Artificial Proto-Modelling:

Building Precursors of a Next Standard Model

(or: statistically learning dispersed signals at the LHC)

Sabine Kraml

LPSC Grenoble

Based on arXiv:2012.12246 with A. Lessa and W. Waltenberger (to appear in JHEP)

PROTOMODEL BUILDER

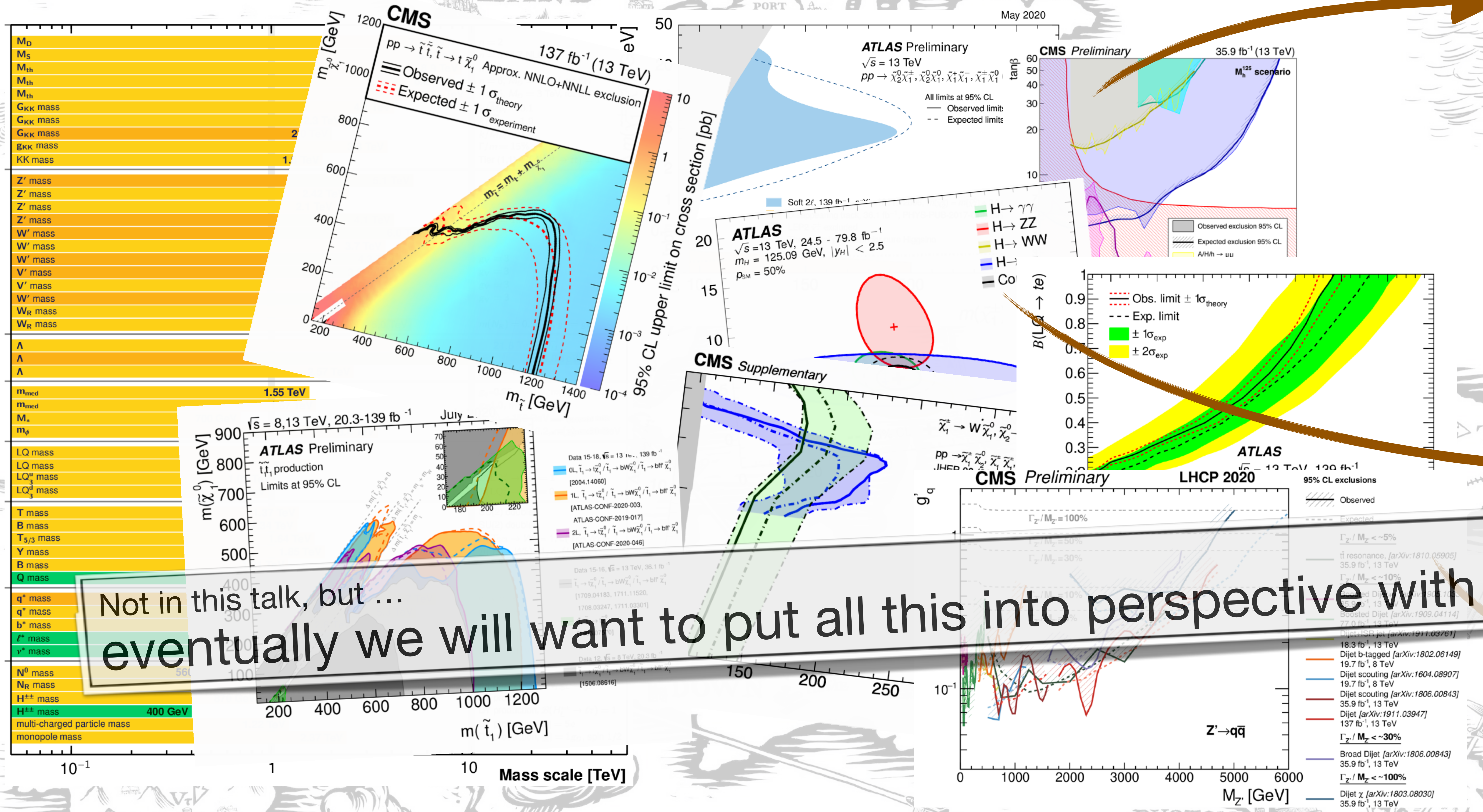
LCWS2021 • global interpretation sessions • 18 March 2021 (virtual)

Motivation

We all believe *some* BSM is out there (LHC, ILC, CLIC, …)

We want a global view of what all the experimental data tell us about new physics.

Dispersed signals* may be hiding in the LHC results ….



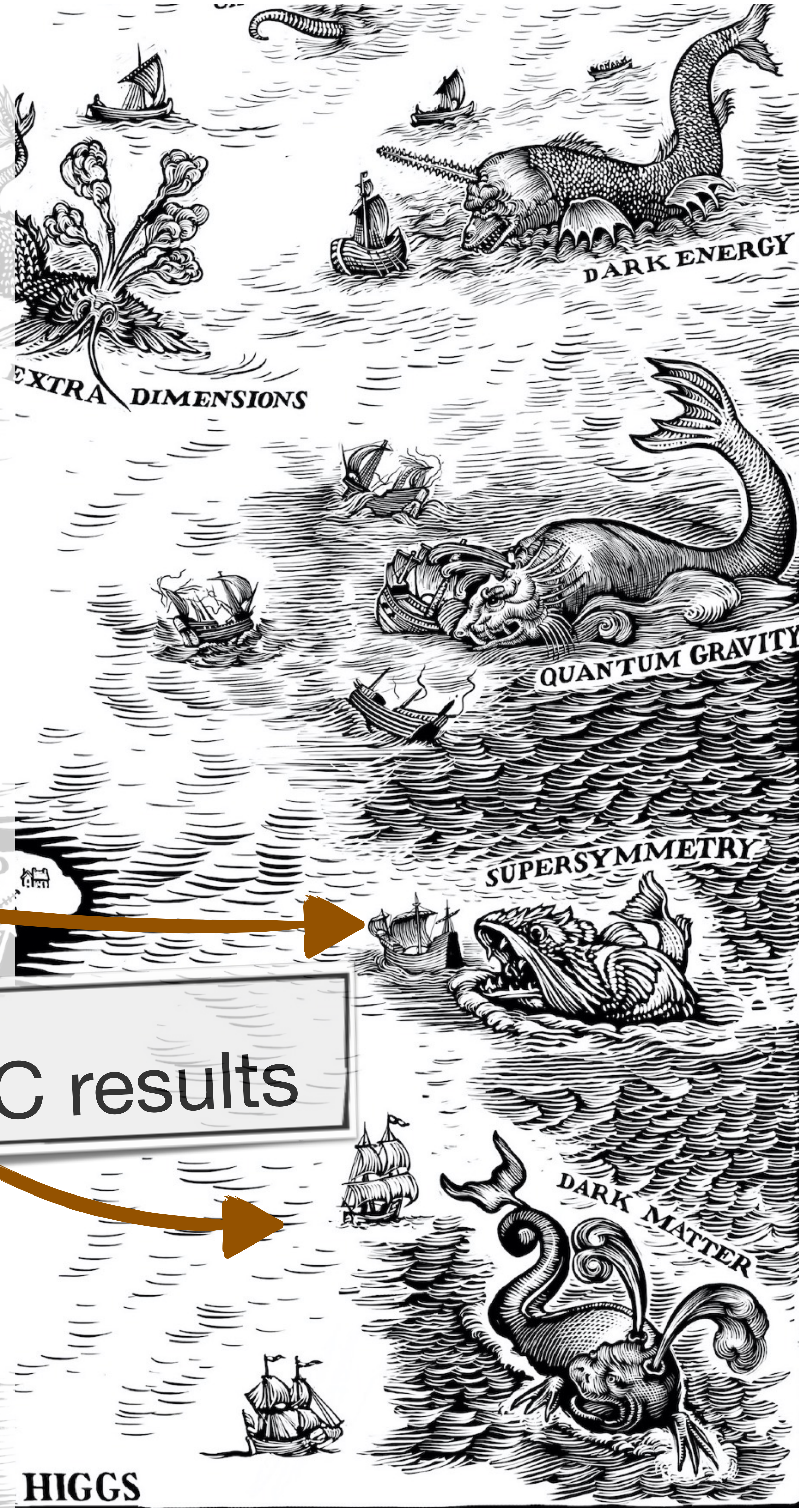


Illustration from Jon Butterworth, A Map of the Invisible

Not in this talk, but …

eventually we will want to put all this into perspective with LC results

* effects of new particles which are spread out over several search regions or final states

Motivation

We all believe *some* BSM is out there (LHC, ILC, CLIC, …)

- LHC currently has no clear sign of new physics; nonetheless there may be dispersed signals* hiding in the slew of data

  * effects of new particles which are spread out over several search regions or final states

- Dispersed signals can be missed by channel-by-channel analyses which each look at only part of the data

- Change of perspective in the quest for BSM; global exploration of the LHC data to complement the analyses in individual final states

- Even in the event of a discovery, given the plethora of possible BSM incarnations, classical hypothesis testing (like for the Higgs) may not be enough to determine the underlying theory.

- Also, measurements at a future e+e- LC are likely to give only part of the full BSM picture; need to be able to put them into perspective with all available information from the LHC *et al.*

Illustration from Jon Butterworth, A Map of the Invisible

Top-down or bottom-up

Construct a BSM Lagrangian

↓

Compute masses, couplings, cross sections, branching ratios

↓

Compare with data

We want to try this

Infer the BSM Lagrangian

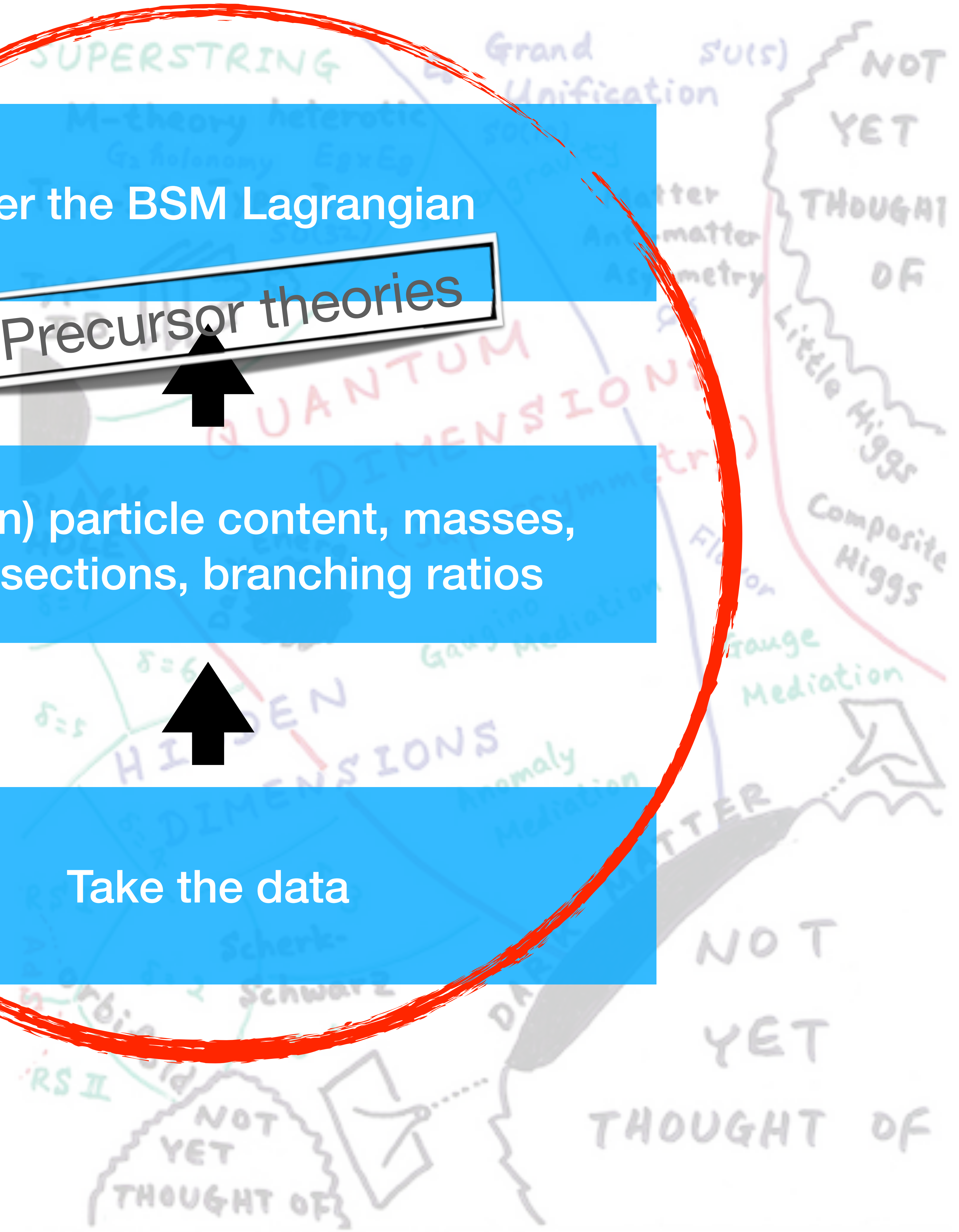
Precursor theories

↑

Fit (learn) particle content, masses, cross sections, branching ratios

↑

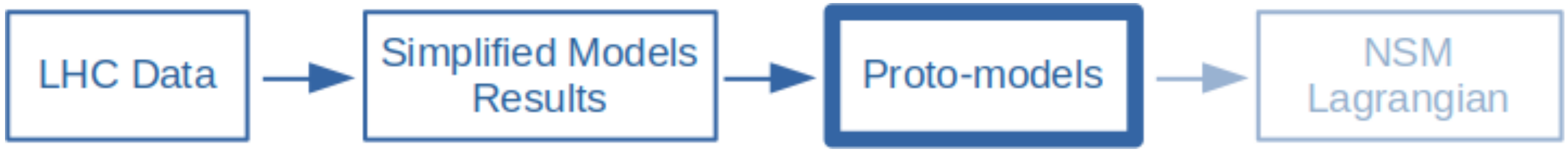
Take the data



from Hitoshi Murayama

Our approach

arXiv:2012.12246
https://smodels.github.io/protomodels/



- Statistical learning algorithm to
  - identify potential dispersed signals in the LHC data
  - fit “proto-models” (new particles, decay modes, signal strengths) to them while remaining compatible with the entirety of LHC results
- Based on simplified model results
  → exploit SModelS functionality and database

In practice

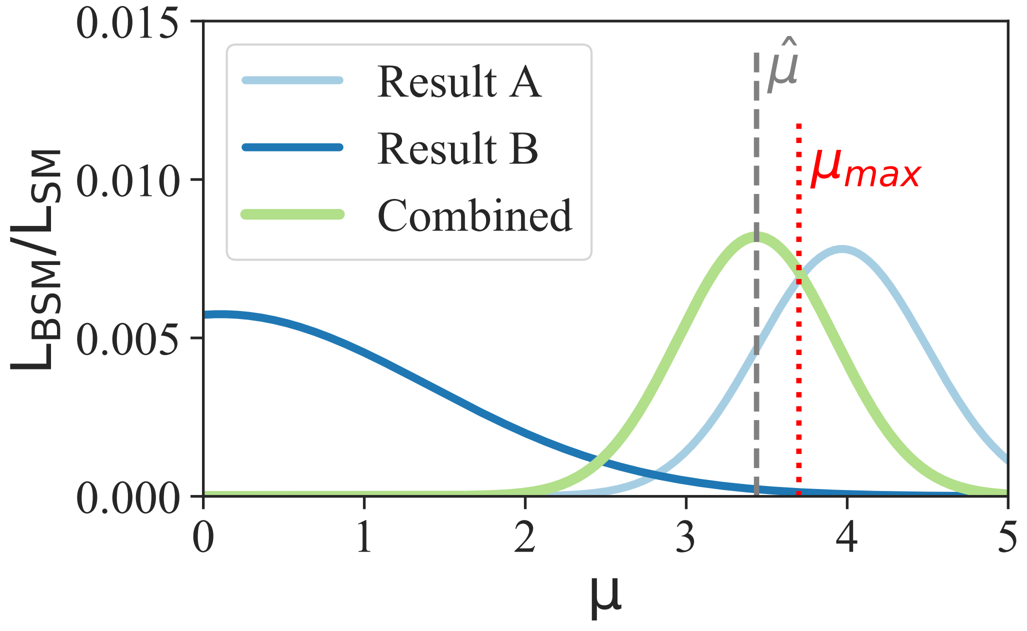
MCMC-like algorithm (dubbed the ***walker***) to comb through proto-model parameter space in order **to identify the models that best fit the data**.

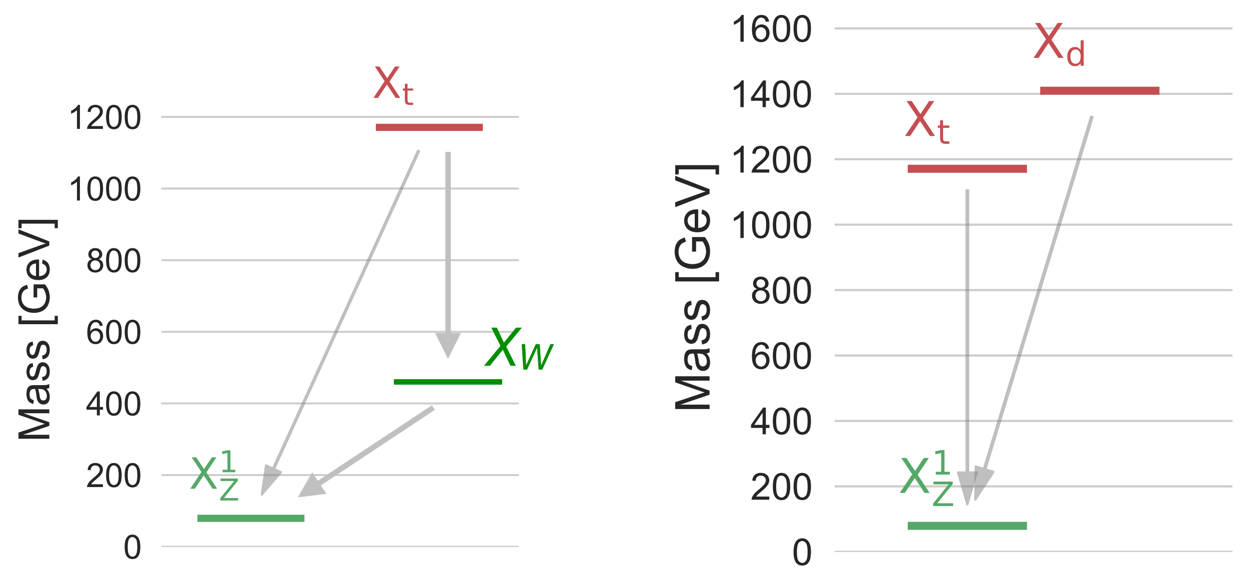
Composed of three blocks, or *machines* which interact with each other:

- Starting with the Standard Model, the ***builder*** creates proto-models, randomly adding or removing new particles and changing any of the proto-model parameters.

- The proto-model is then passed on to the ***critic,*** which checks the model against the database of simplified-model results to determine an upper bound on an overall signal strength (μ_{max}).

- The ***combiner*** identifies all possible combinations of results and constructs a combined likelihood for each subset.





Proto-models are defined by their:

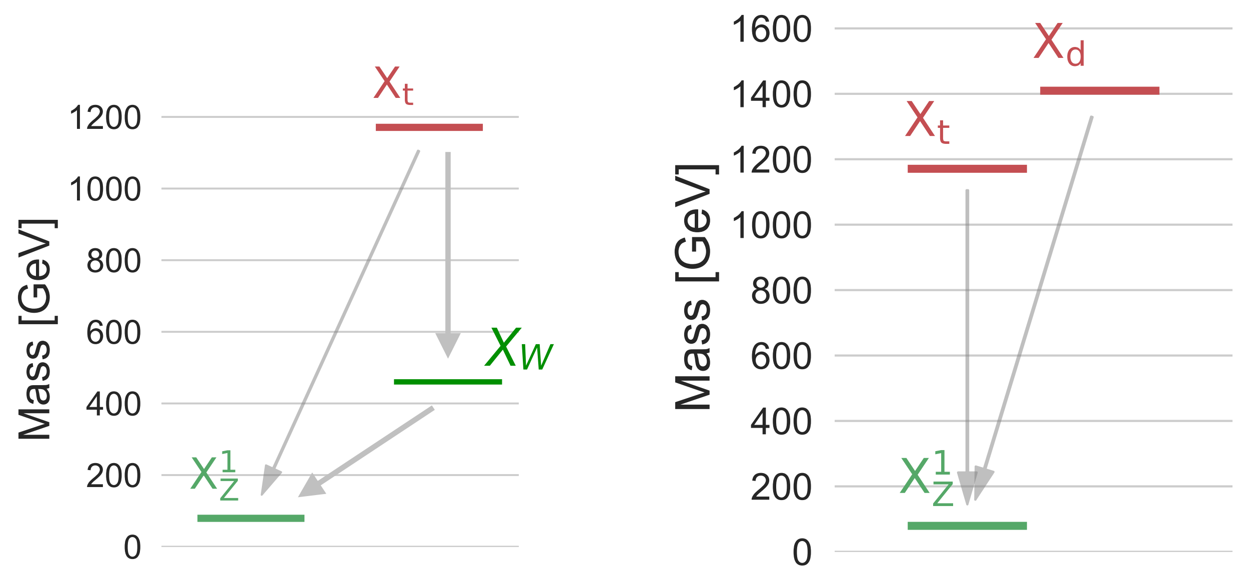
- *Particle content**
- *Masses*
- *Decay modes*
- *Signal strengths*

NB this gives a parameter space of varying dimensionality !

* BSM particles are assumed odd under a Z_2-type symmetry, so they are pair produced and cascade decay to the lightest state

In practice

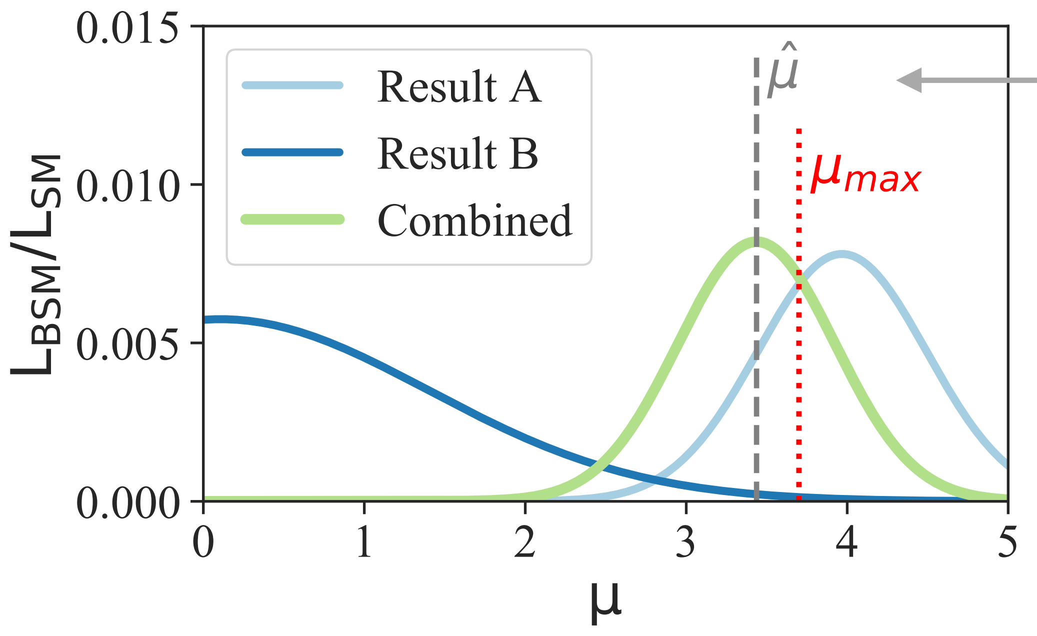
MCMC-like algorithm (dubbed the ***walker***) to comb through proto-model parameter space in order **to identify the models that best fit the data**.



Composed of three blocks, or *machines* which interact with each other:

- Starting with the Standard Model, the ***builder*** creates proto-models, randomly adding or removing new particles and changing any of the proto-model parameters.
- The proto-model is then passed on to the ***critic,*** which checks the model against the database of simplified-model results to determine an upper bound on an overall signal strength (μ_{max}).
- The ***combiner*** identifies all possible combinations of results and constructs a combined likelihood for each subset.

Builder Combiner Critic



Overall signal strength $\hat{\mu}$ chosen to maximise the combined likelihood under the condition $\hat{\mu} < \mu_{max}$.

Proto-model construction rules

for the time being …

- Proto-models are defined by their BSM particle content, the particle masses, production cross sections (signal strengths) and decay branching ratios
- Current assumptions:
  - BSM particles are odd under a Z_2-type symmetry → always produced in pairs and always cascade decay to the lightest state;
  - the lightest BSM particle (LBP) is stable and electrically and color neutral (cosmological constraints, dark matter);
  - except for the LBP, all particles are assumed to decay promptly;
  - only particles with masses within LHC reach are considered part of a specific proto-model.

Possible particle content (spin remains undefined)

- Light quark partners X_q ($q = u, d, c, s$)
- Heavy quark partners X_b, X_t (2 of each)
- Gluon partner X_g (1)
- Electroweak partners X_W (2), X_Z (3)
- Lepton partners X_ℓ, $X_{\nu\ell}$ ($\ell = e, \mu, \tau$)

Cross sections: start from MSSM cross sections, then varied freely via signal strength multipliers

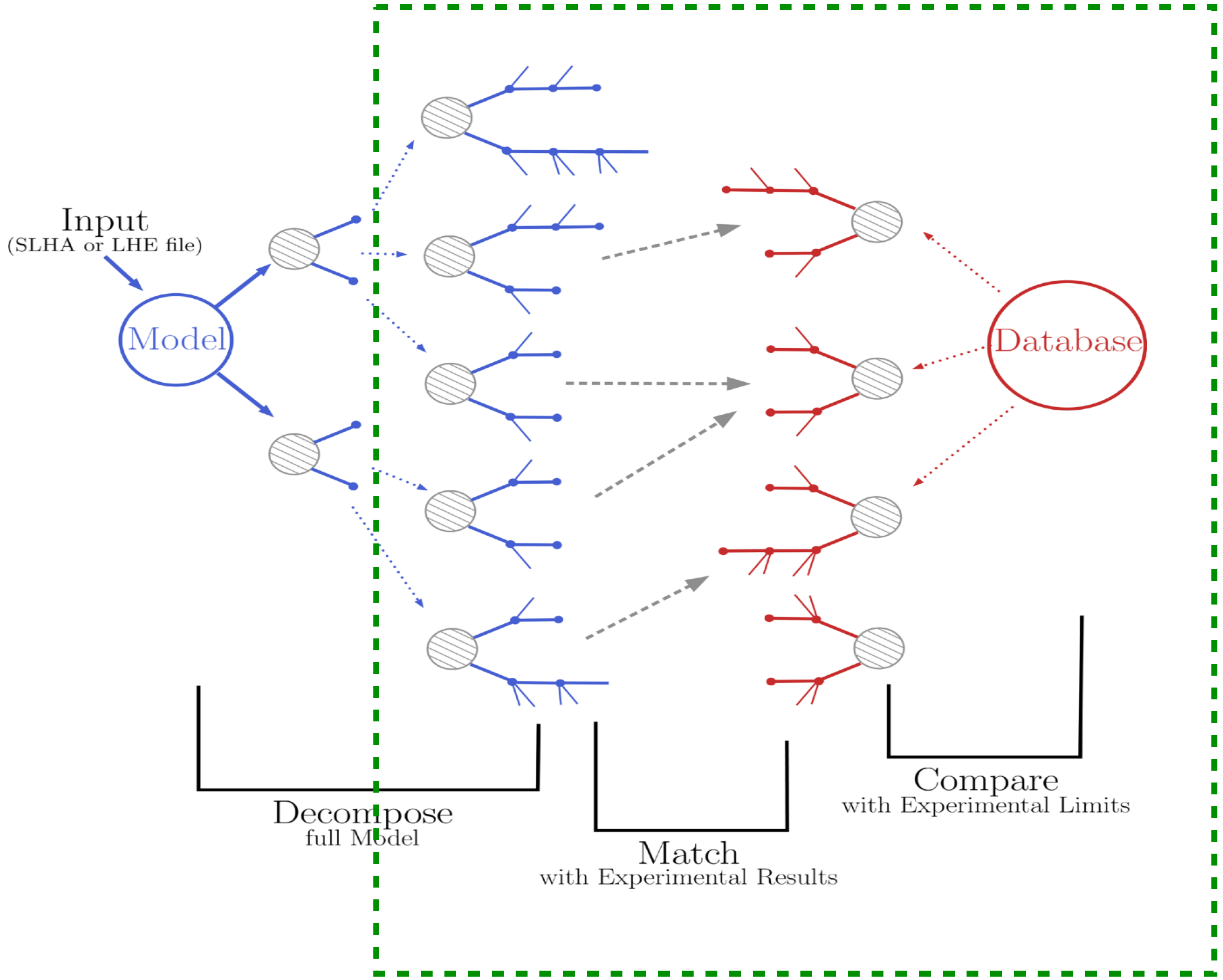
- NB proto-models are just a toolkit, they are *not* intended to be consistent theoretical models

Simplified-model results from SModelS

smodels.github.io

SModelS is a public tool for interpreting simplified-model results from the LHC

- Based on a general procedure to decompose BSM collider signatures featuring a Z_2-like symmetry into simplified-model topologies.
- Large database of simplified-model results (currently 40 ATLAS & 46 CMS searches).
- Very fast b/c no need for MC simulation.

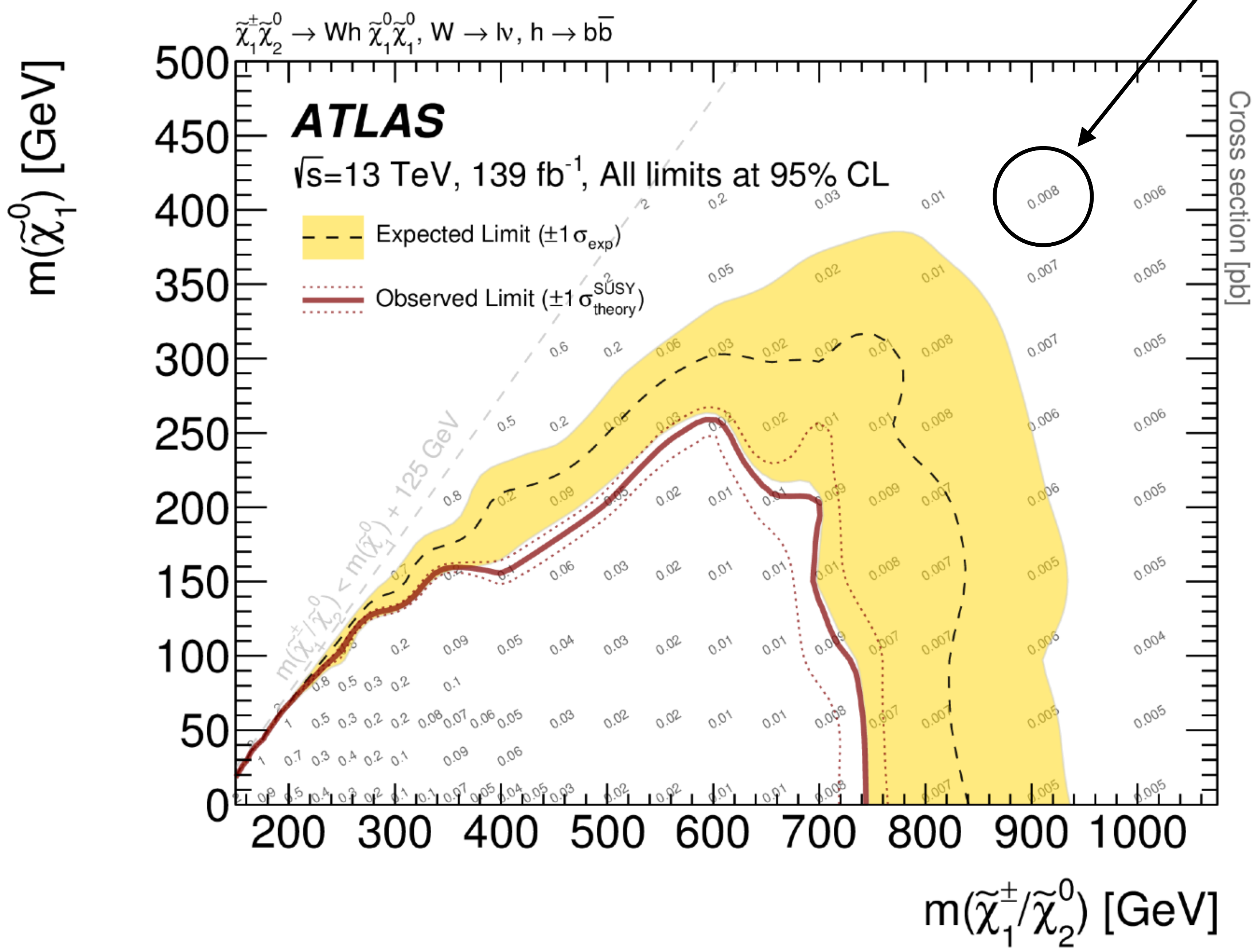


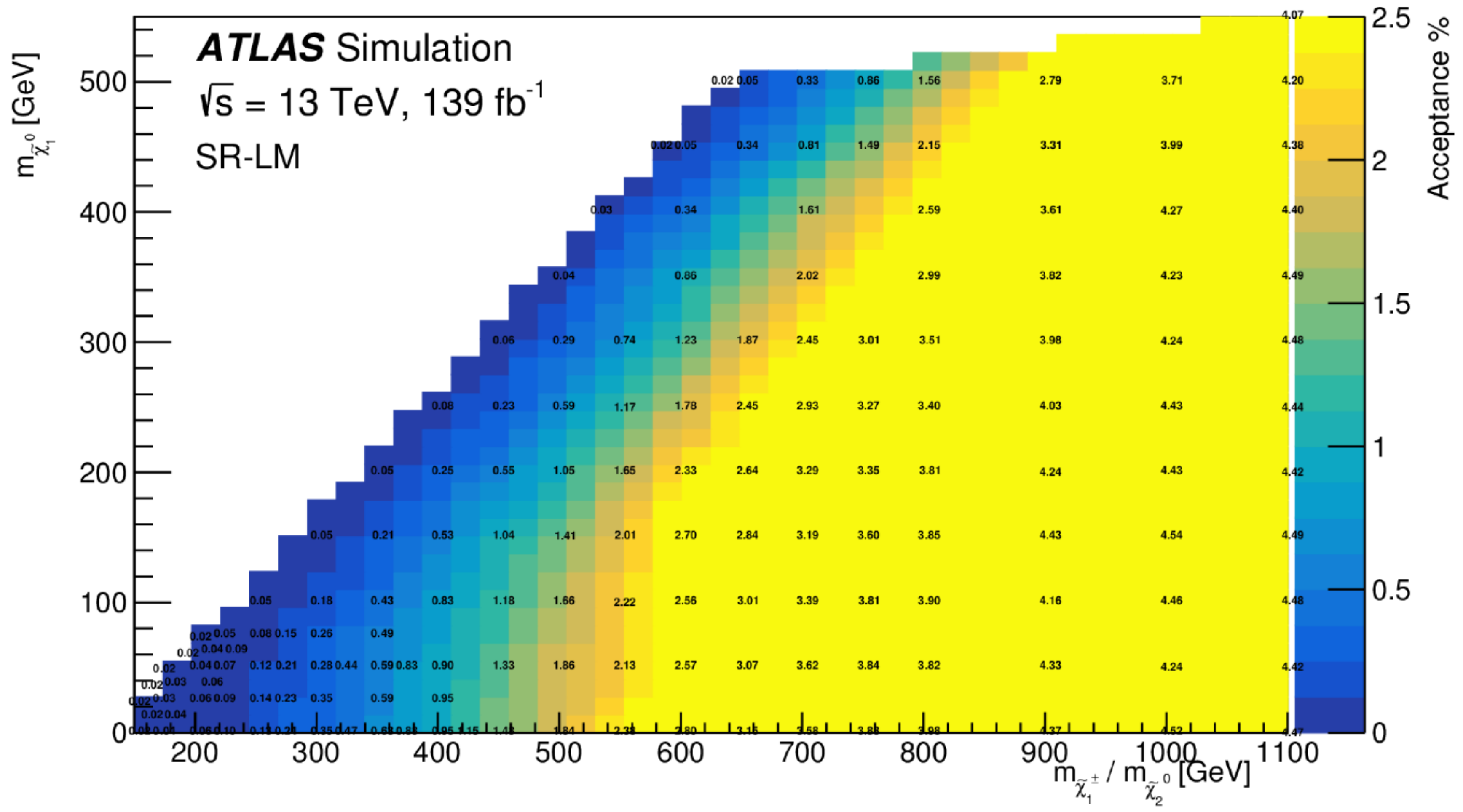
We are exploiting this part of SModelS

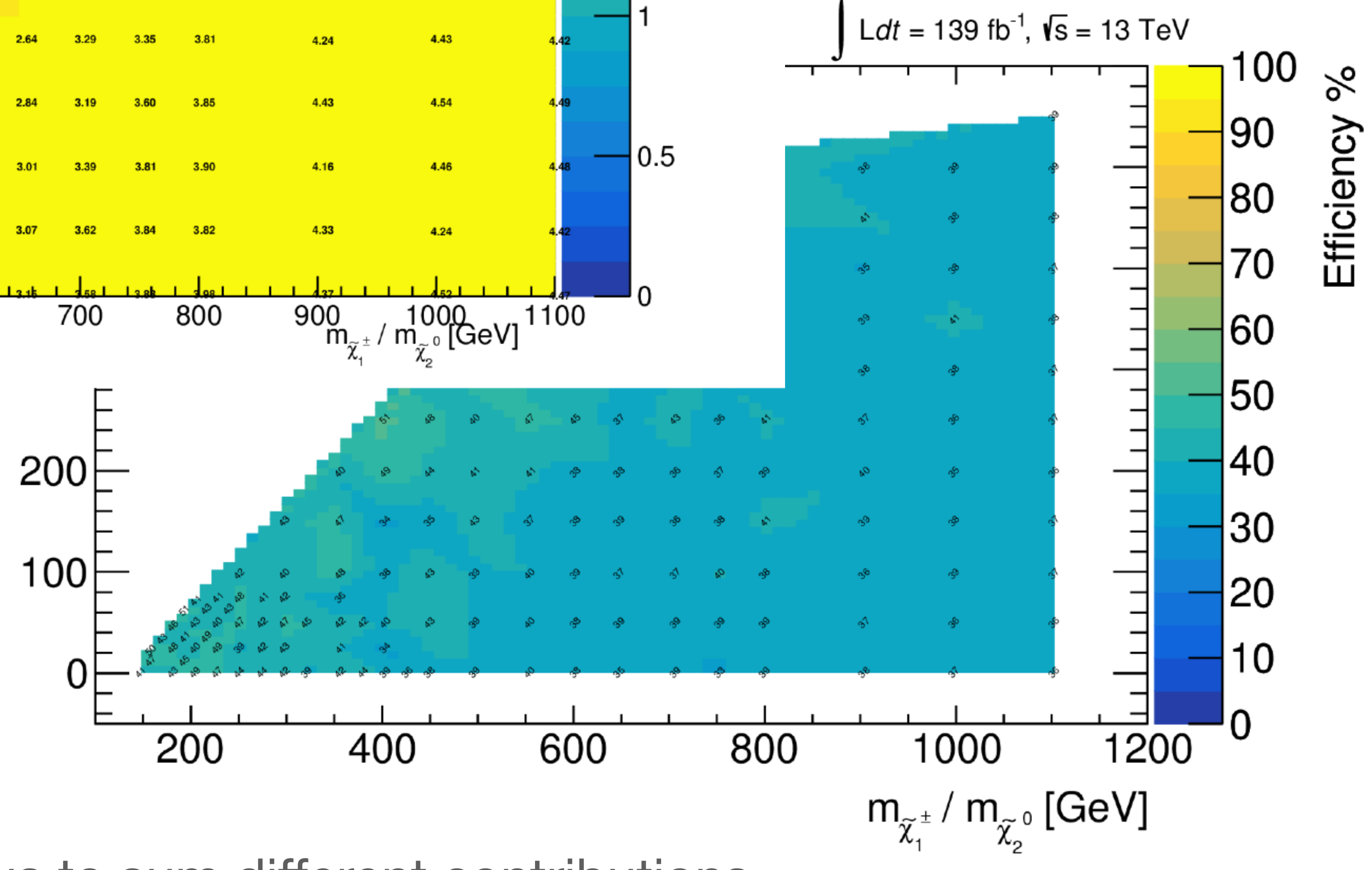
Simplified-model results from SModelS

smodels.github.io

Two types of experimental results: **upper limit maps** and **efficiency (A×ε) maps**

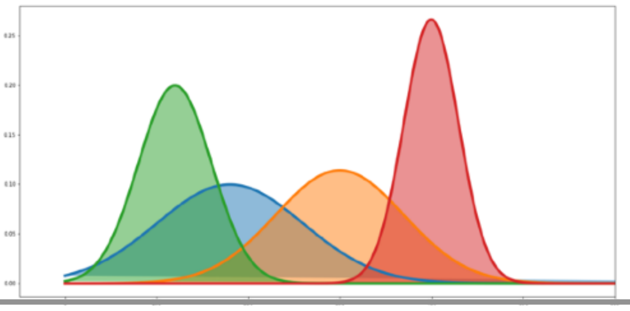






Efficiency map results allow us to sum different contributions to the same signal region, and to ***compute a likelihood***.

Likelihoods for individual analyses



Key to the statistical learning procedure is the construction of a likelihood for the hypothesised signal (the proto-model)

$$\mathrm{L}_{\mathrm{BSM}}(\mu|D) = \mathrm{L}\left(D|\mu + b + \theta\right) p(\theta)$$

μ … signal
b … background
$p(\theta)$ … pdf of nuisances

- For **efficiency-map (EM)** results, we can compute a *simplified likelihood*, assuming *p(θ)* to follow a Gaussian distribution centred around zero and with a variance of *δ^2:*

$$\mathrm{L}_{\mathrm{BSM}}(\mu|D) = \frac{(\mu + b + \theta)^{n_{\mathrm{obs}}} e^{-(\mu+b+\theta)}}{n_{\mathrm{obs}}!} \exp\left(-\frac{\theta^2}{2\delta^2}\right)$$

$\delta^2 = \delta_s^2 + \delta_b^2$
signal+background uncertainties

- CMS sometimes provides a covariance matrix, which allows the combination of signal regions.
- ATLAS has started to provide *full likelihoods* (JSON format), making the above approximation unnecessary.

- For **upper limit (UL) maps**, if observed+expected ULs are available: $\mathrm{L}_{\mathrm{BSM}}$ ~ truncated Gaussian (nb: crude approx!!)

⚠️ If only the observed UL is available → constraint in the form of a step function at the observed 95% CL limit.
Not useful for constructing $\mathrm{L}_{\mathrm{BSM}}$ per se; only used to determine the maximal allowed signal strength $\mu_{\max}$ (in the critic).

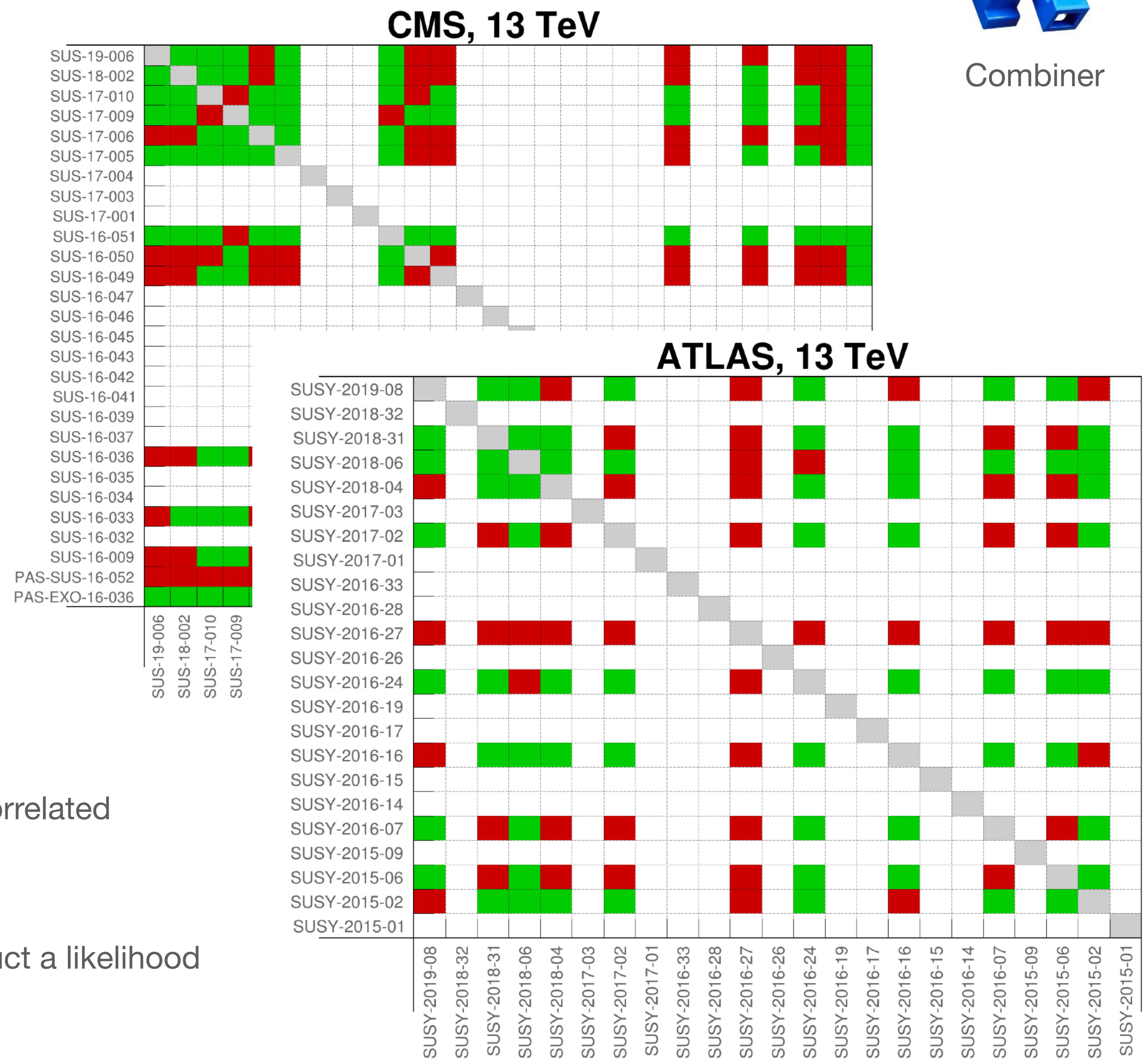
Building a global likelihood (combining analyses)

Combiner

- Global likelihood = product of likelihoods of approximately uncorrelated analyses:
  - from different LHC runs (8 or 13 TeV) and/or
  - from different experiments (ATLAS or CMS) and/or
  - for clearly different signals (e.g., purely hadronic or leptonic)

- A combination "*c*" of analyses is "legal" if 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)

$$\mathrm{L}_c = \prod_{i \in c} \mathrm{L}_i$$



Approx. uncorrelated

Correlated

Can't construct a likelihood

NB: easy to combine with LC results as well

Test statistic

Finally, we need a test statistic that increases for models which better satisfy all the constraints (which includes better fitting potential dispersed signals). Define:

$$K^c := 2\ln \frac{\mathrm{L}^c_{\mathrm{BSM}}(\hat{\mu}) \cdot \pi(\mathrm{BSM})}{\mathrm{L}^c_{\mathrm{SM}} \cdot \pi(\mathrm{SM})} \quad \Rightarrow \quad K := \max_{\forall c \in C} K^c$$

π is a prior, which punishes the proto-model for unneeded complexity (*principle of parsimony*)

$$\pi(\mathrm{BSM}) = \exp\left[-\left(\frac{\mathrm{n_{particles}}}{\mathrm{a}_1} + \frac{\mathrm{n_{BRs}}}{\mathrm{a}_2} + \frac{\mathrm{n_{prod.modes}}}{\mathrm{a}_3}\right)\right], \quad \pi(\mathrm{SM}) \equiv 1$$

where we take $a_1 = 2$, $a_2 = 4$, $a_3 = 8$.

The test statistic thus roughly corresponds to a $\Delta\chi^2$ of the proto-model with respect to the SM, with a penalty for the new degrees of freedom:

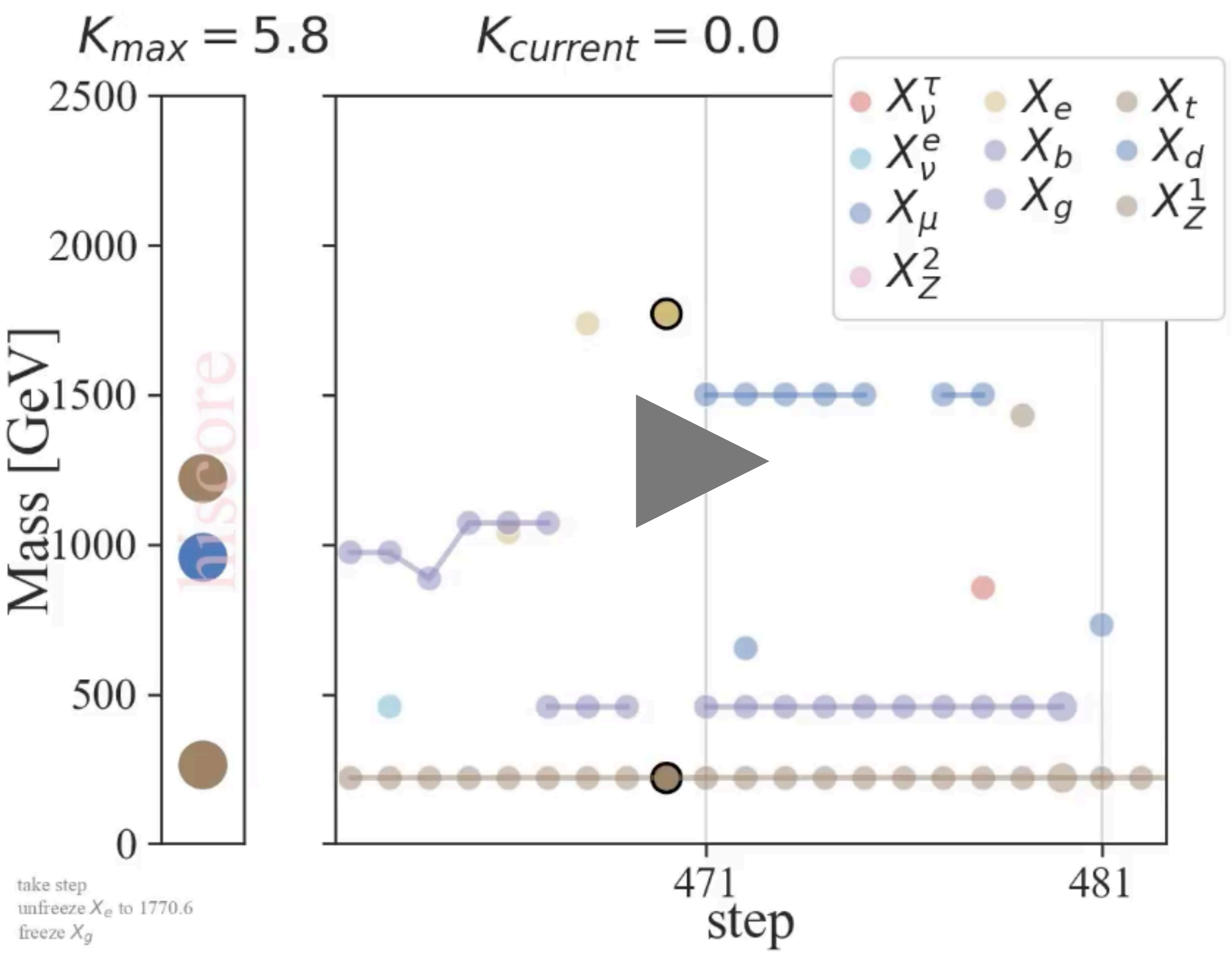
$$K \approx \Delta\chi^2 + 2\ln\pi(\mathrm{BSM})$$

NB: K>0 (K<0) can be interpreted as preference of the proto-model over the SM (and vice-versa)

The Walker

- MCMC (Markov chain) like walk through proto-model parameter space, randomly
  - adding or removing particles
  - changing particle masses
  - changing signal strength multipliers (for production modes)
  - changing decay channels and their branching ratios
- At each step, compute test statistic K
- Step i is reverted with probability

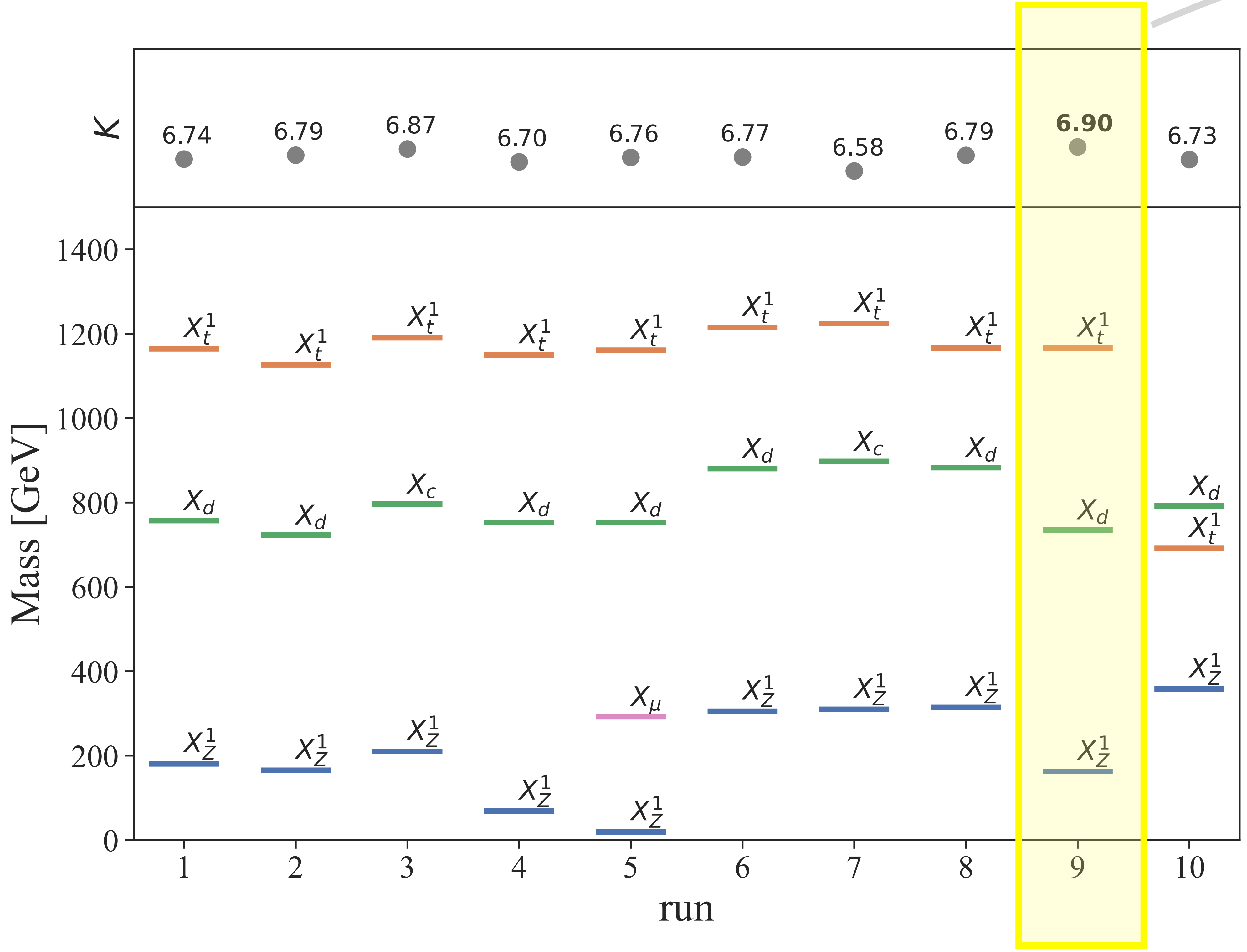
$$\exp\left[\frac{1}{2}(K_i - K_{i-1})\right]$$



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

Results (walking over SModelS v1.2.4 database)

10 “runs”, each consisting of 50 parallel walkers with 1000 steps



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

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

Analysis (all CMS 13 TeV)	Prod	σ_{XX} (fb)	$\sigma^{\mathrm{UL}}_{\mathrm{obs}}$ (fb)	$\sigma^{\mathrm{UL}}_{\mathrm{exp}}$ (fb)	r_{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_{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

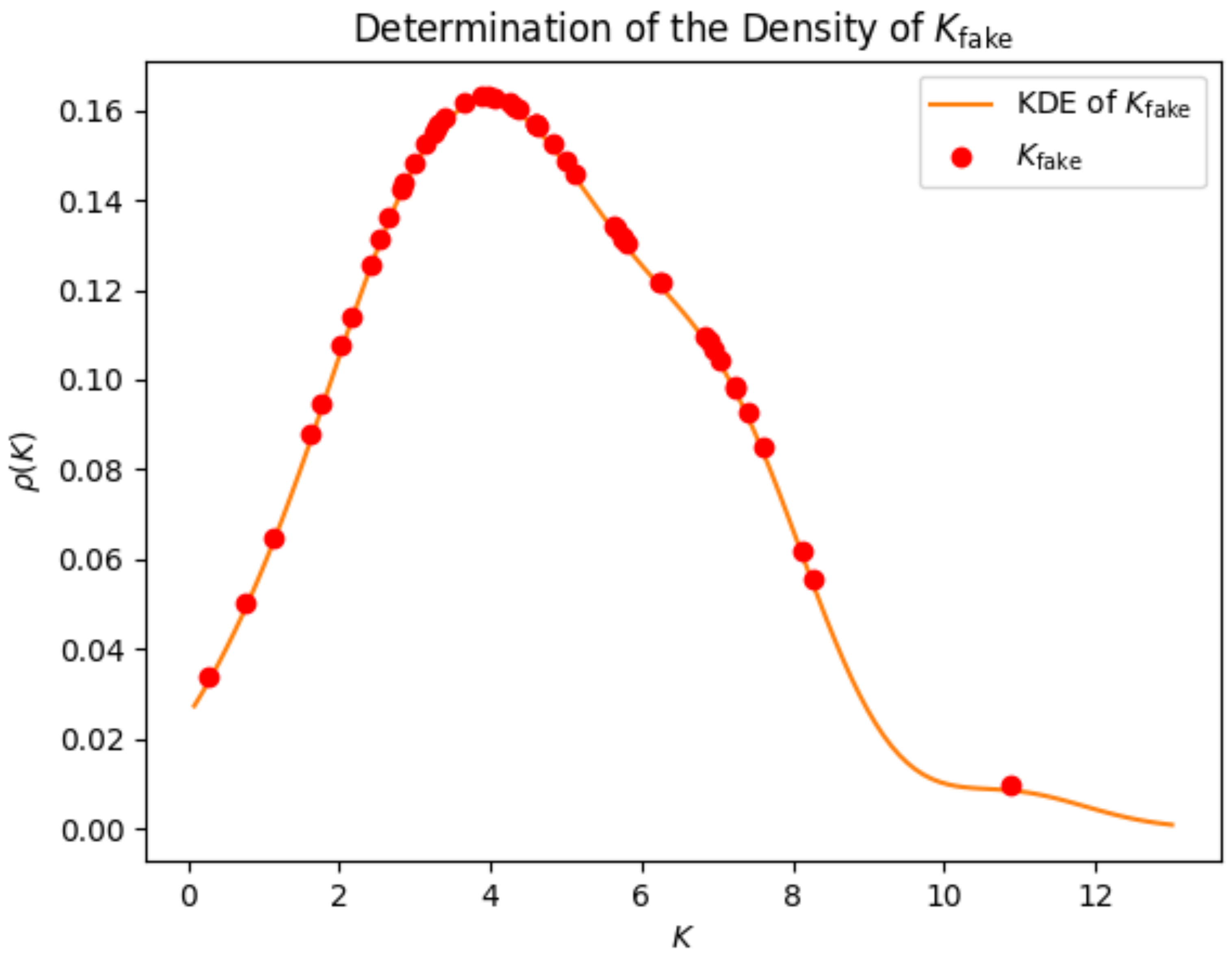
List of the most constraining results for the highest score proto-model

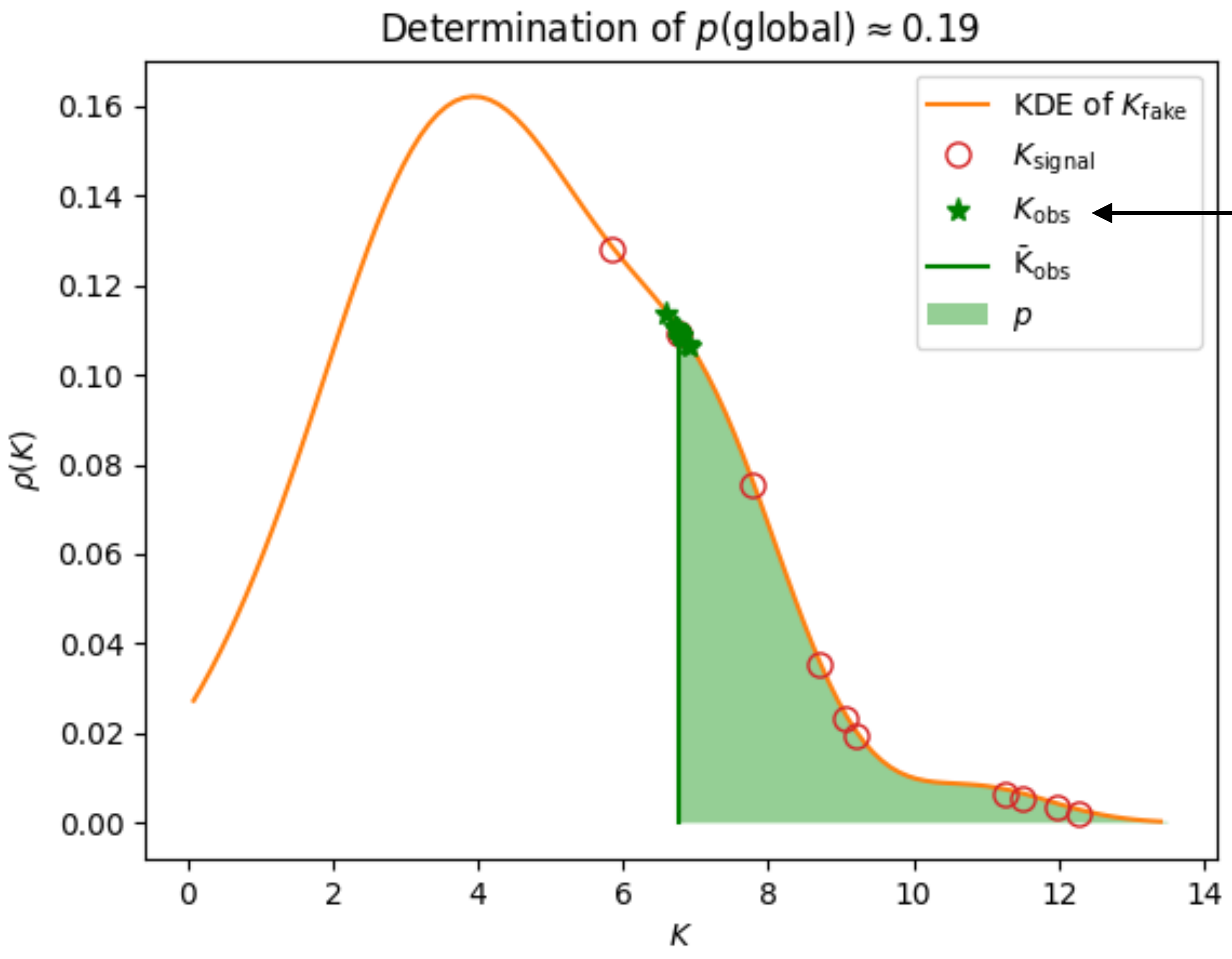
In each run, the high-score model has a top partner and a quark partner with SUSY-like cross sections.
Driven by ~2σ excesses in ATLAS-SUSY-2016-07 (gluino, squark), ATLAS-SUSY-2016-16 (1-lept. stop) and CMS-SUS-16-050 (0-lept. stop) searches.

Testing the SM hypothesis

To mimic the case that **no new physics** is in the data, we produce “fake” SModelS databases by sampling background models (i.e. setting #observed = #expected, sampled within BG uncertainties)

➡ Determine density of K under SM-only hypothesis (K_{fake}) from 50 fake SModelS databases; 50×50 walkers in total





K obtained from the 10 runs over the real SModelS database

➡ Determine global p-value of SM-only hypothesis $$p_{\text{global}} := \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} 1_{[\bar{K}_{\text{obs}},\infty)}(K^i_{\text{fake}}) \approx \int_{\bar{K}_{\text{obs}}}^{\infty} dK\,\rho(K) \approx 0.19$$ (N=50)

Conclusions

- Null results (so far) in the numerous channel-by-channel searches for new particles make it increasingly relevant to change perspective and attempt a more global approach to find out where BSM physics may hide.
- Presented a novel statistical learning algorithm, that identifies potential dispersed signals in the slew of published LHC analyses; the aim is to build candidate proto-models from small excesses in the data, while remaining consistent with all other constraints.
- Procedure may also be relevant for global interpretations including data from a future e+e- LC; ultimate goal is a bottom-up approach to inferring the BSM Lagrangian from the data.

Bottlenecks

- Current realisation is based on simplified-model results in the SModelS database
- limited by the variety and type of simplified-model results available
  - need more (and larger) efficiency-map results
  - need to go beyond SUSY-like simplified models

arXiv:2012.12246 is but a proof-of-principle;
lots of interesting further developments to be done!

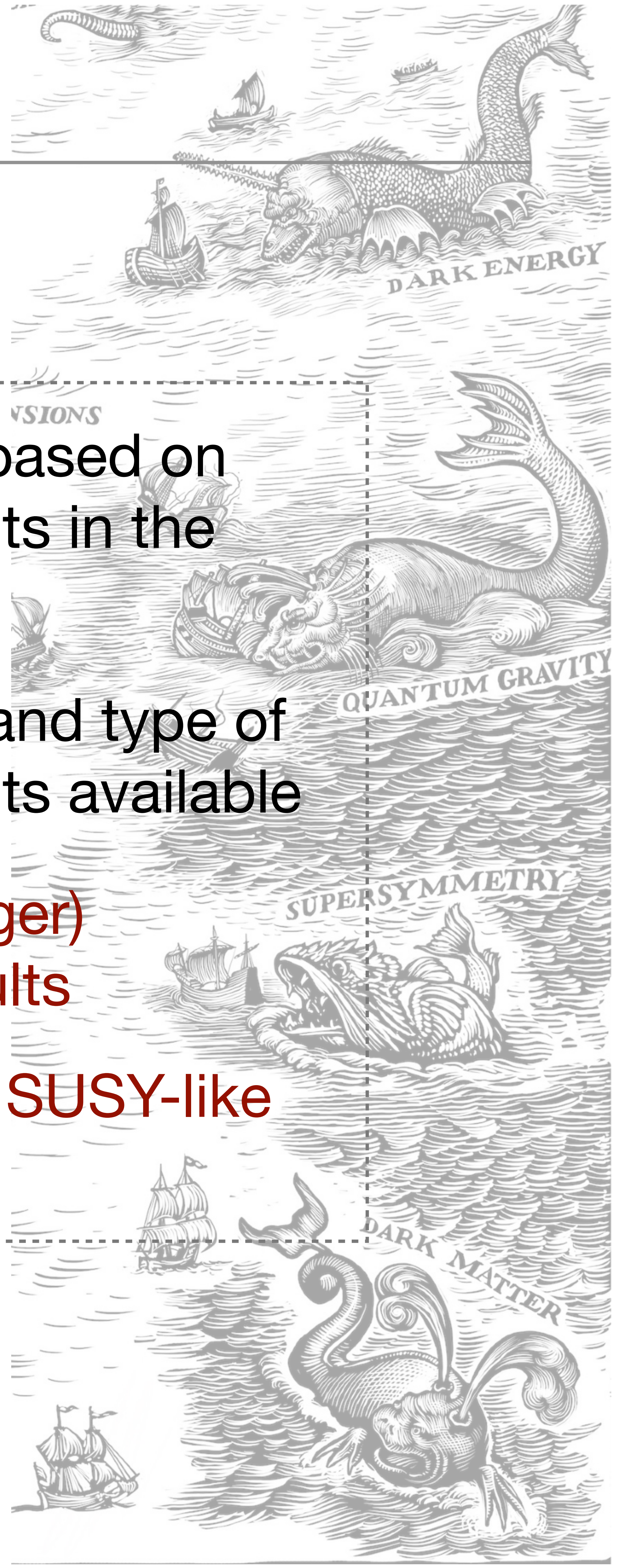


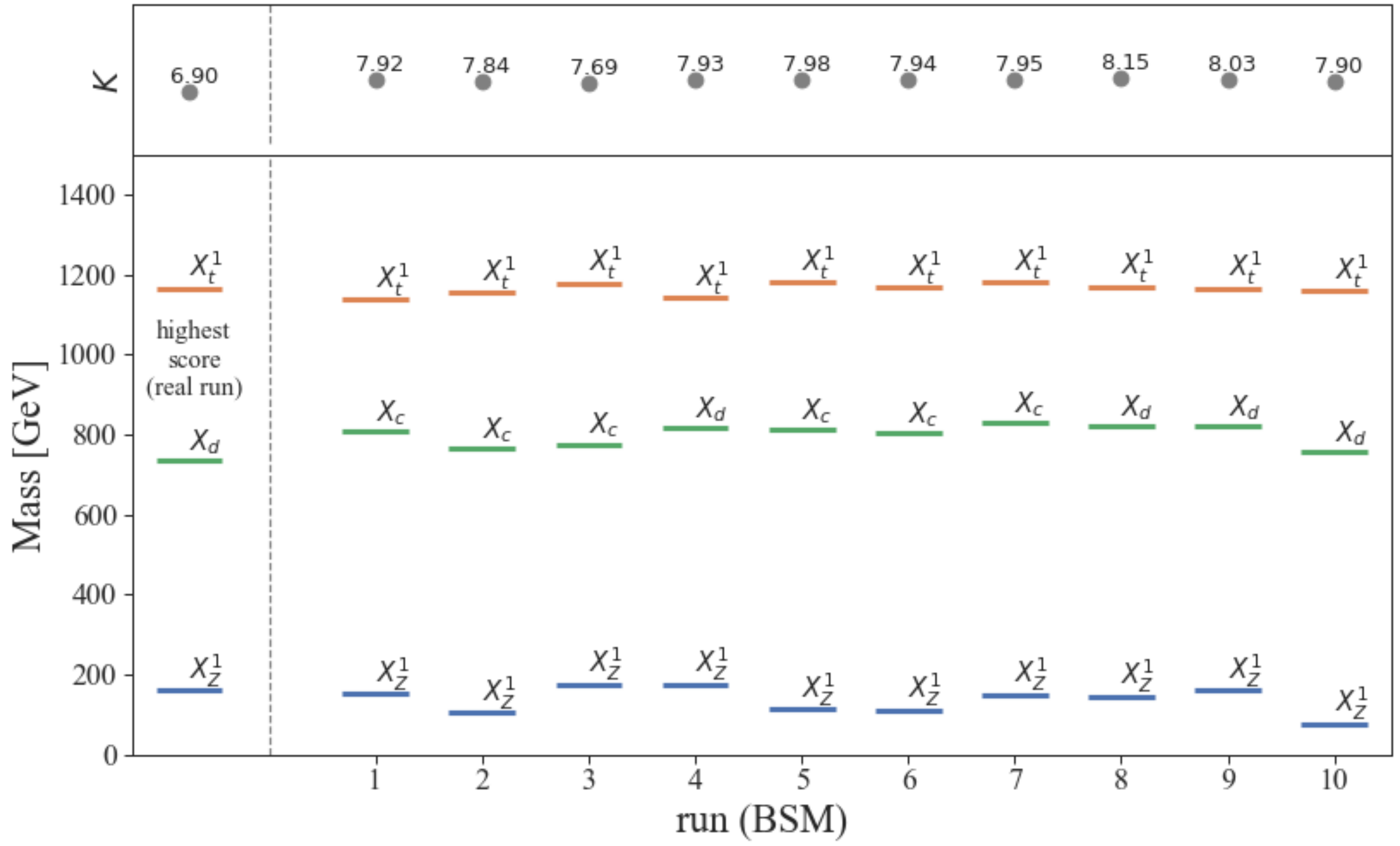
Illustration from Jon Butterworth, A Map of the Invisible

Backup

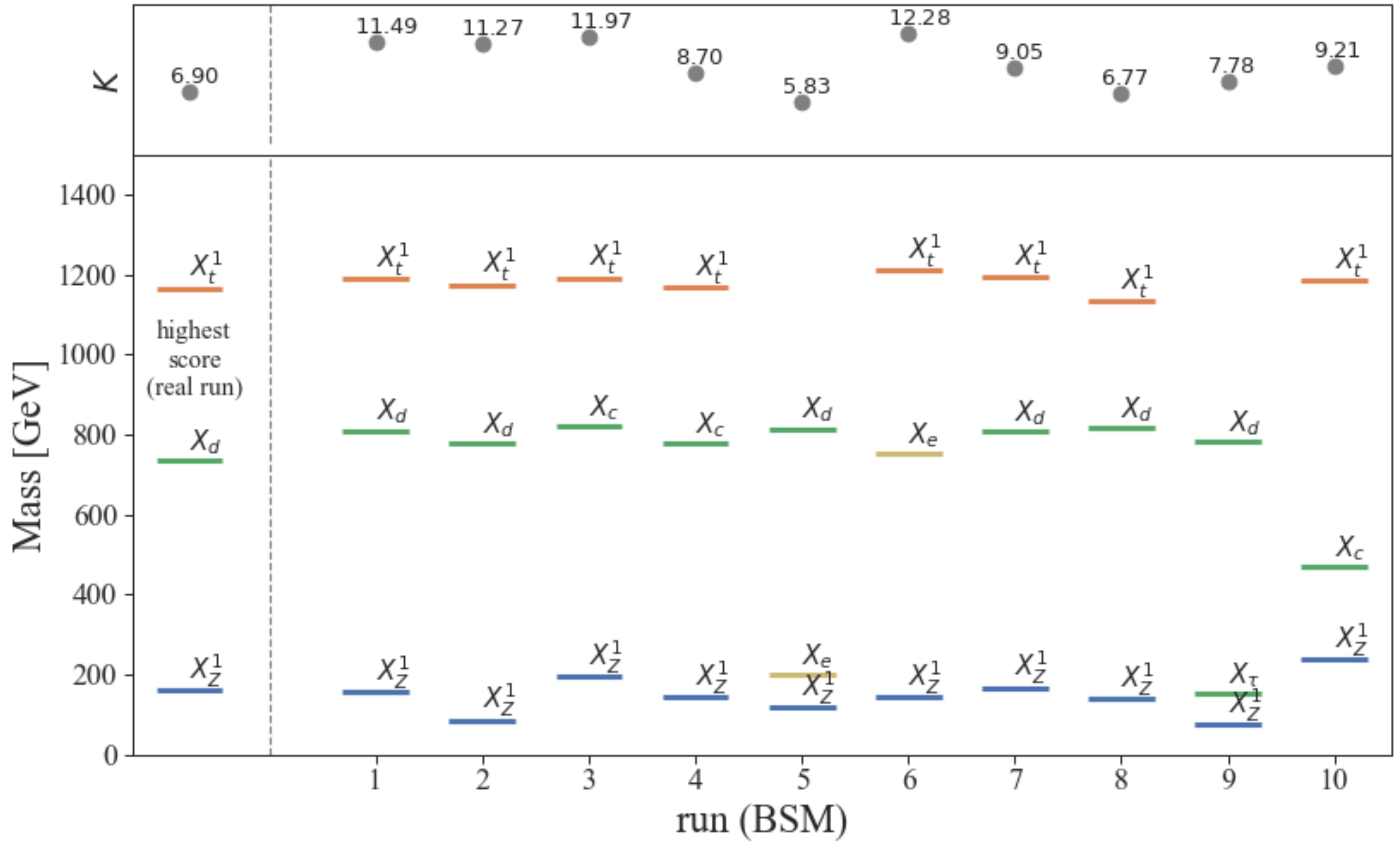
Closure tests: inject a signal

- Replace the observed data with the expected background plus signal yields for the highest-score proto-model with K=6.9; re-run and see whether the injected signal is “rediscovered”
- We do this with and without sampling from the backgrounds

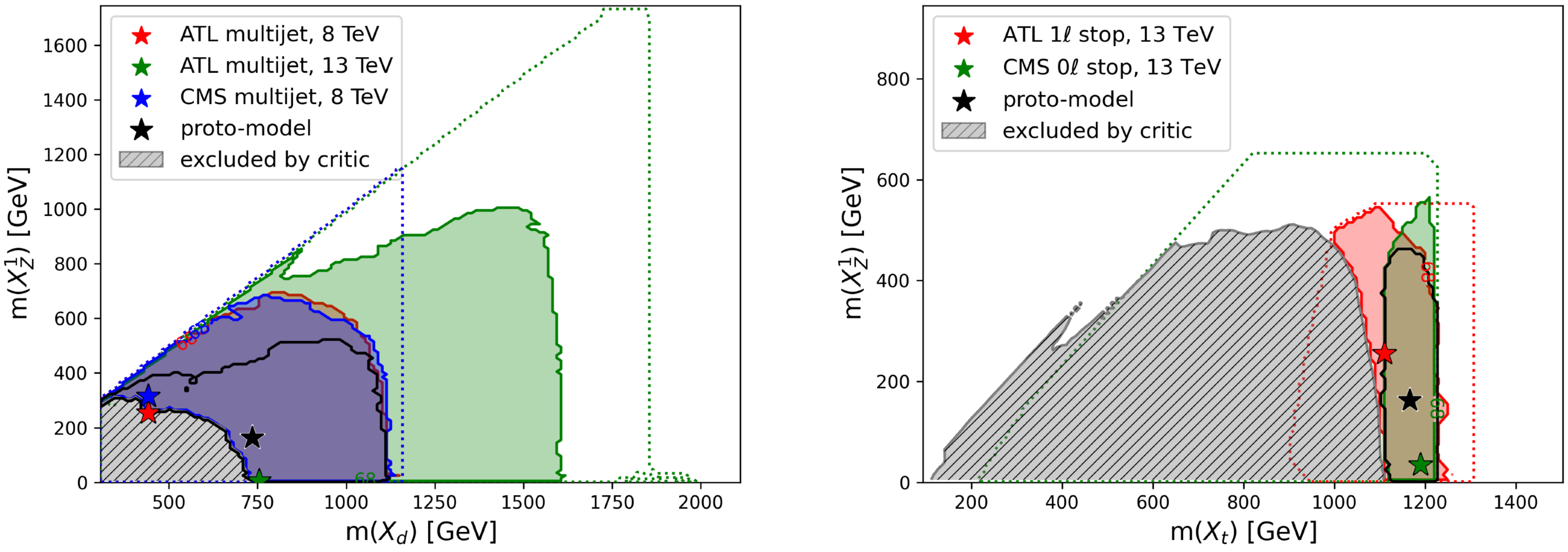
Technical closure test (without sampling)



Full closure test (with sampling)



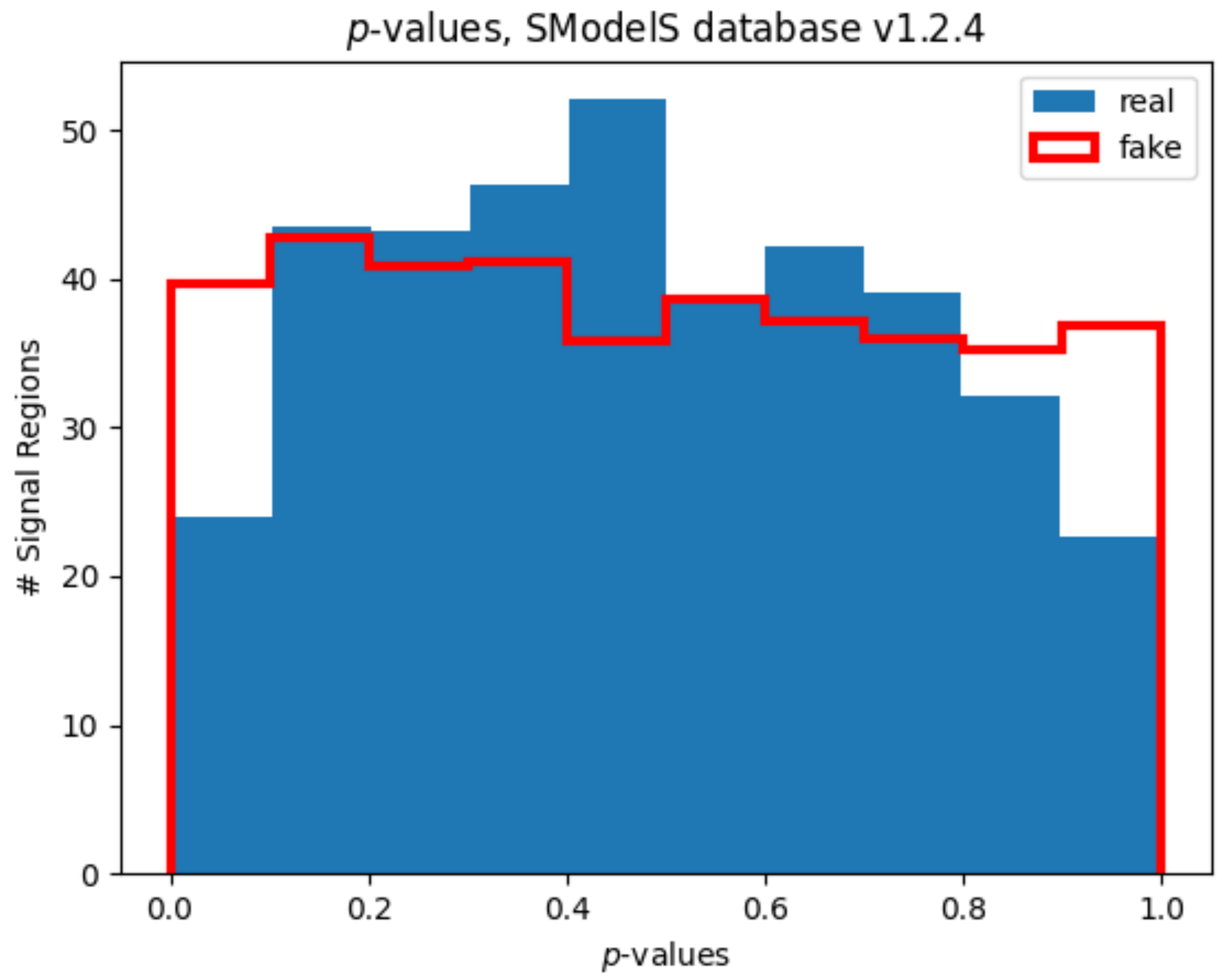
Posterior distributions



68% Bayesian credibility regions (coloured areas, solid contours) in the mass vs. mass planes of the corresponding simplified model. The signal strength multipliers are fixed at the values of the highest-score proto-model with K=6.9. The coloured stars show the preferred points for each individual analysis, while the black stars display the proto-model parameters. **The dashed lines indicate the size of the UL and EM maps provided by the experiments**, thus defining the 100\% credibility regions.

SM backgrounds

- *p*-values computed for all signal regions of the EM results contained in the SModelS database.
- If the observed data is well described by the estimated backgrounds and the uncertainties thereon, one would expect a flat distribution of *p*-values.
- For the fake databases generated under the SM hypothesis, this is the case (red histogram).
- With the real data (blue histogram) the distribution is not flat — this is expected if the background uncertainties are overestimated.



Run2 results in the SModelS database (v1.2.4)

ID	Short Description	$\mathcal{L}$ [fb^{-1}]	UL$_{obs}$	UL$_{exp}$	EM	comb.
ATLAS-SUSY-2015-01 [67]	2 b-jets	3.2	✓			
ATLAS-SUSY-2015-02 [68]	1ℓ stop	3.2	✓		✓	
ATLAS-SUSY-2015-06 [69]	0ℓ + 2–6 jets	3.2			✓	
ATLAS-SUSY-2015-09 [70]	jets + 2 SS or $\geq 3\ell$	3.2	✓			
ATLAS-SUSY-2016-07 [55]	0ℓ + jets	36.1	✓		✓	
ATLAS-SUSY-2016-14 [71]	jets + 2 SS or $\geq 3\ell$	36.1	✓			
ATLAS-SUSY-2016-15 [72]	0ℓ stop	36.1	✓			
ATLAS-SUSY-2016-16 [48]	1ℓ stop	36.1	✓		✓	
ATLAS-SUSY-2016-17 [73]	2 OS leptons	36.1	✓			
ATLAS-SUSY-2016-19 [74]	2 b-jets + τ's	36.1	✓			
ATLAS-SUSY-2016-24 [53]	2–3 ℓ, EWino	36.1	✓		✓	
ATLAS-SUSY-2016-26 [75]	$\geq$ 2 c-jets	36.1	✓			
ATLAS-SUSY-2016-27 [76]	jets + γ	36.1	✓		✓	
ATLAS-SUSY-2016-28 [77]	2 b-jets	36.1	✓			
ATLAS-SUSY-2016-33 [78]	2 OSSF ℓ	36.1	✓			
ATLAS-SUSY-2017-01 [79]	WH(bb), EWino	36.1	✓			
ATLAS-SUSY-2017-02 [80]	0ℓ + jets	36.1	✓	✓		
ATLAS-SUSY-2017-03 [81]	2–3 leptons, EWino	36.1	✓			
ATLAS-SUSY-2018-04 [38]	2 hadronic taus	139.0	✓		✓	JSON
ATLAS-SUSY-2018-06 [82]	3 leptons, EWino	139.0	✓	✓		
ATLAS-SUSY-2018-31 [39]	2b + 2H(bb)	139.0	✓		✓	JSON
ATLAS-SUSY-2018-32 [83]	2 OS leptons	139.0	✓			
ATLAS-SUSY-2019-08 [40]	1ℓ + higgs	139.0	✓		✓	JSON

ID	Short Description	$\mathcal{L}$ [fb^{-1}]	UL$_{obs}$	UL$_{exp}$	EM	comb.
CMS-PAS-EXO-16-036 [84]	HSCP	12.9	✓		✓	
CMS-PAS-SUS-16-052 [34]	ISR jet + soft ℓ	35.9	✓		✓	Cov.
CMS-SUS-16-009 [85]	0ℓ + jets, top tagging	2.3	✓	✓		
CMS-SUS-16-032 [86]	2 b- or 2 c-jets	35.9	✓			
CMS-SUS-16-033 [57]	0ℓ + jets	35.9	✓	✓	✓	
CMS-SUS-16-034 [87]	2 OSSF leptons	35.9	✓			
CMS-SUS-16-035 [88]	2 SS leptons	35.9	✓			
CMS-SUS-16-036 [58]	0ℓ + jets	35.9	✓	✓		
CMS-SUS-16-037 [89]	1ℓ + jets with MJ	35.9	✓			
CMS-SUS-16-039 [90]	2–3ℓ, EWino	35.9	✓			
CMS-SUS-16-041 [91]	jets + $\geq 3\ell$	35.9	✓			
CMS-SUS-16-042 [92]	1ℓ + jets	35.9	✓			
CMS-SUS-16-043 [93]	WH(bb), EWino	35.9	✓			
CMS-SUS-16-045 [94]	jets + H $\to \gamma\gamma$	35.9	✓			
CMS-SUS-16-046 [95]	high-p_T γ	35.9	✓			
CMS-SUS-16-047 [96]	γ + jets, high H_T	35.9	✓			
CMS-SUS-16-049 [97]	0ℓ stop	35.9	✓	✓		
CMS-SUS-16-050 [49]	0ℓ stop, m_{T2}	35.9	✓	✓		
CMS-SUS-16-051 [59]	1ℓ stop	35.9	✓	✓		
CMS-SUS-17-001 [98]	2ℓ stop	35.9	✓			
CMS-SUS-17-003 [99]	2 taus	35.9	✓			
CMS-SUS-17-004 [43]	EWino combination	35.9	✓			
CMS-SUS-17-005 [100]	1ℓ stop, soft	35.9	✓	✓		
CMS-SUS-17-006 [101]	jets + boosted H(bb)	35.9	✓	✓		
CMS-SUS-17-009 [52]	2 OSSF leptons	35.9	✓	✓		
CMS-SUS-17-010 [102]	2ℓ EWino, stop	35.9	✓	✓		
CMS-SUS-18-002 [103]	γ + (b-)jets	35.9	✓	✓		
CMS-SUS-19-006 [15]	0ℓ + jets, MHT	137.0	✓	✓		

Can't construct a likelihood

Run1 results in the SModelS database (v1.2.4)

ID	Short Description	$\mathcal{L}$ [fb^{-1}]	UL$_{\text{obs}}$	UL$_{\text{exp}}$	EM
ATLAS-CONF-2013-007 [104]	2 SS ℓ + (b-)jets	20.7	✓		
ATLAS-CONF-2013-061 [105]	0–1ℓ + 4–7 ($\geq$ 3 b-)jets	20.1	✓		
ATLAS-CONF-2013-089 [106]	2ℓ + jets	20.3	✓		
ATLAS-SUSY-2013-02 [54]	0ℓ + 2–6 jets	20.3	✓		✓
ATLAS-SUSY-2013-04 [107]	0ℓ + 7–10 jets	20.3	✓		✓
ATLAS-SUSY-2013-05 [108]	0ℓ + 2 b-jets	20.1	✓		✓
ATLAS-SUSY-2013-08 [109]	Z + b-jets	20.3	✓		
ATLAS-SUSY-2013-09 [110]	jets + 2 SS or $\geq$ 3ℓ	20.3	✓		
ATLAS-SUSY-2013-11 [111]	2ℓ (e, μ)	20.3	✓		✓
ATLAS-SUSY-2013-12 [112]	3ℓ (e, μ, τ)	20.3	✓		
ATLAS-SUSY-2013-15 [113]	1ℓ + 4 (1 b-)jets	20.3	✓		✓
ATLAS-SUSY-2013-16 [114]	0ℓ + 6 (2 b-)jets	20.1	✓		✓
ATLAS-SUSY-2013-18 [115]	0-1ℓ + $\geq$ 3 b-jets	20.1	✓		✓
ATLAS-SUSY-2013-19 [116]	2 OS ℓ + (b-)jets	20.3	✓		
ATLAS-SUSY-2013-21 [117]	monojet or c-jet	20.3			✓
ATLAS-SUSY-2013-23 [118]	WH ($1\ell + bb$ or $\gamma\gamma$)	20.3	✓		
ATLAS-SUSY-2014-03 [119]	$\geq$ 2(c-)jets	20.3			✓

ID	Short Description	$\mathcal{L}$ [fb^{-1}]	UL$_{\text{obs}}$	UL$_{\text{exp}}$	EM
CMS-EXO-12-026 [120]	HSCP	18.8	✓		
CMS-EXO-13-006 [121]	HSCP	18.8			✓
CMS-PAS-SUS-13-015 [122]	$\geq$ 5(b-)jets, top tag	19.4			✓
CMS-PAS-SUS-13-016 [123]	2 OS ℓ + (b-)jets	19.7	✓		✓
CMS-PAS-SUS-13-018 [124]	1–2 b-jets, M_{CT}	19.4	✓		
CMS-PAS-SUS-13-023 [125]	0ℓ stop	18.9	✓		
CMS-SUS-12-024 [126]	0ℓ + $\geq$ 3 (b-)jets	19.4	✓		✓
CMS-SUS-12-028 [127]	0ℓ + (b-)jets, α_T	11.7	✓	✓	
CMS-SUS-13-002 [128]	$\geq$ 3ℓ (+jets)	19.5	✓	✓	
CMS-SUS-13-004 [129]	$\geq$ 1 b-jet, Razor	19.3	✓		
CMS-SUS-13-006 [130]	multi-ℓ EWino	19.5	✓		
CMS-SUS-13-007 [131]	1ℓ + $\geq$ 2 b-jets	19.3	✓		✓
CMS-SUS-13-011 [132]	1ℓ + $\geq$ 4 (1b-)jets + $\not{E}_T$	19.5	✓		✓
CMS-SUS-13-012 [56]	0ℓ + 3–8 jets, MHT	19.5	✓	✓	✓
CMS-SUS-13-013 [133]	2 SS ℓ + (b-)jets	19.5	✓	✓	✓
CMS-SUS-13-019 [134]	$\geq$ 2 jets, M_{T2}	19.5	✓		
CMS-SUS-14-010 [135]	b-jets + 4 Ws (0–4ℓ)	19.5	✓	✓	
CMS-SUS-14-021 [136]	ISR jet + 1–2 soft ℓ	19.7	✓	✓	

Can't construct a likelihood

ATLAS' full likelihoods

ATL-PHYS-PUB-2019-029 (05 Aug 2019)

- Plain-text serialisation of HistFactory workspaces, JSON format

  - Provides background estimates, changes under systematic variations, and observed data counts at the same fidelity as used in the experiment.

	Description	Modification	Constraint Term c_χ	Input
constrained	Uncorrelated Shape	$\kappa_{scb}(\gamma_b) = \gamma_b$	$\prod_b \text{Pois}\left(r_b = \sigma_b^{-2} \mid \rho_b = \sigma_b^{-2}\gamma_b\right)$	σ_b
	Correlated Shape	$\Delta_{scb}(\alpha) = f_p\left(\alpha \mid \Delta_{scb,\alpha=-1}, \Delta_{scb,\alpha=1}\right)$	$\text{Gaus}\,(a = 0 \mid \alpha, \sigma = 1)$	$\Delta_{scb,\alpha=\pm1}$
	Normalisation Unc.	$\kappa_{scb}(\alpha) = g_p\left(\alpha \mid \kappa_{scb,\alpha=-1}, \kappa_{scb,\alpha=1}\right)$	$\text{Gaus}\,(a = 0 \mid \alpha, \sigma = 1)$	$\kappa_{scb,\alpha=\pm1}$
	MC Stat. Uncertainty	$\kappa_{scb}(\gamma_b) = \gamma_b$	$\prod_b \text{Gaus}\left(a_{\gamma_b} = 1 \mid \gamma_b, \delta_b\right)$	$\delta_b^2 = \sum_s \delta_{sb}^2$
	Luminosity	$\kappa_{scb}(\lambda) = \lambda$	$\text{Gaus}\,(l = \lambda_0 \mid \lambda, \sigma_\lambda)$	$\lambda_0, \sigma_\lambda$
free	Normalisation	$\kappa_{scb}(\mu_b) = \mu_b$		
	Data-driven Shape	$\kappa_{scb}(\gamma_b) = \gamma_b$		

Rate modifications defined in HistFactory for bin b, sample s, channel c.

  - Usage: RooFit, **pyhf**

  - Target: long-term data/analysis preservation, reinterpretation purposes

Why public likelihoods

- The statistical model of an experimental analysis provides the complete mathematical description of that analysis

  $p(o|\alpha)$ relating the observed quantities o to the parameters α

- Given the likelihood, all the standard statistical approaches are available for extracting information from it

- Essential information for any detailed interpretation of experimental results

  = determining the compatibility of the observations with theoretical predictions

Les Houches Recommandations (2012)

3b: When feasible, provide a mathematical description of the final likelihood function in which experimental data and parameters are clearly distinguished, either in the publication or the auxiliary information. Limits of validity should always be clearly specified.

3c: Additionally provide a digitized implementation of the likelihood that is consistent with the mathematical description.

arXiv:1203.2489