Artificial Proto-Modelling:

Building Precursors of a Next Standard Model from Simplified Models Results

PROTOMODEL BUILDER

arXiv:2012.12246

JHEP 03 (2021) 207

vCHEP,
May 2021

HEPHY INSTITUT FÜR HOCHENERGIEPHYSIK

CCNH CENTRO DE CIÊNCIAS NATURAIS E HUMANAS Universidade Federal do ABC

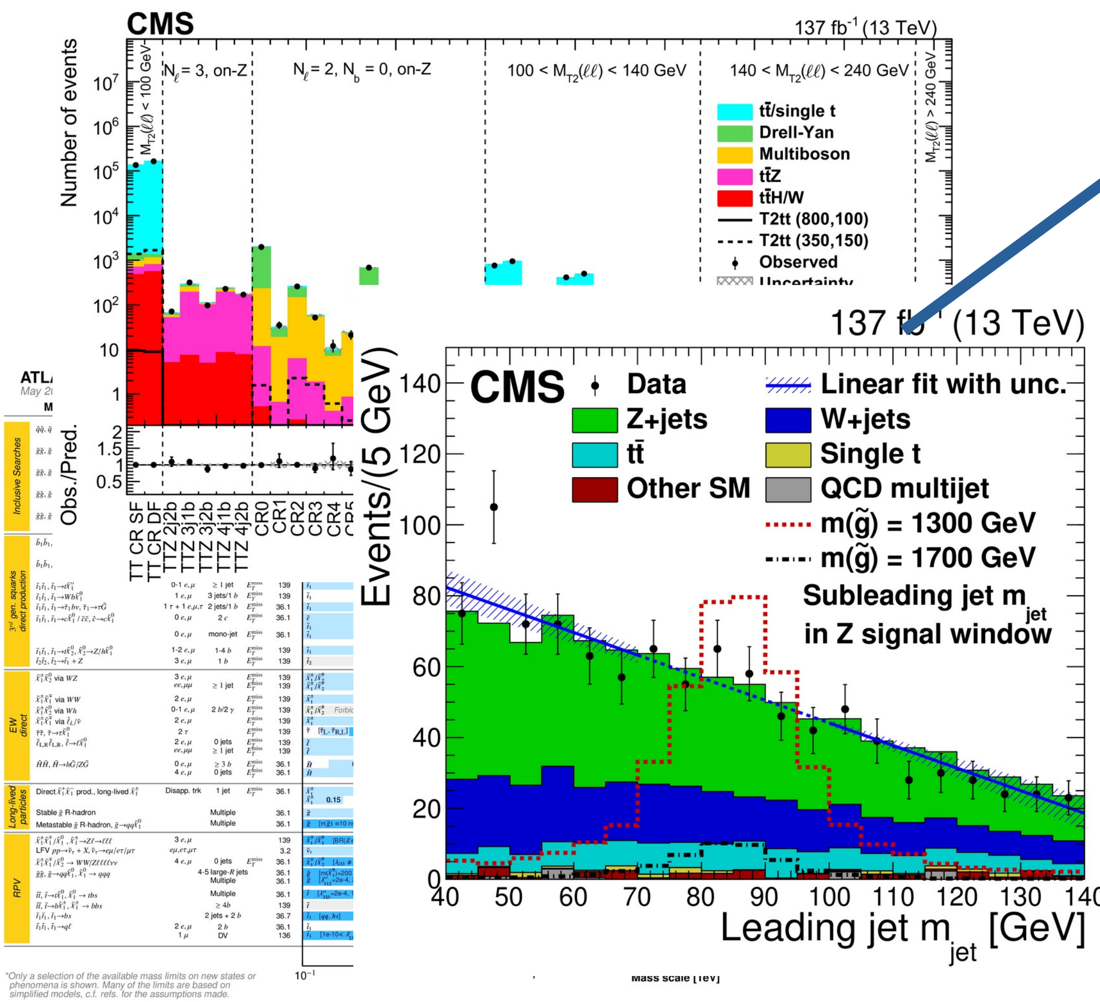
LPSC GRENOBLE | MODANE

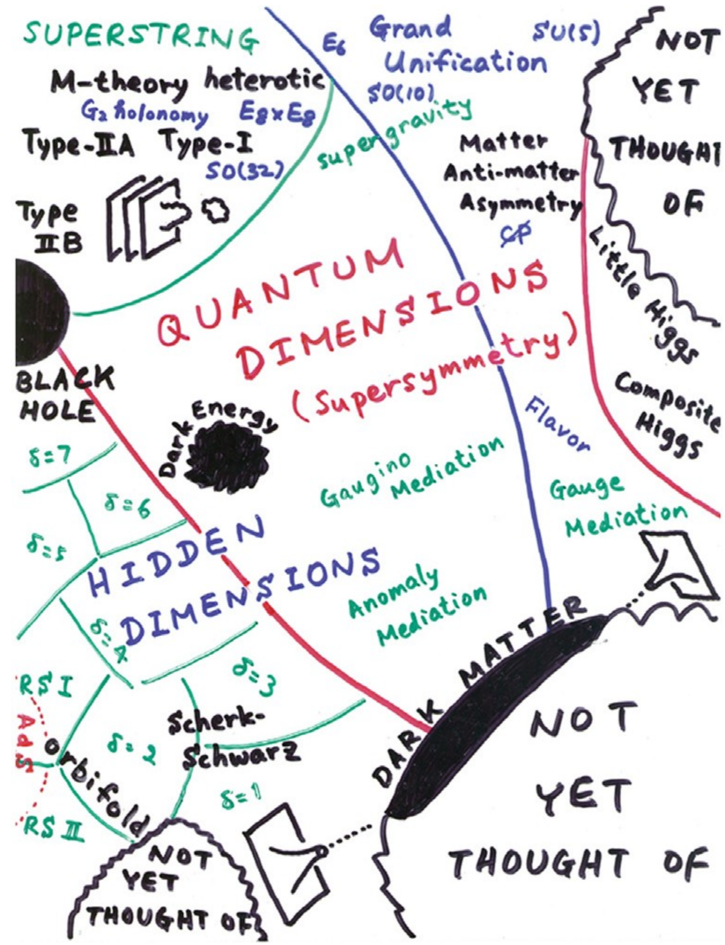
Wolfgang Waltenberger (ÖAW and Uni Vienna),
Andre Lessa (UFABC São Paulo), Sabine Kraml (LPSC Grenoble)

https://smodels.github.io/protomodels

Problem statement

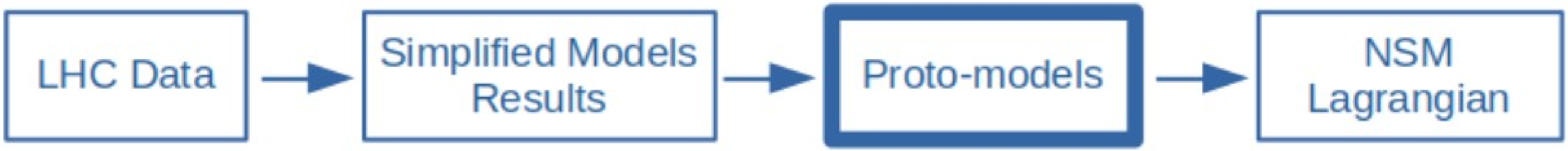
How will we infer the right hypothetical Next Standard Model (NSM) from this deluge of experimental results? Classical hypothesis testing might not anymore do the trick.





→ The Inverse Problem of Particle Physics

Our Approach

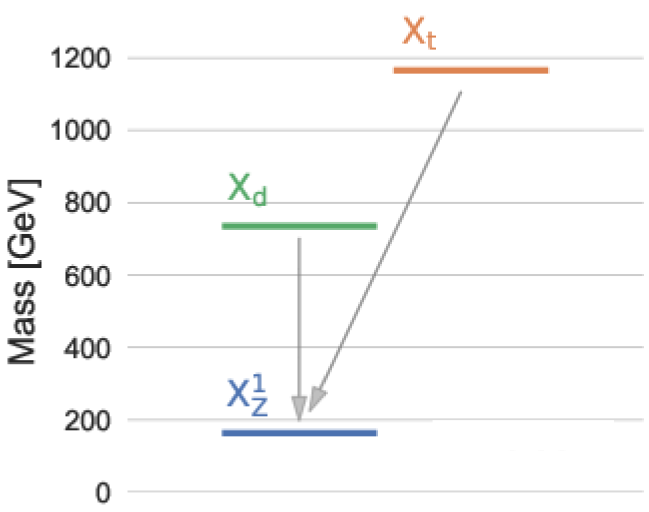


Instead of testing BSM scenarios one-by-one against the experimental data:

- identify dispersed signals in the slew of published LHC analyses
- build candidate “proto-models” from them.

MCMC-like random walk
through ”proto-model” space of:

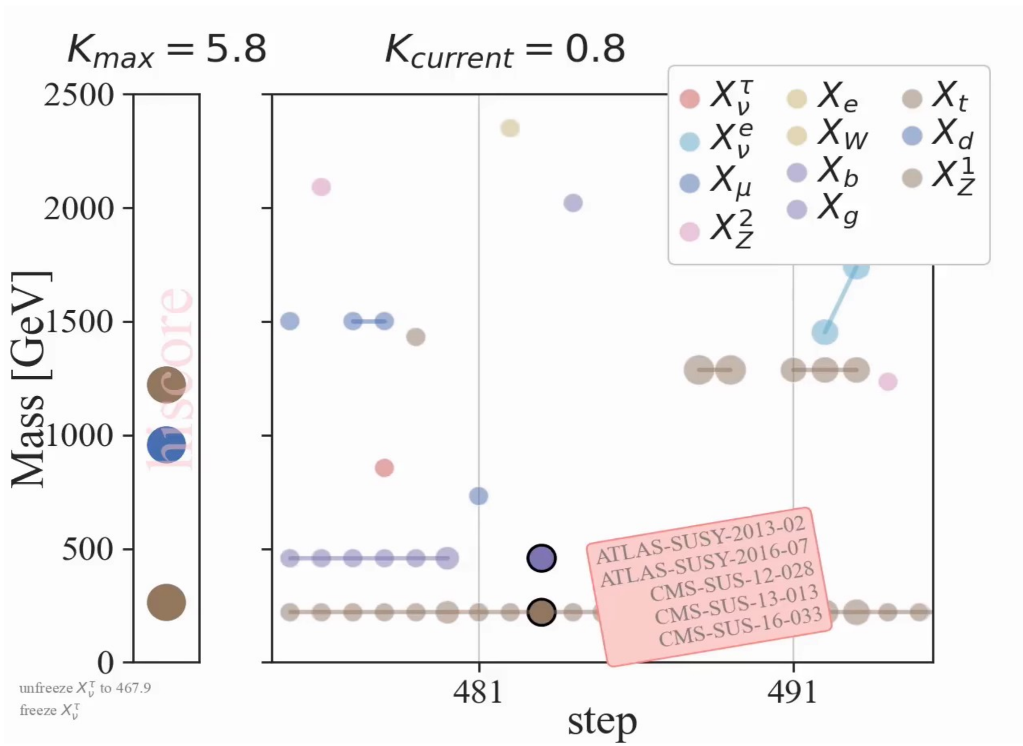
- particle content
- masses
- signal strengths [!]
- branching ratios



Our Approach

a test statistic

Particle spectra



A hiscore protomodel

potential dispersed signals

Random modifications

an MCMC-like walk

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

The Test Statistic

The test statistic **K**^{c} of a protomodel **BSM** for a “complete” set **c** of approximately uncorrelated results

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

Priors of the models used to penalize for model complexity, similar to an AIC.

Joint likelihoods: combining “complete” sets of results that are assumed to be approximately uncorrelated.

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

Input Data

- likelihood computation based on simplified models results in SModelS database
- vast number of efficiency and upper limit maps from 47 CMS and 48 ATLAS publications.

CMS 13 TeV

#	ID	Short Description	Type	$\mathcal{L}$ [fb^{-1}]
1	CMS-PAS-EXO-16-036	hscp search	ul, eff	12.9
2	CMS-PAS-SUS-16-052	soft l, <= 2 jets	ul, eff	35.9
3	CMS-SUS-16-009	multijets + $\not{E}_T$, top tagging	ul	2.3
4	CMS-SUS-16-032	Sbottom and compressed stop	ul	35.9
5	CMS-SUS-16-033	0ℓ + jets + $\not{E}_T$	ul, eff	35.9
6	CMS-SUS-16-034	2 OSSF l's	ul	35.9
7	CMS-SUS-16-035	2 SS l's	ul	35.9
8	CMS-SUS-16-036	0ℓ + jets + $\not{E}_T$	ul	35.9
9	CMS-SUS-16-037	1ℓ + jets + $\not{E}_T$ with MJ	ul	35.9
10	CMS-SUS-16-039	multi-l EWK searches	ul	35.9
11	CMS-SUS-16-041	multi-ls + jets + $\not{E}_T$	ul	35.9
12	CMS-SUS-16-042	1ℓ + jets + $\not{E}_T$	ul	35.9
13	CMS-SUS-16-043	EWK WH	ul	35.9
14	CMS-SUS-16-045	Sbottom to bHbH and H $\rightarrow \gamma\gamma$	ul	35.9
15	CMS-SUS-16-046	γ + $\not{E}_T$	ul	35.9
16	CMS-SUS-16-047	γ + HT	ul	35.9
17	CMS-SUS-16-049	All hadronic stop	ul	35.9
18	CMS-SUS-16-050	0ℓ + top tag	ul	35.9
19	CMS-SUS-16-051	1ℓ stop	ul	35.9
20	CMS-SUS-17-001	Stop search in dil + jets + $\not{E}_T$	ul	35.9
21	CMS-SUS-17-003	2 taus + $\not{E}_T$	ul	35.9
22	CMS-SUS-17-004	EW-ino combination	ul	35.9
23	CMS-SUS-17-005	1ℓ + multijets + $\not{E}_T$, top tagging	ul	35.9
24	CMS-SUS-17-006	jets + boosted H(bb) + $\not{E}_T$	ul	35.9
25	CMS-SUS-17-009	SFOS l's + $\not{E}_T$	ul	35.9
26	CMS-SUS-17-010	2L stop	ul	35.9
27	CMS-SUS-18-002	γ, jets, b-jets+ $\not{E}_T$, top tagging	ul	35.9
28	CMS-SUS-19-006	0ℓ + jets, MHT	ul	137.0

CMS 8 TeV

#	ID	Short Description	Type	$\mathcal{L}$ [fb^{-1}]
			ul	18.8
			eff	18.8
			eff	19.4
		+ $\not{E}_T$	ul, eff	19.7
			ul	19.4
			ul	18.9
		T	ul, eff	19.4
			ul	11.7
			ul	19.5
			ul	19.3
		H	ul	19.5
			ul, eff	19.3
			ul, eff	19.5
			ul, eff	19.5
			ul, eff	19.5
			ul	19.5
17	CMS-SUS-14-010	b-jets + 4 Ws	ul	19.5
18	CMS-SUS-14-021	soft l's, low n_{jets}, high $\not{E}_T$	ul	19.7

ATLAS 13 TeV

#	ID	Short Description	Type	$\mathcal{L}$ [fb^{-1}]
1	ATLAS-SUSY-2015-01	2 b-jets + $\not{E}_T$	ul	3.2
2	ATLAS-SUSY-2015-02	single l stop	ul, eff	3.2
3	ATLAS-SUSY-2015-06	0 l's + 2-6 jets + $\not{E}_T$	eff	3.2
4	ATLAS-SUSY-2015-09	jets + 2 SS l's or >=3 l's	ul	3.2
5	ATLAS-SUSY-2016-07	0ℓ + jets + $\not{E}_T$	ul, eff	36.1
6	ATLAS-SUSY-2016-14	2 SS or 3 l's + jets + $\not{E}_T$	ul	36.1
7	ATLAS-SUSY-2016-15	0ℓ stop	ul	36.1
8	ATLAS-SUSY-2016-16	1ℓ stop	ul, eff	36.1
9	ATLAS-SUSY-2016-17	2 opposite sign l's + $\not{E}_T$	ul	36.1
10	ATLAS-SUSY-2016-19	stops to staus	ul	36.1
11	ATLAS-SUSY-2016-24	2-3 l's + $\not{E}_T$, EWino	ul, eff	36.1
12	ATLAS-SUSY-2016-26	>=2 c jets + $\not{E}_T$	ul	36.1
13	ATLAS-SUSY-2016-27	jets + γ + $\not{E}_T$	ul, eff	36.1
14	ATLAS-SUSY-2016-28	2 b-jets + $\not{E}_T$	ul	36.1
15	ATLAS-SUSY-2016-33	2 OSSF l's + $\not{E}_T$	ul	36.1
16	ATLAS-SUSY-2017-01	EWK WH(bb) + $\not{E}_T$	ul	36.1
17	ATLAS-SUSY-2017-02	0ℓ + jets + $\not{E}_T$	ul	36.1
18	ATLAS-SUSY-2017-03	multi-l EWK searches	ul	36.1
19	ATLAS-SUSY-2018-04	2 hadronic taus	ul, eff	139.0
20	ATLAS-SUSY-2018-06	3 l's EW-ino	ul	139.0
21	ATLAS-SUSY-2018-31	2b + 2H(bb) + $\not{E}_T$	ul, eff	139.0
22	ATLAS-SUSY-2018-32	2 OS l's + $\not{E}_T$	ul	139.0
23	ATLAS-SUSY-2019-08	1ℓ + higgs + $\not{E}_T$	ul, eff	139.0

ATLAS 8 TeV

#	ID	Short Description	Type	$\mathcal{L}$ [fb^{-1}]
		+ $\not{E}_T$	ul	20.7
		$\not{E}_T$	ul	20.1
			ul	20.3
			ul, eff	20.3
		+ $\not{E}_T$	ul, eff	20.3
			ul, eff	20.1
			ul	20.3
			ul	20.3
			ul, eff	20.3
			ul	20.3
		T	ul, eff	20.3
		T	ul, eff	20.1
		+ $\not{E}_T$	ul, eff	20.1
		$\not{E}_T$	ul	20.3
		T	eff	20.3
16	ATLAS-SUSY-2013-23	1ℓ + 2 b-jets (or 2 γs) + $\not{E}_T$	ul	20.3
17	ATLAS-SUSY-2014-03	>= 2(c-)jets + $\not{E}_T$	eff	20.3

- Ideal case: digitized results on HepData
- Sometimes root files on the collaboration's wiki pages
- Otherwise extract information from pdf plots

- Since SModelS v2.0.0: binary version of our database on zenodo
- a text-based human-readable version on github

LIKELIHOODS

Depending on how much information we have, we can construct approximate likelihoods at different levels of “crudeness”

- Only exclusion lines
  If only exclusion lines are given, without upper limits, we can do nothing

- Observed 95% CL upper limits 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

- Efficiency maps
  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

Combos

Likelihoods

BETTER

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

7

The Combiner

Analyses that look at different chunks of LHC data[*] are allowed to be combined.

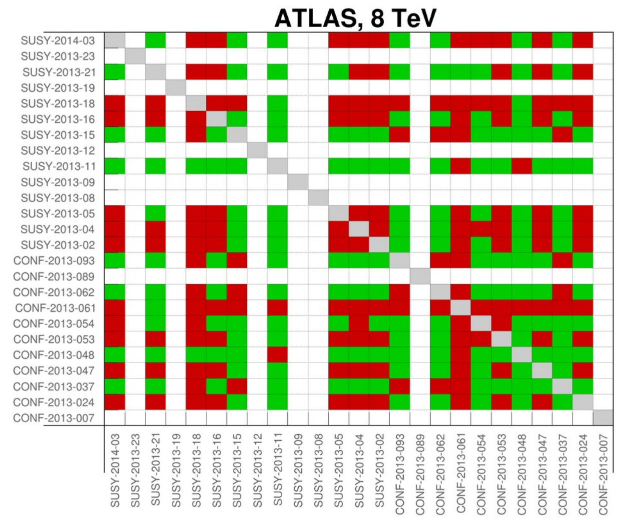
green: approximately uncorrelated → combinable

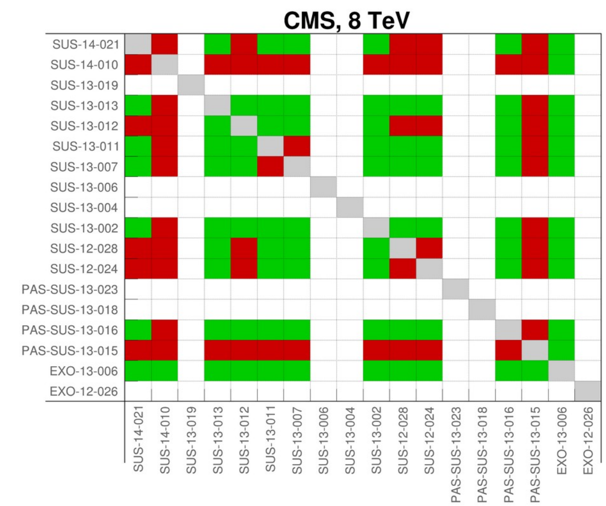
red: correlated, not combinable

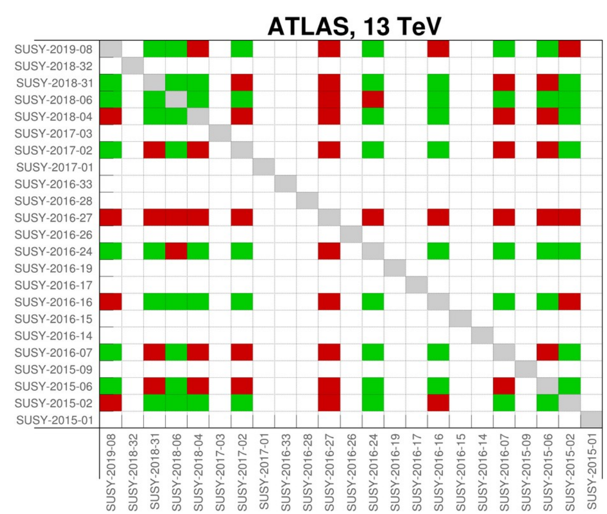
White: cannot construct a likelihood

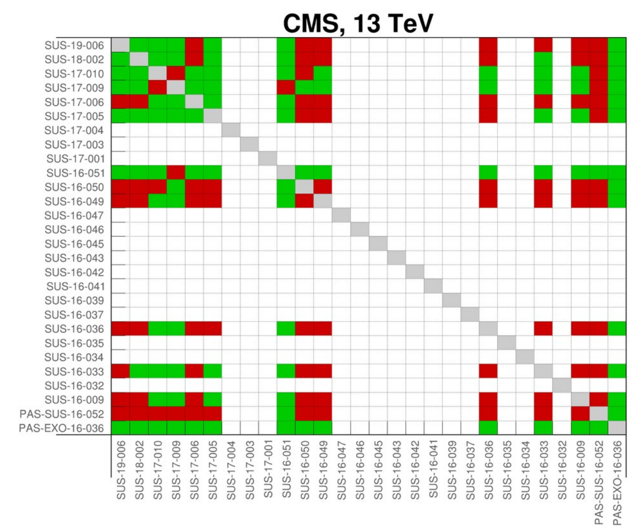
Signal regions within each analysis: correlated

Analyses from different runs or experiments are treated as ~ uncorrelated





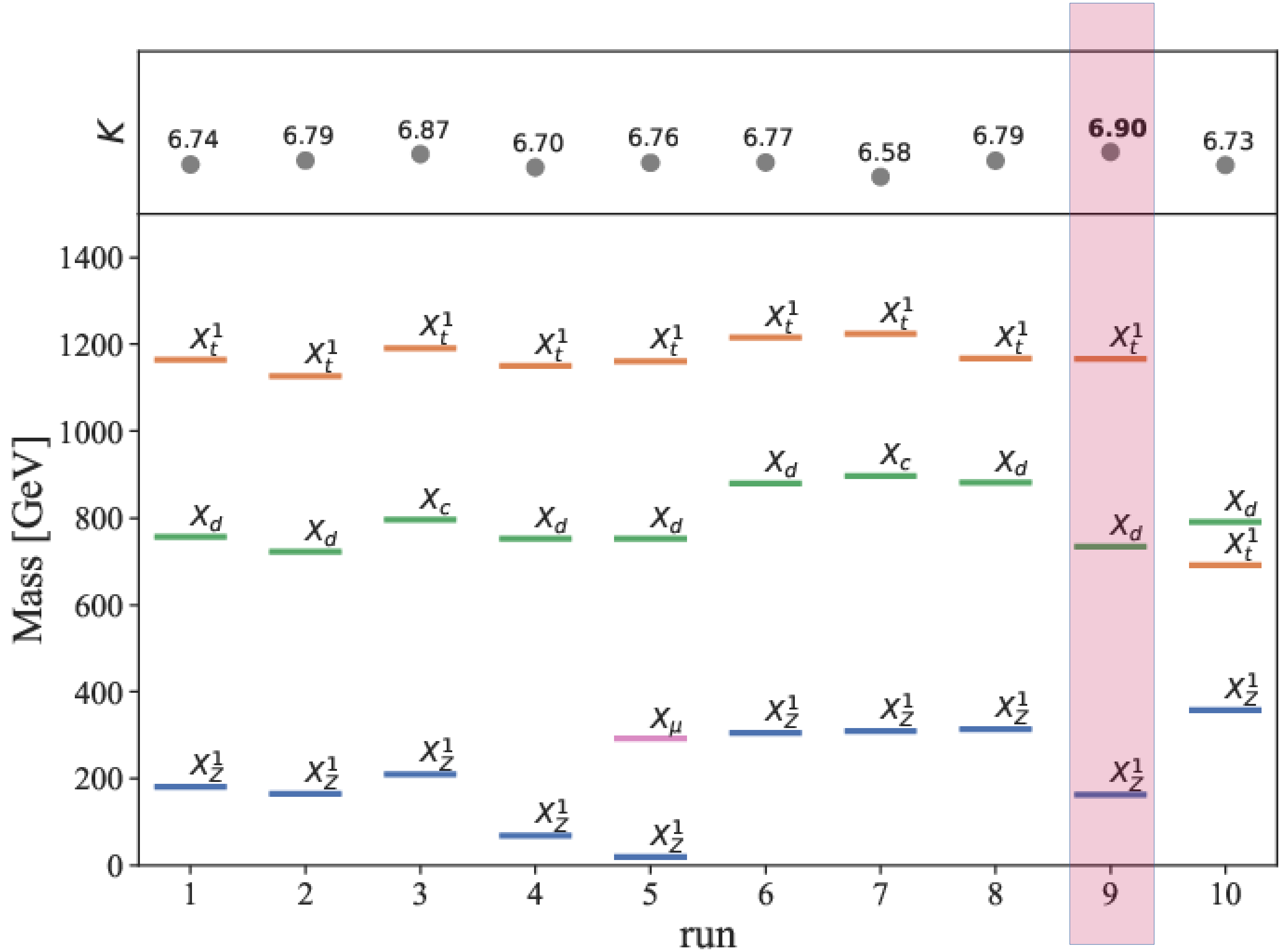




[*] we only look at signal regions, ignore control regions. Combination = multiplication of likelihoods.

Results

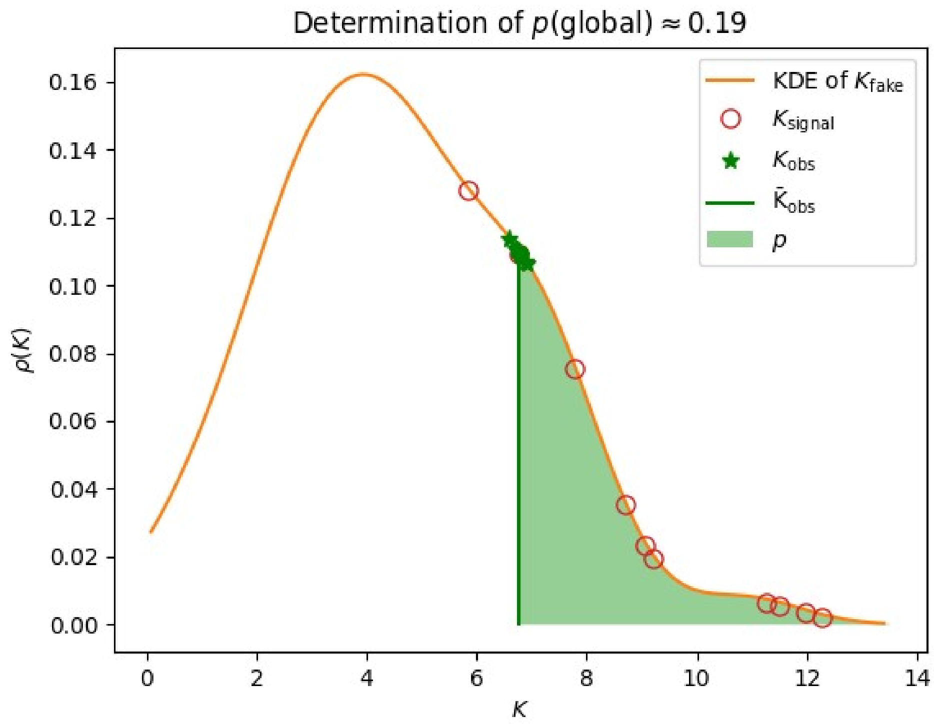
We defined a "run" as 50 parallel walkers, making 1,000 steps each. We performed 10 such runs on the SModelS database:



All 10 runs introduced a top partner as well as a light quark partner. The cross sections are compatible with values expected from the MSSM. The best test statistic was K=6.9.

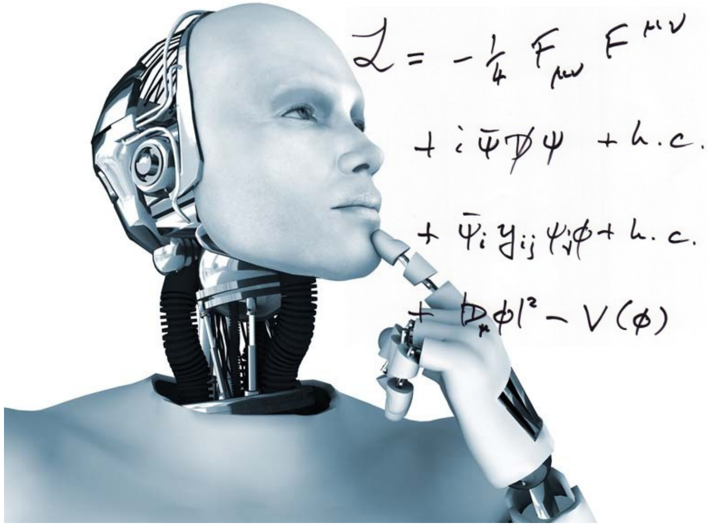
Global p-Value

- We sample from the probability models of the results in our database to synthesize "Standard Model-only" data.
- Running our algorithm over these "fake" data we can compute a global *p*-value for the Standard Model hypothesis: p(global) ~ 0.19



Because we are confident that this quantity is estimated conservatively, we claim to observe a *very mild tension* with the Standard Model hypothesis. No look-elsewhere effect applies.

Conclusions



- Given the current LHC results, we think it is necessary to also take a more global approach at interpretation.

- We propose an automated, bottom-up approach at inferring the prospective Next Standard Model, with “proto-models” – precursor theories – as a next, data-driven step.

- Our prototype presented here builds on ~ 100 CMS and ATLAS simplified models results

- About **1,000,000 CPU-cores * hours** were spent for this first prototype (run on a large “slurm” cluster in Austria)

Theory and model building has arrived at big data and large scale computing

(also lattice QCD, multi-loop/multi-leg calculations, etc etc)

Backup

The Test Statistic Revisited

Remember, the test statistic was:

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

"c" is an index that runs over all "legal" combinations: legal := uncorrelated + "complete"
(results that can be added **have to be** added)

μ denotes an global signal strength multiplier – the production cross sections are free parameters.

$$\hat{\mu} \in [0, \mu_{\mathrm{max}}]$$

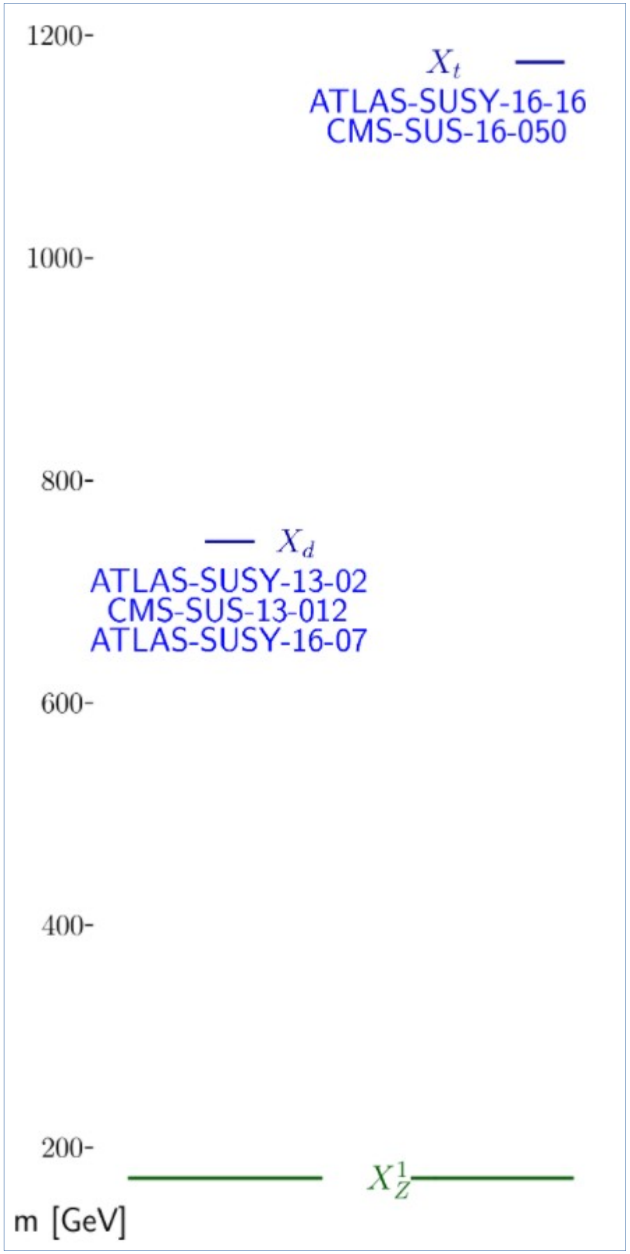
It is maximized in the denominator, but its support is restricted such that no negative results ("exclusions") in the SModelS database are violated (the "critic").

The priors π are constructed to penalize for model complexity:

$$\pi(M) = \exp\left[-\left(\frac{n_{\mathrm{particles}}}{a_1} + \frac{n_{\mathrm{BRs}}}{a_2} + \frac{n_{\mathrm{productionmodes}}}{a_3}\right)\right]$$

which boils down to a criterion that is similar to the Akaike Information Criterion (AIC)

The Hiscore Proto-Model



Analysis	Dataset	Obs	Exp	Z	P	Signal
ATL multijet, 8 TeV [54]	SR6jtp	6	4.9 ± 1.6	$0.4\ \sigma$	X_d	0.25
ATL multijet, 13 TeV [55]	2j_Me ...	611	526 ± 31	$2.2\ \sigma$	X_d	44.18
ATL 1ℓ stop, 13 TeV [48]	tN_high	8	3.8 ± 1	$1.9\ \sigma$	X_t	3.93
CMS multijet, 8 TeV [56]		30.8 fb	19.6 fb	$1.1\ \sigma$	X_d	2.66 fb
CMS 0ℓ stop, 13 TeV [49]		4.5 fb	2.5 fb	$1.6\ \sigma$	X_t	2.62 fb

Table 3: Analyses contributing to the K value of the highest score proto-model

Tension!

Table 3: the dispersed excess

Analysis (all CMS 13 TeV)	Prod	σ_{XX} (fb)	$\sigma^{\rm UL}_{\rm obs}$ (fb)	$\sigma^{\rm UL}_{\rm exp}$ (fb)	$r_{\rm obs}$
CMS multijet, M_{H_T}, 137 fb^{-1} [15]	$(\bar{X}_d, X_d)$	23.96	18.45	21.57	1.30
CMS multijet, M_{H_T}, 137 fb^{-1} [15]	$(\bar{X}_t, X_t)$	2.62	2.04	2.08	1.28
CMS multijet, M_{H_T}, 36 fb^{-1} [57]	$(\bar{X}_d, X_d)$	23.96	19.26	28.31	1.24
CMS multijet, $M_{\rm T2}$, 36 fb^{-1} [58]	$(\bar{X}_d, X_d)$	23.96	26.02	31.79	0.92
CMS 1ℓ stop, 36 fb^{-1} [59]	$(\bar{X}_t, X_t)$	2.62	2.91	4.44	0.90

Table 4: List of the most constraining results for the highest score proto-model. The

Table 4: what is driving the “critic”

Signal strength multipliers: $(\bar{X}_t, X_t) = 1.2; (\bar{X}_d, X_d), (X_d, X_Z^1), (\bar{X}_d, X_Z^1) = 0.49$

Contributions by particles: $X_t : K_{\rm without} = 2.59(59\%), X_d : K_{\rm without} = 3.90(41\%)$
Last updated: Mon Dec 14 20:08:06 2020

https://smodels.github.io/protomodels/2020_PioneerStudy/real9/index.html

The Test Statistic

For every legal combination, we define a test statistic *K*

$$K^c := -2 \ln \frac{\mathrm{L}^c_{\mathbf{SM}} \cdot \pi(\mathbf{SM})}{\mathrm{L}^c_{\mathbf{BSM}}(\hat{\mu}) \cdot \pi(\mathbf{BSM})}$$ Eq. 6

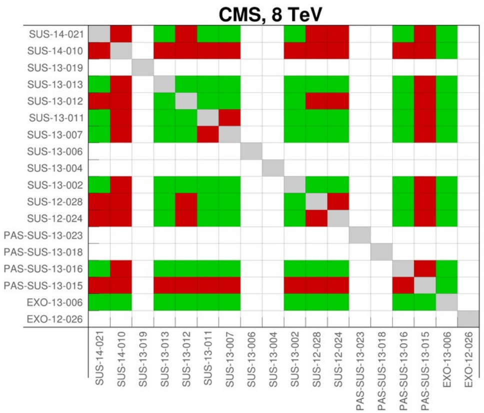
π(BSM) is the prior of the BSM model. We use it to “regularize” the model, i.e. impose the *law of parsimony*:

$$\pi(M) = \exp\left[-\left(\frac{n_{\text{particles}}}{a_1} + \frac{n_{\text{BRs}}}{a_2} + \frac{n_{\text{productionmodes}}}{a_3}\right)\right]$$ Eq. 9

That way, one new particle with one non-trivial branching ratio and two production modes is similar to one degree of freedom in Akaike’s information criterion (the sign is however flipped, and it’s a likelihood ratio), i.e. the test statistic is roughly equivalent to

$$K \approx \Delta\chi^2 - 2n_{\text{particles}}$$

An additional particle will have to increase the “(delta-)chi-square” by approximately two units.



THE COMBINER

we allow the machine
to combine likelihooods.

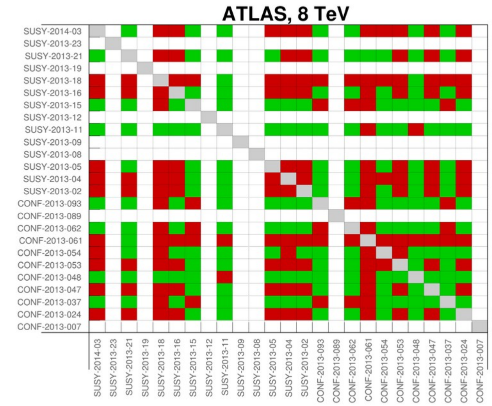
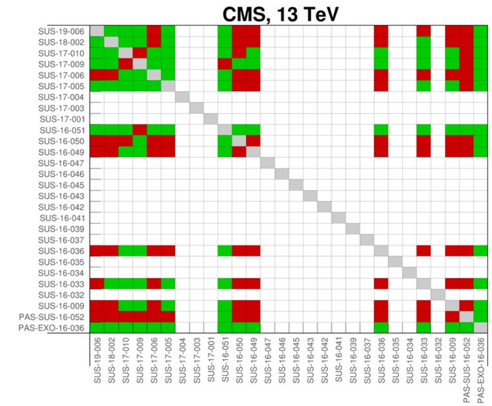


Fig. 2

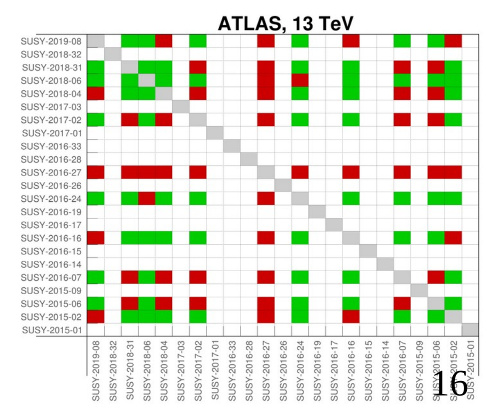
Approximately uncorrelated are analyses that are:

- from different runs, and/or
- from different experiments, and/or
- looking for (clearly) different signatures



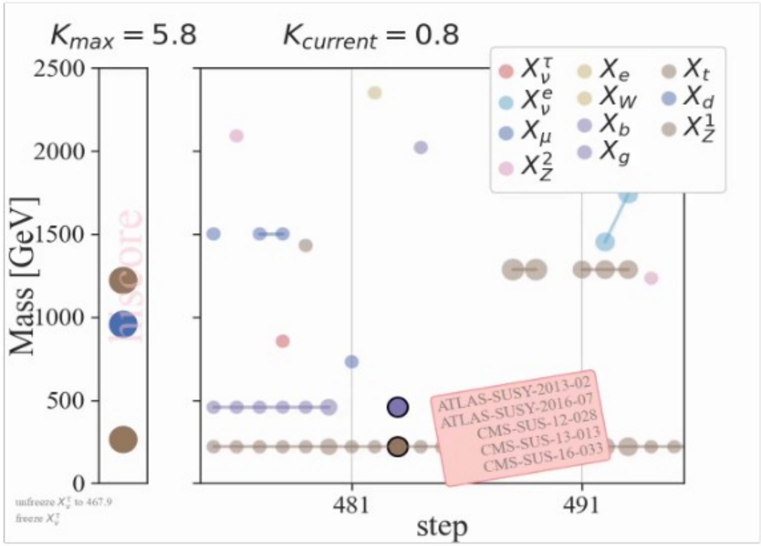
A combination “c” of analyses is “legal” if the following conditions are met:

- all results are mutually uncorrelated (= ”combinable”)
- if a result can be added, it has to be added (any subset of a legal combination is not itself legal)
- combined likelihood: $L_c = \prod_{i \in c} L_i$



The Walker

The Walker takes care of moving in the protomodel space with varying dimensionality by performing the following types of modifications to the protomodel:



- add or remove particles from the protomodel
- change the masses of particles
- change the signal strengths of production modes
- change decay channels and branching ratios

At each step the test statistic *K* is computed. An MCMC-like procedure[*] is then applied in the sense that the step is reverted with a probability of

$$\exp\left[\frac{1}{2}(K_i - K_{i-1})\right]$$

Appendix A.1

if and only if K_i is smaller than K_{i-1}

* (note however, instead of ratios of unnormalized posteriors we have ratios of ratios of unnormalized posteriors)

The Walks

We define a “run” as 50 parallel walks, each taking 1000 steps.

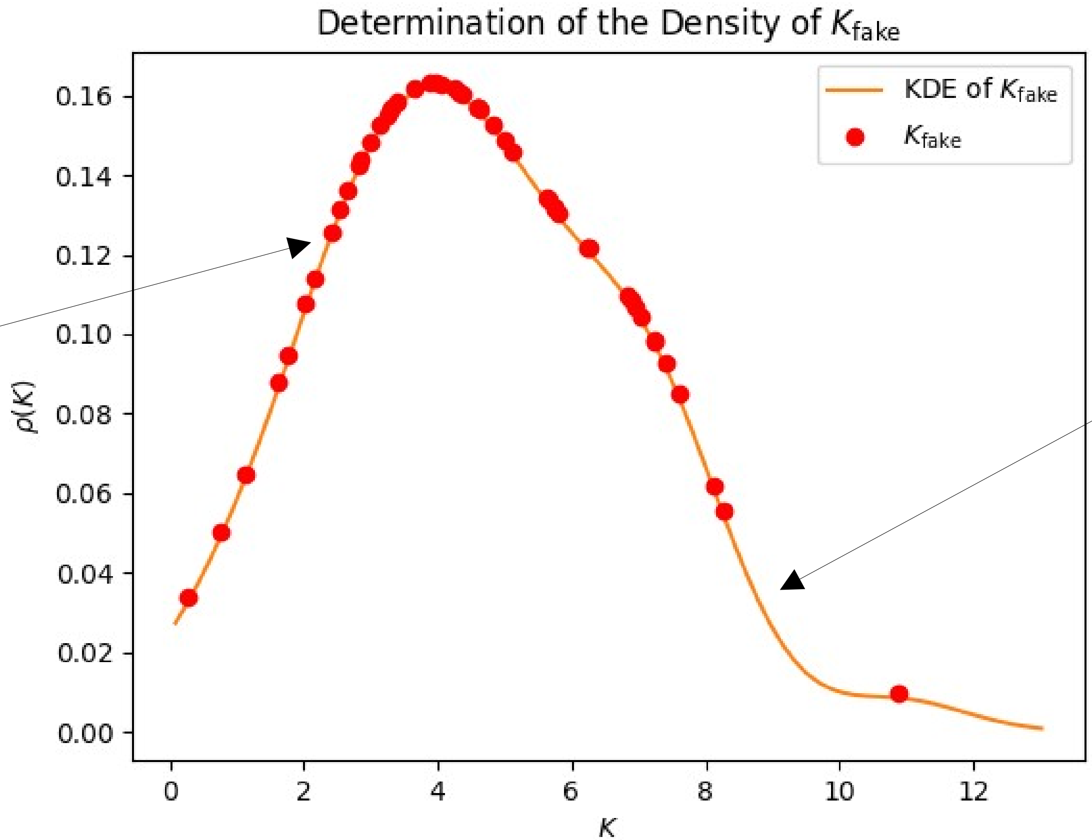
We performed

- 10 runs on the SModelS database (Sec. 5.2)
- 50 runs on fake “Standard Model-like” databases (Sec 5.1)
  to be able to determine a global *p*-value under the SM hypothesis
- 2x10 runs on fake “Signal-like” databases (Sec 5.3)
  to show closure of the method

Walking Over Fake Standard Model Databases

- Produced 50 “fake” SModelS databases by sampling background models
- Corresponds to typical LHC results if no new physics is in data
- Determine 50 “fake” K values by running 50 walkers on each of the 50 databases (50 x 50 walkers in total) → density of K under null SM-only hypothesis

K for one “fake” background-only database.



Density of K estimated via a simple Kernel density estimator.

Sec. 5.1, Fig. 6

Convergence of Method

Are we sure we found the global maximum? When scanning individual protomodel parameters, while fixing the others, it seems so:

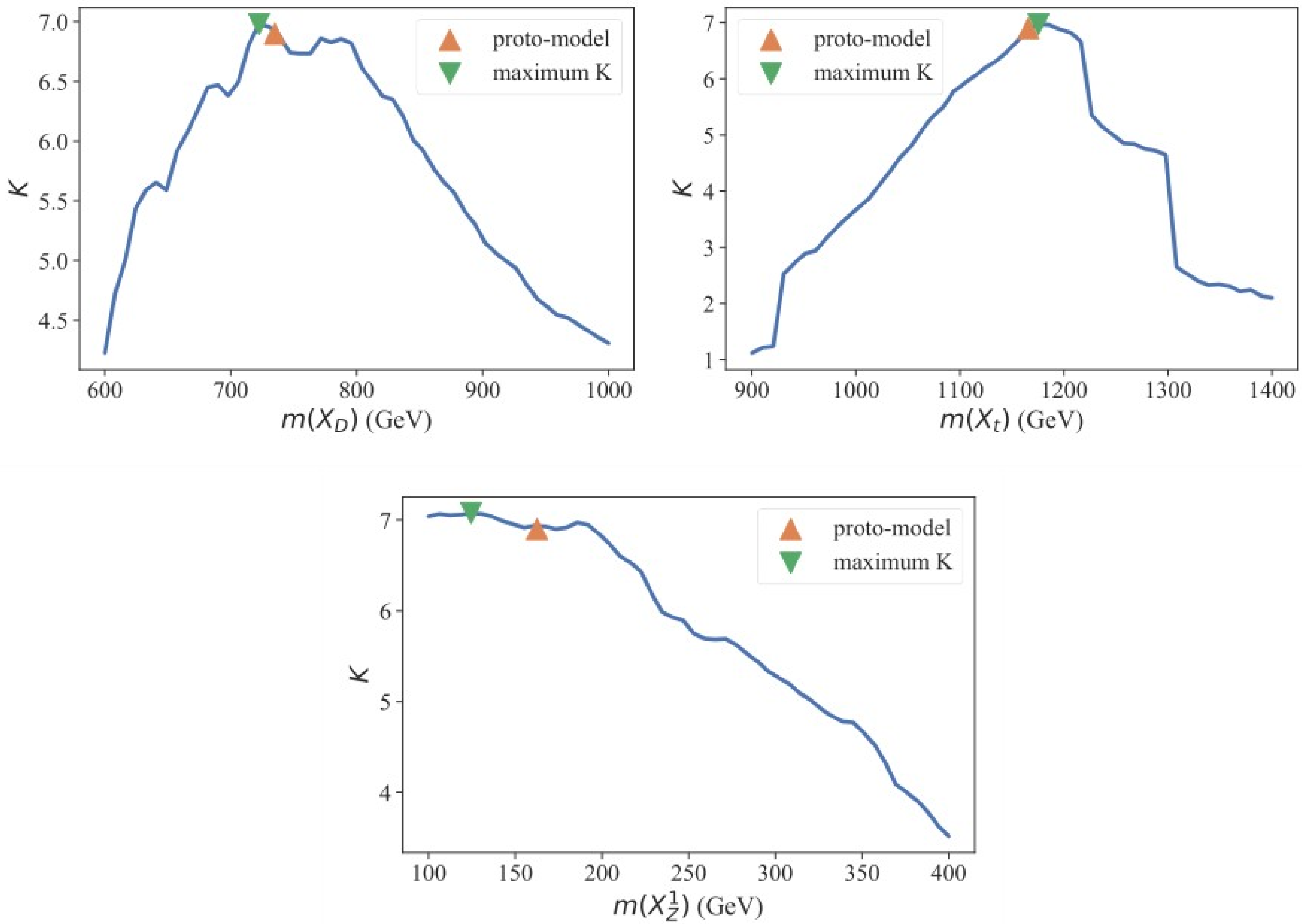
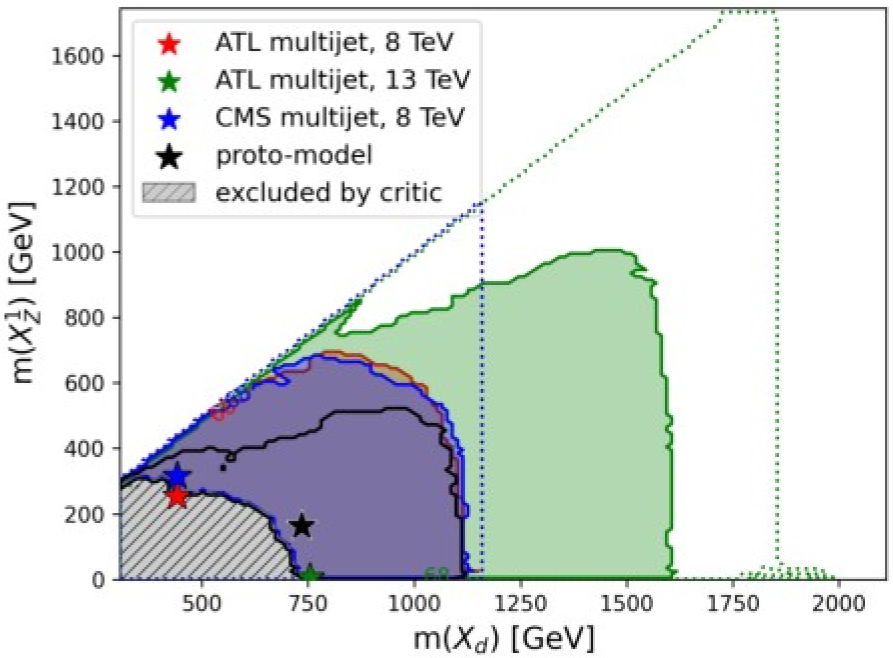
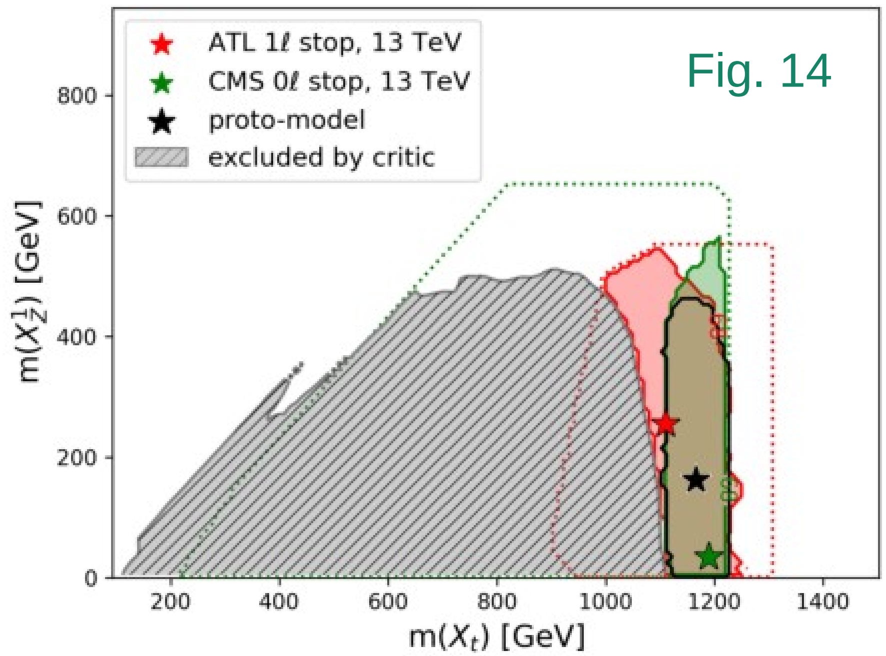


Fig. 12

Mutual (In-)Compatibility





68% Bayesian credibility regions of the particle masses, fixating all other parameters.

- very little handle on the masses
- results suffer from the fact that the efficiency and upper limit maps are limited in the mass ranges (the dashed lines are the limits of the maps). → try and fix in next iteration.
- tension between builder and critic – will understand this better with future, improved, efficiency maps
- Aim for full posteriors in next iteration of this effort

p-Values per Signal Region

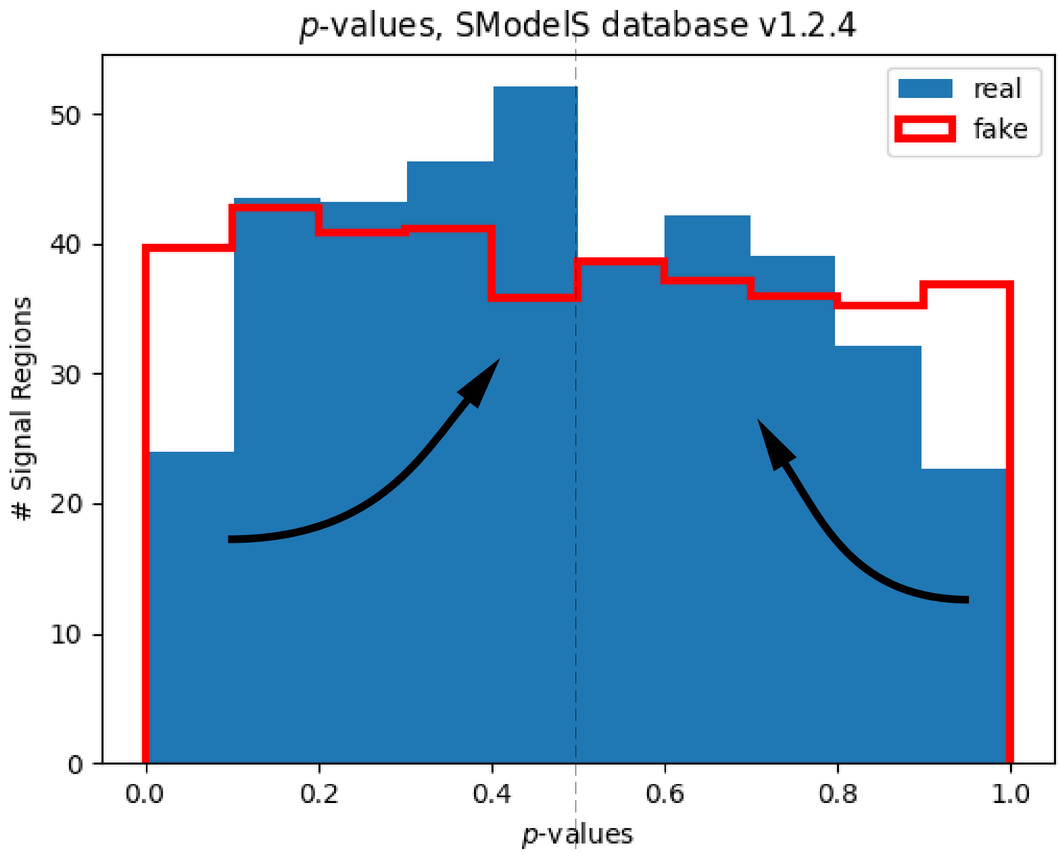


Fig. 7

- *p*-values for signal regions in SModelS database
- errors on background estimated modelled as ("single enveloping") Gaussian
- filtered out regions with expected number of the events < 3.5
- blue area is real data, red line is "fake" BG-only simulated databases
- results compatible with idea that BG errors are conservative, see also arXiv:1410.2270
- slighty more excesses ($p \rightarrow 0$) than underfluctuations ($p \rightarrow 1$)

→ p_{global} is most likely conservative!

Walking Over Databases With Fake Signals

To show closure of our method, we inject the winning protomodel as a signal in fake databases, and see if the algorithm can reconstruct the injected signal.

Sec 5.3

Technical closure test

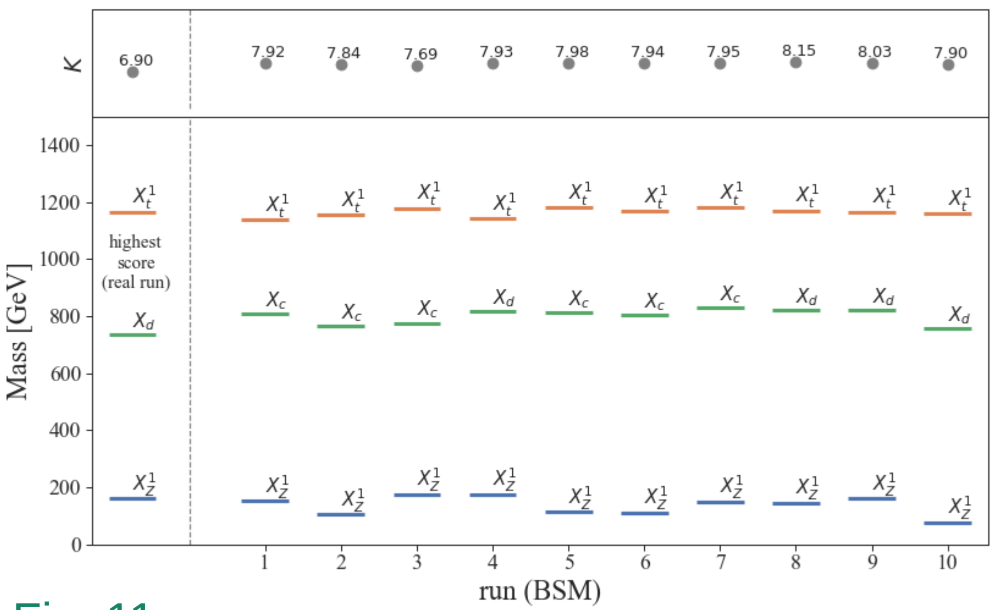


Fig. 11

No sampling of the models for the SRs, i.e. observed events := expected SM + expected signal events

Physics closure test

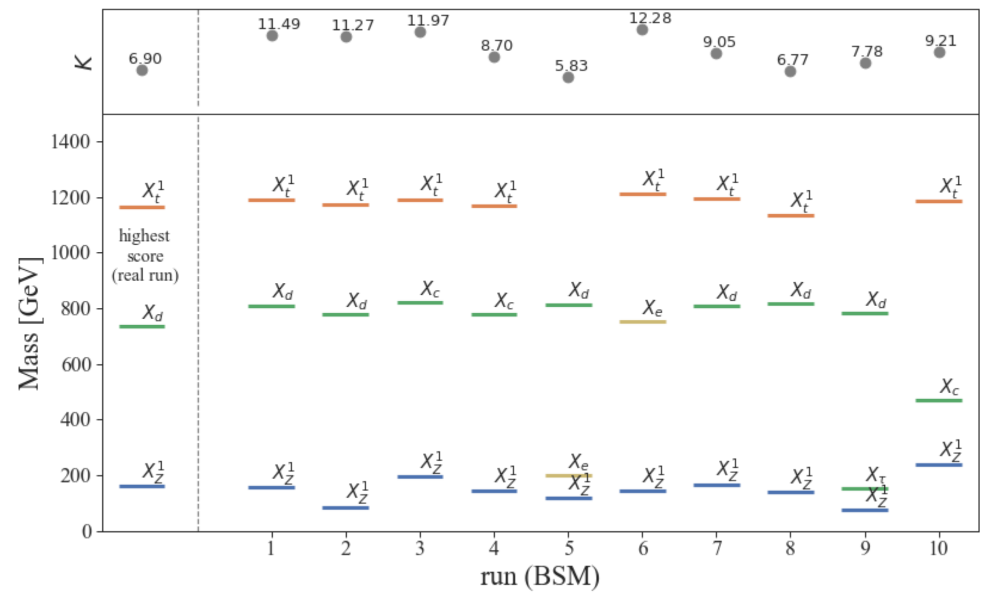


Fig. 10

Sampling turned on

Future Improvements

Improvements of the SModelS database:

- add latest full run-2 CMS and ATLAS publications (Moriond!)
- produce efficiency maps for existing results
- enlarge mass range of older efficiency maps

Improvements in speed:

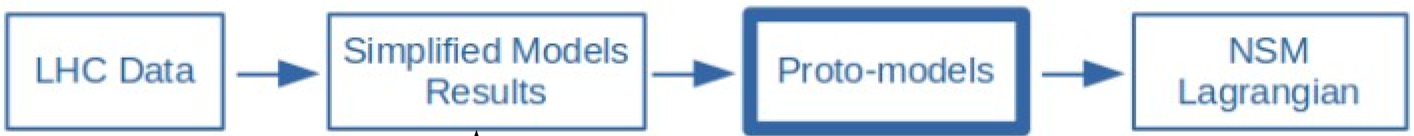
- learn the SModelS database
- make everything differentiable

Improvements in procedure:

- improve the “analyses correlation matrix”, automate the determination
- ponder relationship between proto-models and effective field theories
- connect proto-models with complete theories

Why Differentiable?

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).



described as likelihoods L that are differentiable with respect to the yields y_i

we have started an effort to make SModelS differentiable w.r.t SMS parameters p_j, by learning our entire database:

that's just a sum of simplified models → differentiable!

for individual candidates we can make this differentiable w.r.t fundamental parameters Θ_l, via neural networks, with efforts similar to DeepXS, or "TheoryGANs" [*]:

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

Needless to say, the data pipeline sketched above is not the only feasible one. Differentiability however would be a helpful tool for all possible data pipelines. A similar rationale would apply also to EFTs, Wilson coefficients and data from measurements.

https://arxiv.org/abs/1810.08312

→ Differentiable Inductive Reasoning!

Likelihoods

CMS 8 TeV

#	ID	$\mathcal{L}$ [fb^{-1}]	UL$_{\rm obs}$	UL$_{\rm exp}$	EM
1	CMS-EXO-12-026	18.8	✓		
2	CMS-EXO-13-006	18.8			✓
3	CMS-PAS-SUS-13-015	19.4			✓
4	CMS-PAS-SUS-13-016	19.7	✓		✓
5	CMS-PAS-SUS-13-018	19.4	✓		
6	CMS-PAS-SUS-13-023	18.9	✓		
7	CMS-SUS-12-024	19.4	✓		✓
8	CMS-SUS-12-028	11.7	✓	✓	
9	CMS-SUS-13-002	19.5	✓	✓	
10	CMS-SUS-13-004	19.3	✓		
11	CMS-SUS-13-006	19.5	✓		
12	CMS-SUS-13-007	19.3	✓		✓
13	CMS-SUS-13-011	19.5	✓		✓
14	CMS-SUS-13-012	19.5	✓	✓	✓
15	CMS-SUS-13-013	19.5	✓	✓	✓
16	CMS-SUS-13-019	19.5	✓		
17	CMS-SUS-14-010	19.5	✓	✓	
18	CMS-SUS-14-021	19.7	✓	✓	

CMS 13 TeV

#	ID	$\mathcal{L}$ [fb^{-1}]	UL$_{\rm obs}$	UL$_{\rm exp}$	EM	comb.
1	CMS-PAS-EXO-16-036	12.9	✓		✓	
2	CMS-PAS-SUS-16-052	35.9	✓		✓	Cov.
3	CMS-SUS-16-009	2.3	✓	✓		
4	CMS-SUS-16-032	35.9	✓			
5	CMS-SUS-16-033	35.9	✓	✓	✓	
6	CMS-SUS-16-034	35.9	✓			
7	CMS-SUS-16-035	35.9	✓			
8	CMS-SUS-16-036	35.9	✓	✓		
9	CMS-SUS-16-037	35.9	✓			
10	CMS-SUS-16-039	35.9	✓			
11	CMS-SUS-16-041	35.9	✓			
12	CMS-SUS-16-042	35.9	✓			
13	CMS-SUS-16-043	35.9	✓			
14	CMS-SUS-16-045	35.9	✓			
15	CMS-SUS-16-046	35.9	✓			
16	CMS-SUS-16-047	35.9	✓			
17	CMS-SUS-16-049	35.9	✓	✓		
18	CMS-SUS-16-050	35.9	✓	✓		
19	CMS-SUS-16-051	35.9	✓	✓		
20	CMS-SUS-17-001	35.9	✓			
21	CMS-SUS-17-003	35.9	✓			
22	CMS-SUS-17-004	35.9	✓			
23	CMS-SUS-17-005	35.9	✓	✓		
24	CMS-SUS-17-006	35.9	✓	✓		
25	CMS-SUS-17-009	35.9	✓	✓		
26	CMS-SUS-17-010	35.9	✓	✓		
27	CMS-SUS-18-002	35.9	✓	✓		
28	CMS-SUS-19-006	137.0	✓	✓		

See Tables 6 and 8

Color coding same as in previous slide

LIKELIHOODS

ATLAS 13 TeV

#	ID	$\mathcal{L}$ [fb^{-1}]	UL$_{\text{obs}}$	UL$_{\text{exp}}$	EM	comb.
1	ATLAS-SUSY-2015-01	3.2	✓			
2	ATLAS-SUSY-2015-02	3.2	✓		✓	
3	ATLAS-SUSY-2015-06	3.2			✓	
4	ATLAS-SUSY-2015-09	3.2	✓			
5	ATLAS-SUSY-2016-07	36.1	✓		✓	
6	ATLAS-SUSY-2016-14	36.1	✓			
7	ATLAS-SUSY-2016-15	36.1	✓			
8	ATLAS-SUSY-2016-16	36.1	✓		✓	
9	ATLAS-SUSY-2016-17	36.1	✓			
10	ATLAS-SUSY-2016-19	36.1	✓			
11	ATLAS-SUSY-2016-24	36.1	✓		✓	
12	ATLAS-SUSY-2016-26	36.1	✓			
13	ATLAS-SUSY-2016-27	36.1	✓		✓	
14	ATLAS-SUSY-2016-28	36.1	✓			
15	ATLAS-SUSY-2016-33	36.1	✓			
16	ATLAS-SUSY-2017-01	36.1	✓			
17	ATLAS-SUSY-2017-02	36.1	✓	✓		
18	ATLAS-SUSY-2017-03	36.1	✓			
19	ATLAS-SUSY-2018-04	139.0	✓		✓	JSON
20	ATLAS-SUSY-2018-06	139.0	✓	✓		
21	ATLAS-SUSY-2018-31	139.0	✓		✓	JSON
22	ATLAS-SUSY-2018-32	139.0	✓			
23	ATLAS-SUSY-2019-08	139.0	✓		✓	JSON

CMS 13 TeV

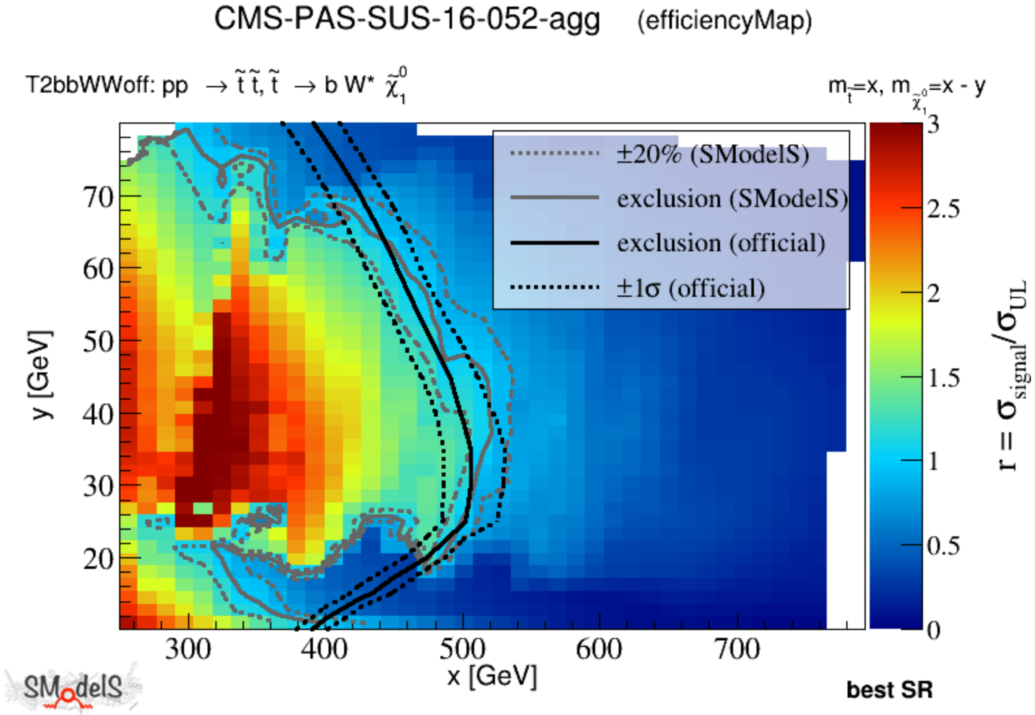
#	ID	$\mathcal{L}$ [fb^{-1}]	UL$_{\text{obs}}$	UL$_{\text{exp}}$	EM	comb.
1	CMS-PAS-EXO-16-036	12.9	✓		✓	
2	CMS-PAS-SUS-16-052	35.9	✓		✓	Cov.
3	CMS-SUS-16-009	2.3	✓	✓		
4	CMS-SUS-16-032	35.9	✓			
5	CMS-SUS-16-033	35.9	✓	✓	✓	
6	CMS-SUS-16-034	35.9	✓			
7	CMS-SUS-16-035	35.9	✓			
8	CMS-SUS-16-036	35.9	✓	✓		
9	CMS-SUS-16-037	35.9	✓			
10	CMS-SUS-16-039	35.9	✓			
11	CMS-SUS-16-041	35.9	✓			
12	CMS-SUS-16-042	35.9	✓			
13	CMS-SUS-16-043	35.9	✓			
14	CMS-SUS-16-045	35.9	✓			
15	CMS-SUS-16-046	35.9	✓			
16	CMS-SUS-16-047	35.9	✓			
17	CMS-SUS-16-049	35.9	✓	✓		
18	CMS-SUS-16-050	35.9	✓	✓		
19	CMS-SUS-16-051	35.9	✓	✓		
20	CMS-SUS-17-001	35.9	✓			
21	CMS-SUS-17-003	35.9	✓			
22	CMS-SUS-17-004	35.9	✓			
23	CMS-SUS-17-005	35.9	✓	✓		
24	CMS-SUS-17-006	35.9	✓	✓		
25	CMS-SUS-17-009	35.9	✓	✓		
26	CMS-SUS-17-010	35.9	✓	✓		
27	CMS-SUS-18-002	35.9	✓	✓		
28	CMS-SUS-19-006	137.0	✓	✓		

See Tables 5 and 8

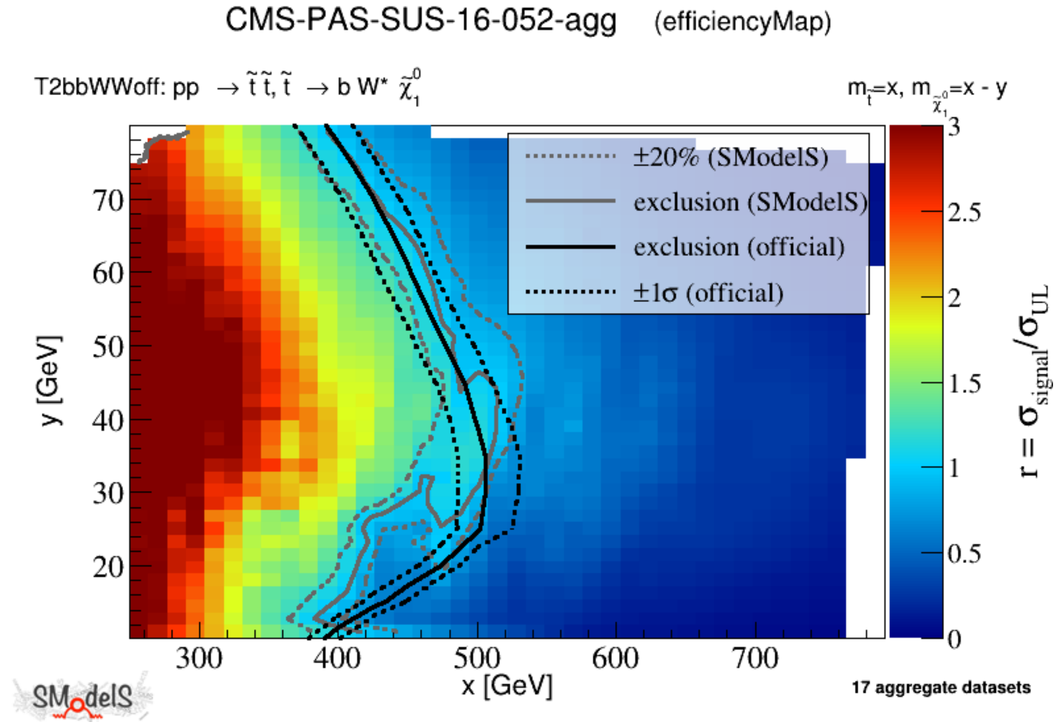
Color coding same as in previous slide

Likelihoods

Limit without combination of signal regions



Limit with combination of signal regions



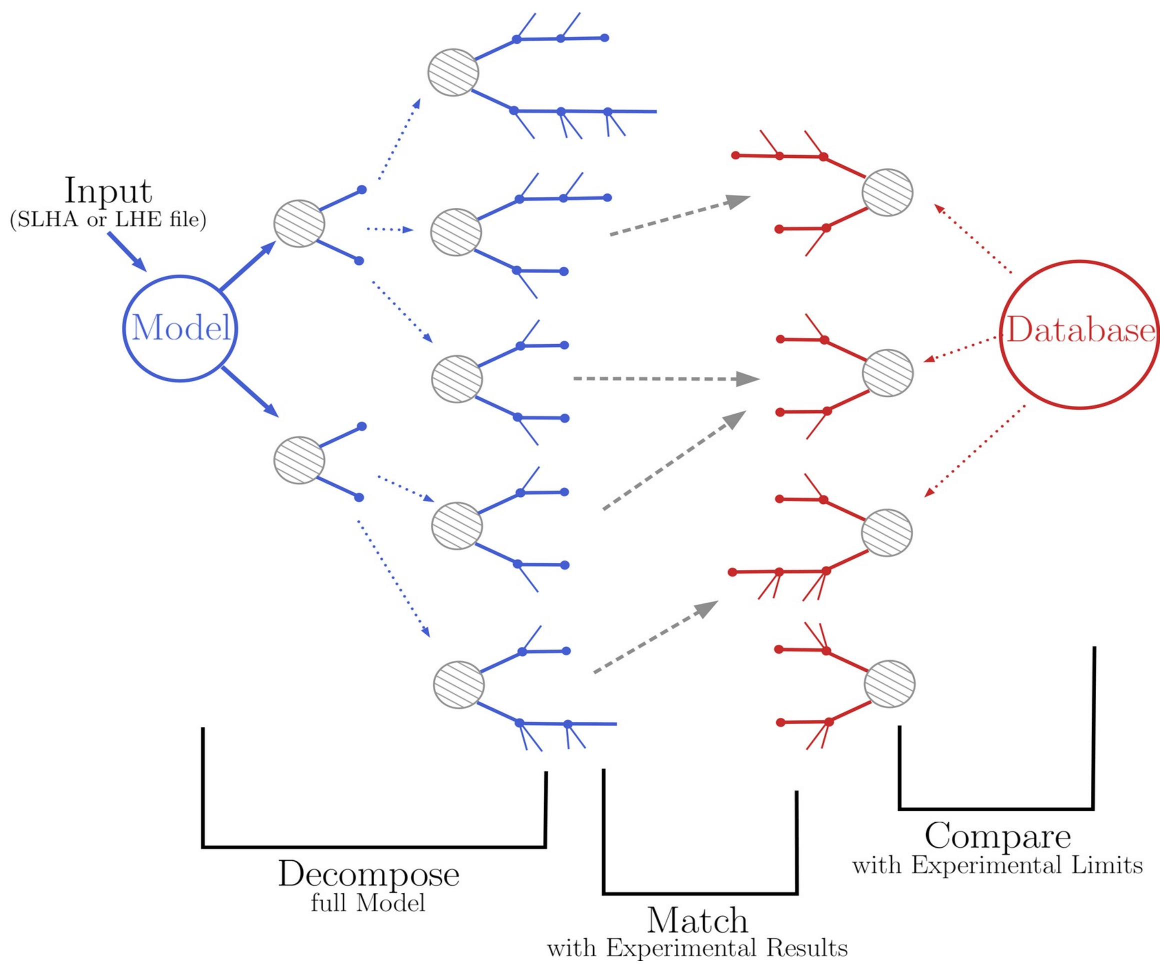
CMS-NOTE-2017-001

arXiv:1809.05548

JHEP 04 (2019) 064

SModelS – a decomposer and a database

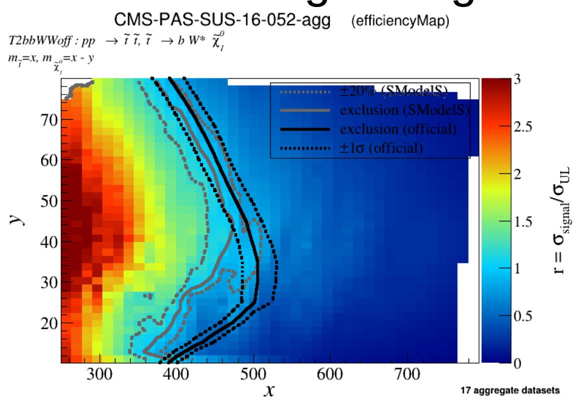
We decompose full theories into SMS topologies, and match them against our database.
Depending on how much information we have access to, we can do different things.



SModelS – a decomposer and a database

efficiency maps with simplified likelihoods

Around 2017/18, CMS started to publish simplified likelihoods for a handful of analyses, making it possible for outsiders to combine signal regions. Until then, SModelS has never been able to combine SRs.



Simplified likelihood, v1: All nuisances summarized in a single "all enveloping" multivariate Gaussian that "connects" all signal regions (which are Poissonian counting variables):

CMS-NOTE-2017-001

$$\mathcal{L}_S(\mu, \boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{(\mu \cdot s_i + b_i + \theta_i)^{n_i} e^{-(\mu \cdot s_i + b_i + \theta_i)}}{n_i!} \cdot \exp\left(-\frac{1}{2}\boldsymbol{\theta}^T \mathbf{V}^{-1} \boldsymbol{\theta}\right)$$

Simplified likelihood, v2: a skewness term is added to allow for asymmetrical distributions.

$$L_{\mathrm{S}}(\alpha, \theta) = \prod_{I=1}^{P} \mathrm{Pr}\left(n_I^{\mathrm{obs}} \,\middle|\, n_{s,I}(\alpha) + a_I + b_I\theta_I + c_I\theta_I^2\right) \cdot \frac{\mathrm{e}^{-\frac{1}{2}\theta^{\mathrm{T}}\rho^{-1}\theta}}{\sqrt{(2\pi)^P}}$$

arXiv:1809.05548
JHEP 04 (2019) 064

proto-models