


Wolfgang Waltenberger (MBI and Uni Wien)

Imperial College London, Fundmental Physics Seminar, January 2026

image courtesy of Jon Butterworth, Chris Wormell and chatgpt

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:

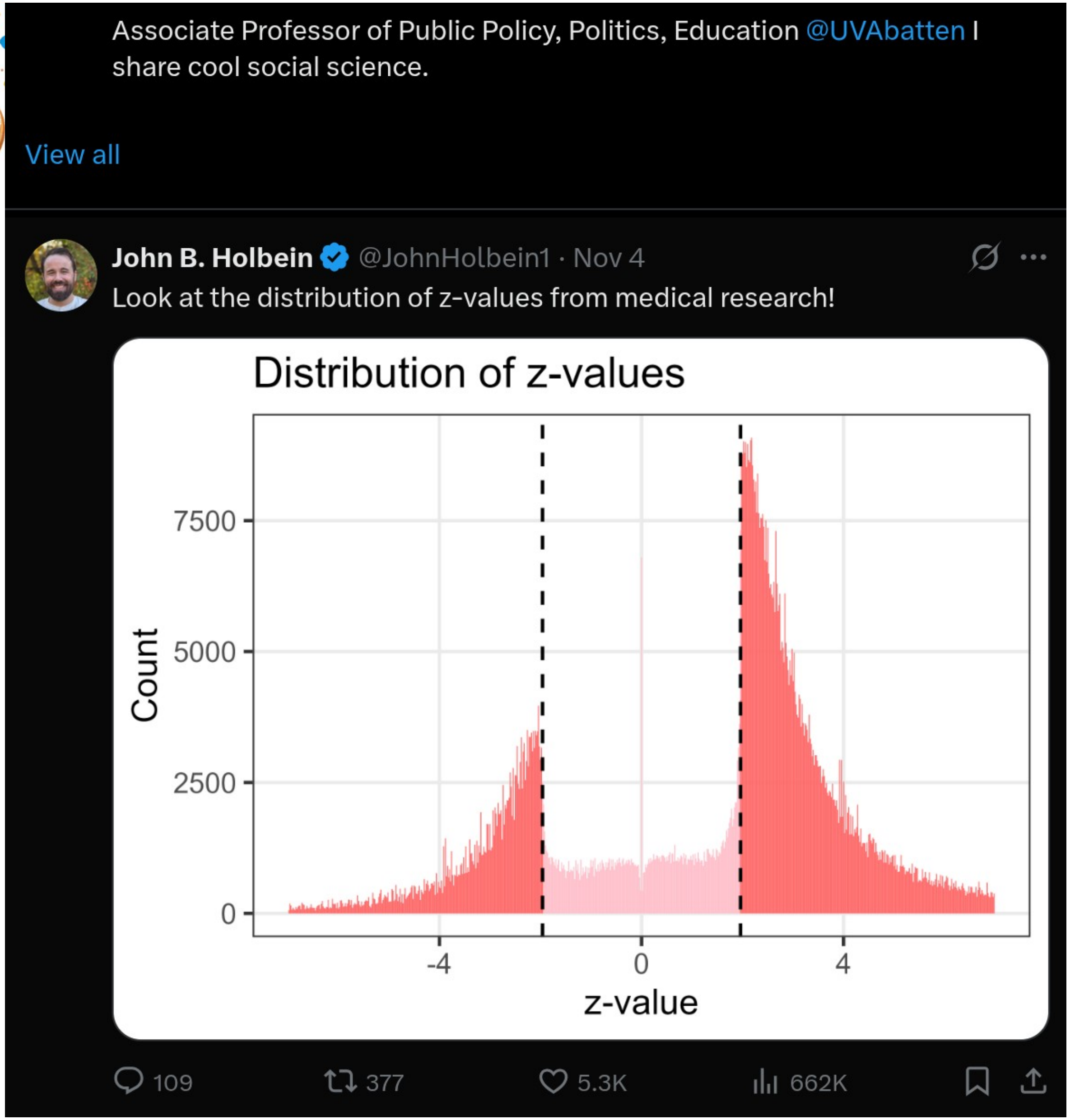


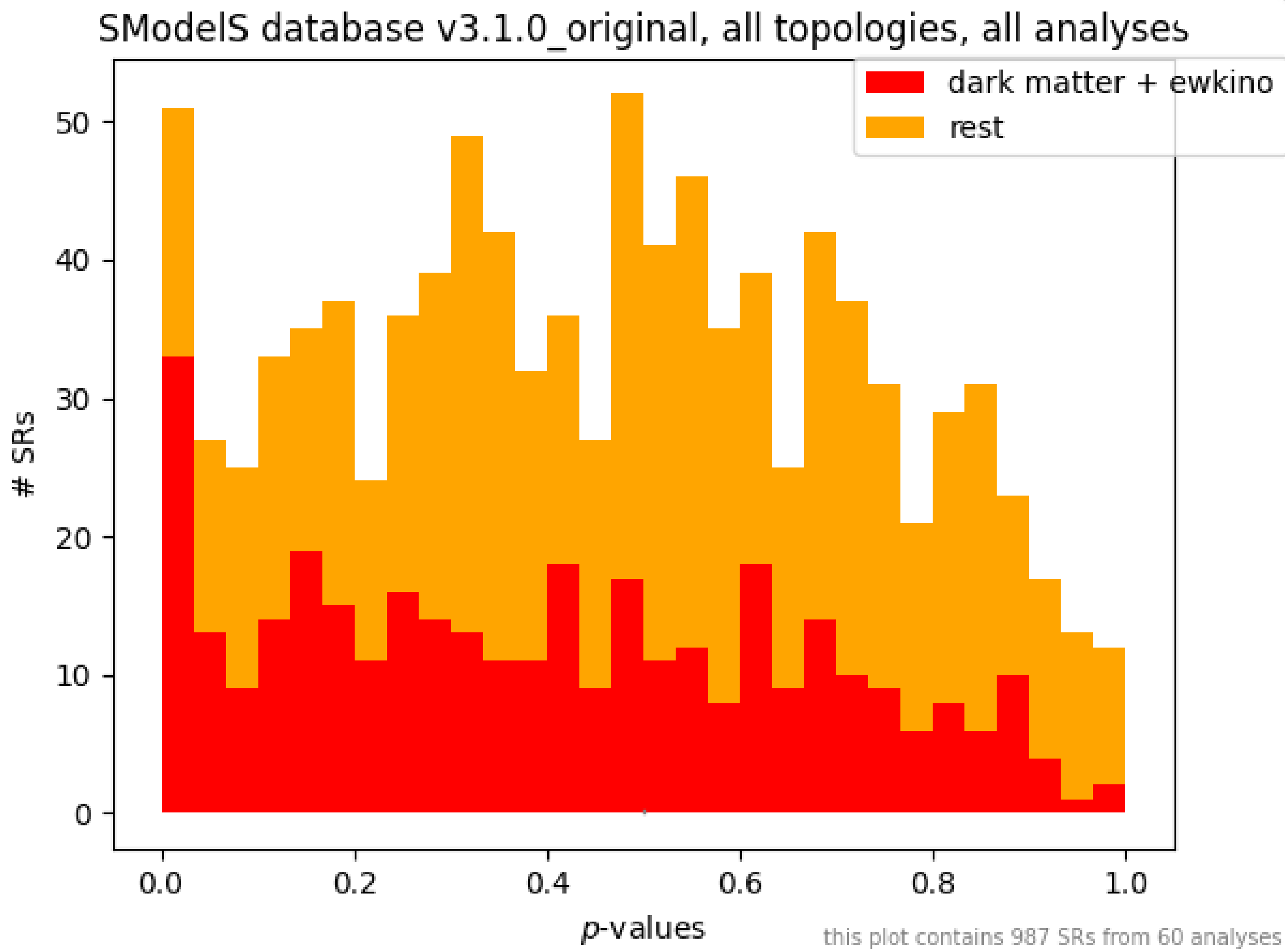
Distribution of p-values of all the signal regions we have in our database, Standard Model hypothesis. If no new physics is in the data, it should follow a standard normal (dashed black). Excesses are at + infinity

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.

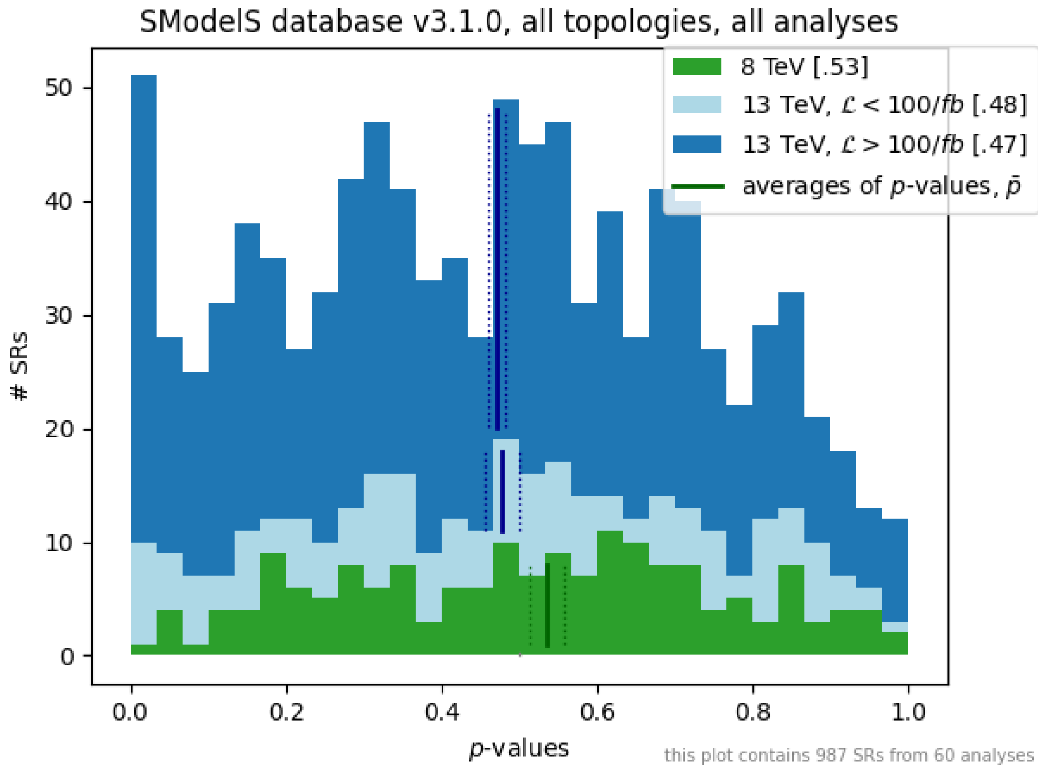
(a similar plot posted
For medical research.

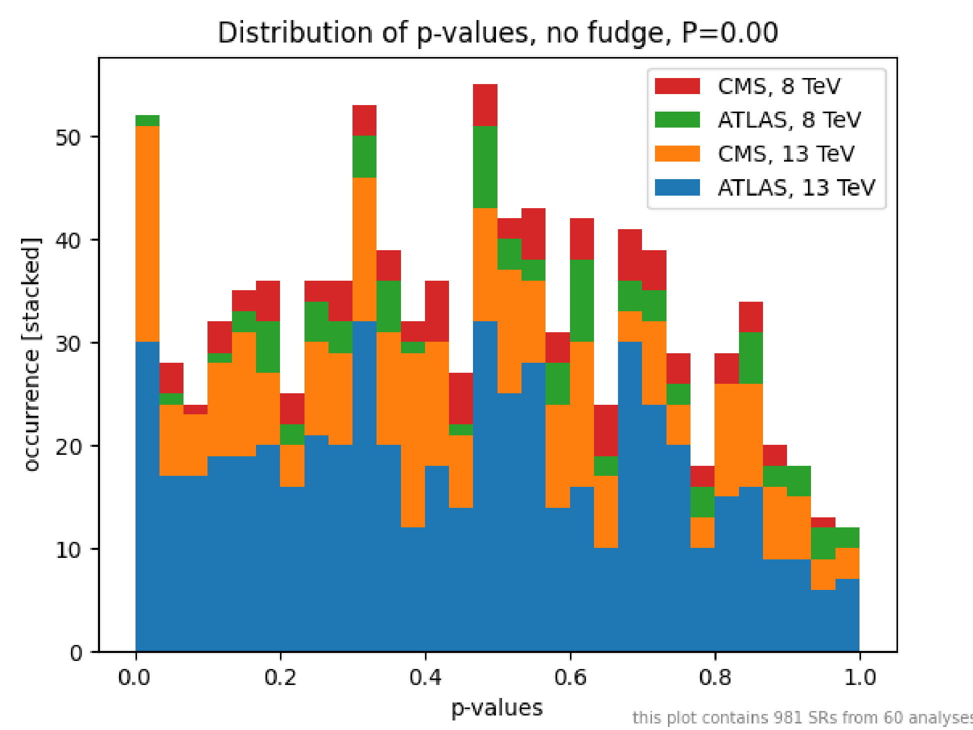
No offense meant, medical researchers, we are very happy you are doing all the hard research)





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


SModelS database v3.1.0, all topologies, all analyses
8 TeV [.53]
13 TeV, $\mathcal{L} < 100/fb$ [.48]
13 TeV, $\mathcal{L} > 100/fb$ [.47]
averages of p-values, $\bar{p}$
SRs
0
10
20
30
40
50
0.0
0.2
0.4
0.6
0.8
1.0
p-values
this plot contains 987 SRs from 60 analyses

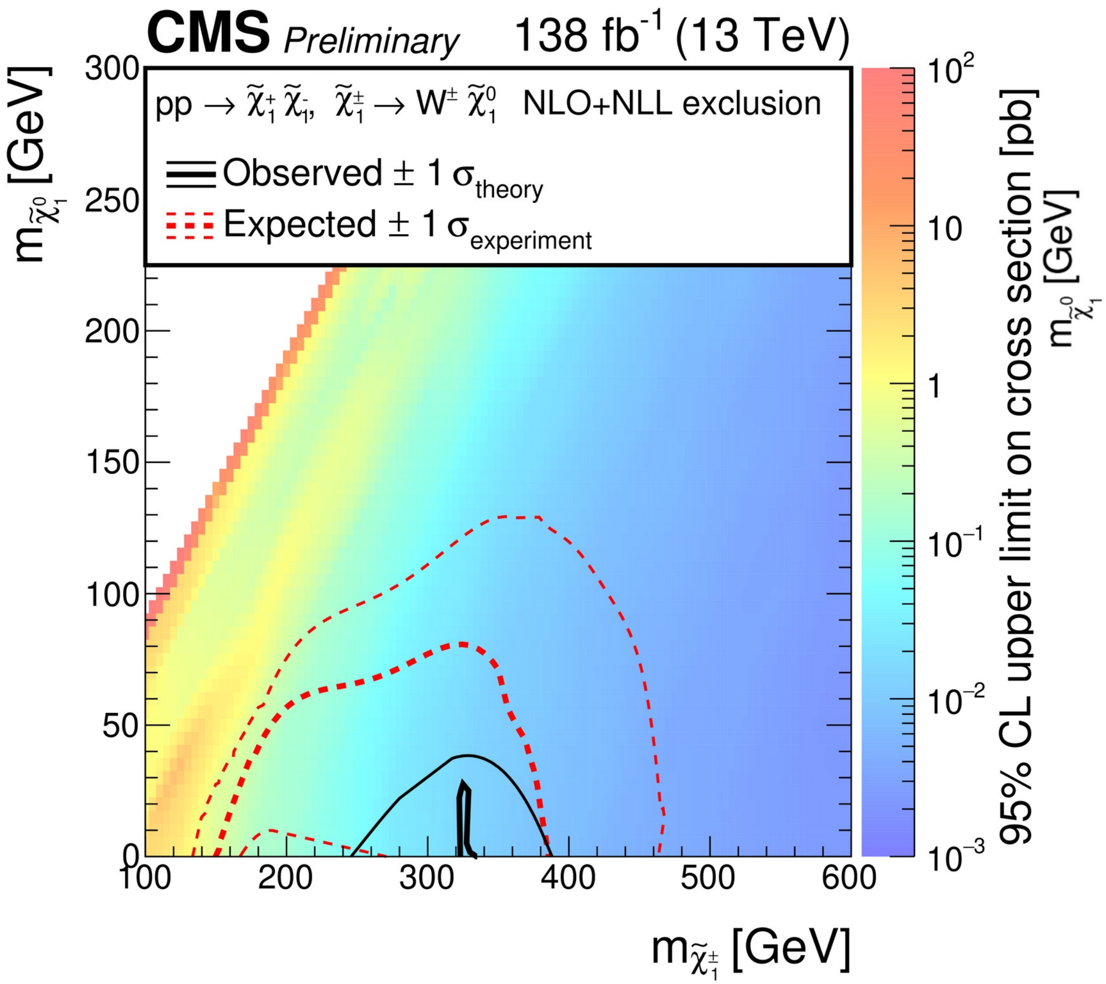

Distribution of p-values, no fudge, P=0.00
CMS, 8 TeV
ATLAS, 8 TeV
CMS, 13 TeV
ATLAS, 13 TeV
occurrence [stacked]
0
10
20
30
40
50
0.0
0.2
0.4
0.6
0.8
1.0
p-values
this plot contains 981 SRs from 60 analyses

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

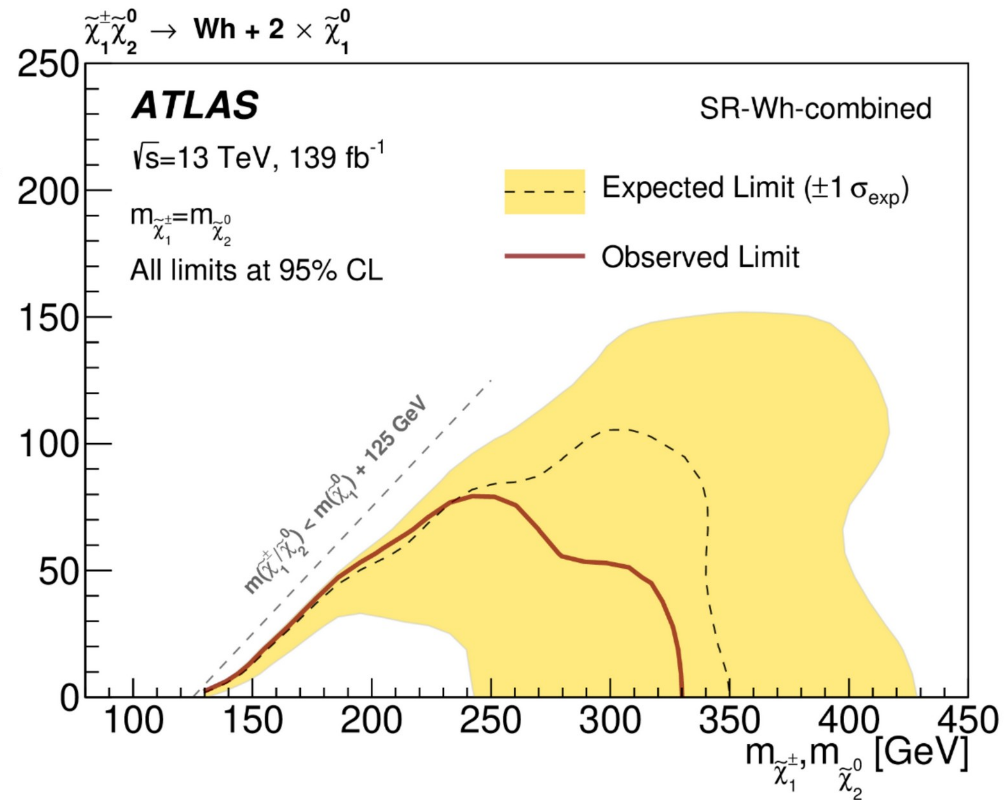
Results of about 50% of electroweakino/dark matter searches are not In the database (either not yet, or technical obstacles).

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

Results Not (yet) in SModelS

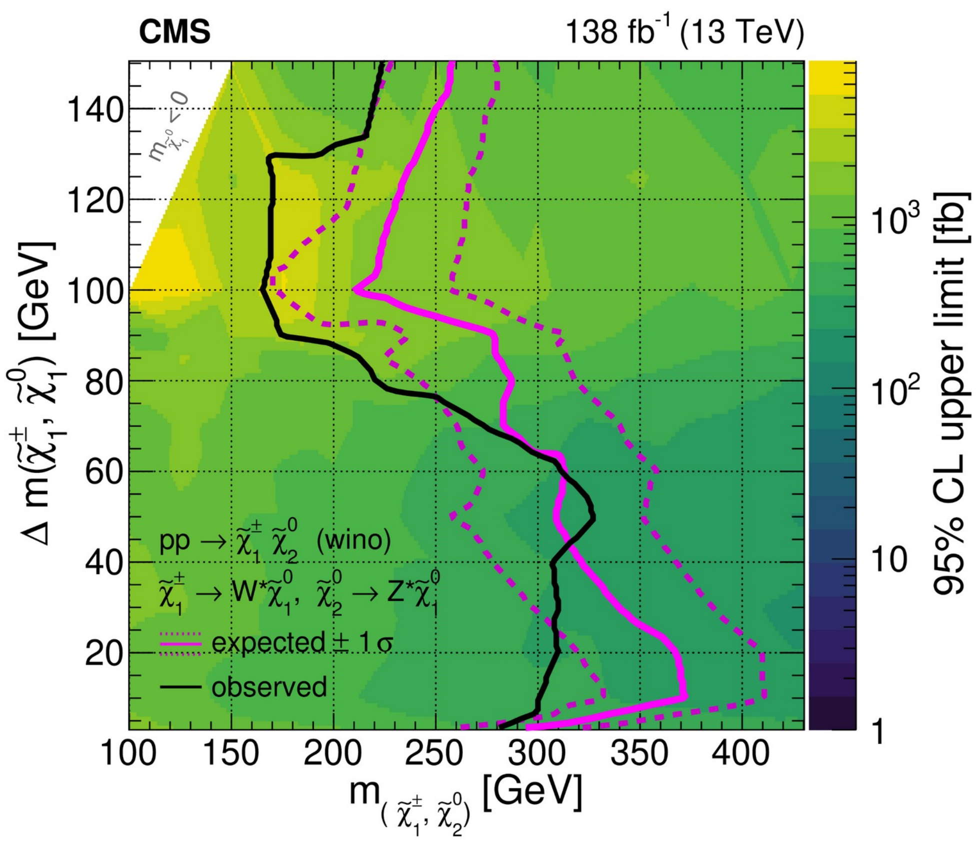


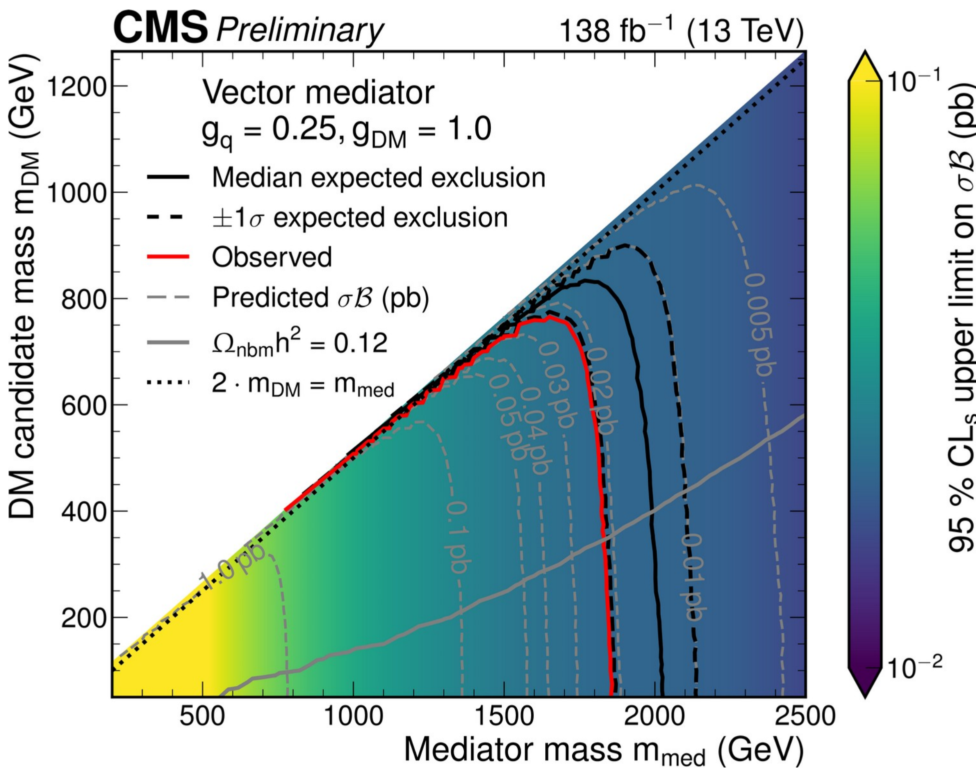
CMS-SUS-23-002, chargino pair production



ATLAS-SUSY-2023-03,
Charginos + neutralinos

Results Not (yet) in SModelS

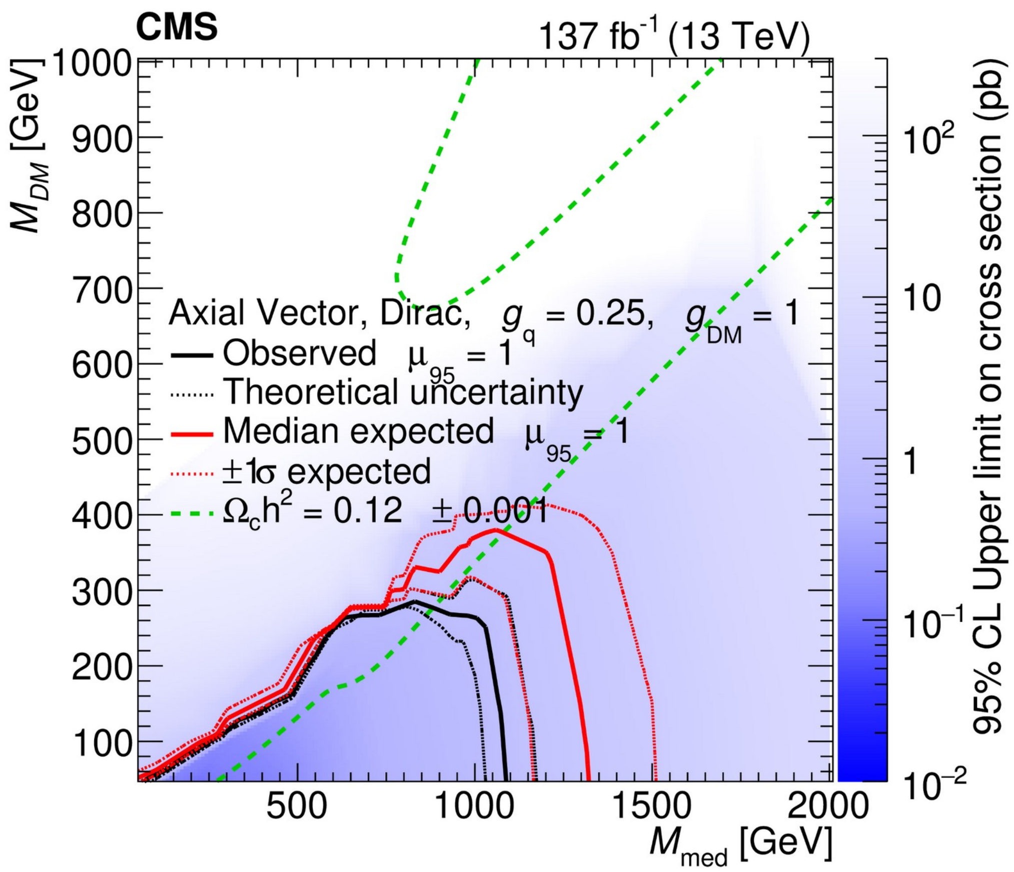




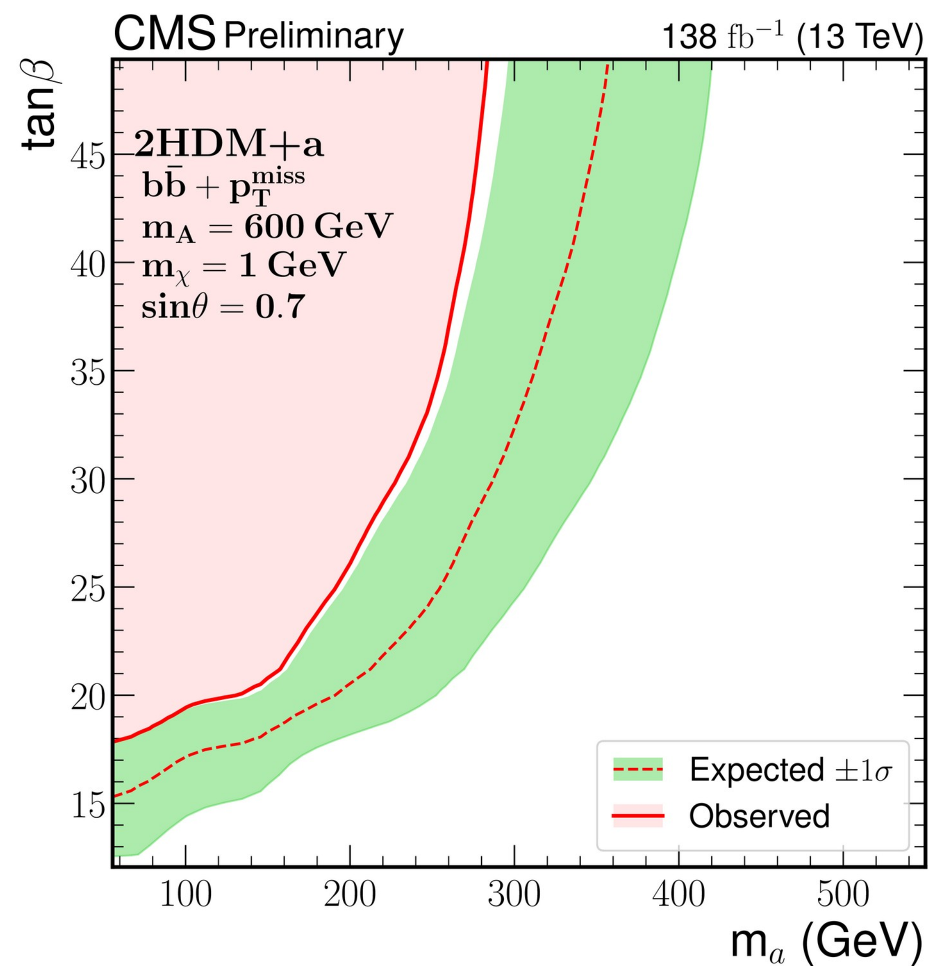
CMS-SUS-23-003, chargino neutralino production

CMS-SUS-23-004, mono-top

Results Not (yet) in SModelS

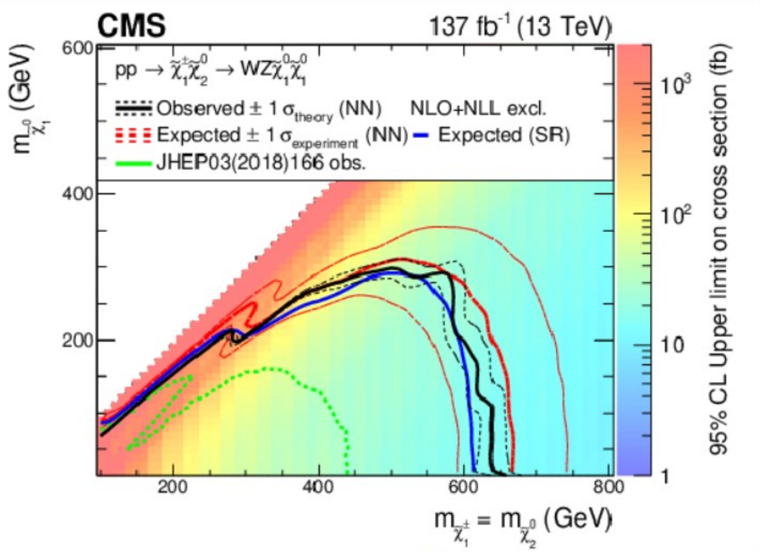


CMS-SUS-23-016, MET + gamma

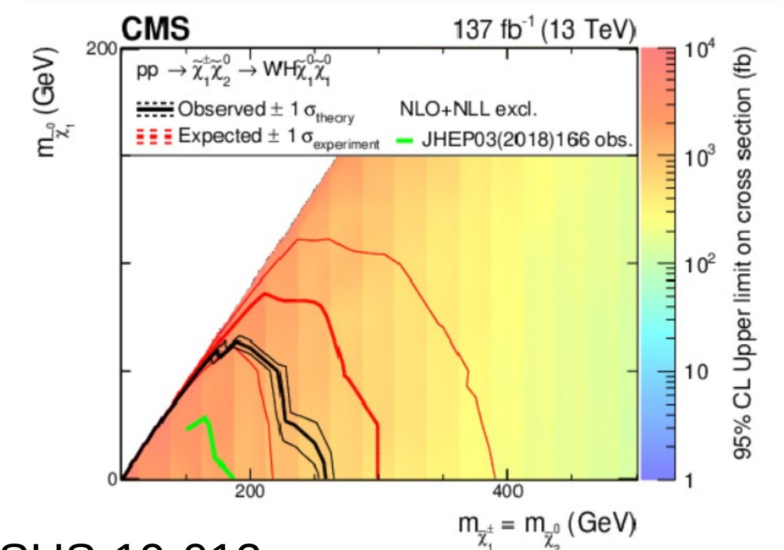


CMS-SUS-23-008, MET+bb

Results Not (yet) in SModelS

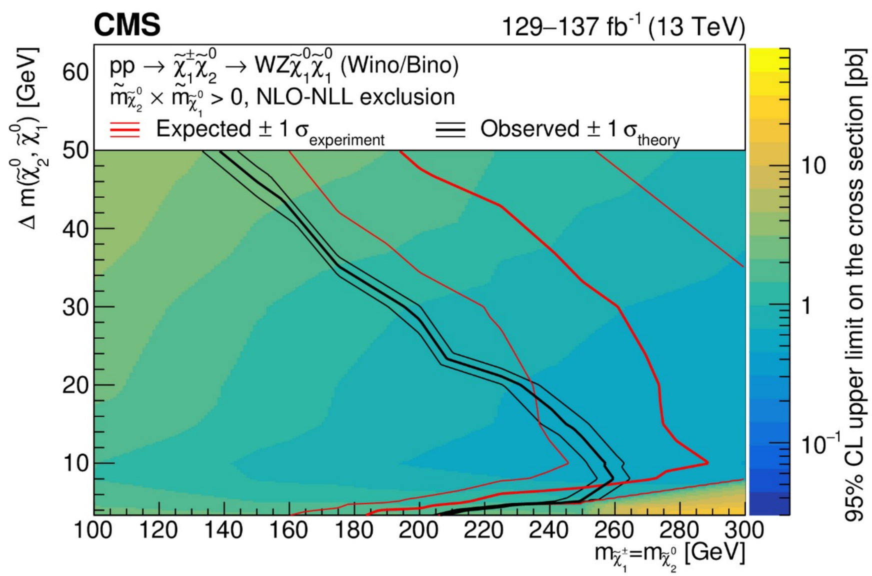


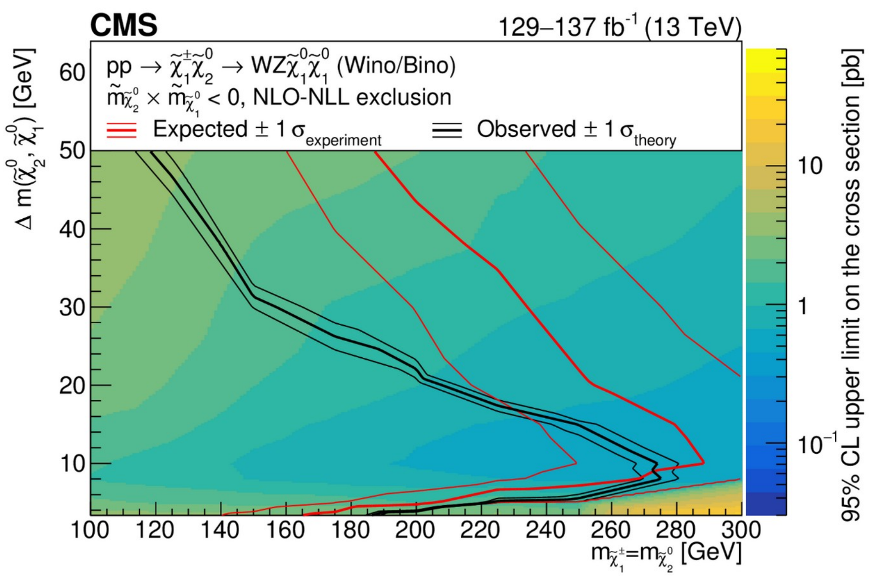
png pdf



png pdf

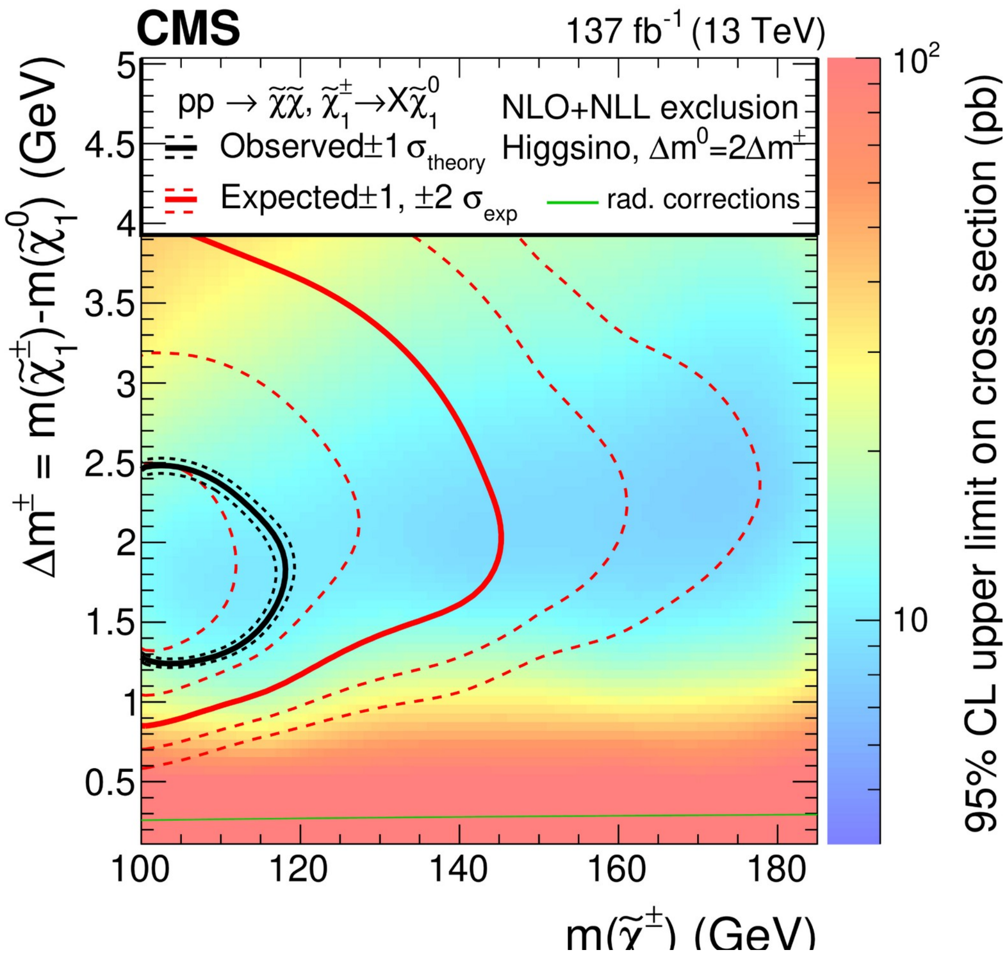
CMS-SUS-19-012, ewkinos

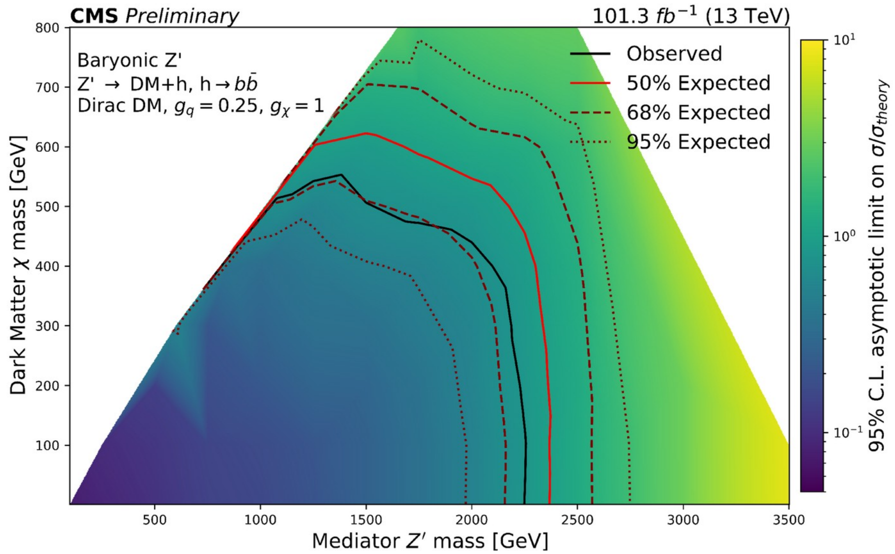




CMS-SUS-18-004, ewkinos

Results Not (yet) in SModelS





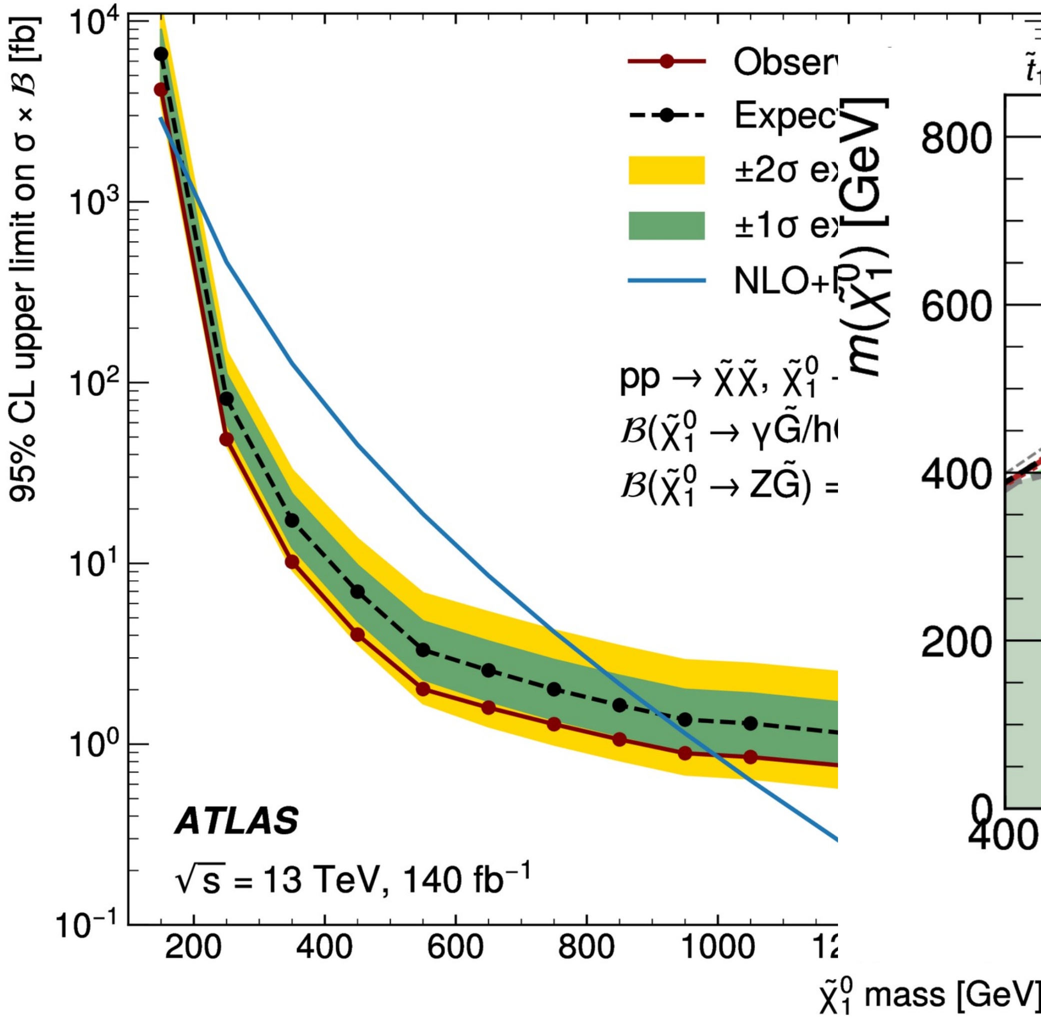
CMS-SUS-24-003

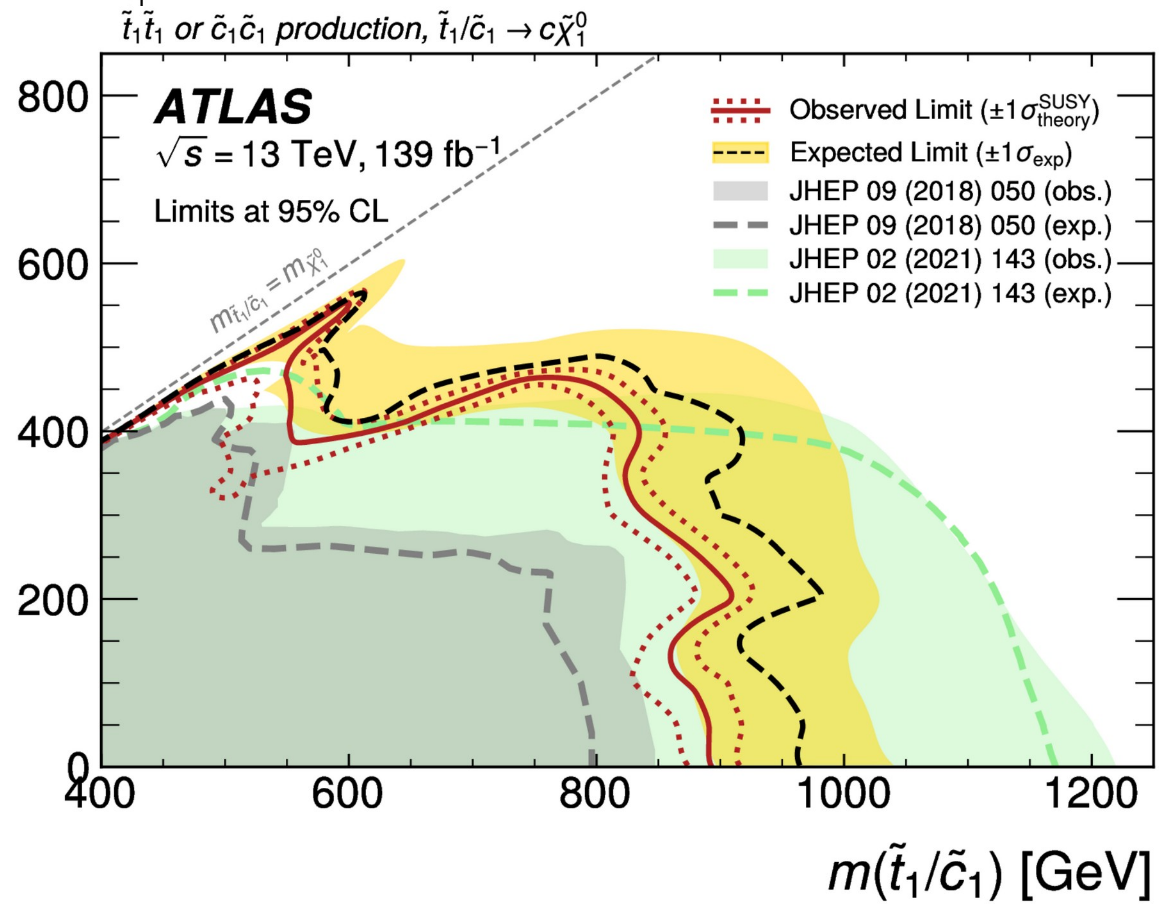
Ewkinos, small mass splitting

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

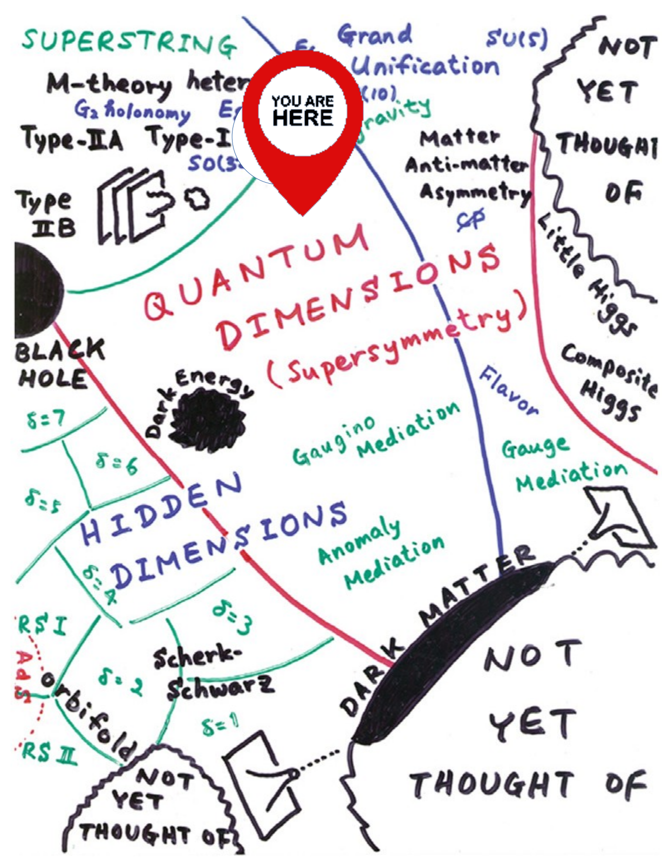
It is crucial for theorists to be conservatively radical:

*) 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 gobal 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!

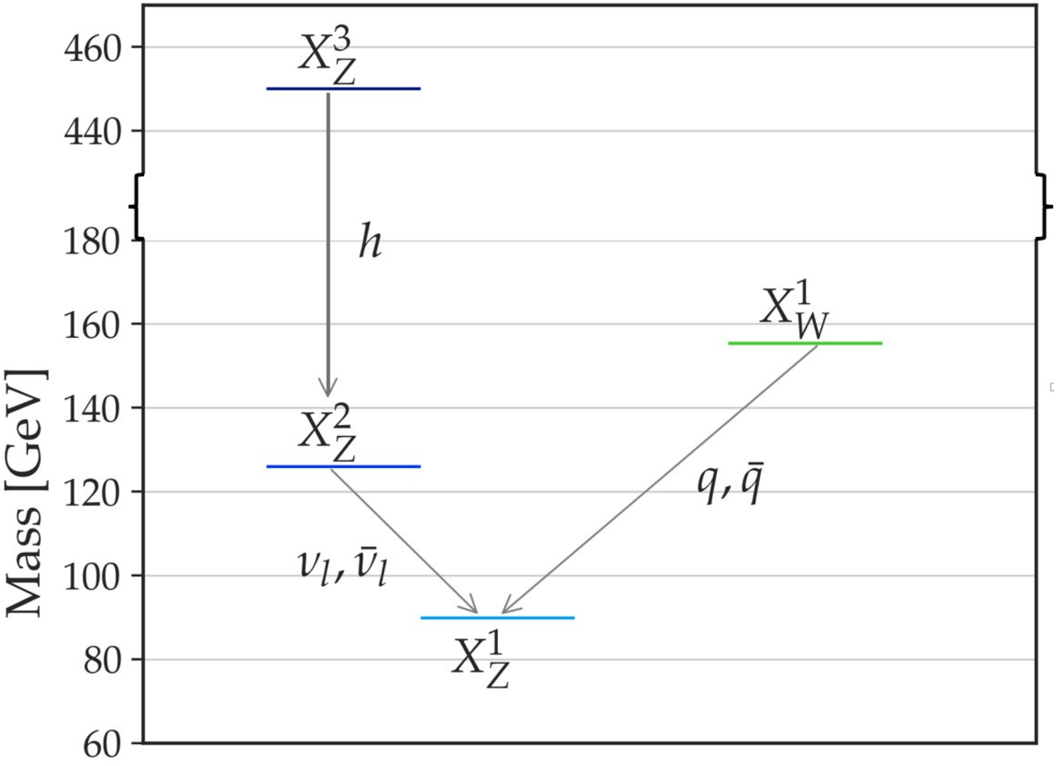


Let’s try and find a simple prototype BSM theory that explains the excessive number of excesses!

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"

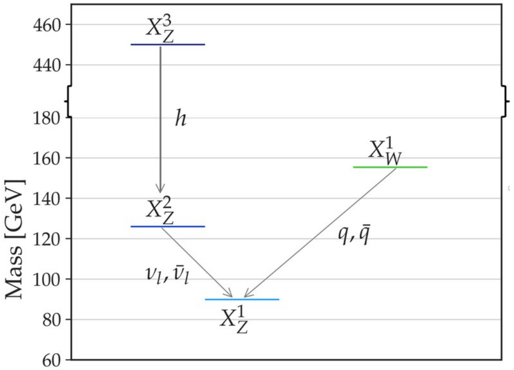
(We introduced this notion of a "protomodel". It's an incomplete BSM model, driven by data, not by theory considerations. LHC production cross sections are free parameters, BSM particle spins are left unspecified)



X^i_Z = "neutralino-like"

X^i_W = "chargino-like"

Why do we consider this to be an interesting protomodel?



Analysis Name	Type	Dataset	Topos	Observed	Expected	Approx σ	Particles
CMS-EXO-20-004	comb	-	TChiISR, TChiZISRqq	2e+02 fb	1e+02 fb	4.6 σ	$X^1{}_W$, $X^1{}_Z$, $X^2{}_Z$
ATLAS-EXOT-2018-06	em	EM10	TChiISR, TChiZISRqq	413	359.00 +/- 10.00	2.5 σ	$X^1{}_W$, $X^1{}_Z$, $X^2{}_Z$
CMS-SUS-20-004	comb	-	TChiHH	0.3 fb	0.2 fb	2.1 σ	$X^3{}_Z$, $X^2{}_Z$

*) It contextualizes three of the biggest excesses we are seeing in the databases, that are considered mutually independent.

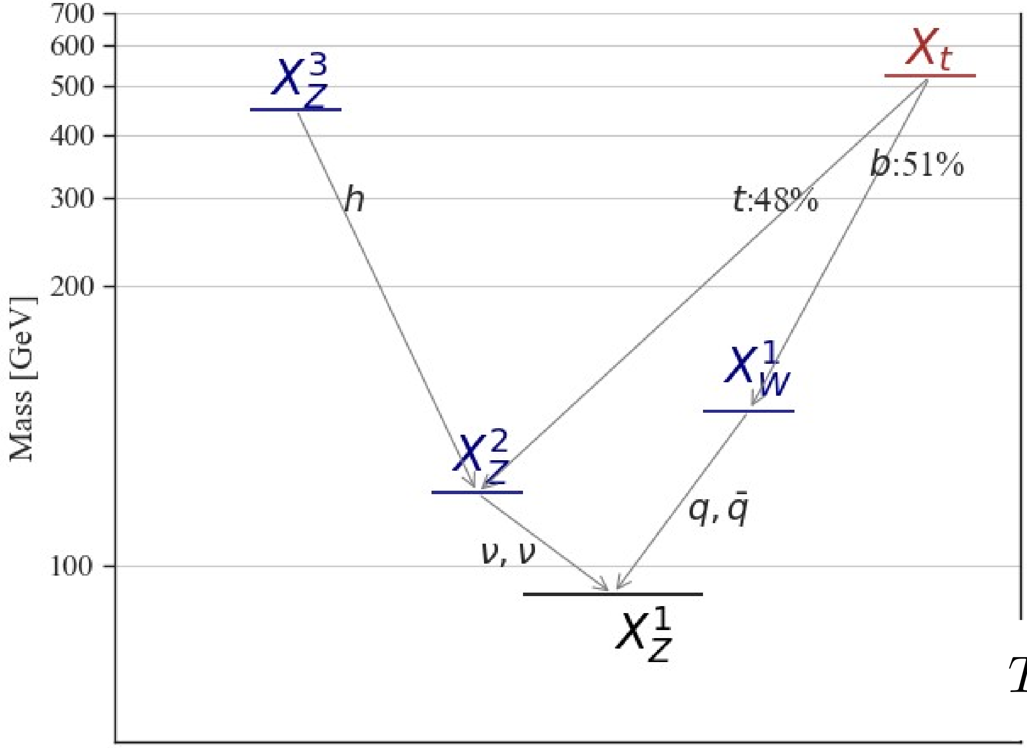
*) we can thus approximately combine the likelihoods of these three analyses – by multiplying them.

*) Another result exists, but is correlated, so we cannot add it in that fashion (but it would also exhibit an excess)

$$T_L := 2\ln\left(\frac{L^c_{\mathrm{BSM}}(\hat{\mu})\cdot\pi_{\mathrm{BSM}}}{L^c_{\mathrm{SM}}\cdot\pi_{\mathrm{SM}}}\right) = 33.69 = 5.8^2$$

"local 5.8 sigma" (if ndf=1) – with a GIGANTIC look elsewhere effect!

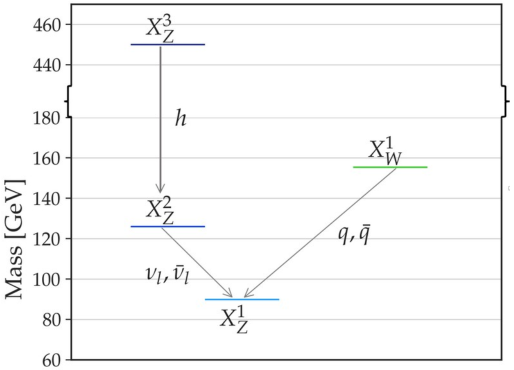
Here is another such protomodel



$$T_L := 2\ln\left(\frac{L^c_{\mathrm{BSM}}(\hat{\mu})\cdot\pi_{\mathrm{BSM}}}{L^c_{\mathrm{SM}}\cdot\pi_{\mathrm{SM}}}\right) = 39.06 = 6.25^2$$

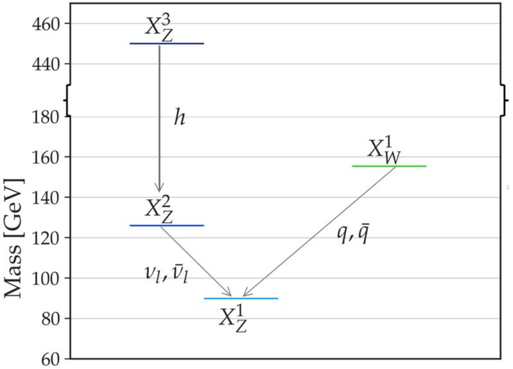
Analysis Name	Type	Dataset	Topos	Observed	Expected	Approx σ	Particles
CMS-EXO-20-004	comb	-	TChiISR, TChiZISRqq	2e+02 fb	1e+02 fb	4.6 σ	X^1_W, X^1_Z, X^2_Z
ATLAS-EXOT-2018-06	em	EM10	TChiISR, TChiZISRqq	413	359.00 +/- 10.00	2.5 σ	X^1_W, X^1_Z, X^2_Z
ATLAS-SUSY-2018-12	em	SRBTT	T2tt	67	46.00 +/- 7.00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

But let's focus on the simpler protomodel



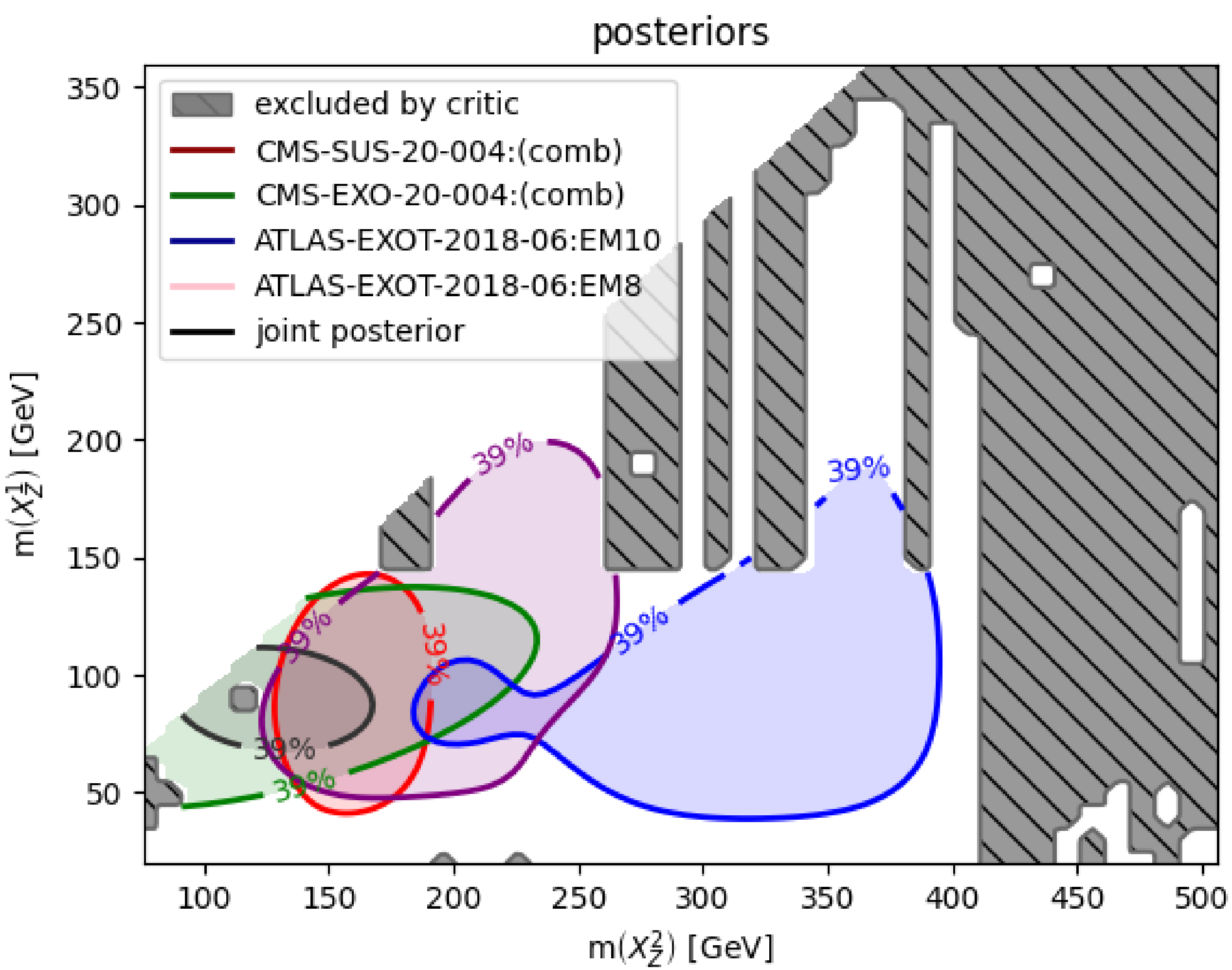
Analysis Name	Type	Dataset	Topos	Observed	Expected	Approx σ	Particles
CMS-EXO-20-004	comb	-	TChiISR, TChiZISRqq	2e+02 fb	1e+02 fb	4.6 σ	$X^1{}_W$, $X^1{}_Z$, $X^2{}_Z$
ATLAS-EXOT-2018-06	em	EM10	TChiISR, TChiZISRqq	413	359.00 +/- 10.00	2.5 σ	$X^1{}_W$, $X^1{}_Z$, $X^2{}_Z$
CMS-SUS-20-004	comb	-	TChiHH	0.3 fb	0.2 fb	2.1 σ	$X^3{}_Z$, $X^2{}_Z$

Do we consider these excesses to be mutually compatible?

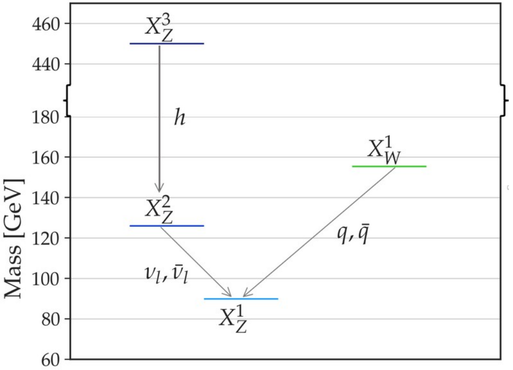


Analysis Name	Type	Dataset	Topos	Observed	Expected	Approx σ	Particles
CMS-EXO-20-004	comb	-	TChiISR, TChiZISRqq	2e+02 fb	1e+02 fb	4.6 σ	X^1_W, X^1_Z, X^2_Z
ATLAS-EXOT-2018-06	em	EM10	TChiISR, TChiZISRqq	413	359.00 +/- 10.00	2.5 σ	X^1_W, X^1_Z, X^2_Z
CMS-SUS-20-004	comb	-	TChiHH	0.3 fb	0.2 fb	2.1 σ	X^3_Z,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

Combining Likelihoods

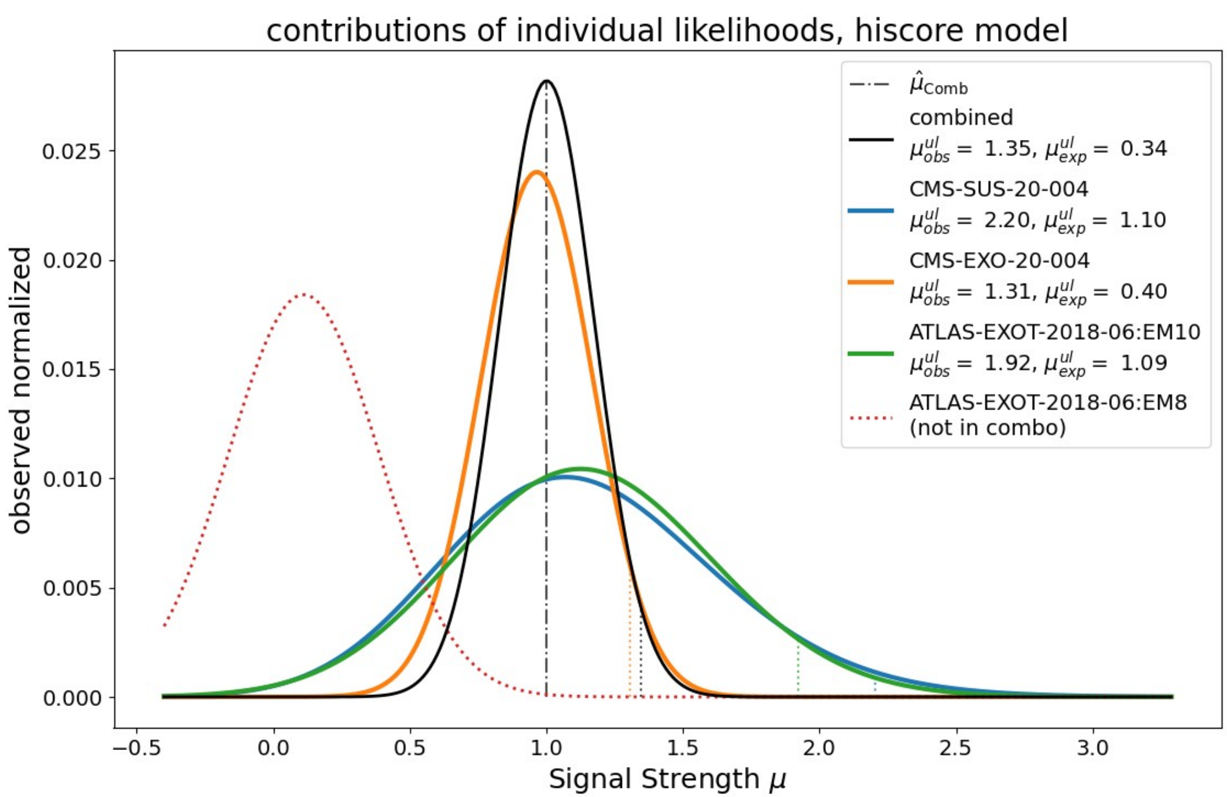


Likelihoods as functions of particle masses



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

Combining Likelihoods



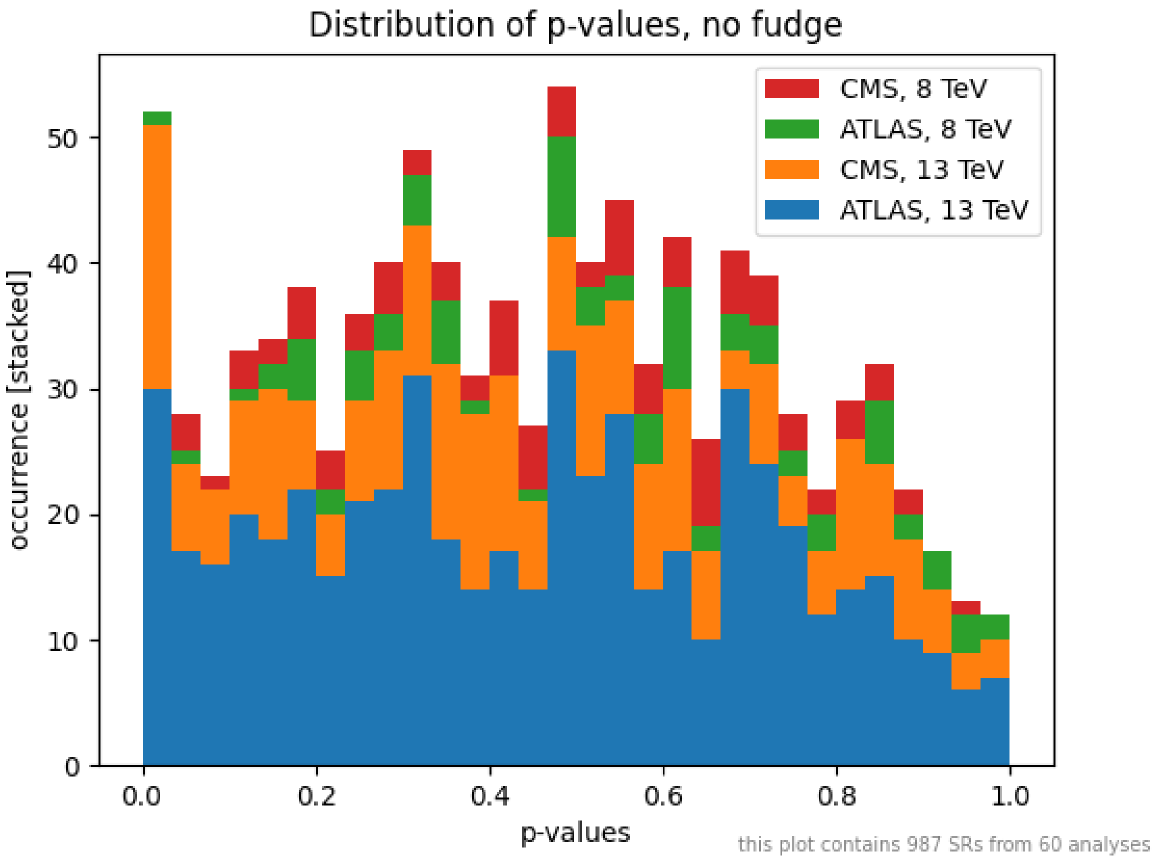
Likelihoods as functions of an overall signal strength multiplier

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

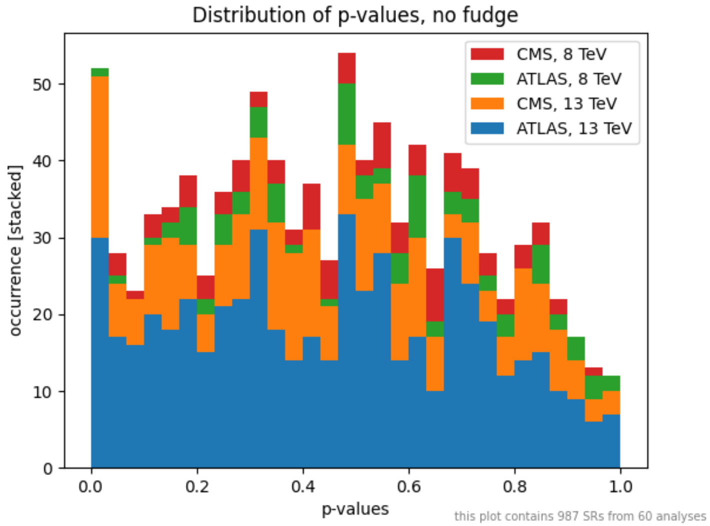
$$T_L := 2 \ln \left(\frac{L^c_{\text{BSM}}(\hat{\mu}) \cdot \pi_{\text{BSM}}}{L^c_{\text{SM}} \cdot \pi_{\text{SM}}} \right) = 33.69 = 5.8^2$$

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!

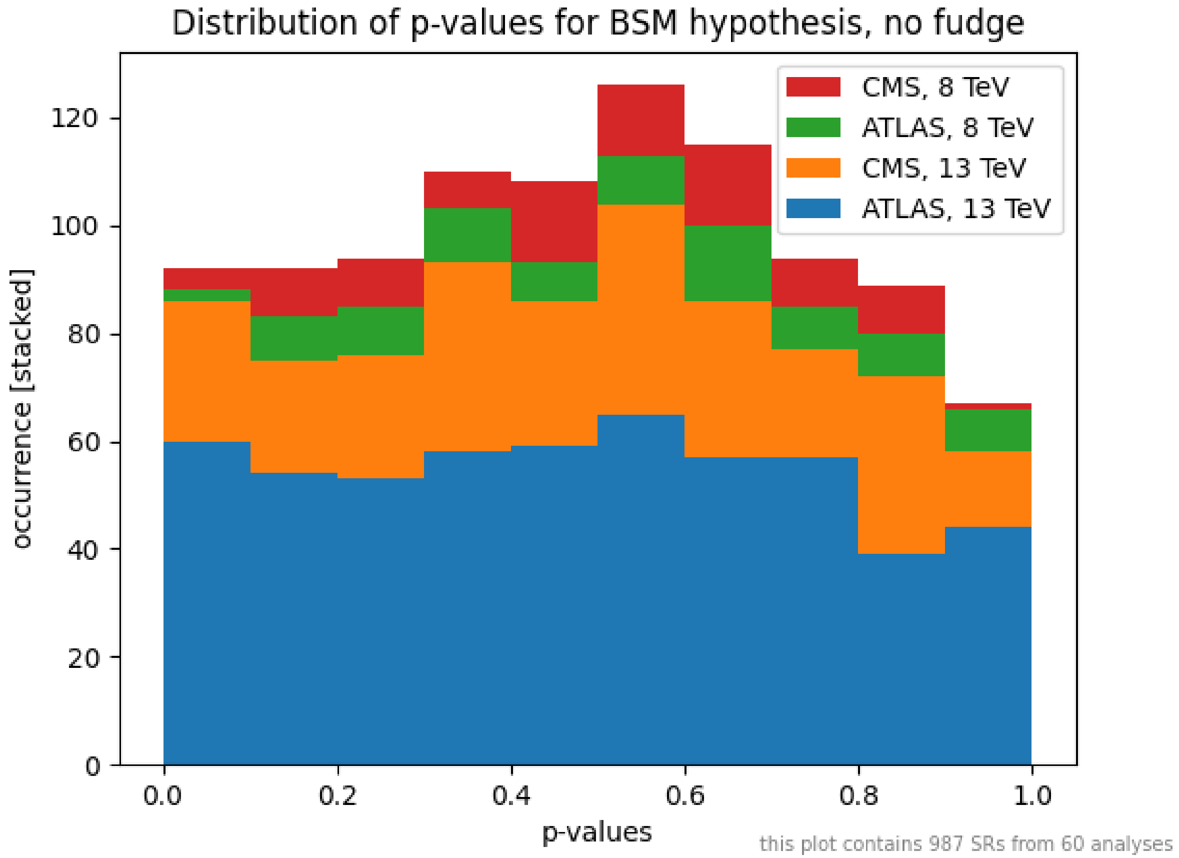
For the SM hypothesis:



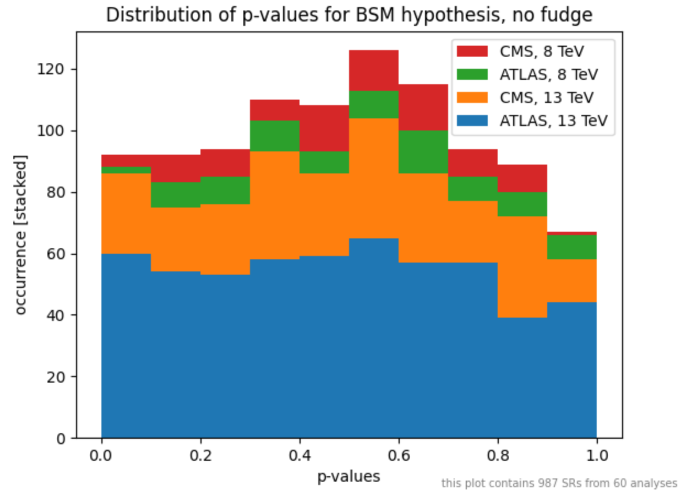
p-values for
The SM hypothesis



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

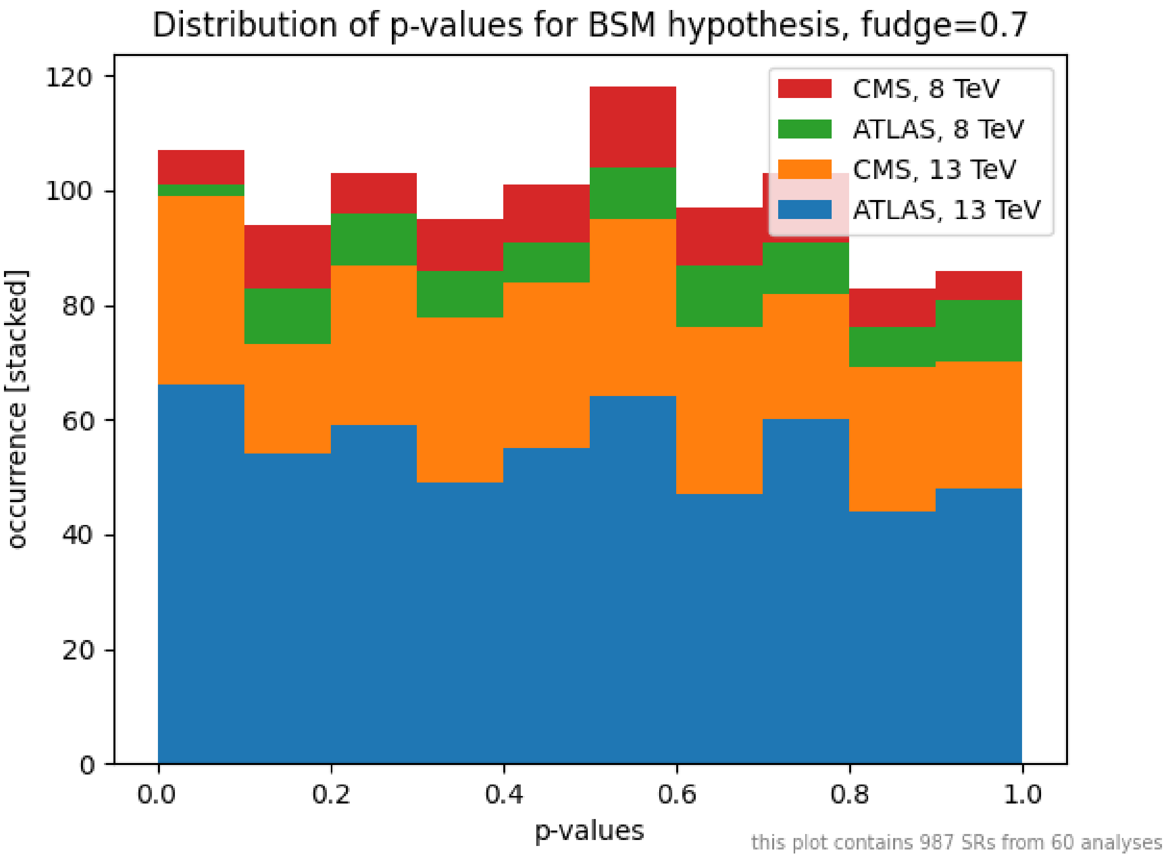


Instead of the SM, let’s compute the p-values for the BSM hypothesis:

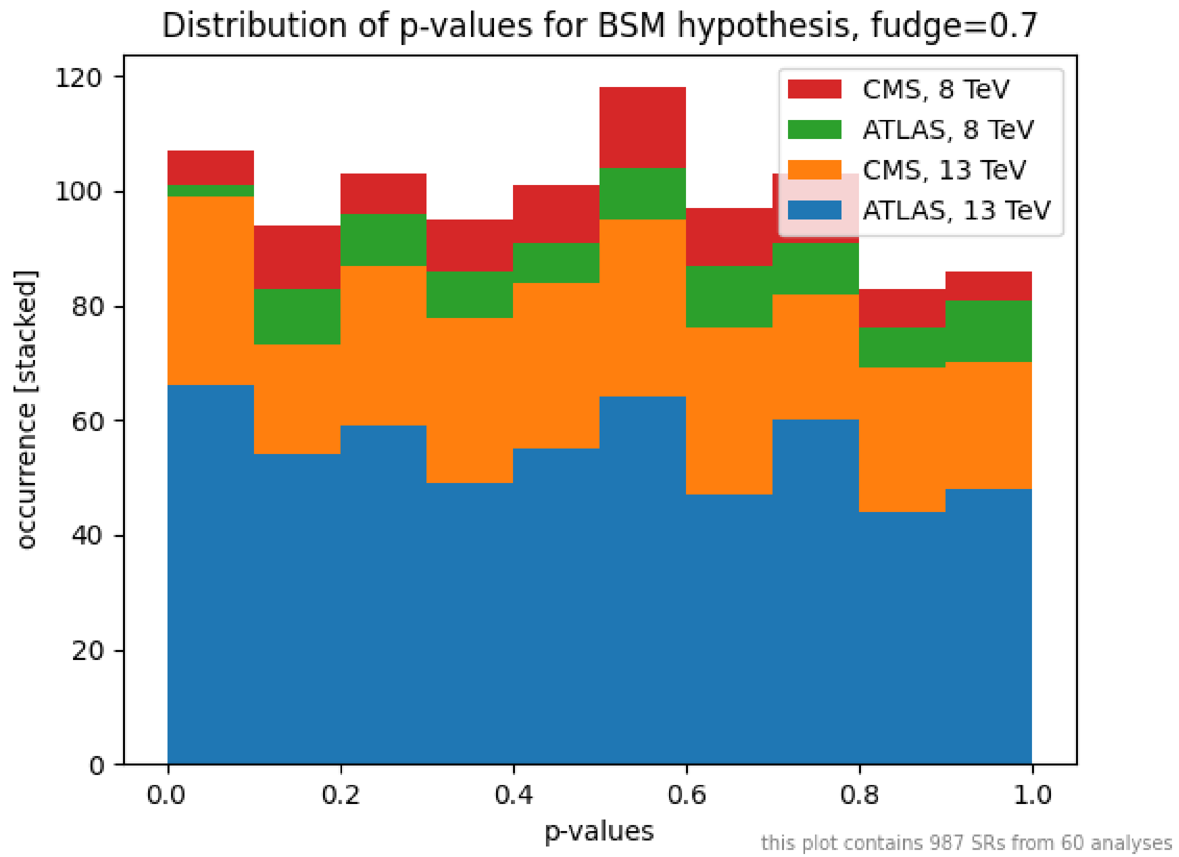


The convex shape makes sense if we assume the systematic errors to be conservatively estimated (as they should).

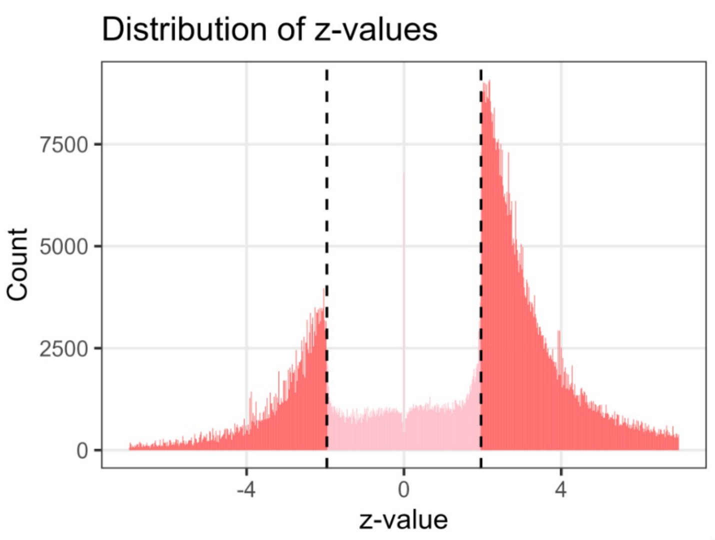
Here is the distribution of the P-values for the protomodel hypothesis, assuming all **systematic** errors to be ~ 40 % overestimated



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

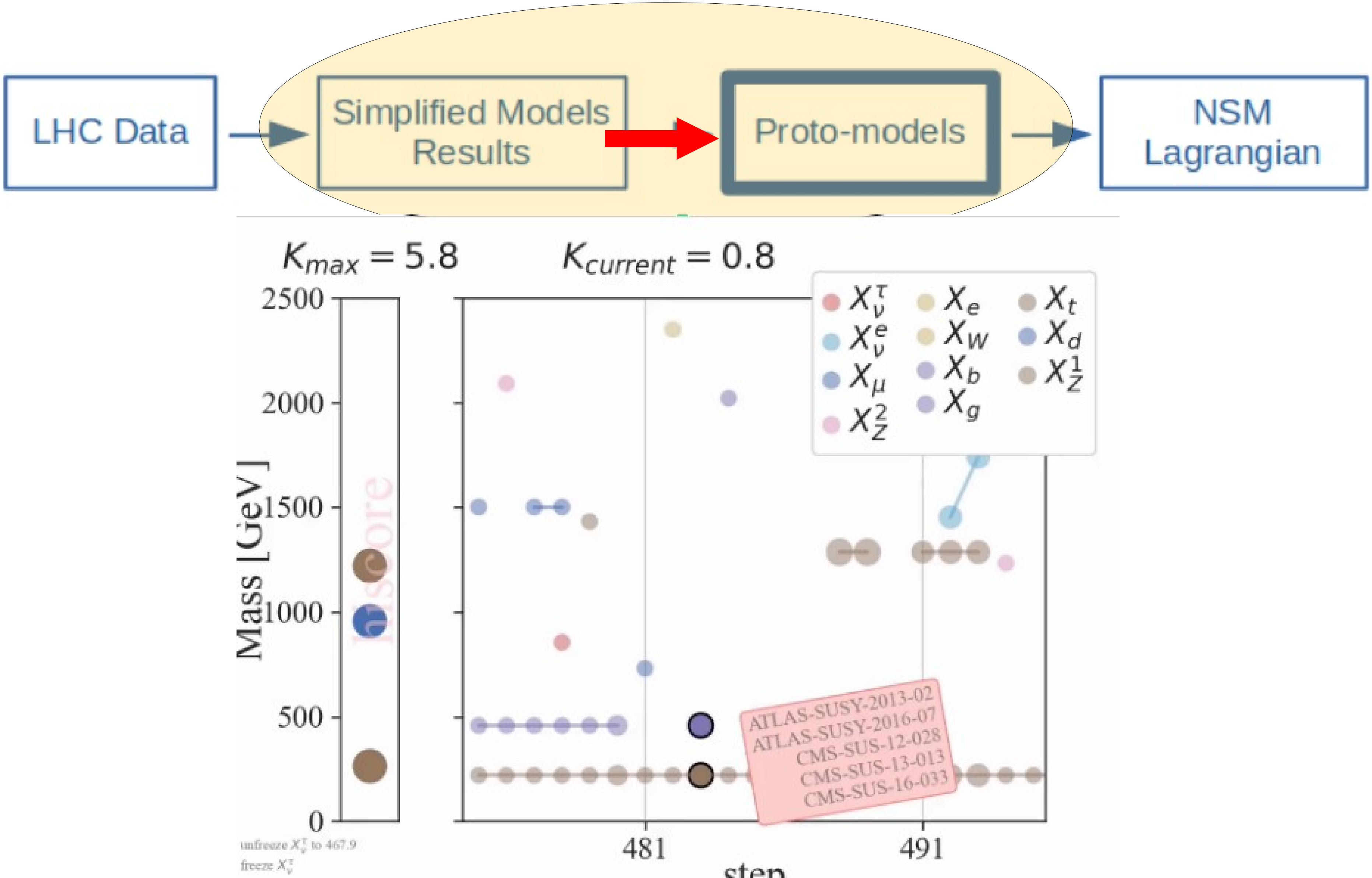


(This feels like nothing short of a miracle. Remember the medical research plot?)



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(BSM)/L(SM) – not likelihoods

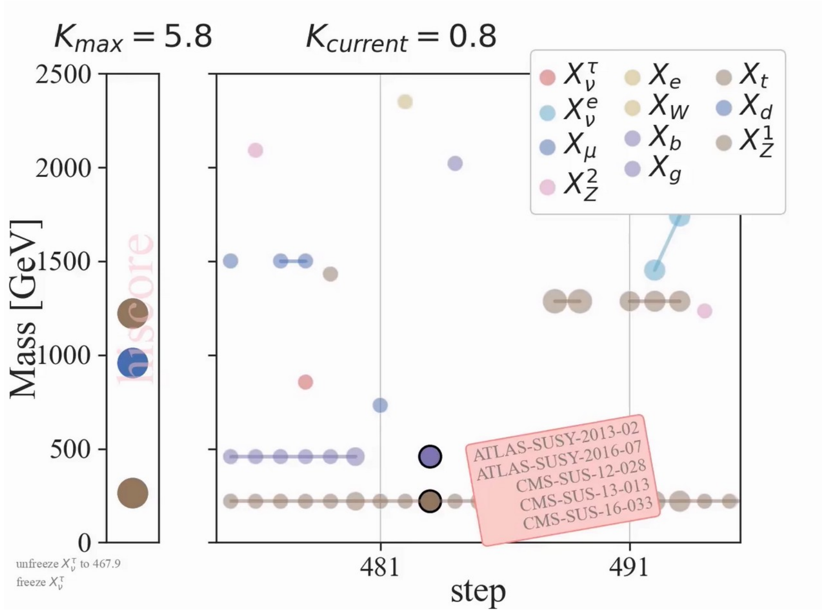
*) we thus look for that legal combination that **maximally violates The SM hypothesis**

*) we added the notion of "critics" that ensure we remain consisent with all negative results in our database

*) we think of production cross sections as free parameters

*) we published a prototype v1 in 2020:

https://arxiv.org/pdf/2012.12246



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

Wolfgang Waltenberger,[a,b,1] André Lessa,[c] Sabine Kraml[d]

[a] Institut für Hochenergiephysik, Österreichische Akademie der Wissenschaften, Nikolsdorfer Gasse 18, A-1050 Wien, Austria

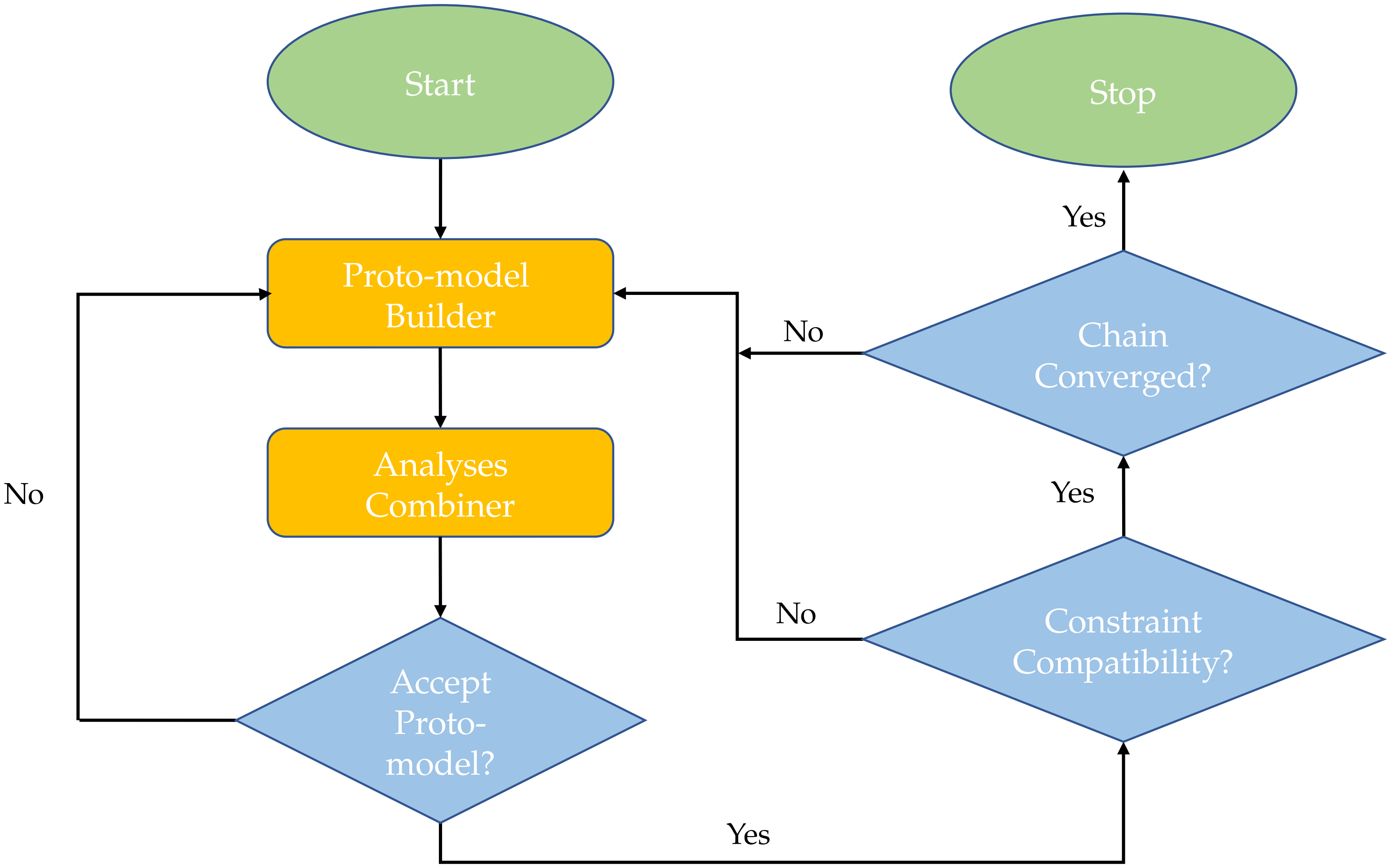
[b] University of Vienna, Faculty of Physics, Boltzmanngasse 5, A-1090 Wien, Austria

[c] Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, Santo André, SP - Brasil

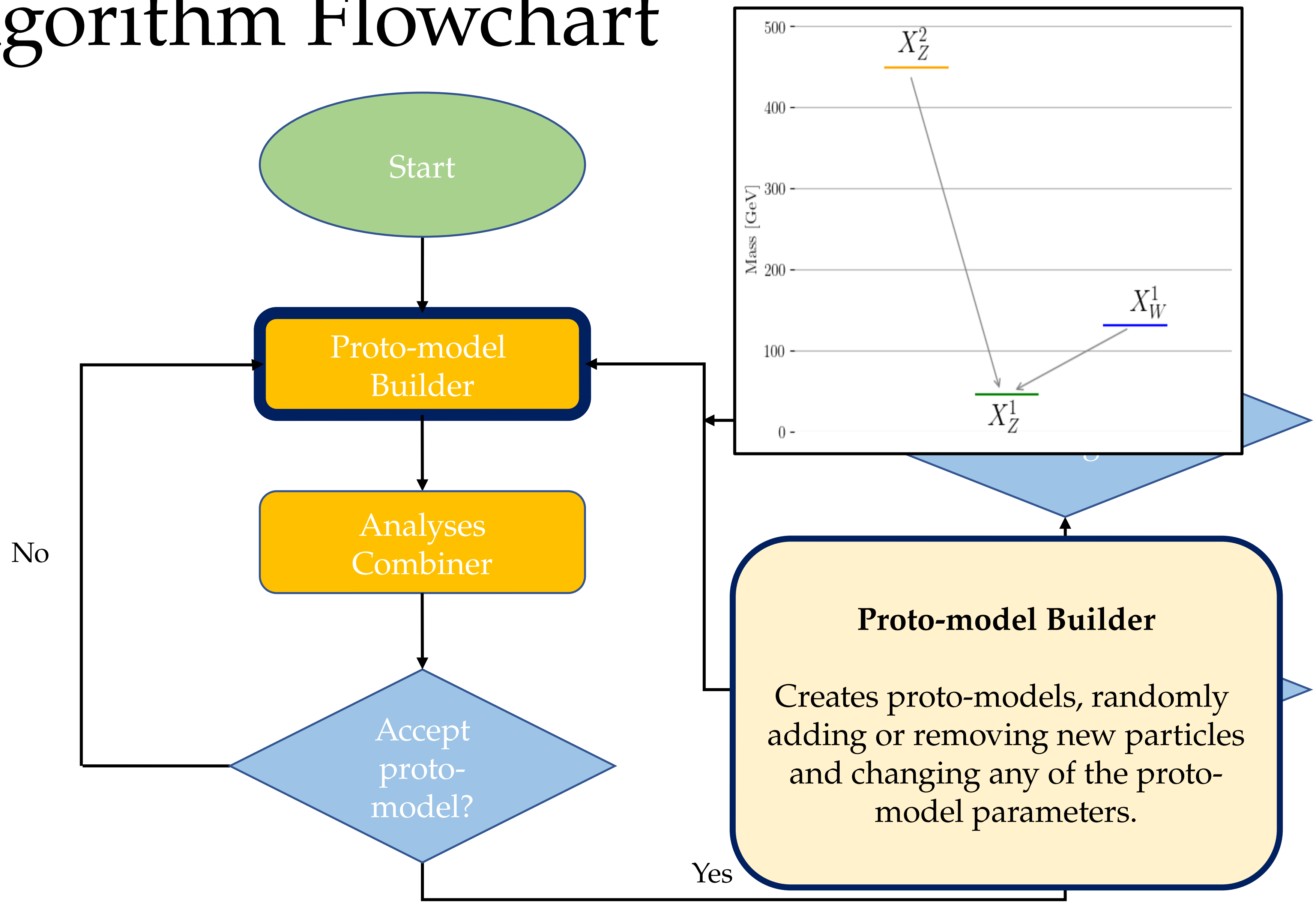
[d] Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3, 53 Avenue des Martyrs, F-38026 Grenoble, France

E-mail: wolfgang.waltenberger@oeaw.ac.at, andre.lessa@ufabc.edu.br, sabine.kraml@lpsc.in2p3.fr

Algorithm Flowchart



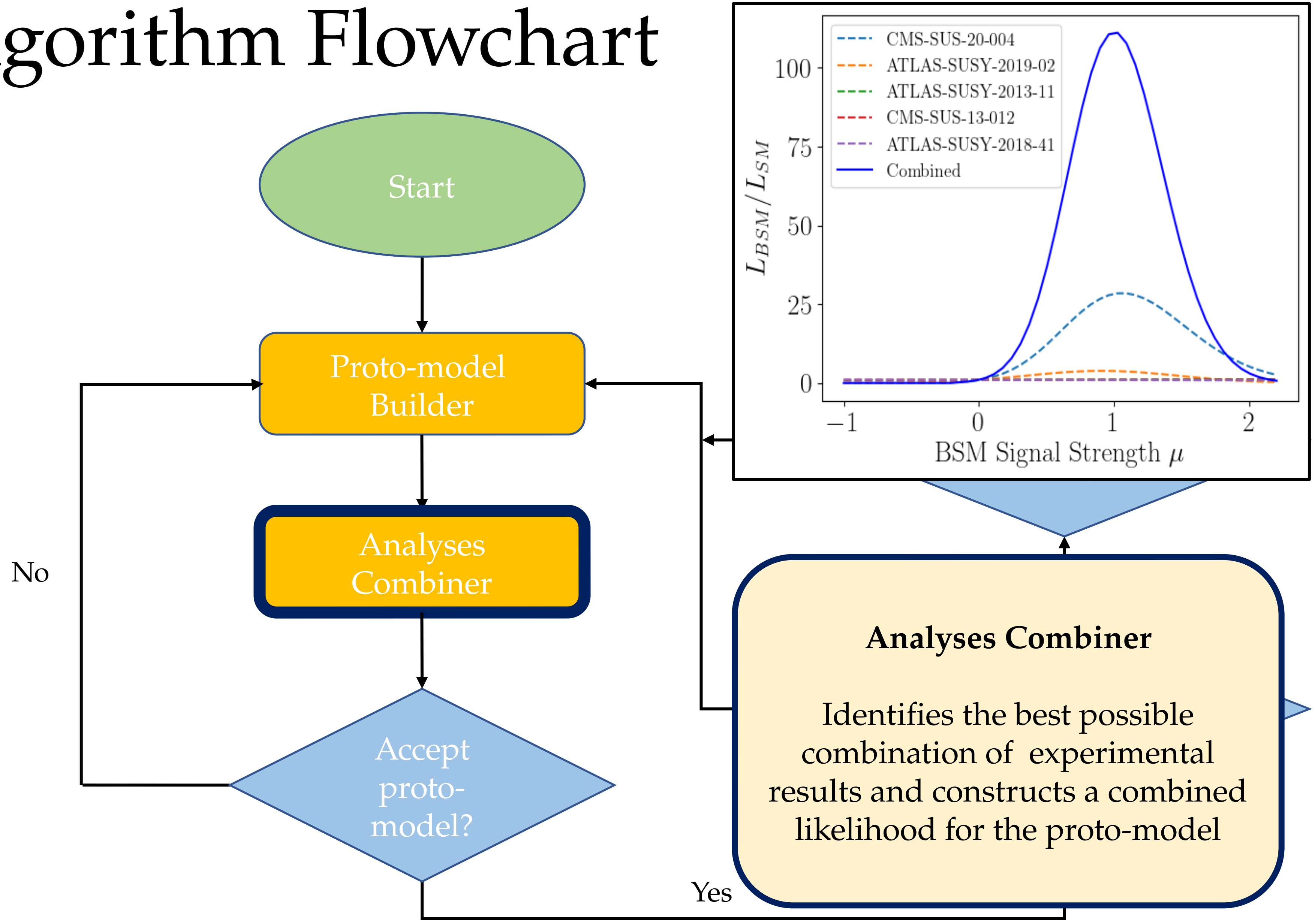
Algorithm Flowchart



Proto-model Builder

Creates proto-models, randomly adding or removing new particles and changing any of the proto-model parameters.

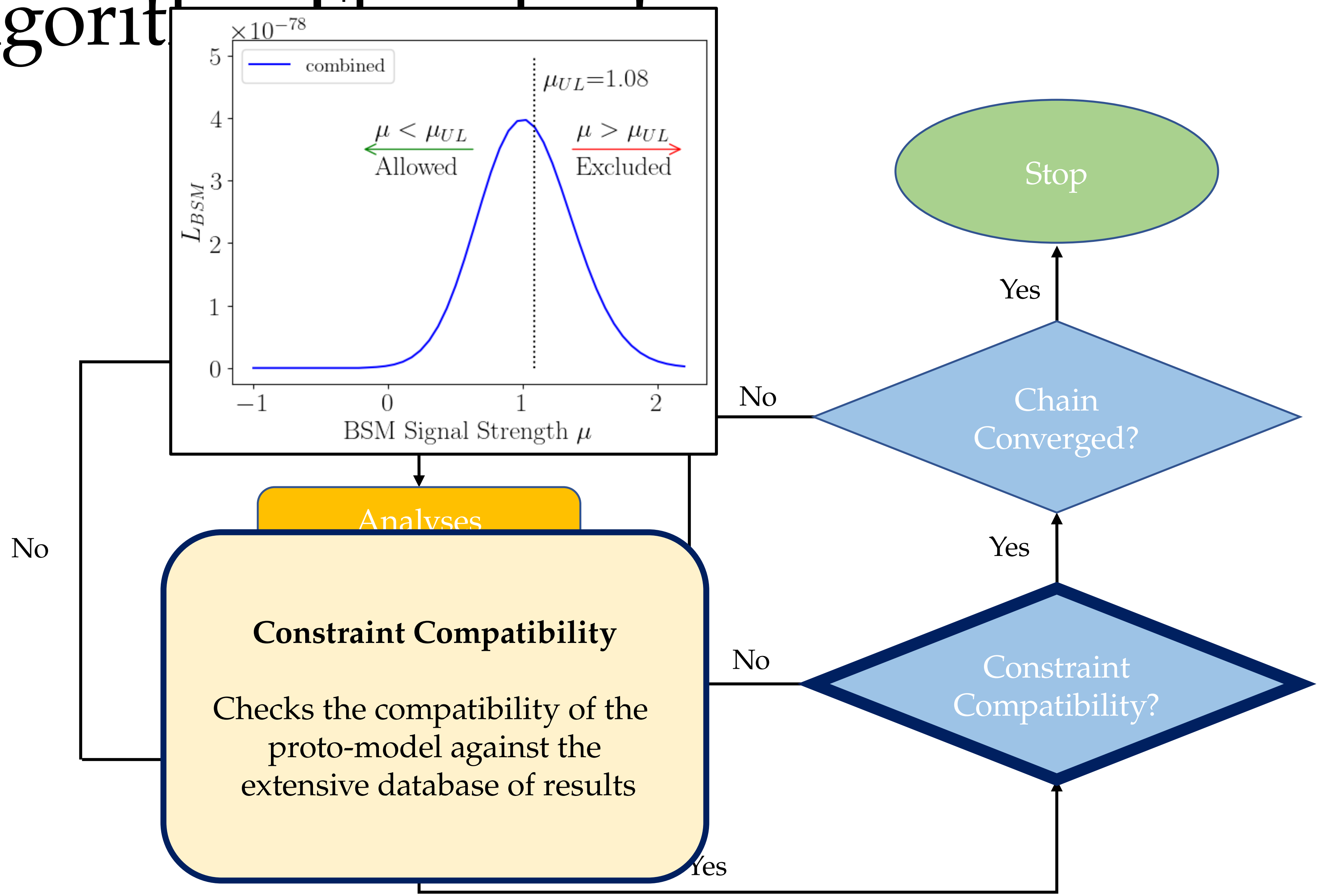
Algorithm Flowchart



Analyses Combiner

Identifies the best possible combination of experimental results and constructs a combined likelihood for the proto-model

Algorit



Constraint Compatibility

Checks the compatibility of the proto-model against the extensive database of results

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).

Quantifies violation of Standard Model hypothesis

Penalizes for model complexity (among other things)

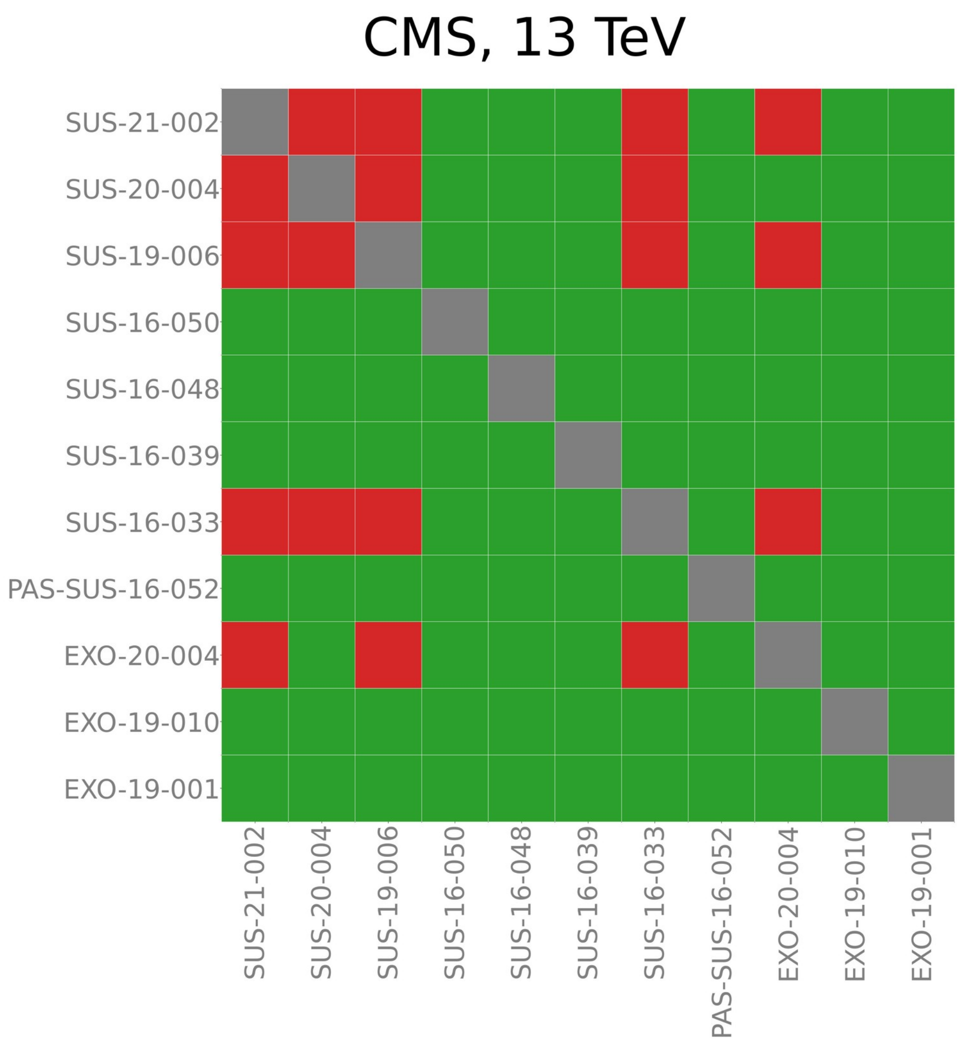
$$K^c := -2 \ln \frac{\mathrm{L}^c_{\mathbf{SM}} \cdot \pi(\mathbf{SM})}{\mathrm{L}^c_{\mathbf{BSM}}(\hat{\mu}) \cdot \pi(\mathbf{BSM})}$$

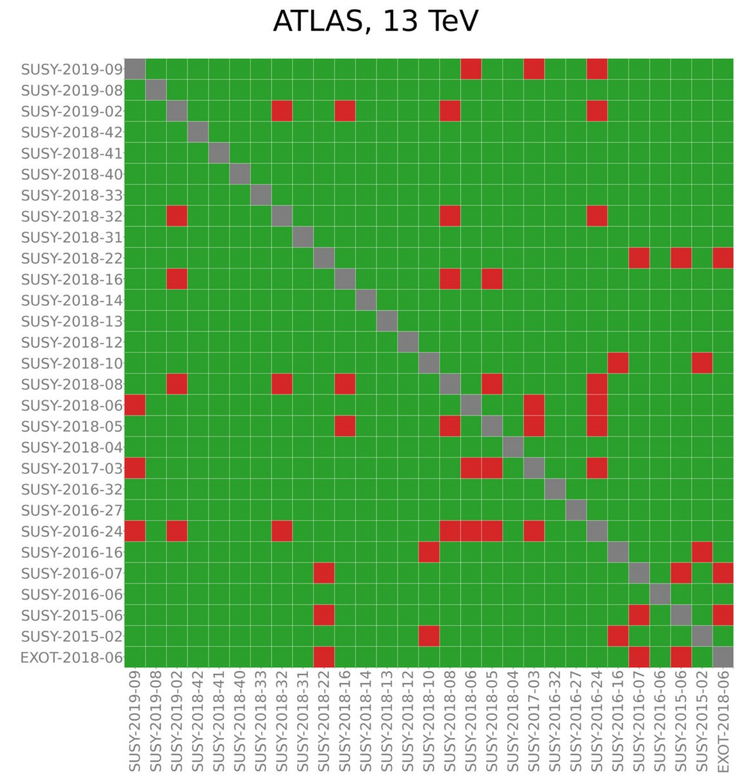
Main penalty is similar to Akaike Information Criterion:

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

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





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

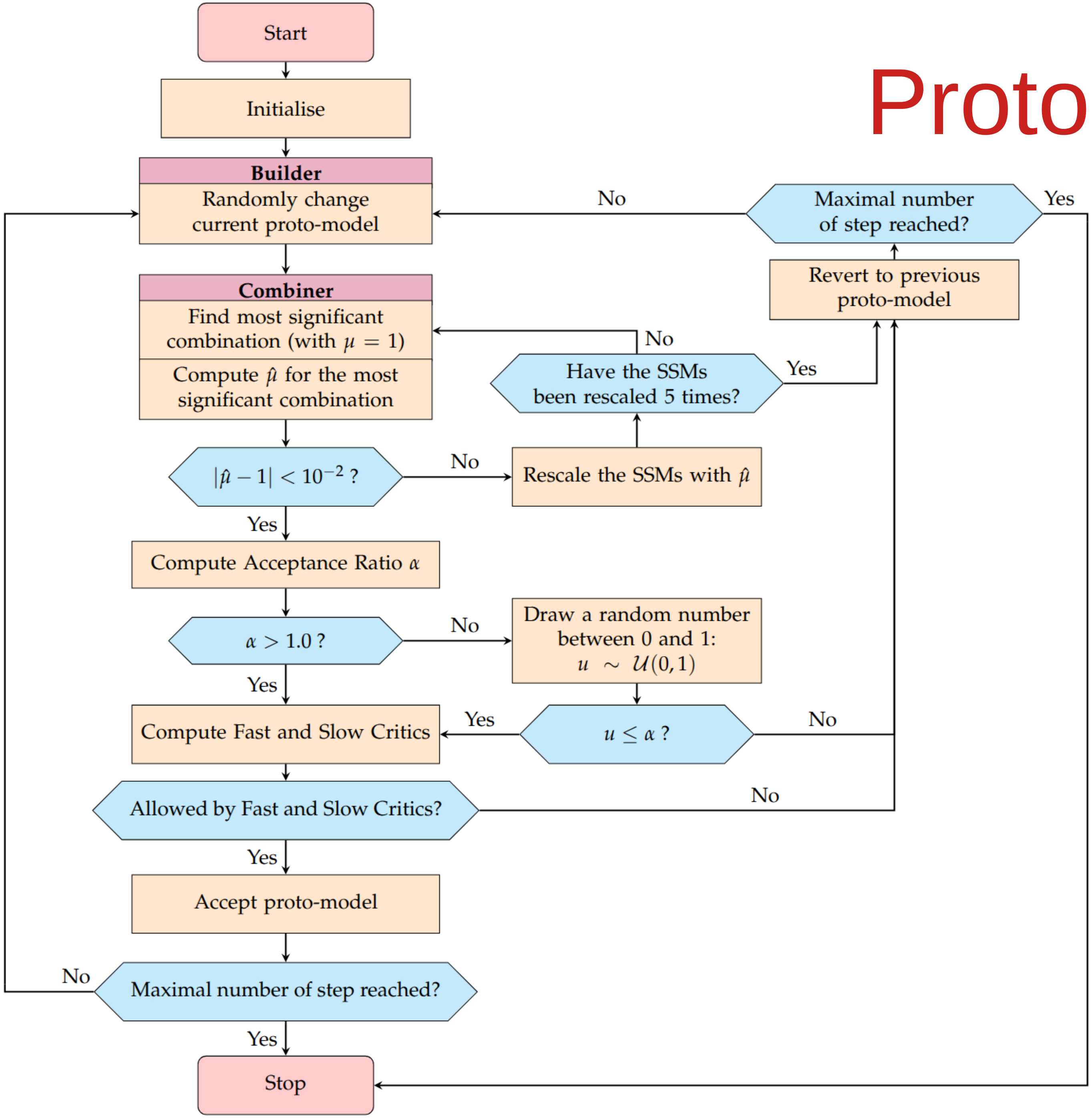
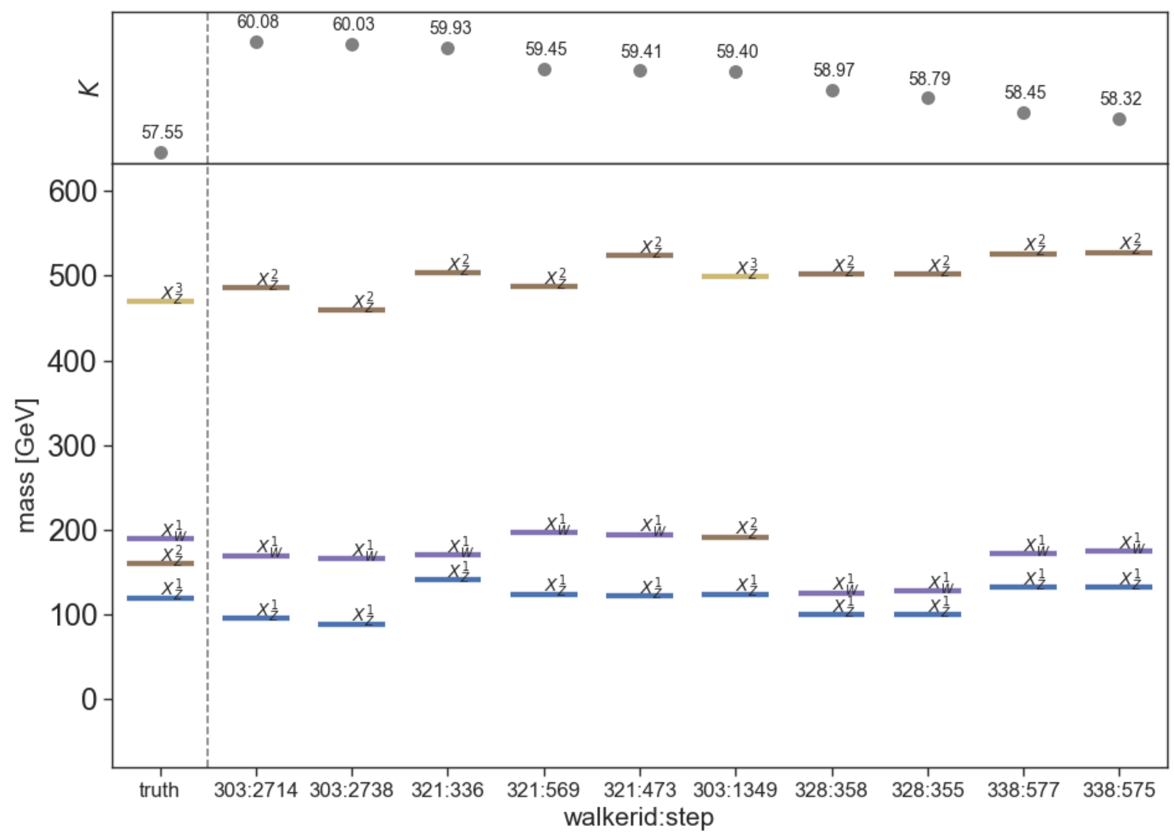


FIGURE 6.2: Simplified flowchart of the new code architecture.

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

Closure test

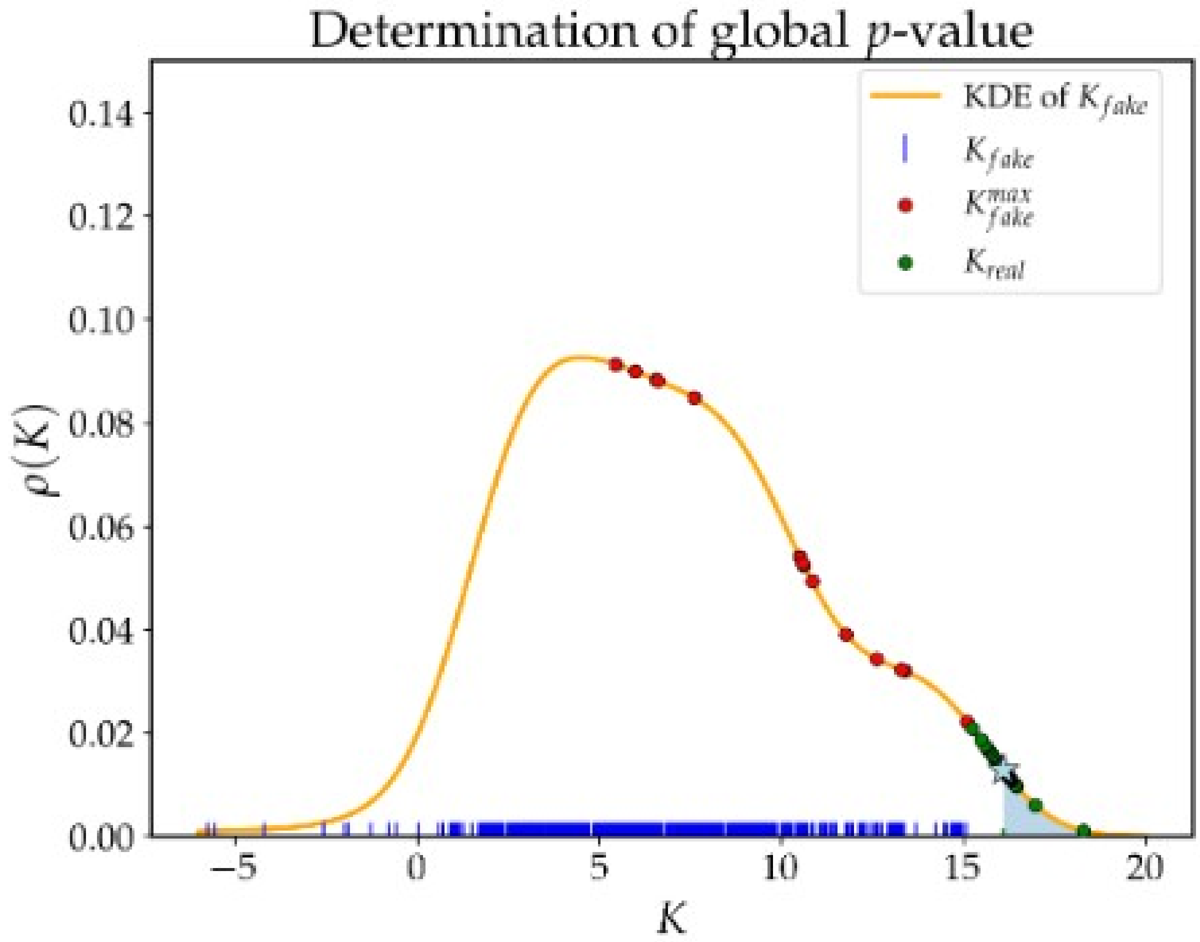
If we create fake data with an injected signal, then run our algorithm over it, do we get the truth back out?



For protomodels similar to our hiscore model, we do.

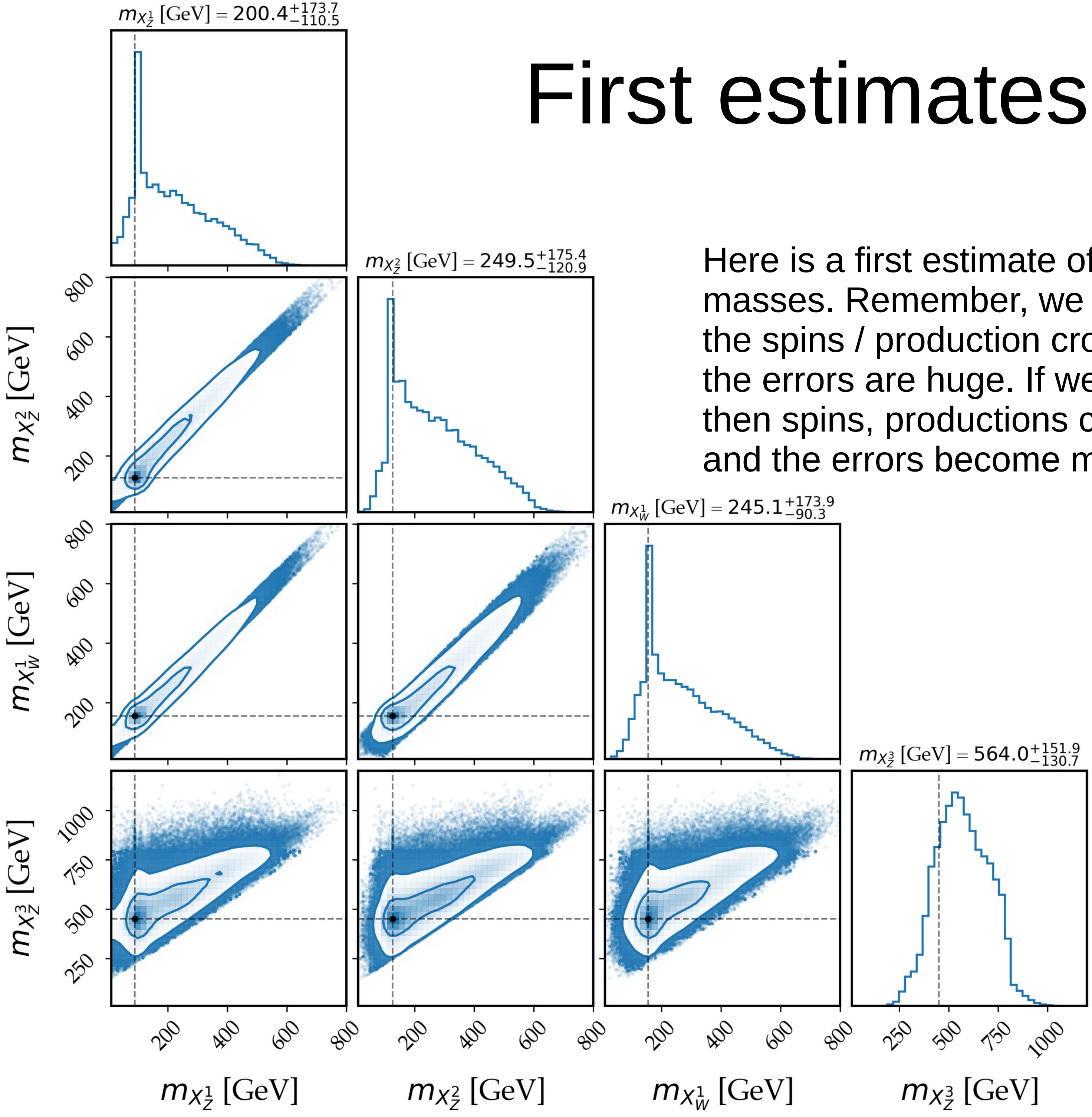
global p-value

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 asssume 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*[1], Benjamin Fuks†[1], Mark D. Goodsell‡[1], and Taylor Murphy§[2]

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

E. Bagnaschi,[a] M. Chakraborti,[b] S. Heinemeyer[c,*] and I. Saha[d]

partners still allowed with masses as low as a few hundred GeV. Over the last years, searches for the "golden channel", $pp \to \tilde{\chi}_2^0\tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 Z^{(*)}\ \tilde{\chi}_1^0 W^{\pm(*)}$ show consistent excesses between CMS and ATLAS in the 2 soft-lepton and 3 soft-lepton plus $\not{E}_T$ 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

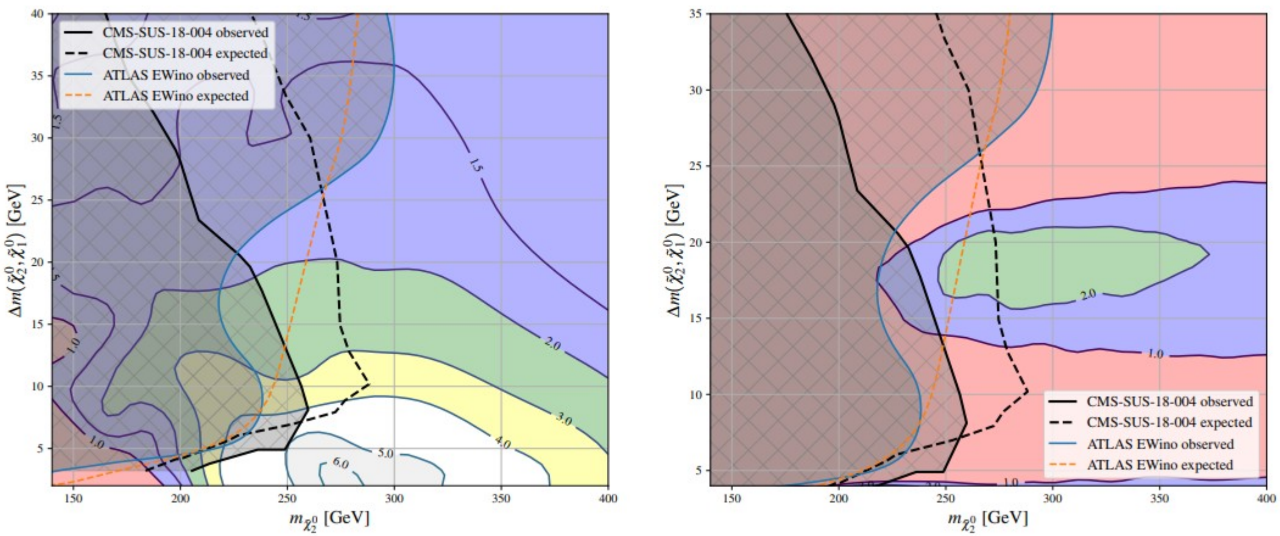
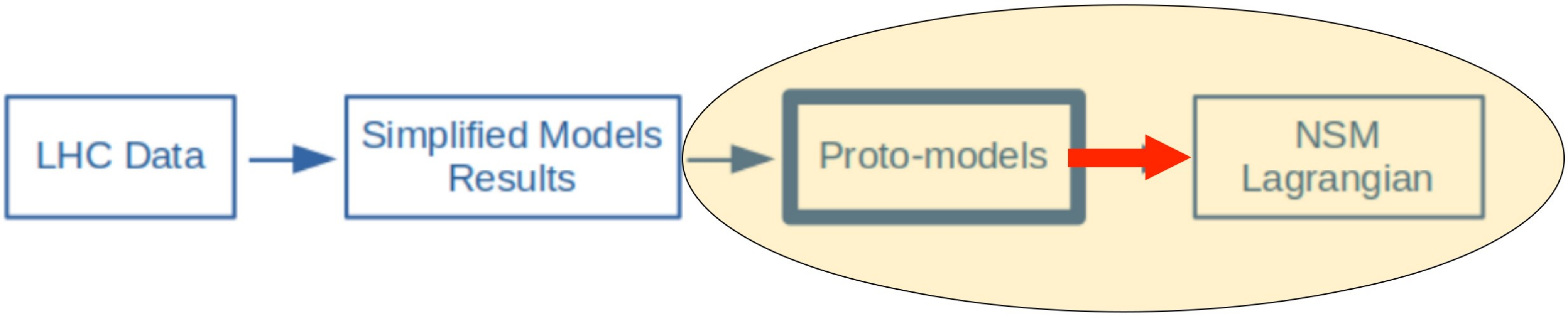


Figure 4: Results for CMS monojet searches (left) and ATLAS soft-lepton+E_T^{miss} (right) in the wino-bino scenario, overlaid with soft-lepton excluded regions reported by ATLAS and CMS. Top: Expected and observed limits. Bottom: Contours of Bayes factor B_{10}.

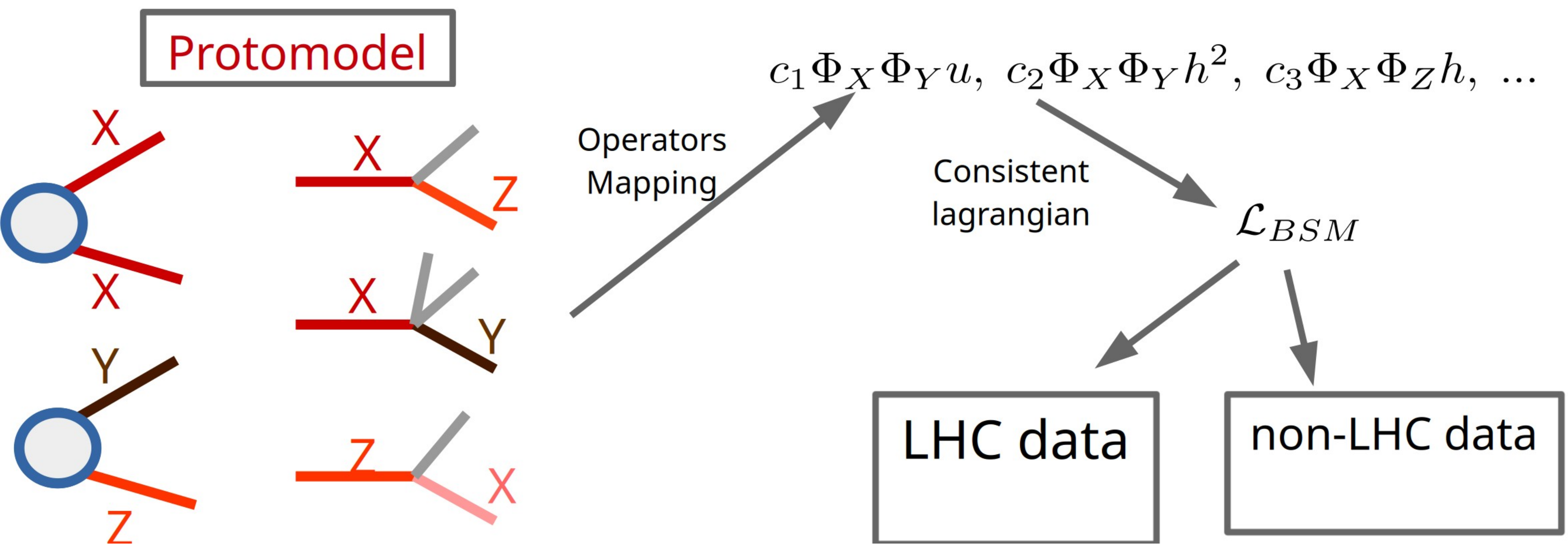
Protomodels – what will we do with them?

(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) – lot's 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?

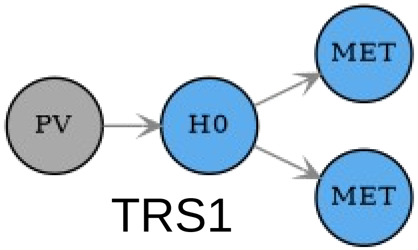
TODO

- 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
- Identicicification of compatible UV complete theories

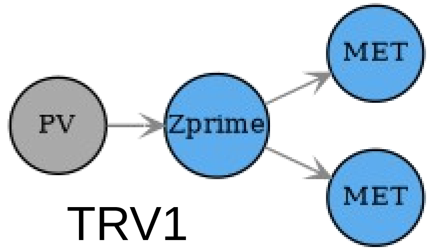
Thank you!

Naively: (*without* assuming conservatism) we would expect
<= 960 * Φ(2.40) ≈ 8 SRs.
We have 28 SRs with Z > 2.40, 4 SRs with Z < - 2.40

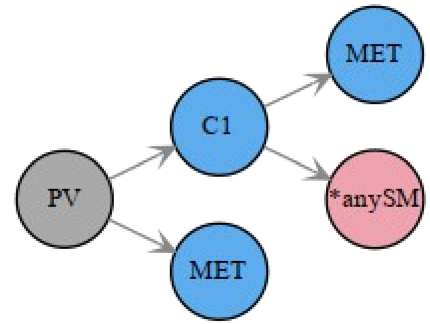
List Of SRs with Z > 2.40



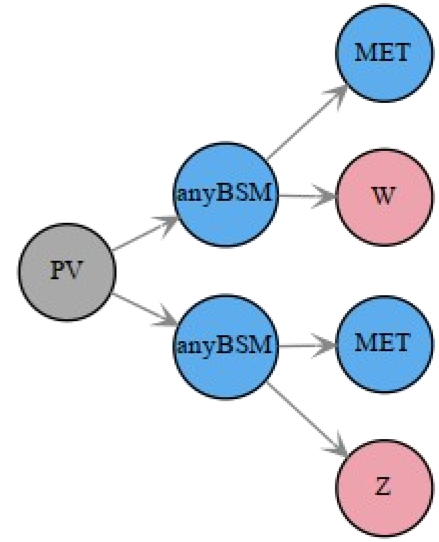
TRS1



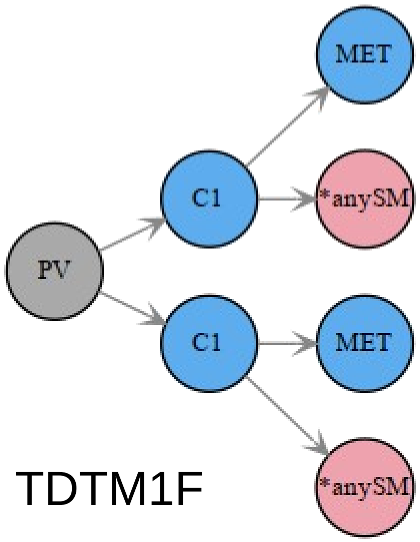
TRV1



TDTM2F

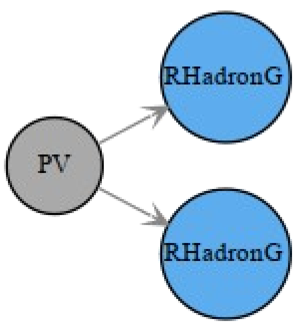


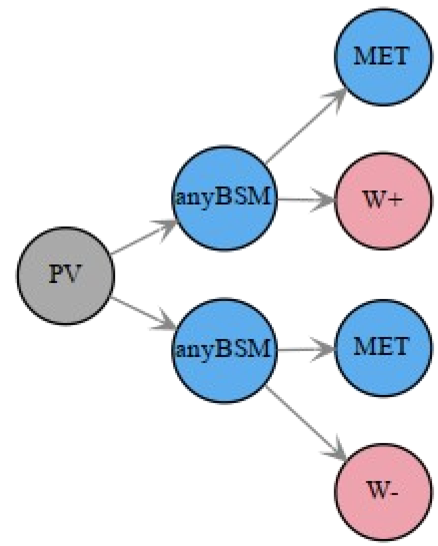
TChiWZ(off)



TDTM1F

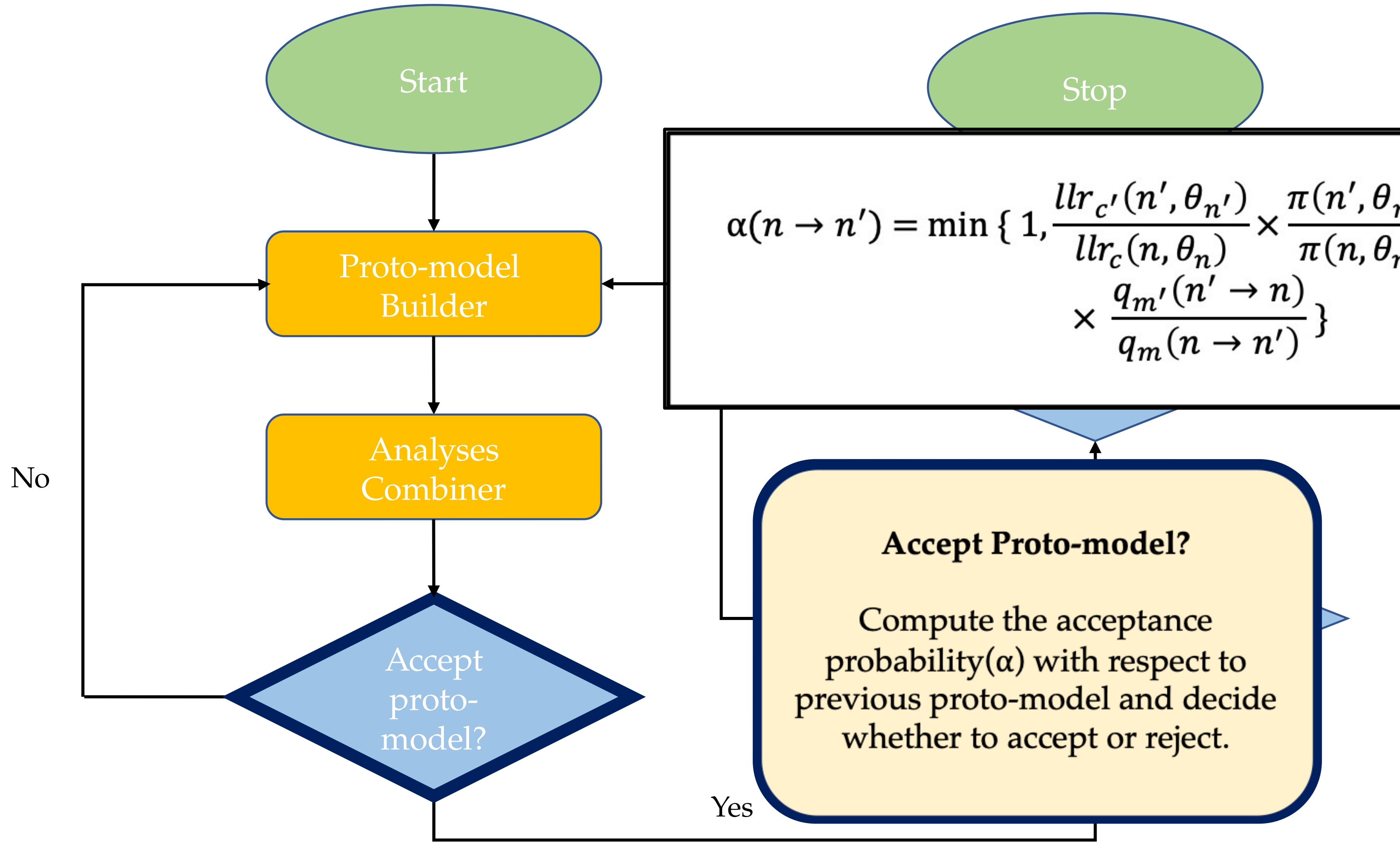
Z	analysis	SR	topo	obsN	expected
4.47	CMS-EXO-20-004	SR_2018_8	TChiISR ...	8639	7836.98±152.03
4.31	CMS-EXO-20-004	SR_2018_4	TChiISR ...	35833	33037.13±618.12
4.11	CMS-EXO-20-004	SR_2018_6	TChiISR ...	14266	13209.01±250.28
3.85	CMS-EXO-20-004	SR_2018_9	TChiISR ...	5406	4914.64±101.04
3.80	ATLAS-SUSY-2018-42	SR1400_High_long	T1DispT ...	7	0.70±0.40
3.78	CMS-EXO-20-004	SR_2018_1	TChiISR ...	191829	178470.12±3615.00
3.72	CMS-EXO-20-004	SR_2018_2	TChiISR ...	104522	97428.42±1896.81
3.63	CMS-EXO-20-004	SR_2018_3	TChiISR ...	59398	55592.19±1059.45
3.46	ATLAS-SUSY-2018-42	SR1500_High_long	TDTM1FT ...	6	0.60±0.40
3.35	CMS-EXO-20-004	SR_2018_11	TChiISR ...	2671	2428.39±55.08
3.30	ATLAS-SUSY-2018-16-hino	SR_ewk_2l_med	TChiWWo ...	170	112.40±12.60
3.21	ATLAS-SUSY-2018-42	SR1300_High_long	TDTM1FT ...	6	0.80±0.40
3.21	ATLAS-SUSY-2018-16	SR_ewk_2l_ee_high_d	TChiWZoff	16	4.80±1.60
3.17	CMS-EXO-20-004	SR_2018_12	TChiISR ...	1613	1441.83±35.64
3.17	ATLAS-SUSY-2018-42	SR1600_High_long	T1DispT ...	5	0.54±0.35
3.01	ATLAS-SUSY-2018-42	SR1200_High_long	T1DispT ...	6	0.90±0.50
3.01	CMS-EXO-20-004	SR_2018_7	TChiISR ...	11730	10991.22±213.74
2.93	ATLAS-SUSY-2017-03	SR3l_ISR	TChiWZ	12	3.90±1.00
2.83	ATLAS-SUSY-2018-42	SR1800_High_long	T1DispTRHadGM1	4	0.44±0.32
2.75	CMS-EXO-20-004	SR_2018_5	TChiISR ...	21854	20714.72±391.42
2.56	ATLAS-SUSY-2019-09	SRlow_0Jd	TChiWZoff	77	54.00±4.00
2.51	CMS-SUS-20-004	SR11	T5HHTChiHH	4	0.06±0.11
2.51	CMS-EXO-20-004	SR_2018_10	TChiISR ...	3285	3065.69±66.79
2.49	ATLAS-SUSY-2018-42	SR2400_High_long	T1DispTRHadGM1	3	0.29±0.28
2.47	ATLAS-EXOT-2018-06	EM10	TChiISR ...	413	359.00±10.00
2.45	ATLAS-SUSY-2018-42	SR2600_High_long	T1DispTRHadGM1	3	0.25±0.31
2.44	ATLAS-SUSY-2018-42	SR2000_High_long	T1DispTRHadGM1	3	0.36±0.25
2.42	ATLAS-SUSY-2018-42	SR2200_High_long	T1DispTRHadGM1	3	0.33±0.28





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

Algorithm Flowchart



Our Inverse Problem

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

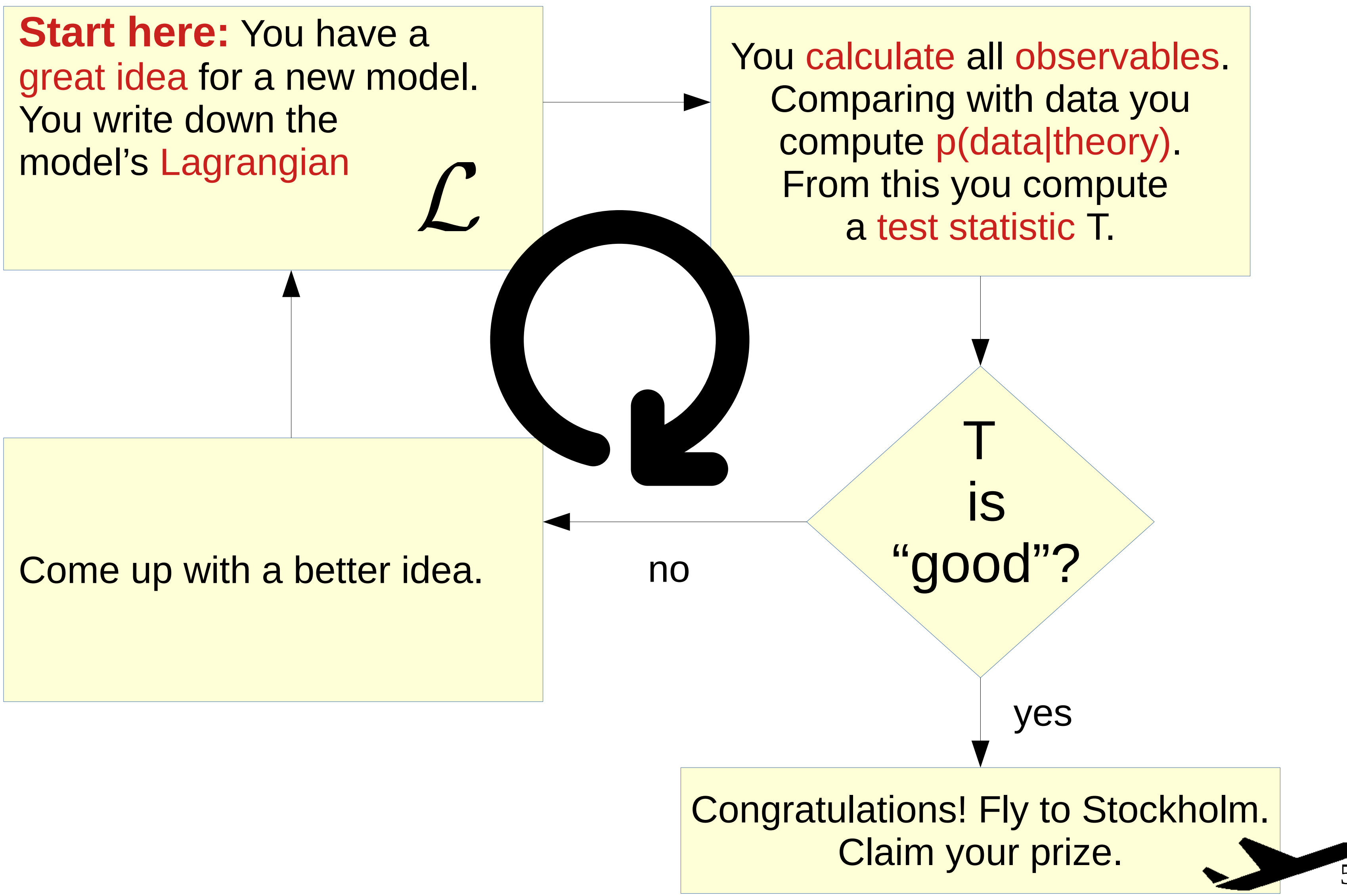
- In addition to lots of other information, we need statistical models (in one way or another)

- we can either start by:
  postulating a theory, then fit its parameters
  ("top-down")

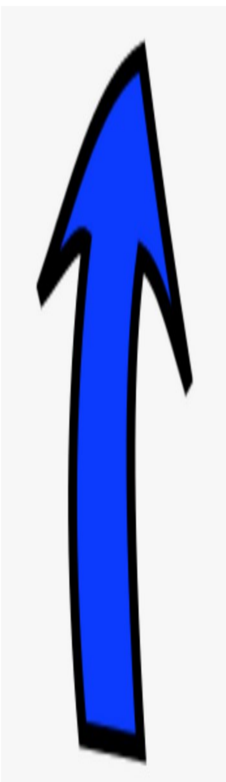
- or:
  try to build our model directly from data
  ("bottom-up")

Top-down versus bottom-up

Top-Down:



Top-down versus bottom-up



Bottom-Up:

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

$\mathcal{L}$

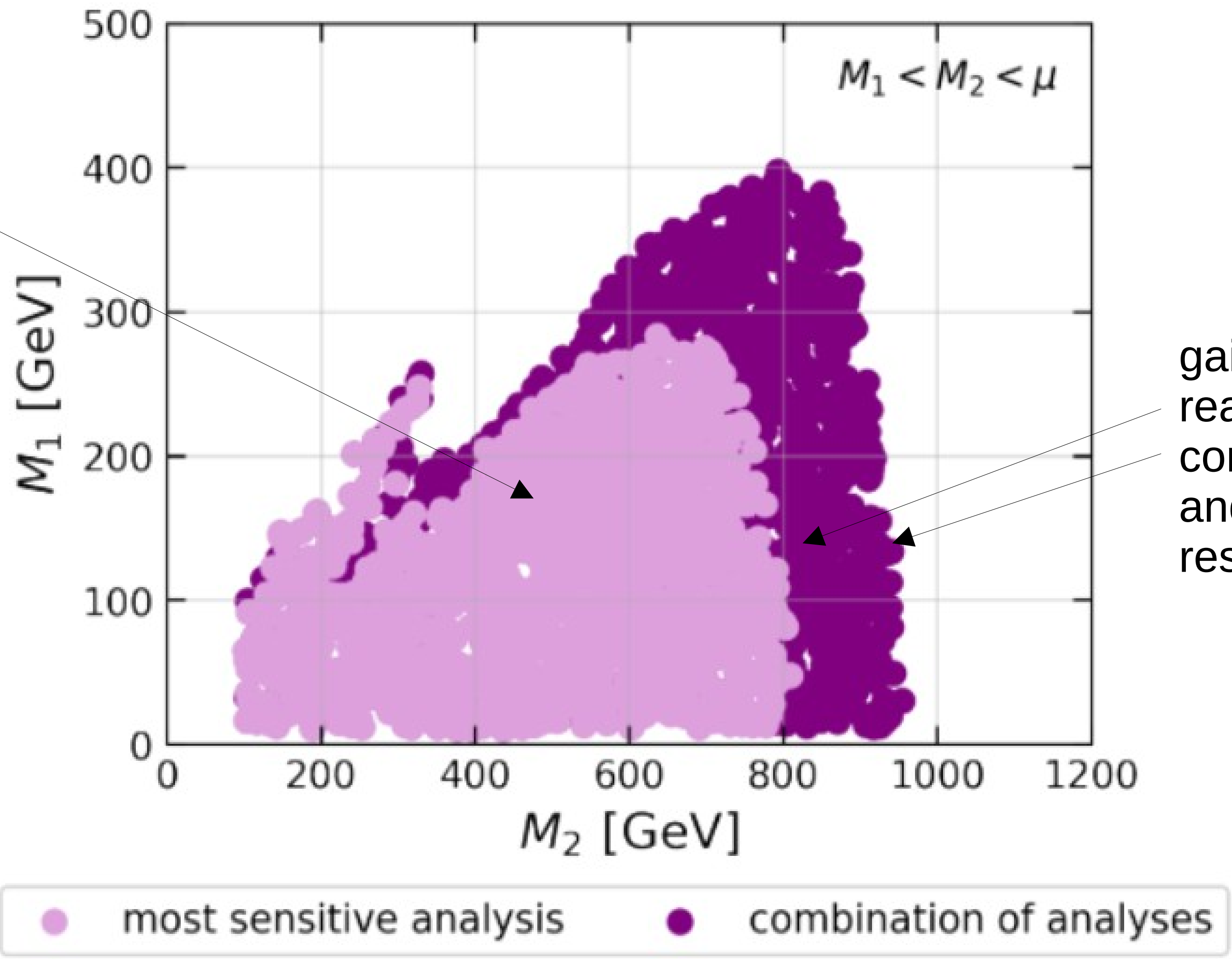
From the descriptions you try and construct precursor theories to the NSM that describe everything you really know about TeV-scale (and below) physics

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").

Top-down: example

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

Excluded
Model points

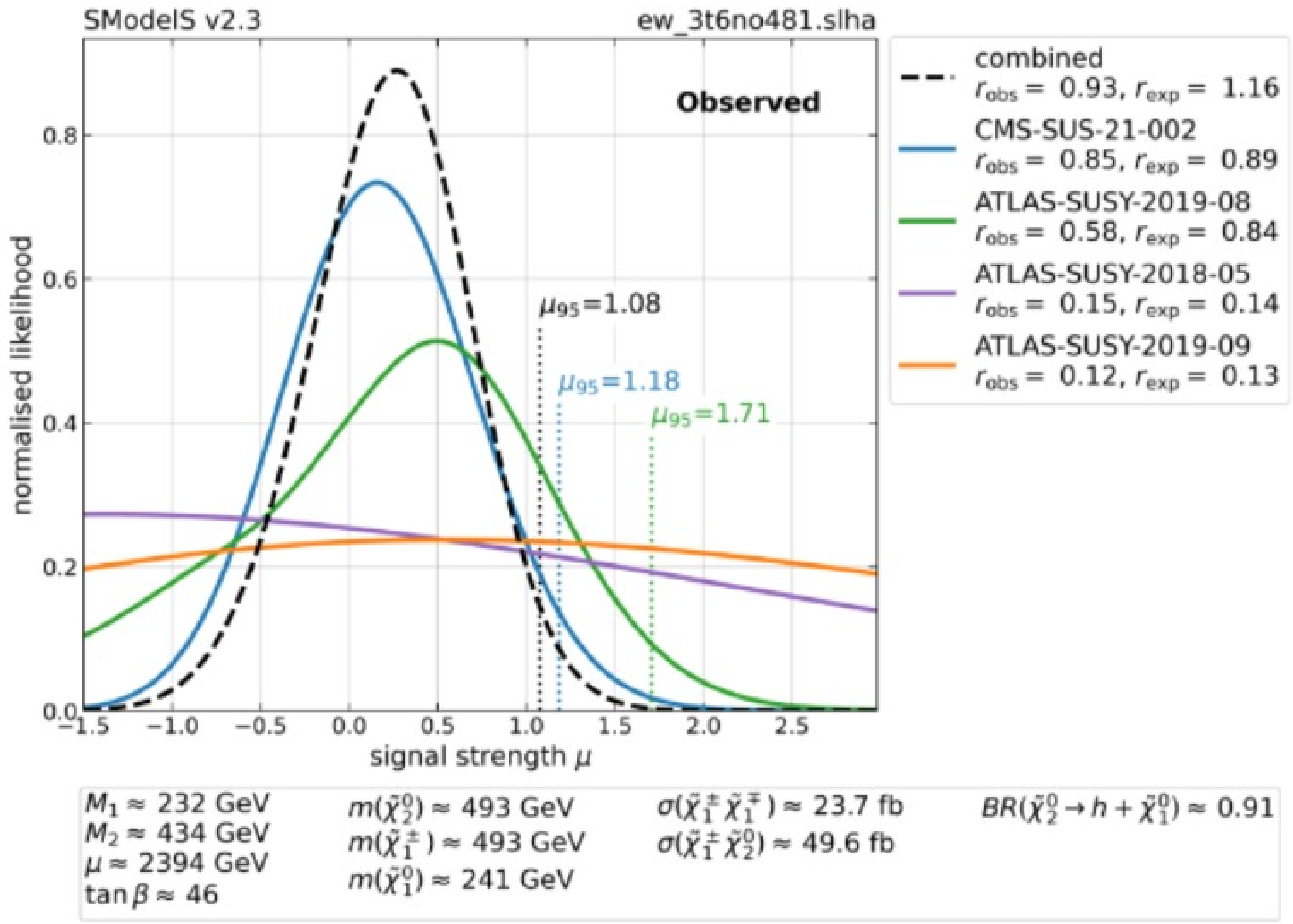


gain in exclusion reach by combining (CMS and ATLAS) results

https://arxiv.org/abs/2312.16635

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).

https://arxiv.org/abs/2312.16635

Statistical Models in the SModelS Database

ATLAS, Run-2, Full Lumi only:

ID	Short Description	$\mathcal{L}$ [fb^{-1}]	UL$_{obs}$	UL$_{exp}$	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 τ	139.0	✓		✓	Pyhf
ATLAS-SUSY-2018-05	2ℓ + jets, EWK	139.0	✓		✓	Pyhf
ATLAS-SUSY-2018-06	3ℓ, EWK	139.0	✓	✓	✓	
ATLAS-SUSY-2018-08	2 OS ℓ	139.0	✓		✓	
ATLAS-SUSY-2018-09	2 SS ℓ	139.0	✓			
ATLAS-SUSY-2018-10	1ℓ + jets	139.0	✓		✓	
ATLAS-SUSY-2018-12	0ℓ + 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 soft ℓ + 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-jets + 2 h	139.0	✓		✓	Pyhf
ATLAS-SUSY-2018-32	2 OS ℓ	139.0	✓		✓	Pyhf
ATLAS-SUSY-2018-40	2 b-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 soft ℓ, EWK	139.0	✓		✓	SLv1
ATLAS-SUSY-2019-08	$1\ell + h(bb)$, EWK	139.0	✓		✓	Pyhf
ATLAS-SUSY-2019-09	3ℓ, 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

Combos

Likelihoods

BETTER

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

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 ineffciencies, etc etc) are described via “auxiliary measurements” –> keeping things frequentist (if we wish)!

Statistics Standardization comittee 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.:

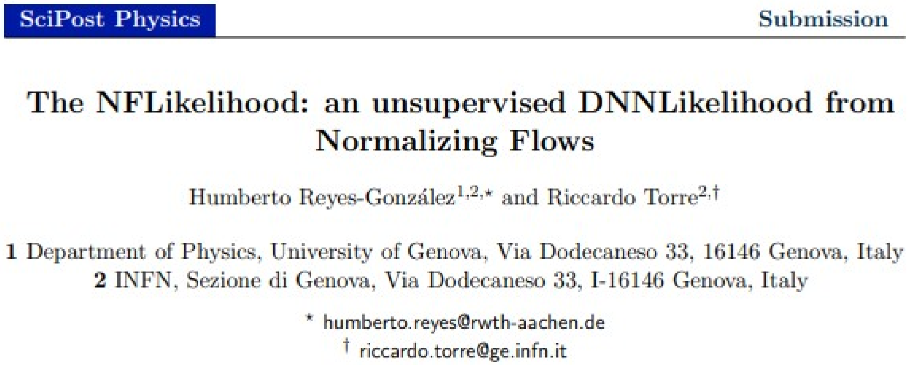

CERN-TH-2019-187

The DNNLikelihood: enhancing likelihood distribution with Deep Learning

Andrea Coccaro[a], Maurizio Pierini[b], Luca Silvestrini[b,c], and Riccardo Torre[a,b]

[a] INFN, Sezione di Genova, Via Dodecaneso 33, I-16146 Genova, Italy
[b] CERN, 1211 Geneva 23, Switzerland
[c] 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:

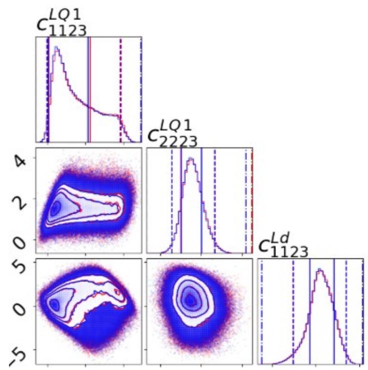

SciPost Physics Submission

The NFLikelihood: an unsupervised DNNLikelihood from Normalizing Flows

Humberto Reyes-González[1,2,*] and Riccardo Torre[2,†]

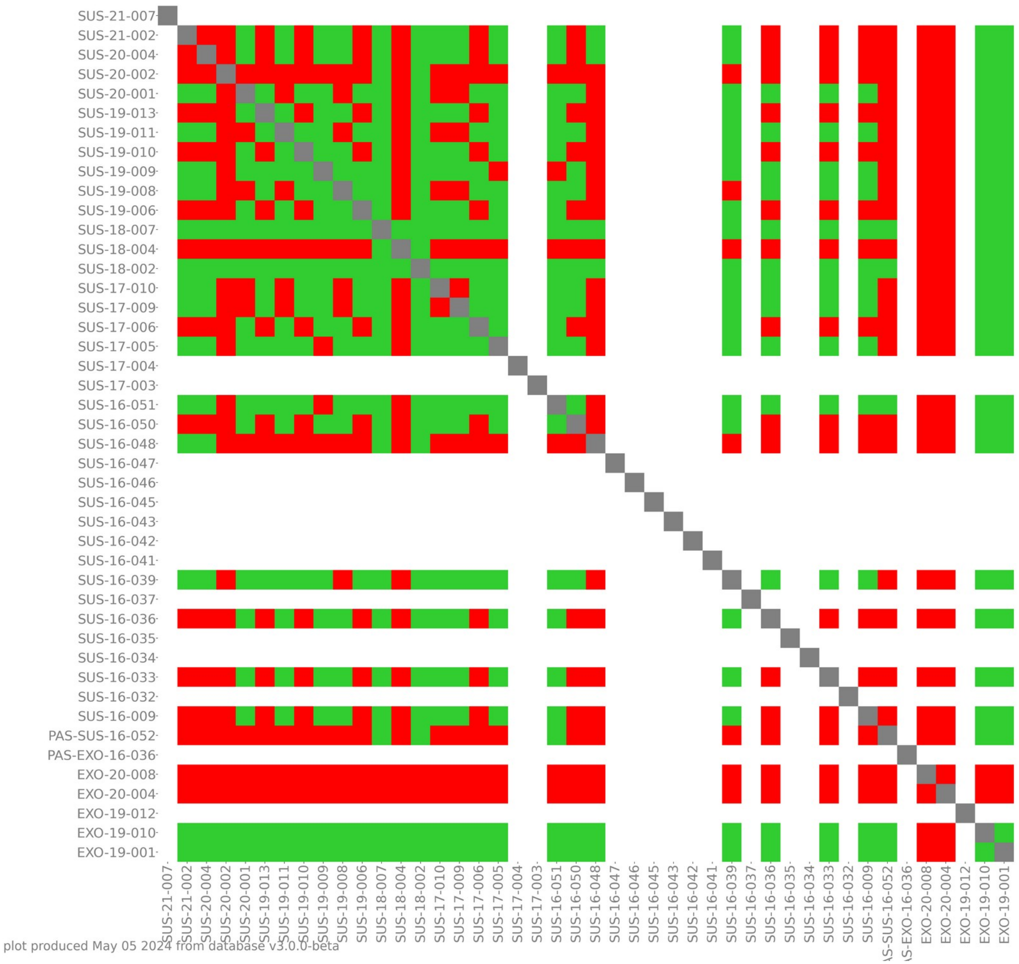
1 Department of Physics, University of Genova, Via Dodecaneso 33, 16146 Genova, Italy
2 INFN, Sezione di Genova, Via Dodecaneso 33, I-16146 Genova, Italy

* humberto.reyes@rwth-aachen.de
† 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

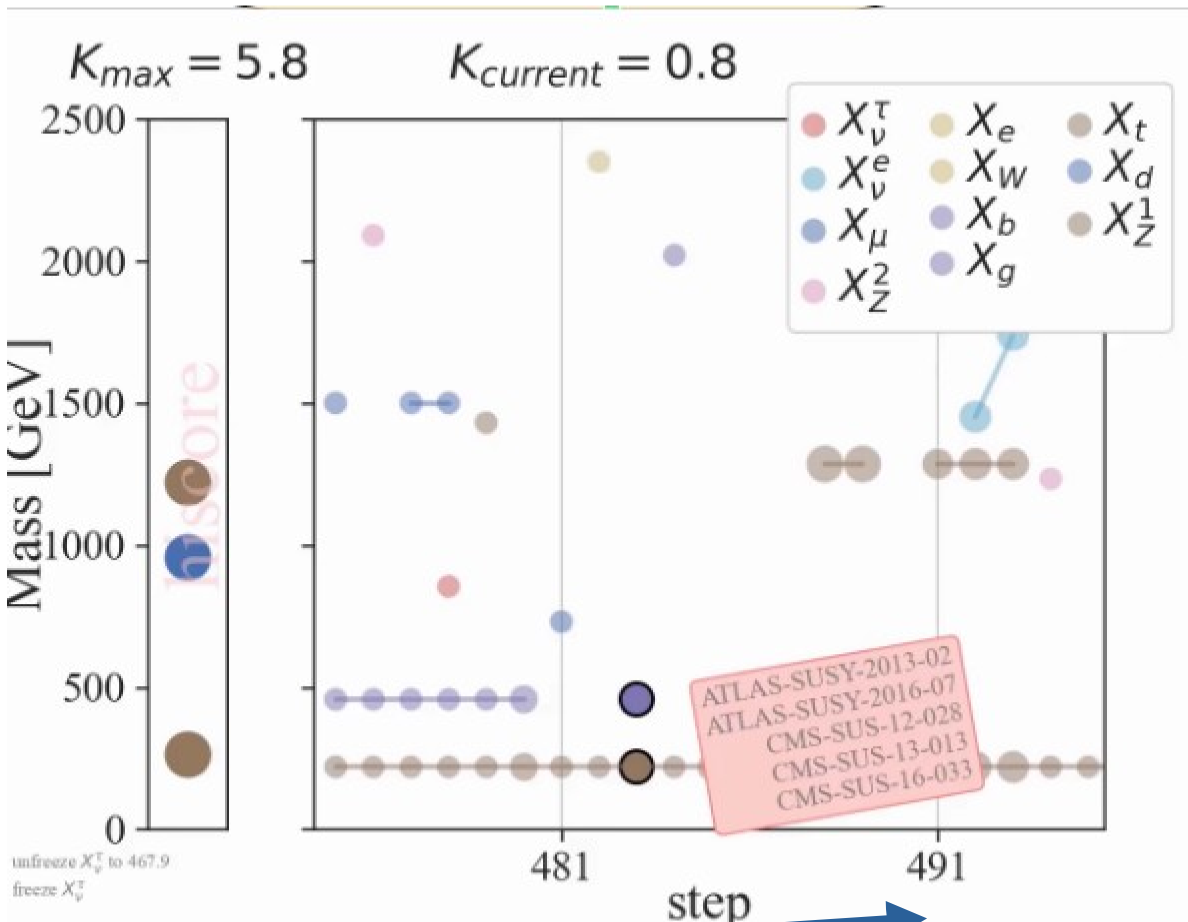
In order to facilitate reproduction of an entire analysis – to e.g. apply it to a new model, or to study analysis overap (previous slide), the analysis needs to be documented in great detail:

- 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

the test statistic

Particle spectra



A hiscore protomodel

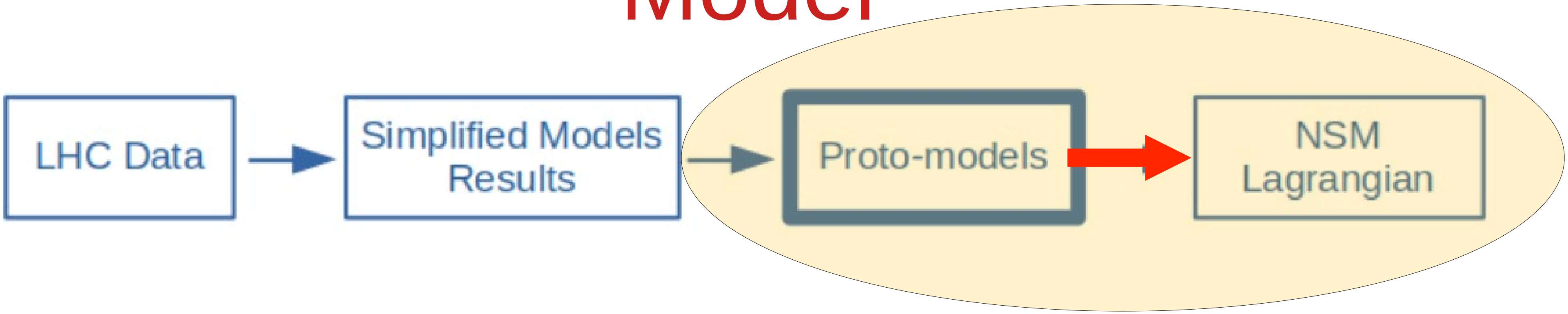
potential dispersed signals

Random modifications

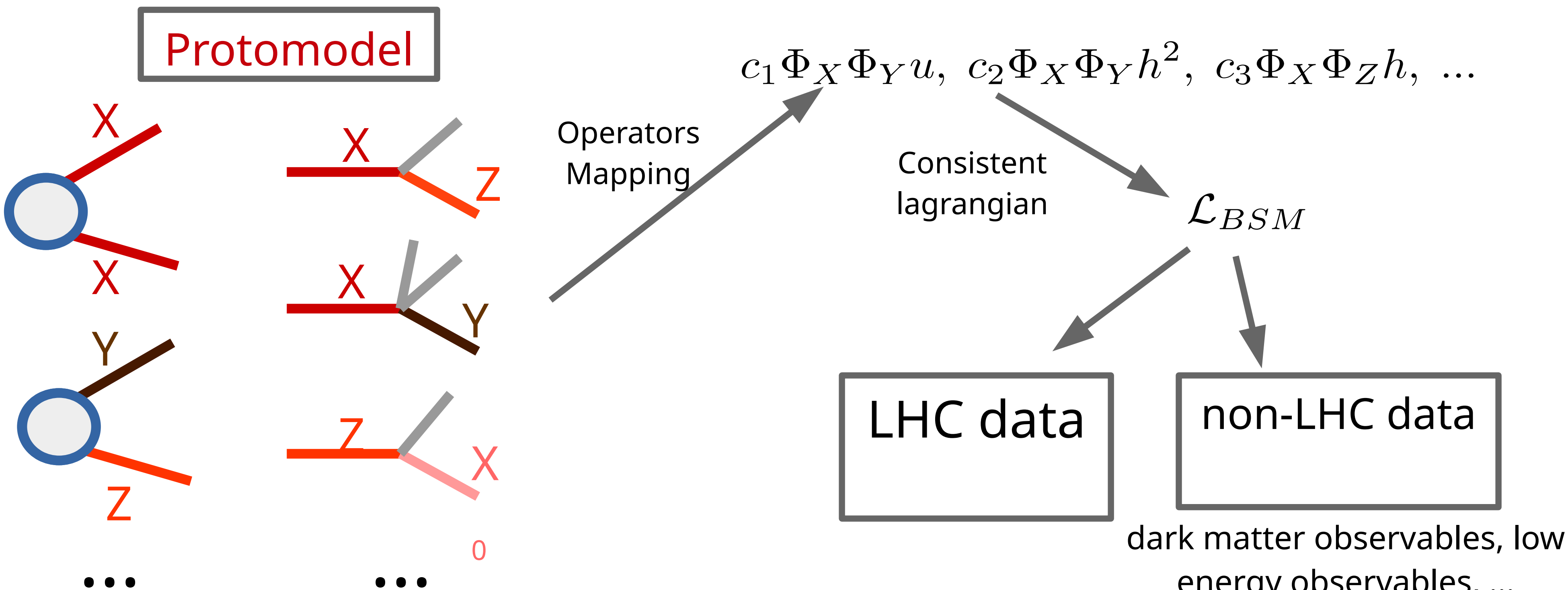
an MCMC-like walk

https://smodels.github.io/protomodels/videos

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 indispensible 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).

LHC Data → Simplified Models Results → **Proto-models** → NSM Lagrangian

- LHC Data: described as likelihoods L that are differentiable with respect to the yields y_i
- Simplified Models Results: SModelS is – in principle -- differentiable
- Proto-models: that's just a sum of simplified models → differentiable!
- NSM Lagrangian: for individual candidates we can make this differentiable w.r.t fundamental parameters Θ_l, via neural networks

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

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