Feedback on use of public likelihoods

ATLAS SUSY and Exotics workshop - 24 Sep 2020 (virtual)

Sabine Kraml
LPSC Grenoble

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)

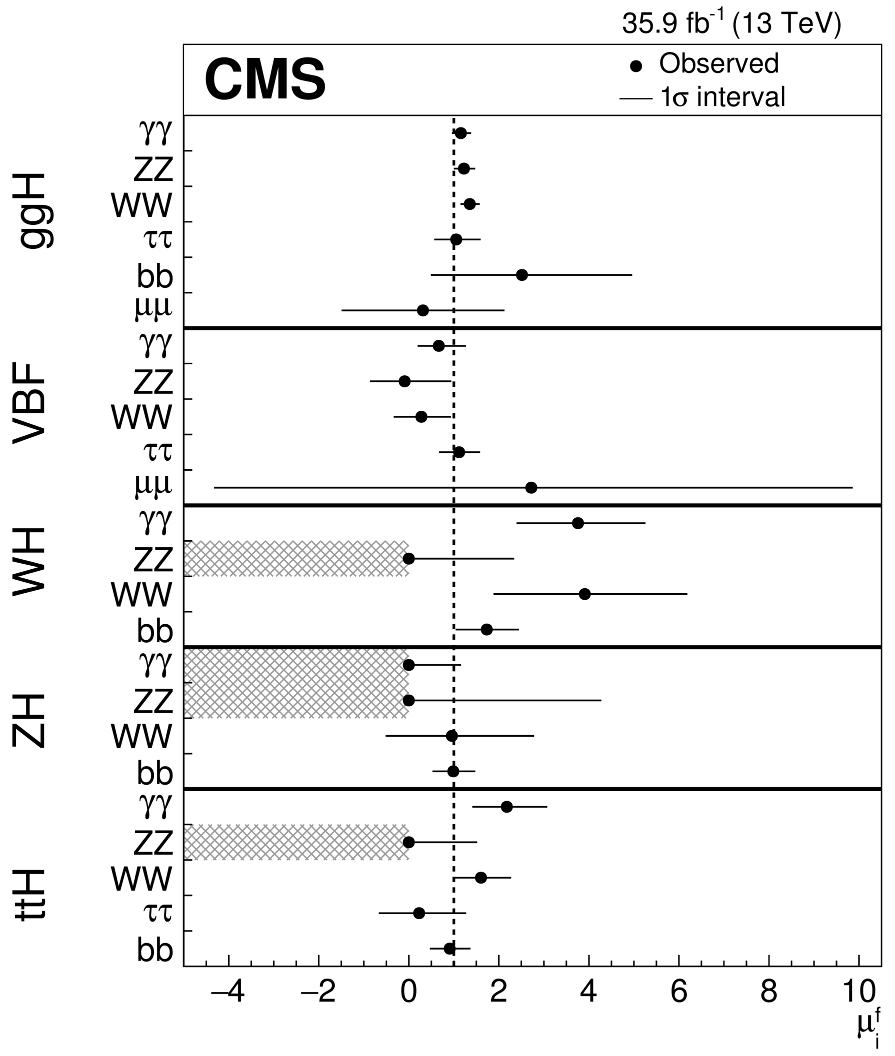
3b: When feasible, provide a mathematical description of the final likelihood function in which experimental data and parameters are clearly distinguished, either in the publication or the auxiliary information. Limits of validity should always be clearly specified.

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

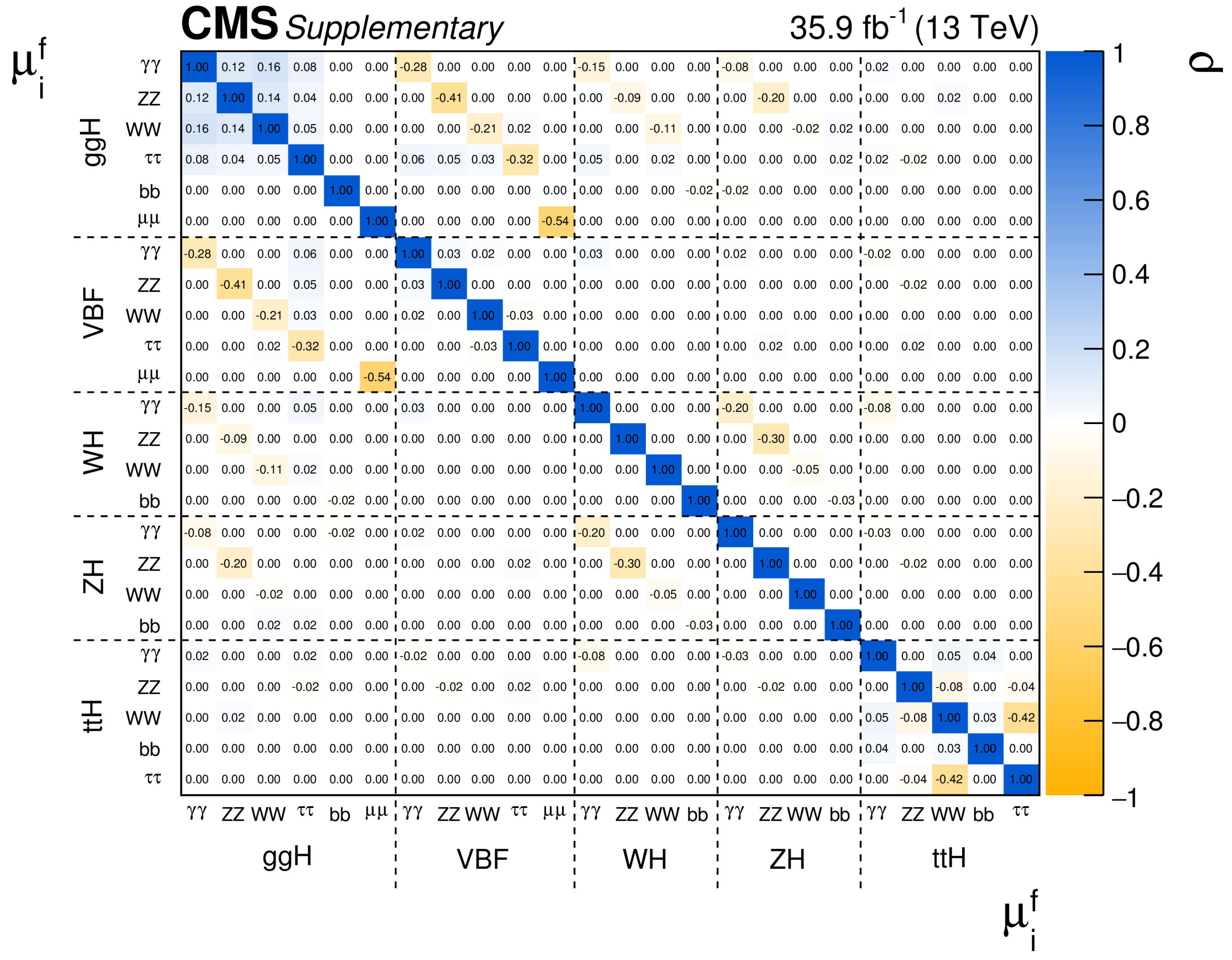
arXiv:1203.2489

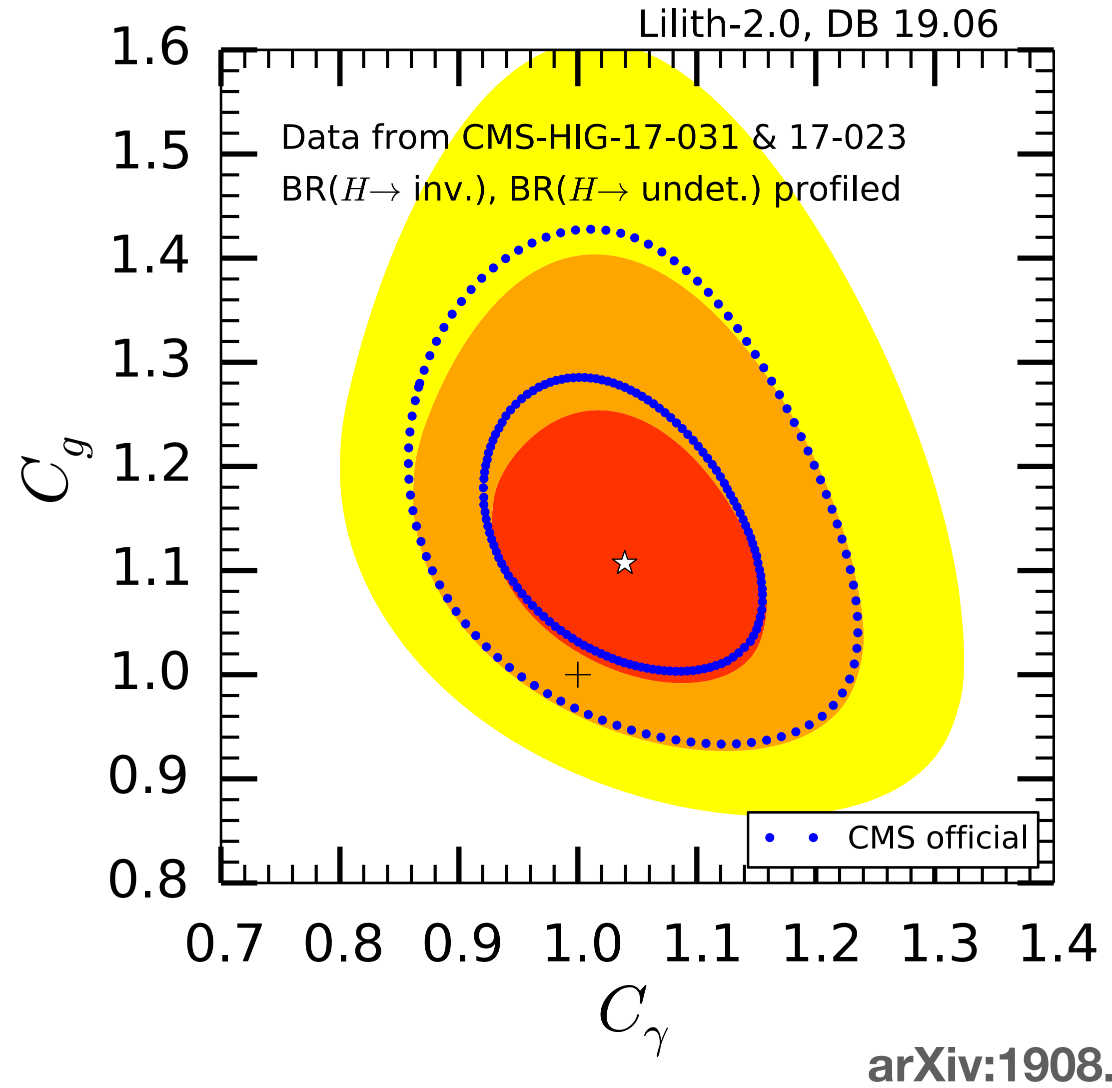
So far: O±δO plus correlations (sometimes)

- Simplified likelihood, Gaussian approximation
    - e.g., Higgs measurements, channel-by-channel correlation matrix



CMS-HIG-17-031





arXiv:1908.03952

«Correlation data […] has proven excellent for stabilising and ensuring better statistical definition in global fits as well as avoiding either overly conservative or over-enthusiastic interpretations.»

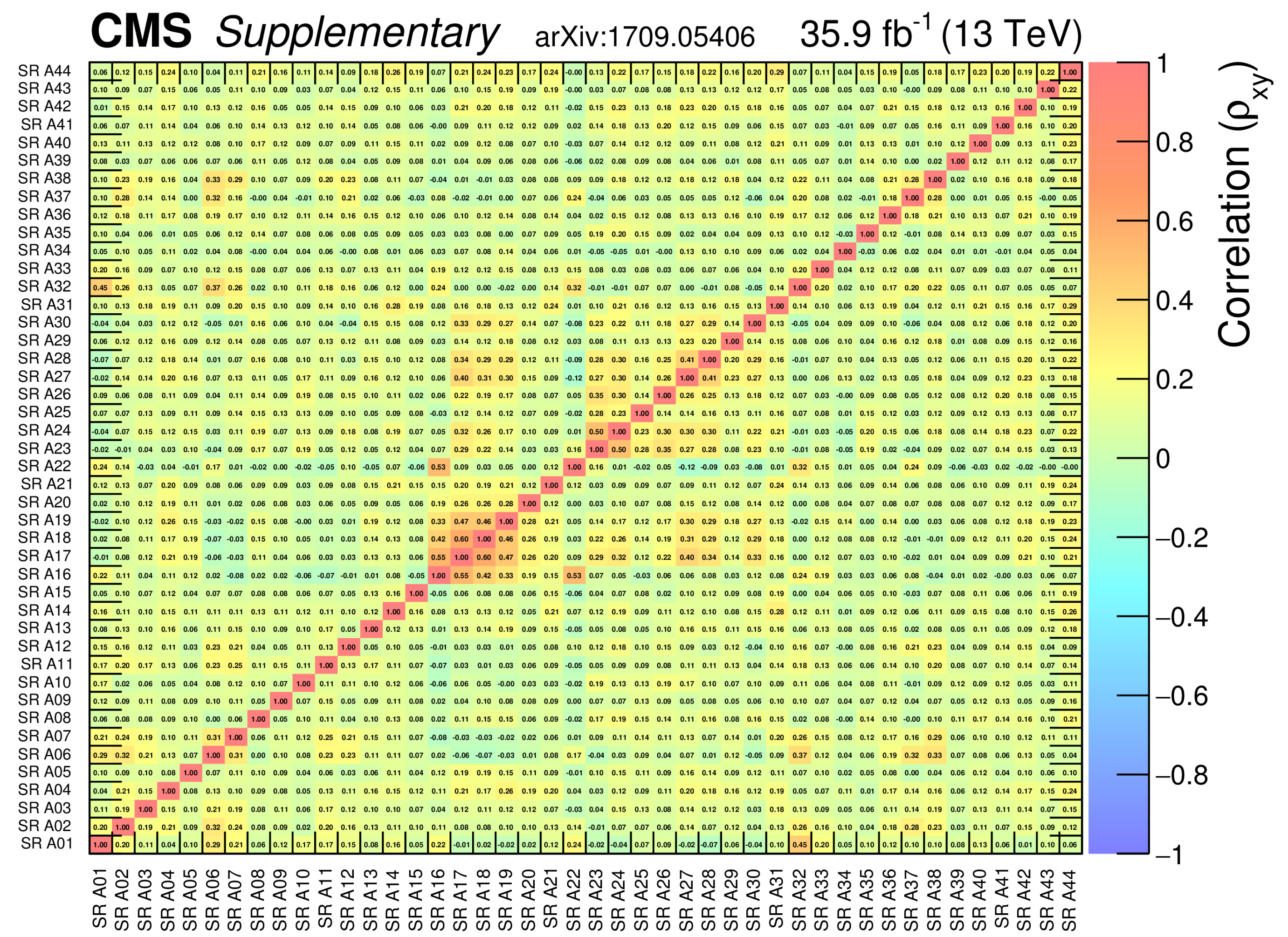
Reinterpretation Forum Report, 2003.07868

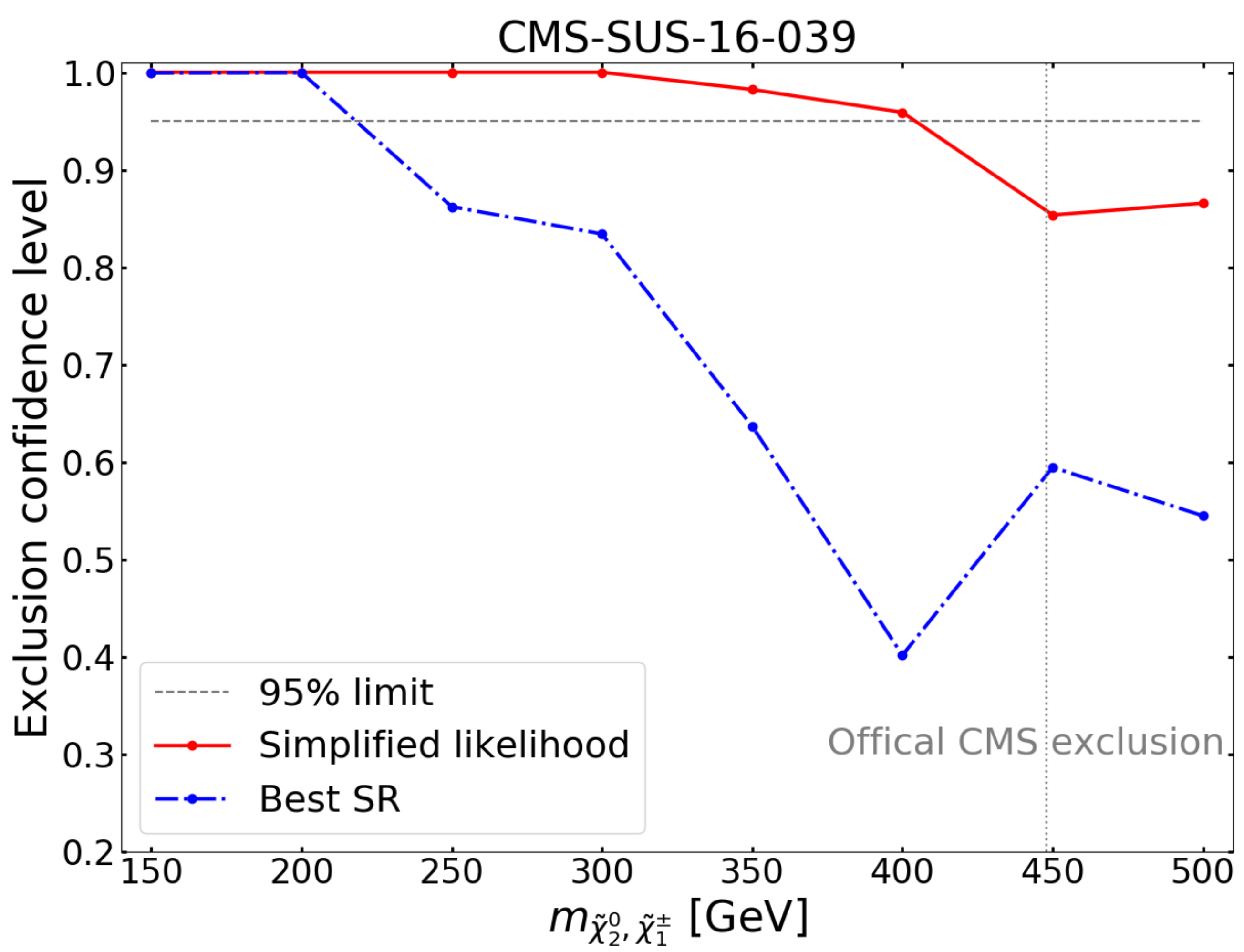
So far: O±δO plus correlations (sometimes)

- Simplified likelihood, Gaussian approximation
  - CMS SUSY group: covariance matrices for combination of signal regions

M_T (GeV)	p_T^{miss} (GeV)	$M_{\ell\ell} < 75$ GeV (exp.)	$M_{\ell\ell} < 75$ GeV (obs.)	$75 \leq M_{\ell\ell} < 105$ GeV (exp.)	$75 \leq M_{\ell\ell} < 105$ GeV (obs.)	$M_{\ell\ell} \geq 105$ GeV (exp.)	$M_{\ell\ell} \geq 105$ GeV (obs.)
0 – 100	50 – 100	185 ± 22	186	2180 ± 260	2278	121 ± 14	123
	100 – 150	35 ± 6	34	440 ± 70	429	32 ± 5	32
	150 – 200	9.3 ± 2.2	11	129 ± 28	123	11.6 ± 2.6	4
	200 – 250	3.3 ± 1.0	1	48 ± 10	37	2.9 ± 0.8	6
	250 – 400			42 ± 9	38		
	400 – 550	4.0 ± 1.0	5	8.5 ± 2.1	5	3.7 ± 1.0	5
	≥550			2.6 ± 0.8	2		
100 – 160	50 – 100	50 ± 8	60	390 ± 50	391	32 ± 5	17
	100 – 150	15 ± 4	19	72 ± 19	61	9.6 ± 2.4	9
	150 – 200	1.9 ± 0.6	1	10 ± 4	9	2.4 ± 0.7	0
	≥200	0.8 ± 0.4	3	4.9 ± 1.9	8	1.0 ± 0.4	2
≥160	50 – 100	13.0 ± 2.8	16	37 ± 9	35	9.4 ± 2.4	9
	100 – 150	11.9 ± 3.2	17	21 ± 8	17	6.6 ± 2.1	3
	150 – 200	3.1 ± 1.2	4	8.9 ± 3.1	7	3.1 ± 1.0	0
	200 – 250	2.1 ± 0.8	3	3.6 ± 1.3	5		
	250 – 400	0.9 ± 0.4	1	4.1 ± 1.6	3	2.5 ± 0.8	0
	≥400			1.0 ± 0.5	1		

CMS-SUS-16-039 (EWino search)





Contribution 15, LH 2019 BSM WG report
arXiv:2002.12220

«Correlation data […] has proven excellent for stabilising and ensuring better statistical definition in global fits as well as avoiding either overly conservative or over-enthusiastic interpretations.»

Reinterpretation Forum Report, 2003.07868

So far: O±δO plus correlations (sometimes)

- Simplified likelihood, Gaussian approximation
    - Extremely useful, but only an approximation, lots of relevant information is lost.
    - One problem is that possible non-Gaussian tails are ignored.
      Not an issue if the uncertainties are small. However, if the uncertainties are large the likelihood should be modelled in more detail.

See also A. Buckley et al, 1809.05548
(Gaussian approximation with a skew)

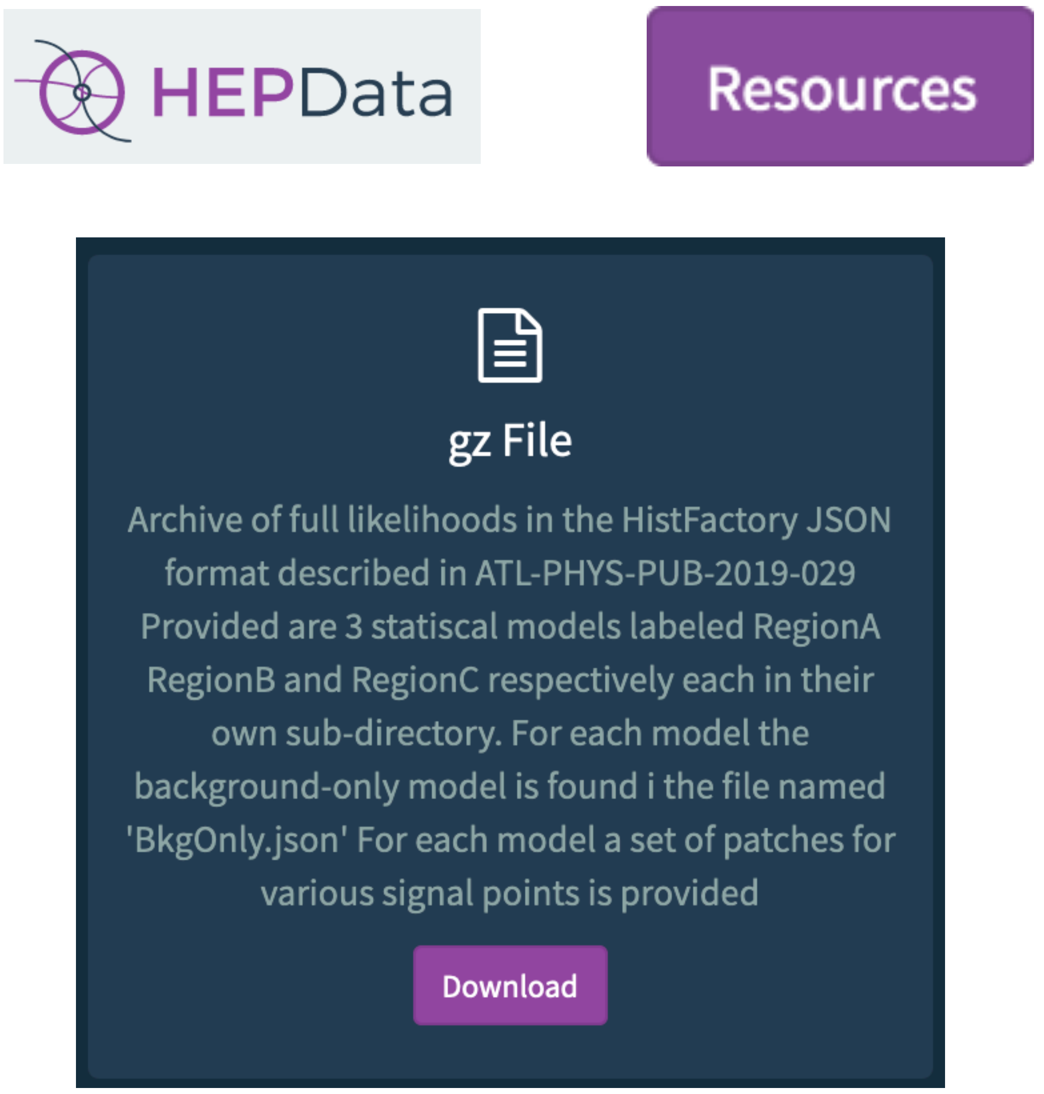
Now: full likelihoods !!

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	$\kappa_{scb}(\gamma_b) = \gamma_b$	$\prod_b \text{Pois}\left(r_b = \sigma_b^{-2} \middle\vert \rho_b = \sigma_b^{-2}\gamma_b\right)$	σ_b
	Correlated Shape	$\Delta_{scb}(\alpha) = f_p\left(\alpha \middle\vert \Delta_{scb,\alpha=-1}, \Delta_{scb,\alpha=1}\right)$	$\text{Gaus}\,(a = 0 \vert\, \alpha, \sigma = 1)$	$\Delta_{scb,\alpha=\pm 1}$
	Normalisation Unc.	$\kappa_{scb}(\alpha) = g_p\left(\alpha \middle\vert \kappa_{scb,\alpha=-1}, \kappa_{scb,\alpha=1}\right)$	$\text{Gaus}\,(a = 0 \vert\, \alpha, \sigma = 1)$	$\kappa_{scb,\alpha=\pm 1}$
	MC Stat. Uncertainty	$\kappa_{scb}(\gamma_b) = \gamma_b$	$\prod_b \text{Gaus}\left(a_{\gamma_b} = 1 \middle\vert \gamma_b, \delta_b\right)$	$\delta_b^2 = \sum_s \delta_{sb}^2$
	Luminosity	$\kappa_{scb}(\lambda) = \lambda$	$\text{Gaus}\,(l = \lambda_0 \vert\, \lambda, \sigma_\lambda)$	$\lambda_0, \sigma_\lambda$
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*.

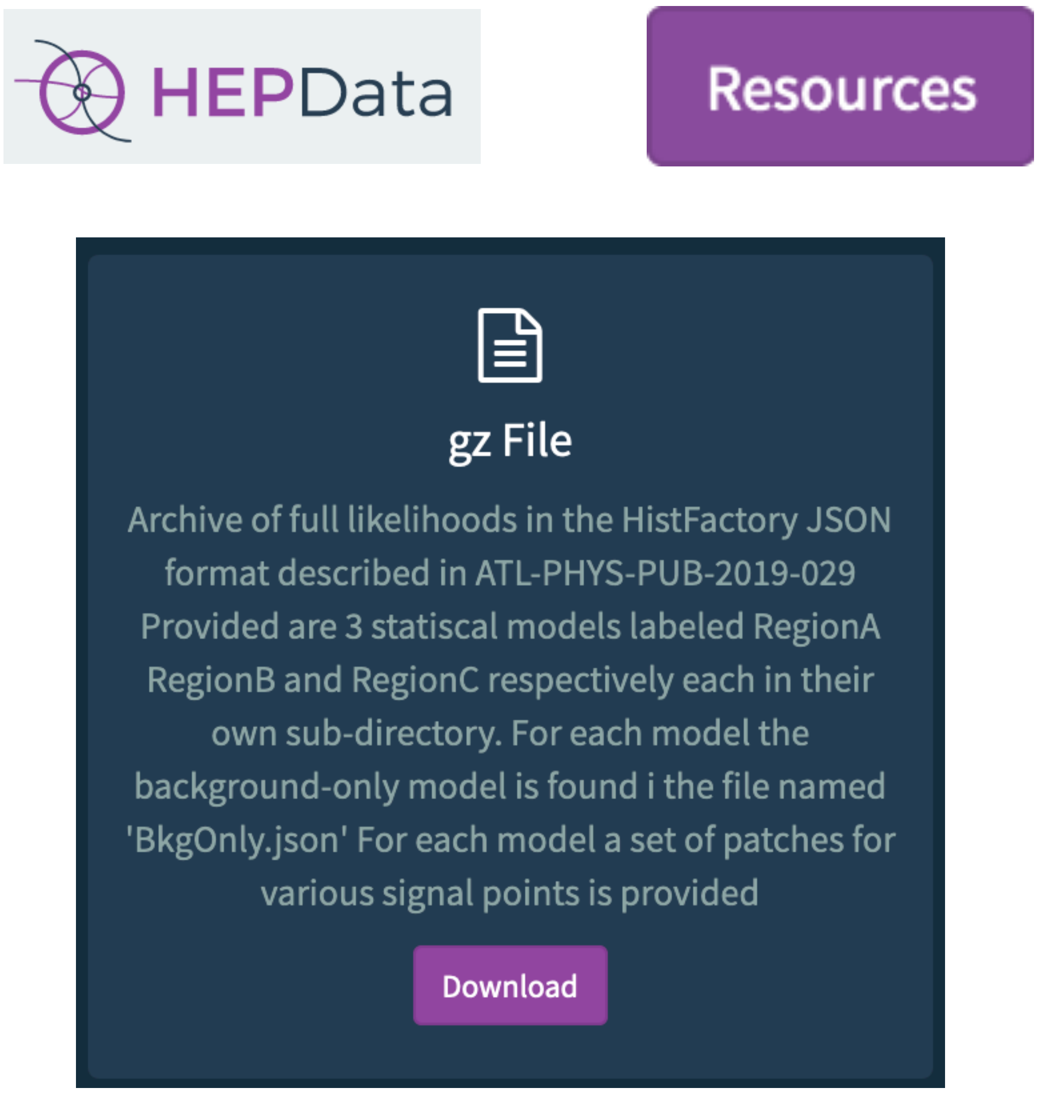


  - Usage: RooFit, **pyhf**
  - Target: long-term data/analysis preservation, reinterpretation purposes

Now: 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	$\Delta_{scb}(\alpha) = f_p\left(\alpha \mid \Delta_{scb,\alpha=-1}, \Delta_{scb,\alpha=1}\right)$	Gaus $(a = 0 \mid \alpha, \sigma = 1)$	$\Delta_{scb,\alpha=\pm1}$
	Normalisation Unc.	$\kappa_{scb}(\alpha) = g_p\left(\alpha \mid \kappa_{scb,\alpha=-1}, \kappa_{scb,\alpha=1}\right)$	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	$\kappa_{scb}(\lambda) = \lambda$	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

So far available for 4/12 SUSY analyses with 139 fb^{-1}

SUSY-2018-31 (1908.03122)	multi-b sbottom: 2b+2H(bb)
SUSY-2018-04 (1911.06660)	stau search, 2 hadr. taus
SUSY-2019-08 (1909.09226)	1 lept. + H(bb), EW-ino
SUSY-2018-06 (1912.08479)	3 lept. EW-ino

Reinterpretation Forum Report 2020

"…. In fact, many of the data products discussed here, such as signal/background yields and correlations, are used by the various external reinterpretation packages to construct likelihoods. Whilst extremely useful, the likelihoods constructed from these products are however always only an approximation to the true underlying experimental likelihood. The reinterpretation workflow can be greatly facilitated and rendered much more precise if the original likelihood of the analysis is published in full. We strongly encourage the movement towards the publication of full experimental likelihoods wherever possible."

"ATLAS has recently started to do this using a JSON serialisation of the likelihood […] The provision of this full likelihood information is much appreciated and we hope that it will become a standard, as it **greatly improves the quality of any reinterpretation**."

Reinterpretation of LHC Results for New Physics: Status and Recommendations after Run 2
arXiv:2003.07868, SciPost Phys. 9, 022 (2020)

What’s the status of usage of the ATLAS pyhf/JSON likelihoods
in public reinterpretation tools?

G. Alguero, **J. Araz**, B. Fuks, SK, W. Waltenberger
Contr. 15, LH 2019 BSM WG report, arXiv:2002.12220

Usage in MadAnalysis 5

MA5-pyhf interface established within the LH PhysTeV 2019 workshop

- the relevant JSON files must be located in the same analysis folder as the recast code (done automatically at the time of the PAD installation).
- The `analysis.info` file must include new <pyhf> elements specifying the names of the JSON files together with the corresponding channels (ensembles of SRs) and the regions they include, as defined in the JSON files.

```
<pyhf id="RegionA">
  <name>atlas_susy_2018_031_SRA.json</name>
    <regions>
      <channel name="SR_meff">SRA_L SRA_M SRA_H</channel>
      <channel name="VRtt_meff"> </channel>
      <channel name="CRtt_meff"> </channel>
    </regions>
</pyhf>
```

G. Alguero, **J. Araz**, B. Fuks, SK, W. Waltenberger
Contr. 15, LH 2019 BSM WG report, arXiv:2002.12220

Usage in MadAnalysis 5

MA5-pyhf interface established within the LH PhysTeV 2019 workshop

- the relevant JSON files must be located in the same analysis folder as the recast code (done automatically at the time of the PAD installation).
- The `analysis.info` file must include new <pyhf> elements specifying the names of the JSON files together with the corresponding channels (ensembles of SRs) and the regions they include, as defined in the JSON files.

```
<pyhf id="RegionA">
  <name>atlas_susy_2018_031_SRA.json</name>
    <regions>
      <channel name="SR_meff">SRA_L SRA_M SRA_H</channel>
      <channel name="VRtt_meff"> </channel>
      <channel name="CRtt_meff"> </channel>
    </regions>
</pyhf>
```

Output in `CLs_output_summary.dat` file

```
<set> <tag> <SR> <best?> <exp> <obs> <CLs> ||

defaultset atlas_susy_2018_031 [pyhf]-RegionA-profile 1 0.0020761 0.0013821 0.9959  ||
defaultset atlas_susy_2018_031 [pyhf]-RegionB-profile 0 -1        -1        0.0000  ||
defaultset atlas_susy_2018_031 [pyhf]-RegionC-profile 0 0.1139845 0.1846410 0.0293  ||
```

G. Alguero, **J. Araz**, B. Fuks, SK, W. Waltenberger
Contr. 15, LH 2019 BSM WG report, arXiv:2002.12220

Usage in MadAnalysis 5

MA5-pyhf interface established within the LH PhysTeV 2019 workshop

- the relevant JSON files must be located in the same analysis folder as the recast code (done automatically at the time of the PAD installation).
- The `analysis.info` file must include new <pyhf> elements specifying the names of the JSON files together with the corresponding channels (ensembles of SRs) and the regions they include, as defined in the JSON files.

```
<pyhf id="RegionA">
  <name>atlas_susy_2018_031_SRA.json</name>
    <regions>
      <channel name="SR_meff">SRA_L SRA_M SRA_H</channel>
      <channel name="VRtt_meff"> </channel>
      <channel name="CRtt_meff"> </channel>
    </regions>
</pyhf>
```

However, this is not public yet,
relevant recast codes are still being validated
☞ hopefully in v1.9 release

Output in `CLs_output_summary.dat` file

```
<set> <t                    > <obs> <CLs> ||

defaults       susy_2018_031 [pyhf]-RegionA-profile 1 0.0020761 0.0013821 0.9959  ||
defaultse  atlas_susy_2018_031 [pyhf]-RegionB-profile 0 -1        -1        0.0000  ||
defaultset atlas_susy_2018_031 [pyhf]-RegionC-profile 0 0.1139845 0.1846410 0.0293  ||
```

PhD student who did the technical implementation, validation, testing, etc.

Usage in SModelS

Gaël Alguero, SK, Wolfgang Waltenberger,
arXiv:2009.01809

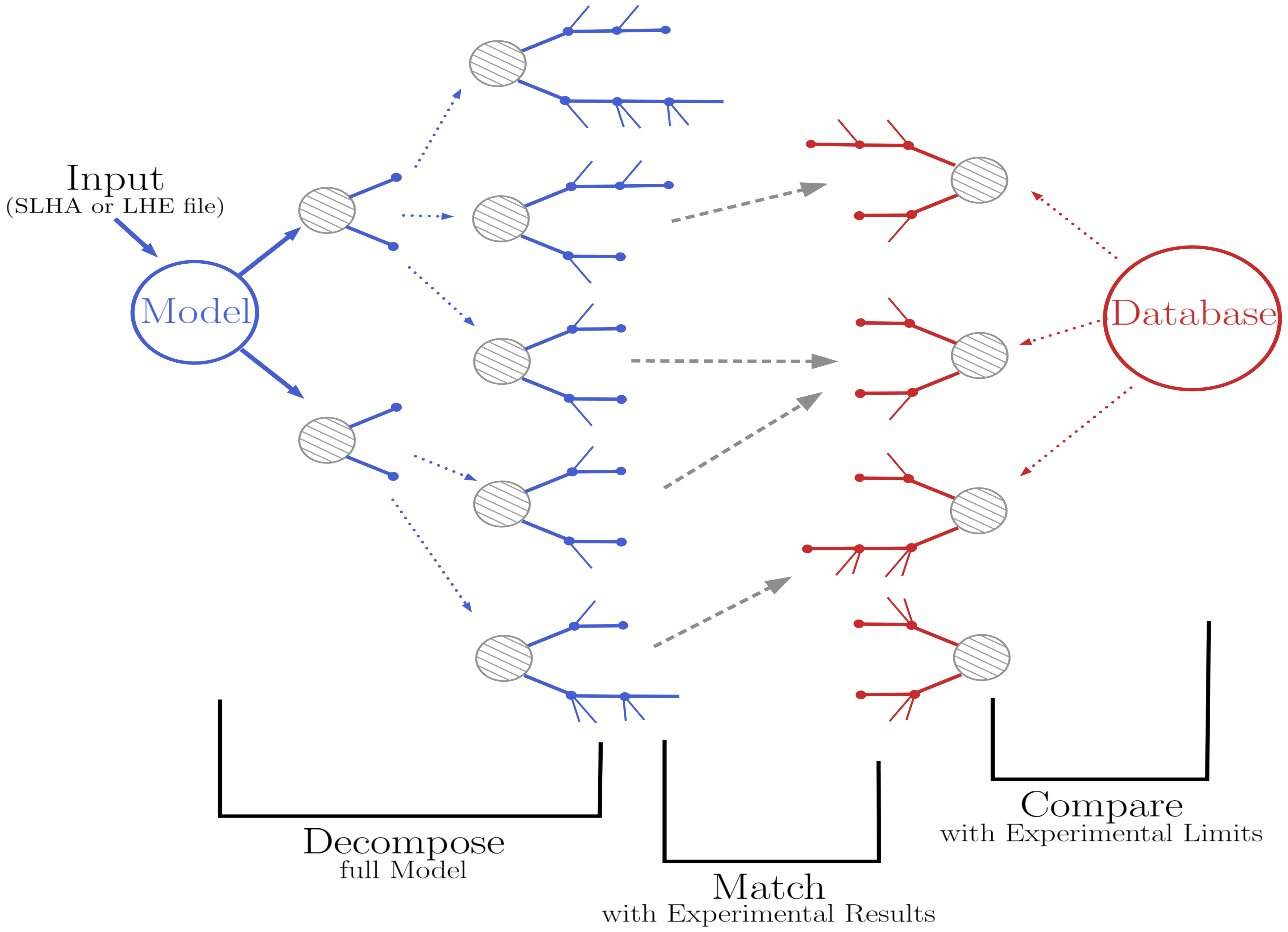
New version, v1.2.4, including support of pyhf released Sep. 3rd, 2020

Usage in SModelS

smodels.github.io

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

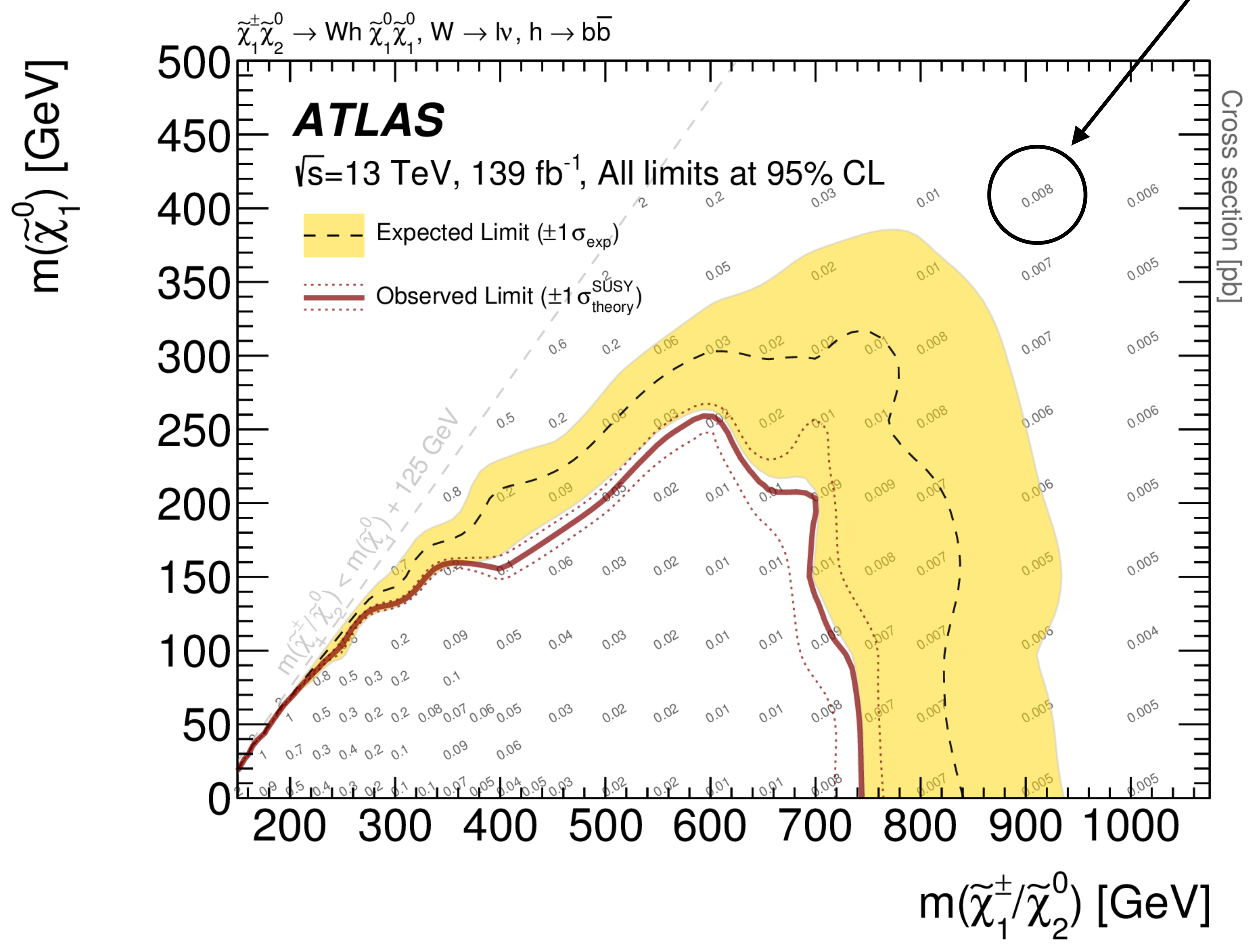
- Based on a general procedure to decompose BSM collider signatures featuring a Z_2-like symmetry into simplified-model topologies.
- Large database of simplified-model results (currently 40 ATLAS & 46 CMS searches).
- Very fast b/c no need for MC simulation.

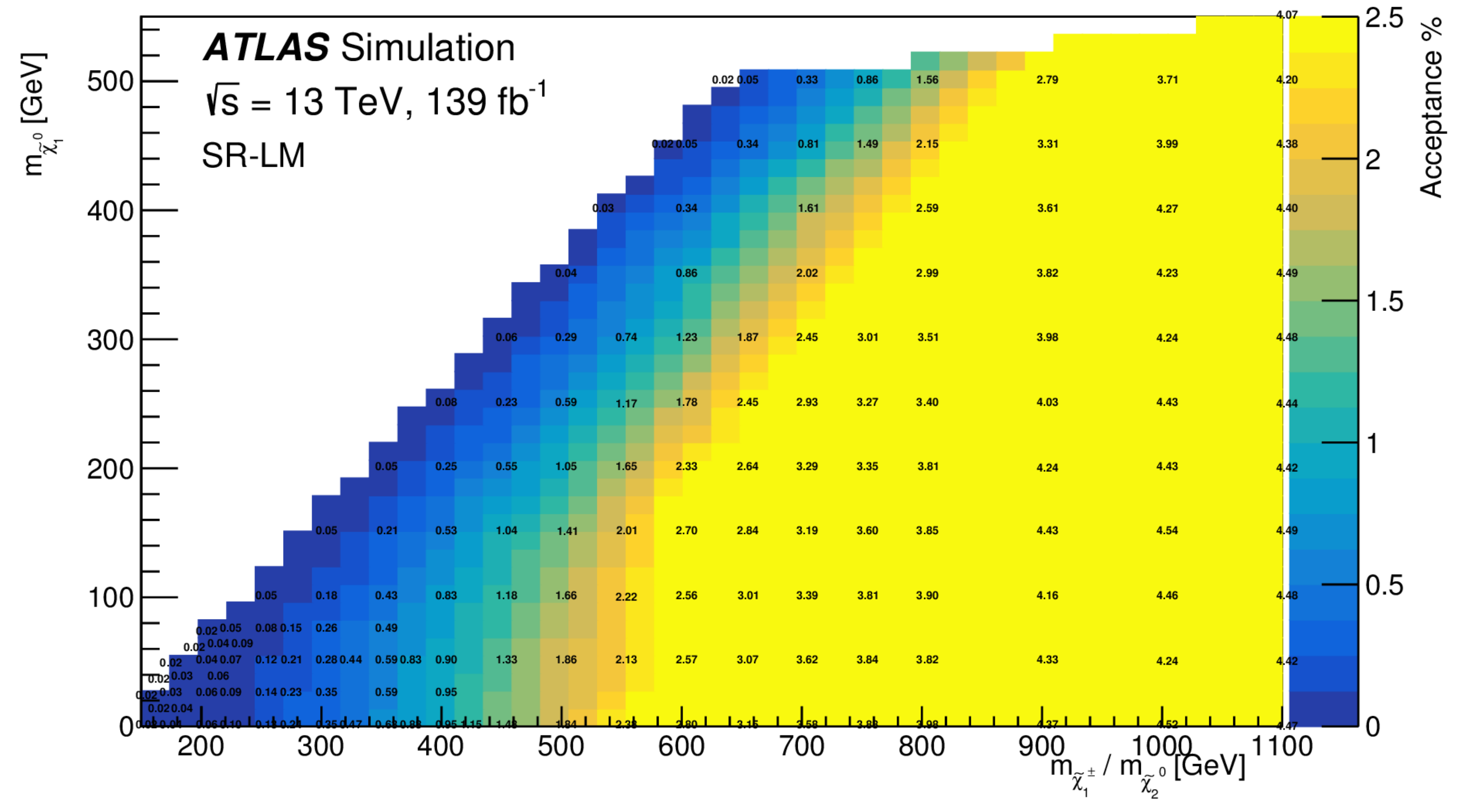


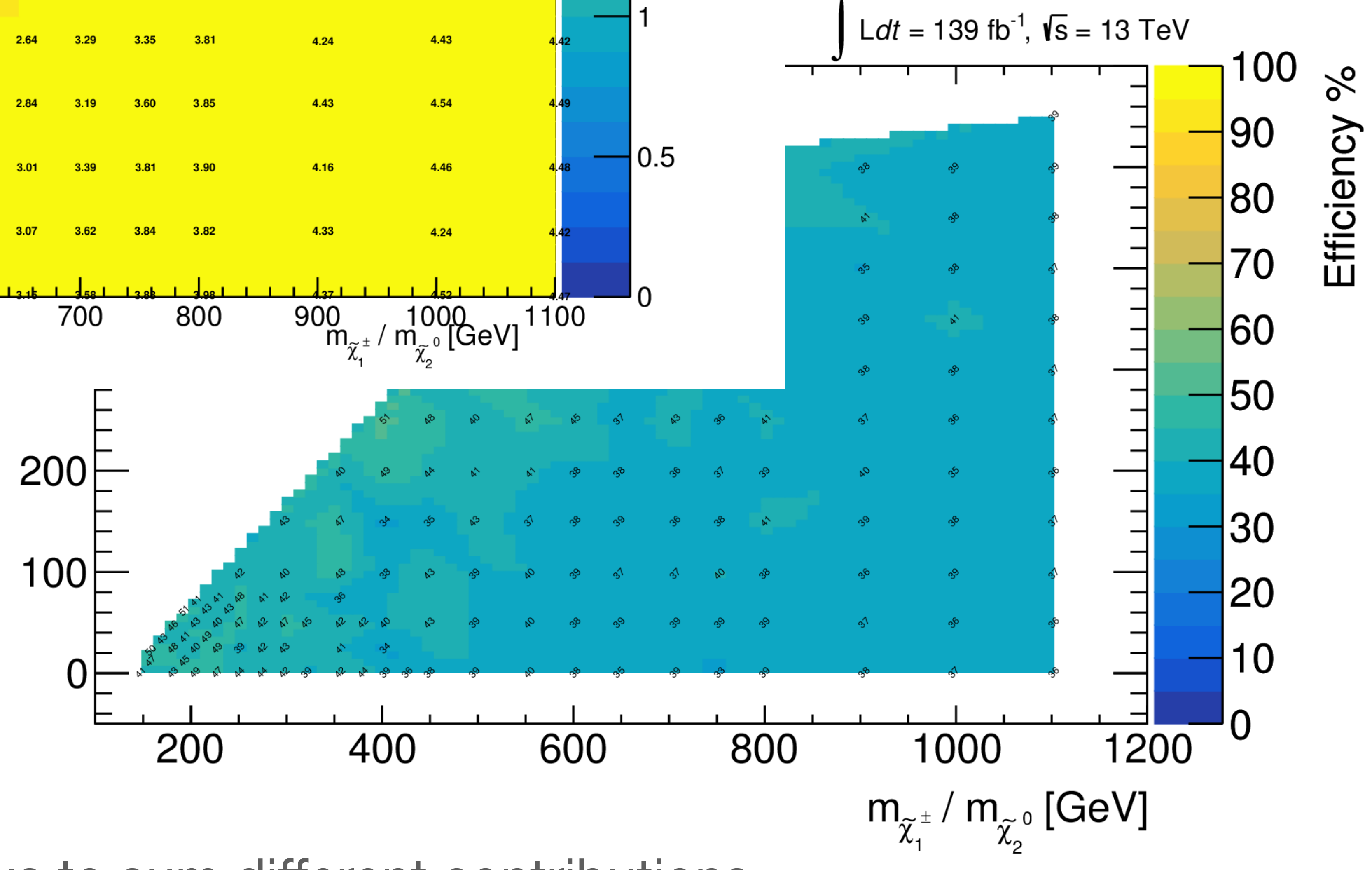
Usage in SModelS

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***.

SModelS-pyhf interface

Gaël Alguero, SK, Wolfgang Waltenberger,
arXiv:2009.01809

- Available from SModelS v1.2.4 onward (released Sep. 3rd, 2020)
- The interfacing of pyhf to SModelS consists of two parts:
  - addition of an independent module `tools/pyhfInterface.py`
  - changes brought to `experiment/datasetObj.py`
- Can be turned on/off by setting

```
combineSR = True/False
```

in the `parameters.ini` file *)

`PyhfData` class:

Storing and handling of the information related to the JSON files and input signal predictions.

Collects information in the workspaces such as the number of SRs, and the paths to the SR `samples` where the BSM predictions are to be written.

The VRs and CRs are assumed not to contribute and removed from the workspaces.

`PyhfUpperLimitComputer` class:

For inferring the upper limits given the `PyhfData` information

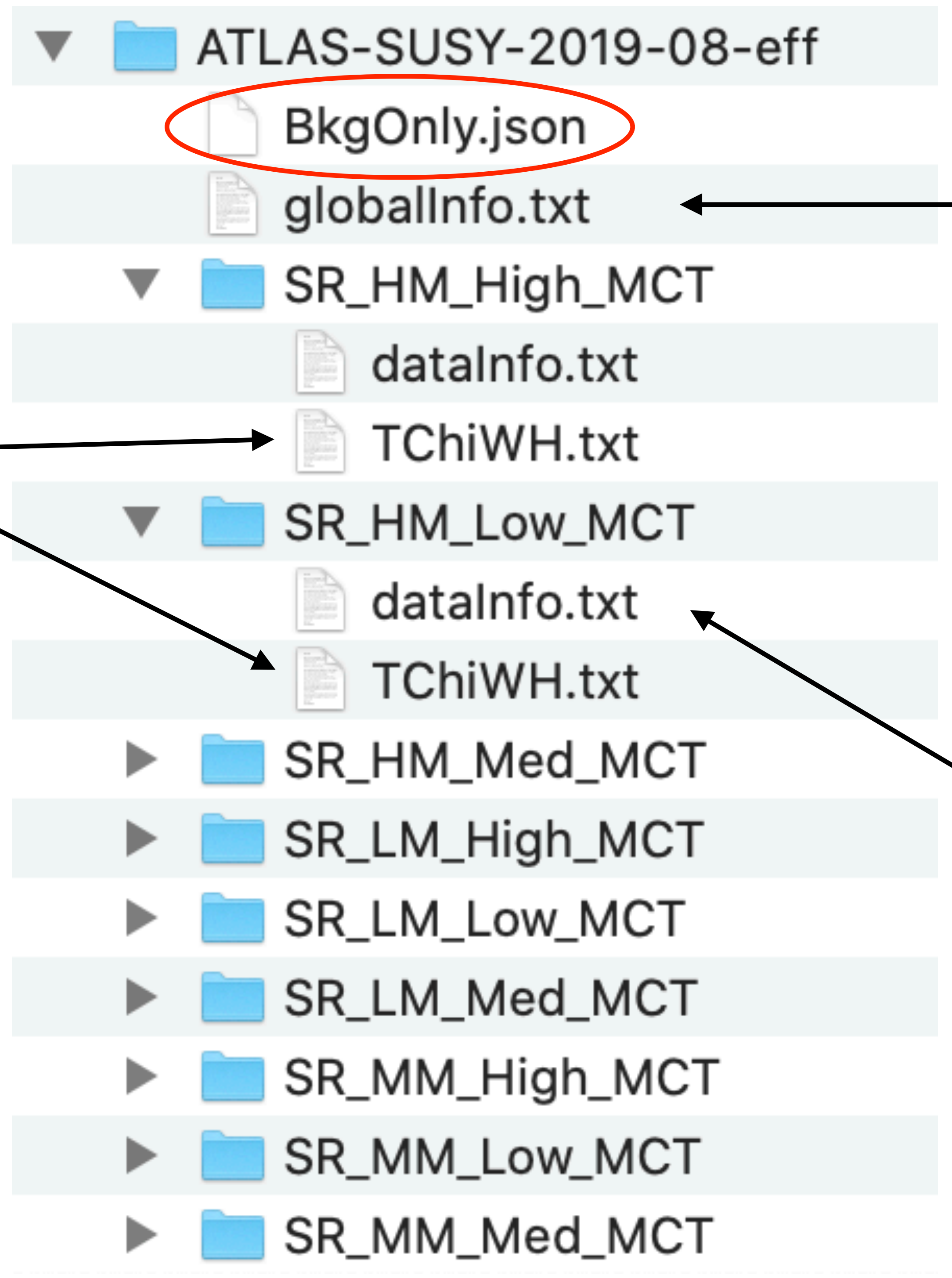
*) The same flag also turns on the SR combination in the simplified likelihood approach for CMS efficiency map results, for which a covariance matrix is available.

SModelS-pyhf interface

Gaël Alguero, SK, Wolfgang Waltenberger,
arXiv:2009.01809

- Available from SModelS v1.2.4 onward (released Sep. 3rd, 2020)
- Implementation in the database:

The actual A×ε maps for each signal region



id: ATLAS-SUSY-2019-08
sqrts: 13*TeV
lumi: 139.0/fb
url: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/SUSY-2019-08/
arxiv: arXiv:1909.09226
contact: atlas-phys-susy-conveners@cern.ch
implementedBy: Gael Alguero
lastUpdate: 2020/8/25
datasetOrder: "SR_HM_Low_MCT", "SR_HM_Med_MCT", "SR_HM_High_MCT", "SR_MM_Low_MCT", "SR_MM_Med_MCT", "SR_MM_High_MCT", "SR_LM_Low_MCT", "SR_LM_Med_MCT", "SR_LM_High_MCT"
jsonFiles: {"BkgOnly.json" : ["SR_HM_Low_MCT", "SR_HM_Med_MCT", "SR_HM_High_MCT", "SR_MM_Low_MCT", "SR_MM_Med_MCT", "SR_MM_High_MCT", "SR_LM_Low_MCT", "SR_LM_Med_MCT", "SR_LM_High_MCT"]}

dataId: SR_HM_Low_MCT
dataType: efficiencyMap
observedN: 6
expectedBG: 4.1
bgError: 1.9
upperLimit: 0.0688*fb
expectedUpperLimit: 0.0505*fb

SModelS-pyhf interface

Gaël Alguero, SK, Wolfgang Waltenberger,
arXiv:2009.01809

- Available from SModelS v1.2.4 onward (released Sep. 3rd, 2020)
- Output

```
Input status: 1
Decomposition output status: 1 #decomposition was successful
# Input File: ../smodels-utils/slha/TChiWH_750_100_750_100.slha
# maxcond = 0.2
# minmassgap = 5.
# ncpus = 1
# sigmacut = 0.01
# Database version: 1.2.4
============================================================================
#Analysis  Sqrts  Cond_Violation  Theory_Value(fb)  Exp_limit(fb)  r  r_expected

 ATLAS-SUSY-2019-08  1.30E+01    0.0  9.798E-02  1.035E-01  9.471E-01  1.260E+00
 Signal Region:  (combined)
 Txnames:  TChiWH
 Chi2, Likelihood =  6.930E-01  6.671E-45
----------------------------------------------------------------------------
 ATLAS-SUSY-2019-08  1.30E+01    0.0  4.325E-02  4.670E-02  9.262E-01  1.614E+00
 Signal Region:  SR_HM_High_MCT
 Txnames:  TChiWH
 Chi2, Likelihood =  3.993E+00  1.933E-02
```

$$pp \to \tilde{\chi}_1^\pm \tilde{\chi}_2^0 \to W^\pm h + E_T^{\rm miss}$$

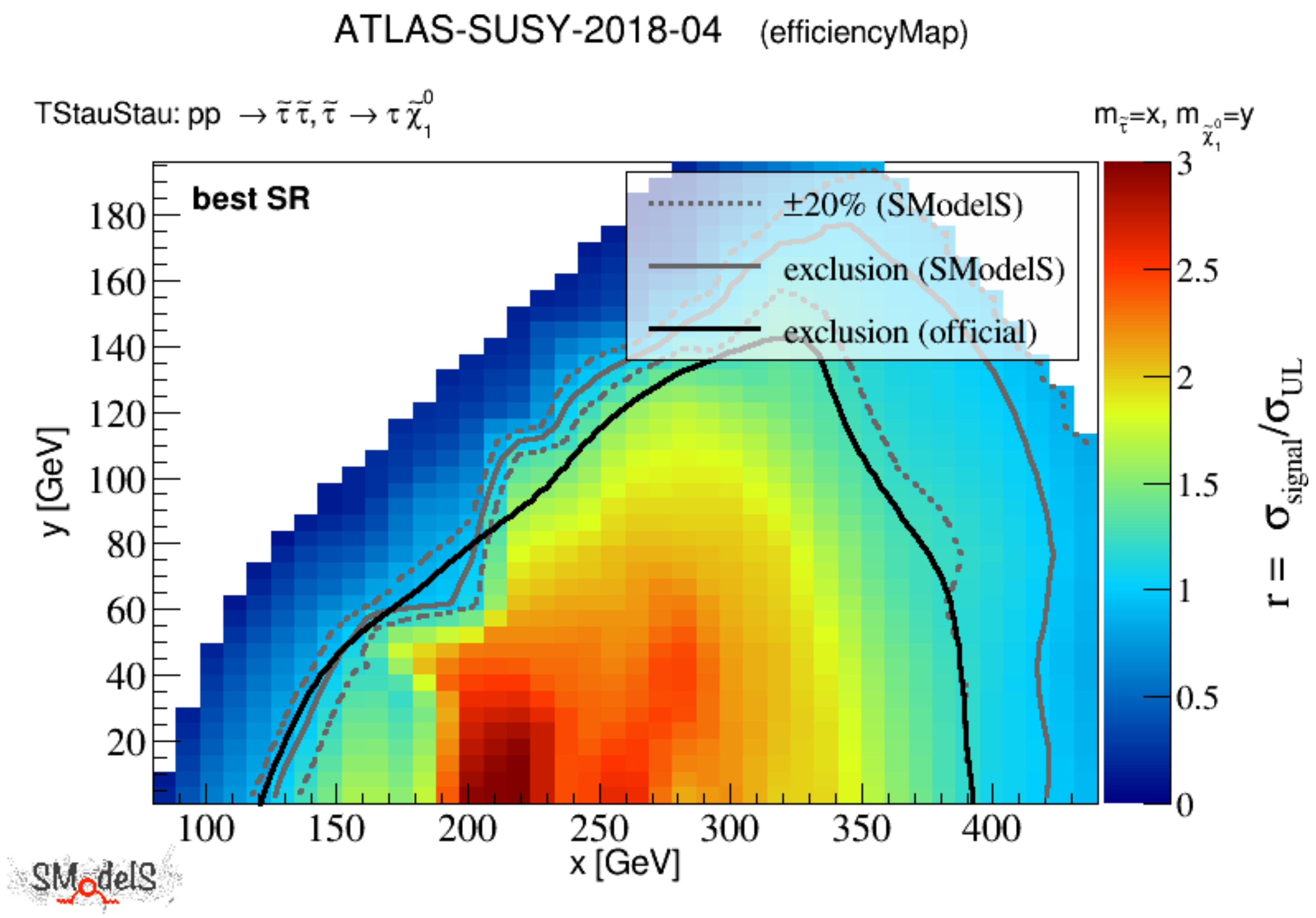
$$m_{\tilde{\chi}_1^\pm, \tilde{\chi}_2^0} = 750 \text{ GeV}$$

$$m_{\tilde{\chi}_1^0} = 100 \text{ GeV}$$

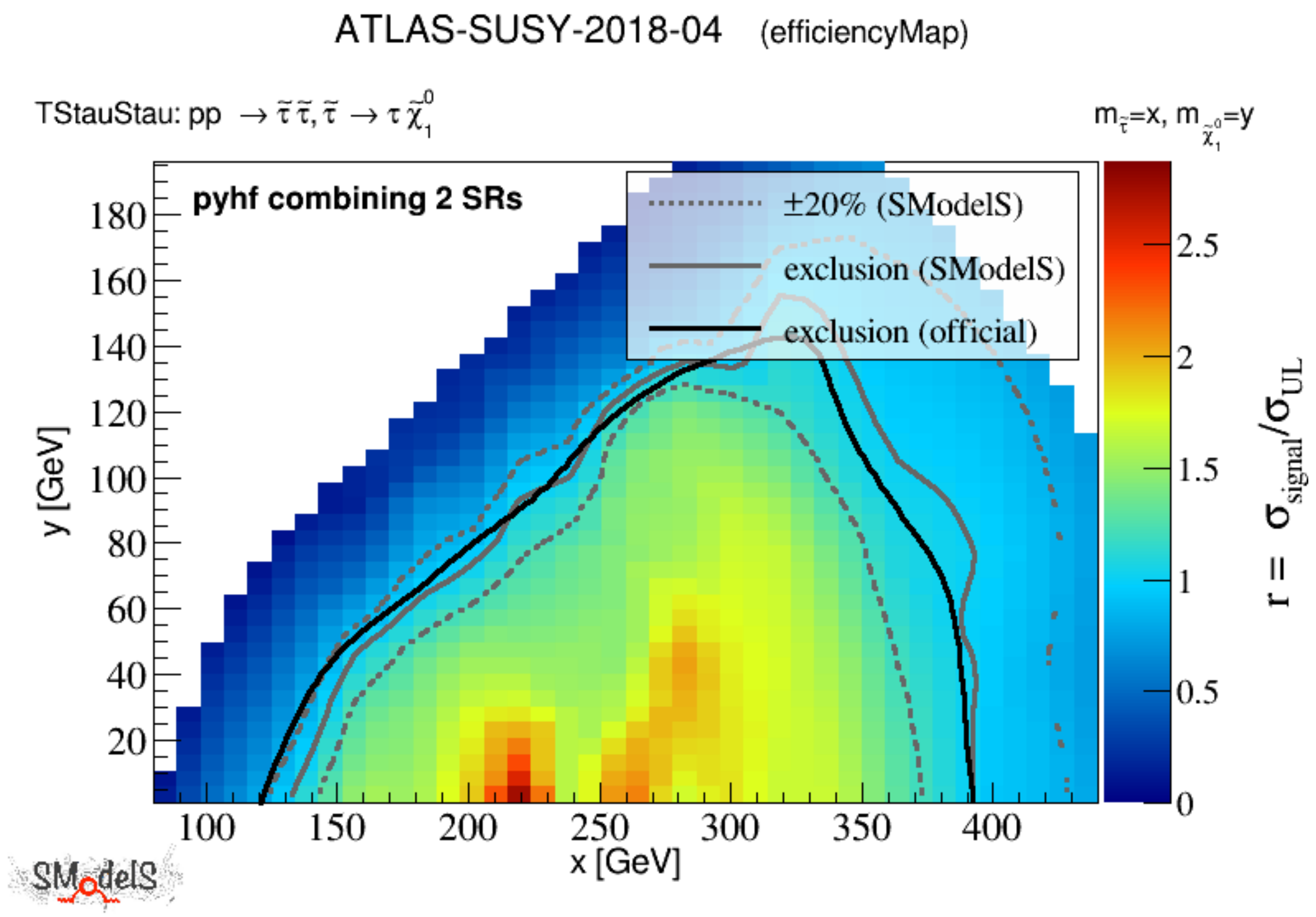
Validation & impact

Gaël Alguero, SK, Wolfgang Waltenberger,
arXiv:2009.01809

- ATLAS-SUSY-2018-04: TStauStau



Best SR: over exclusion



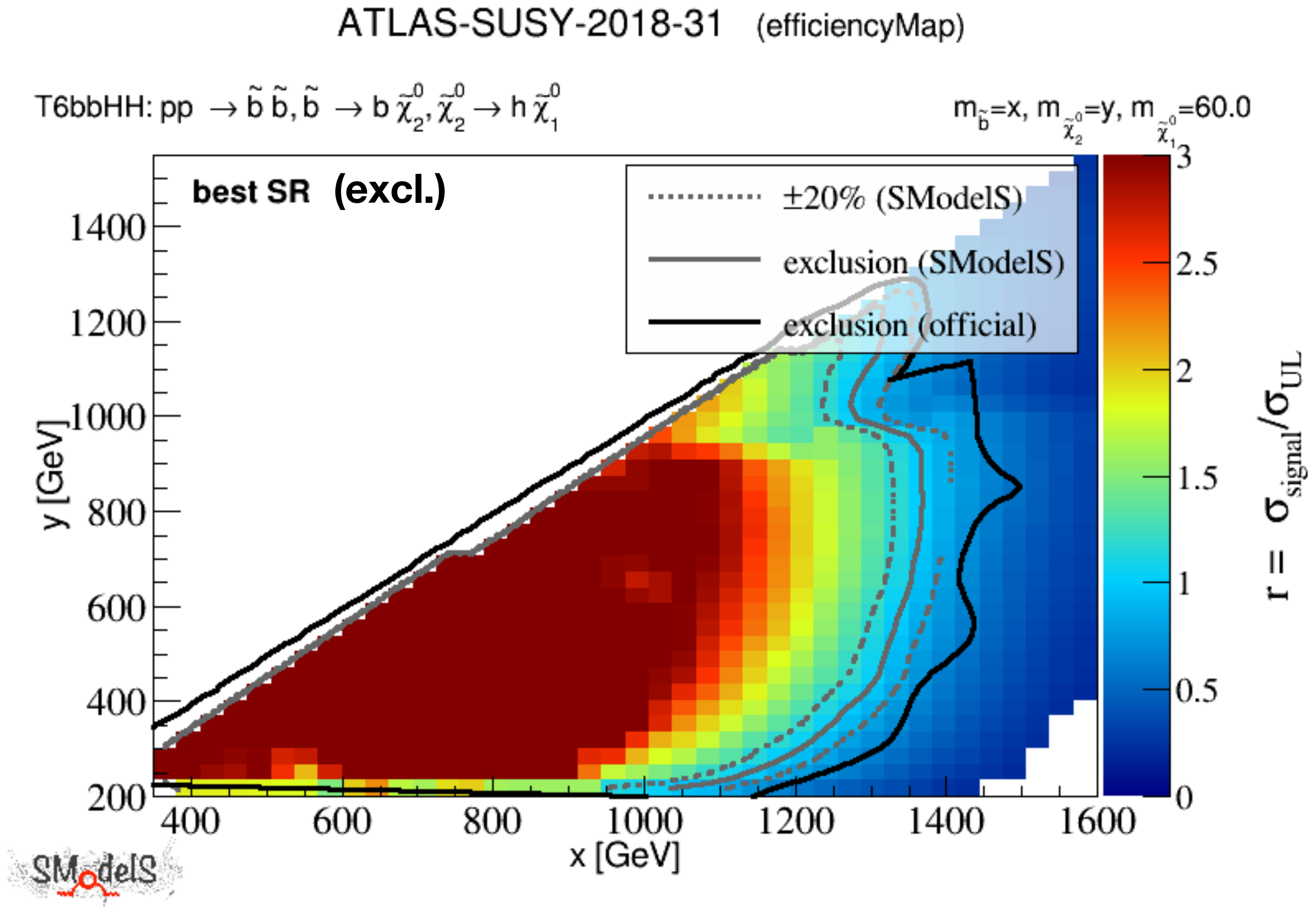
Full likelihood: very good agreement
with official ATLAS result

The remaining small difference is probably due to the (interpolated) A×ε values from the simplified model efficiency maps not exactly matching the “true”ones of the experimental analysis.

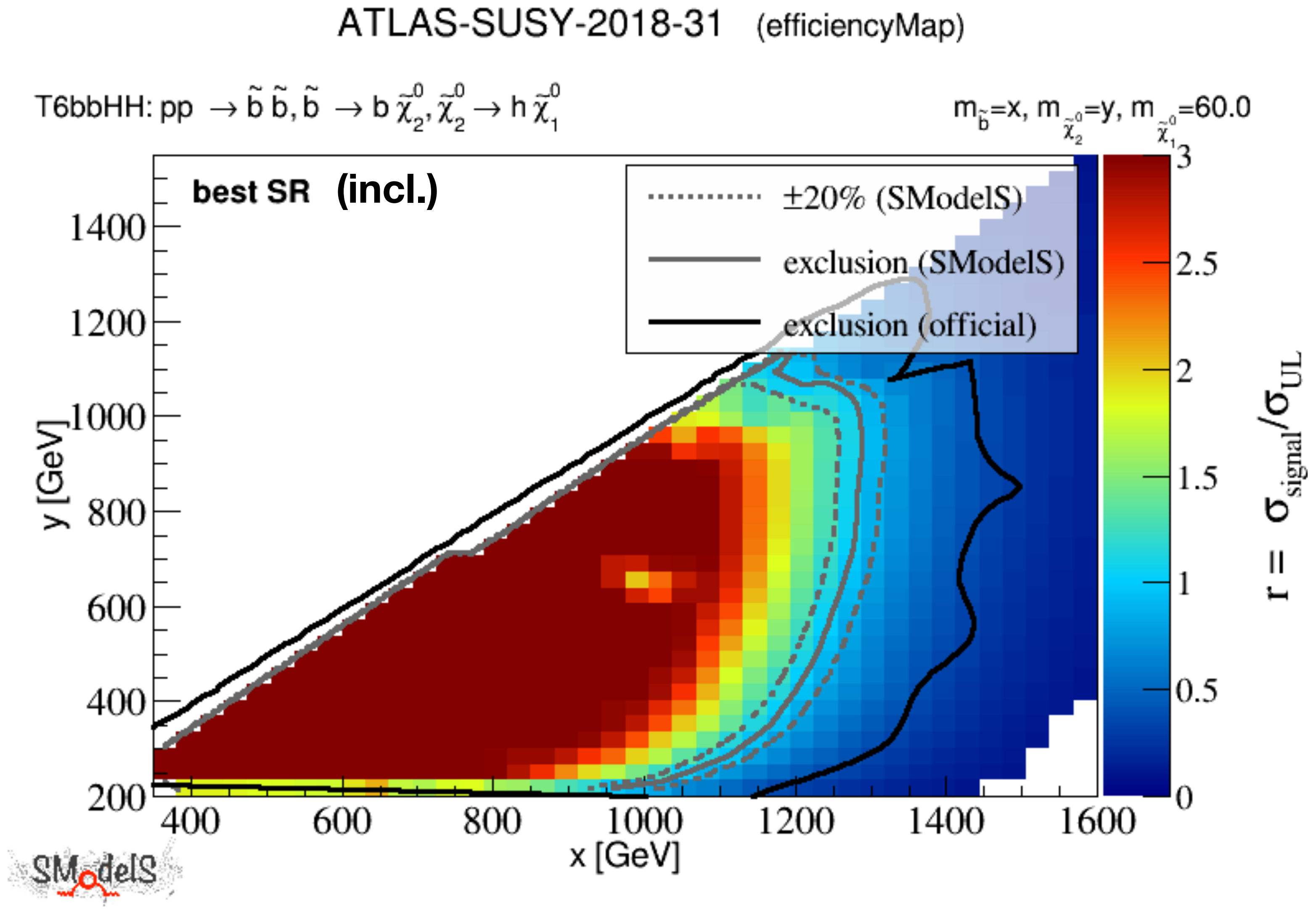
Validation & impact

Gaël Alguero, SK, Wolfgang Waltenberger,
arXiv:2009.01809

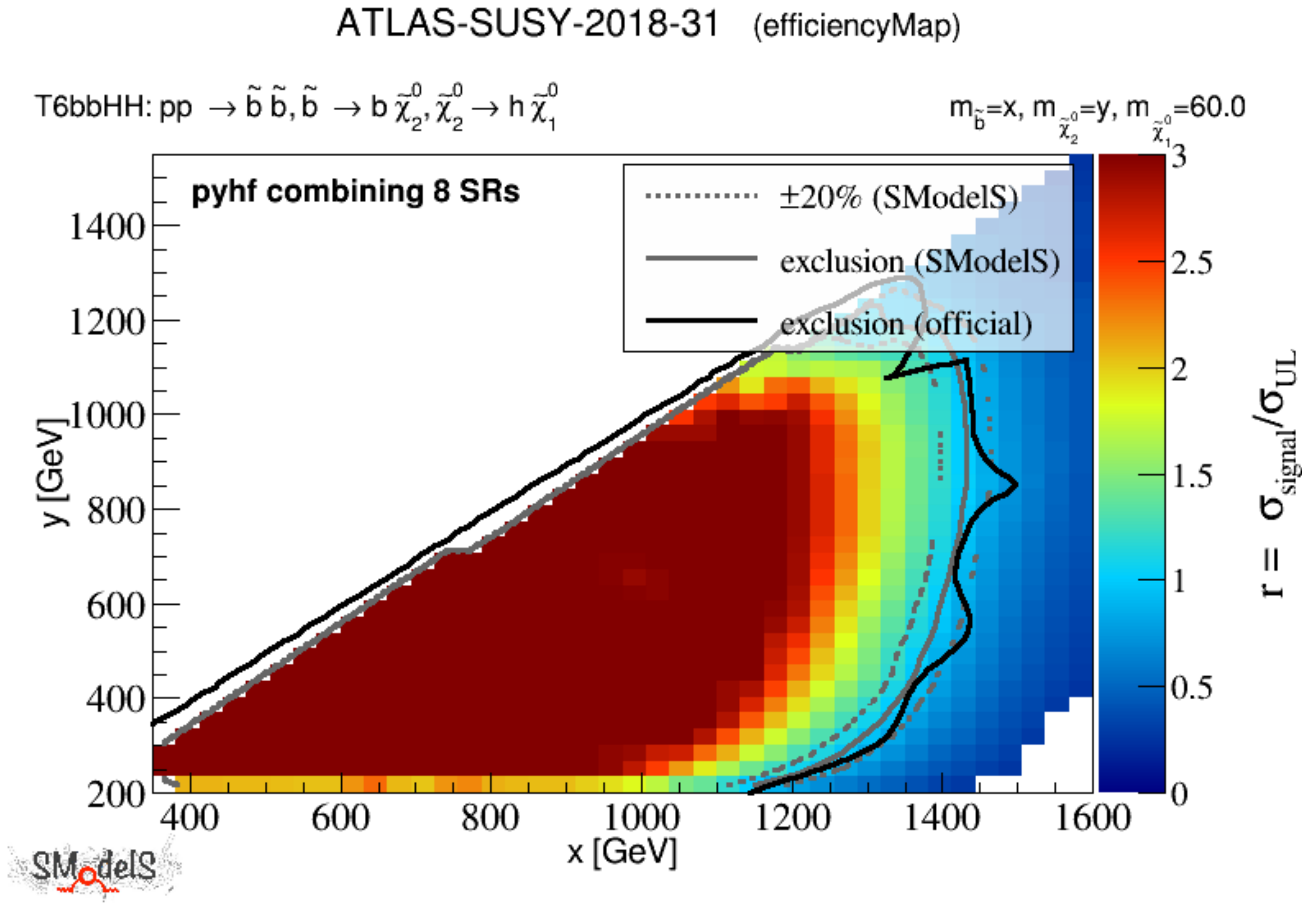
- ATLAS-SUSY-2018-31: T6bbHH, 8 excl. (3 SRA, 1 SRB, 4 SRC), 3 incl. SRs



Best exclusive SR: under exclusion



Best inclusive SR: more under exclusion

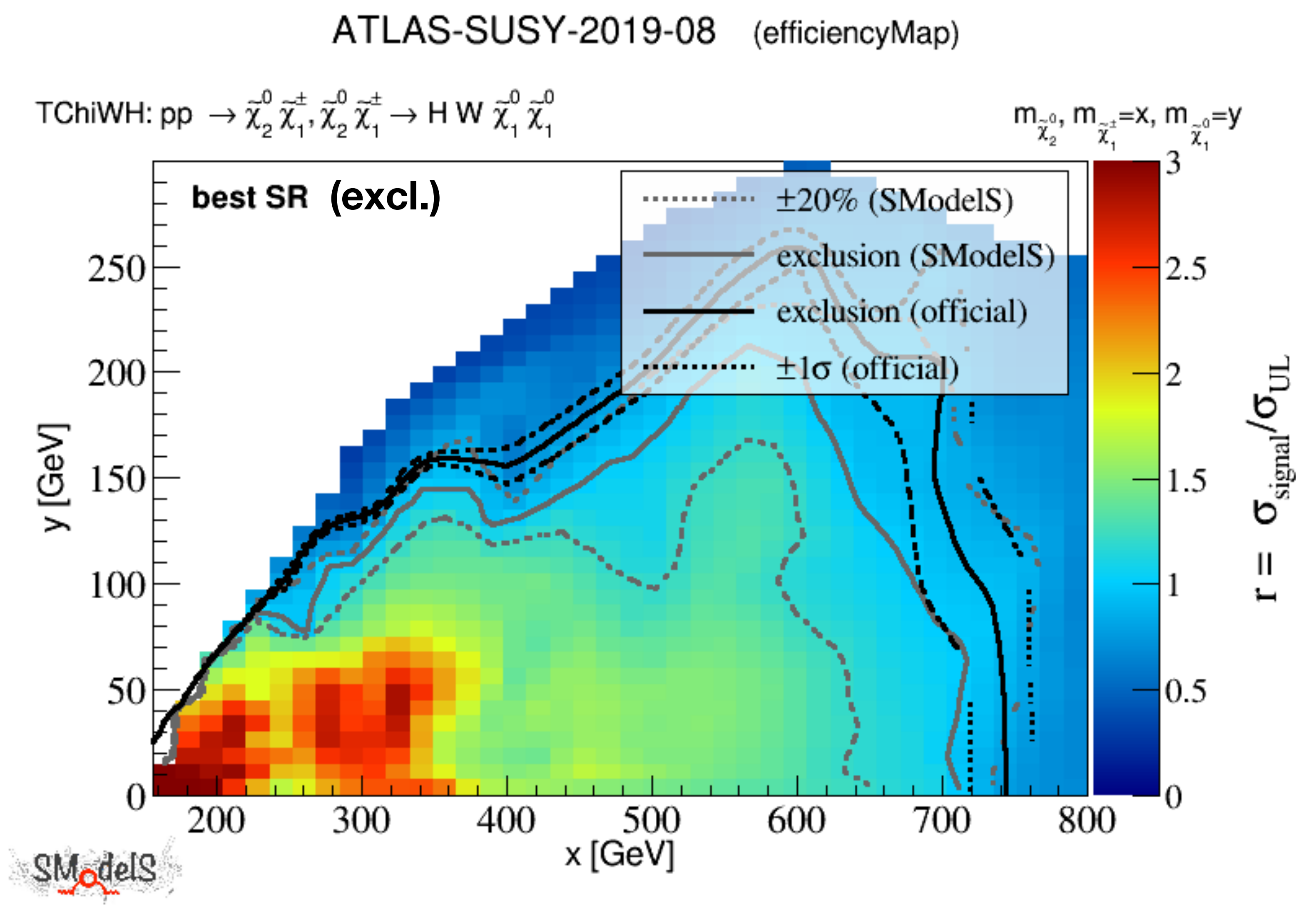


Full likelihood: very good agreement with official ATLAS result

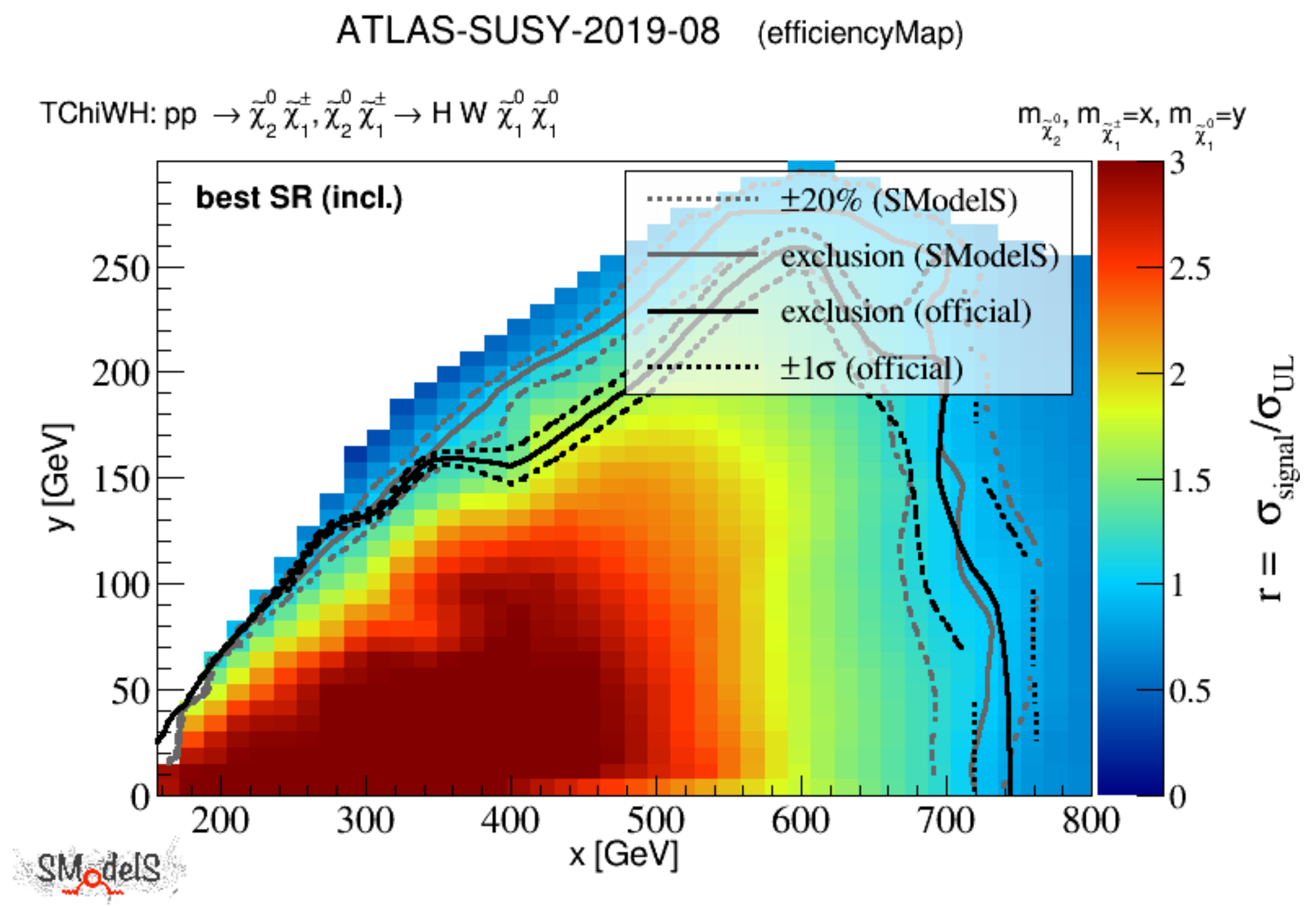
Validation & impact

Gaël Alguero, SK, Wolfgang Waltenberger,
arXiv:2009.01809

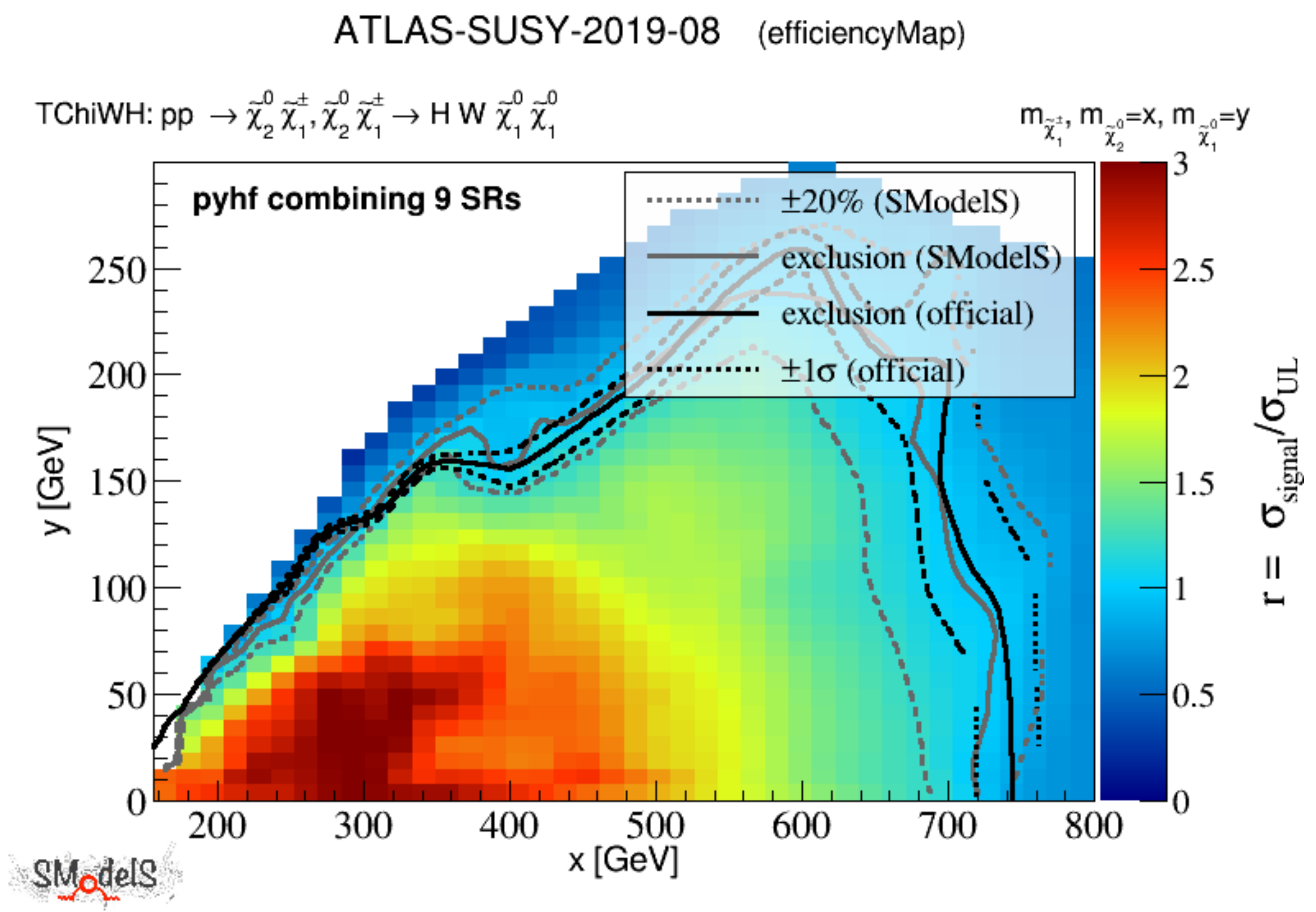
- ATLAS-SUSY-2019-08: TChiWH, 9 excl., 3 incl. SRs



Best exclusive SR: under exclusion

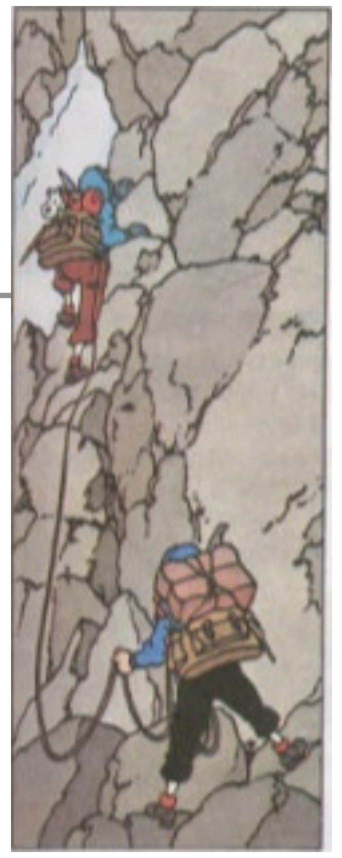


Best inclusive SR: over exclusion for large LSP masses



Full likelihood: excellent agreement with official ATLAS result

Comments



1. Sample names, order of signal regions

Example: SUSY-2019-08: 1lept. + H(bb), EWK

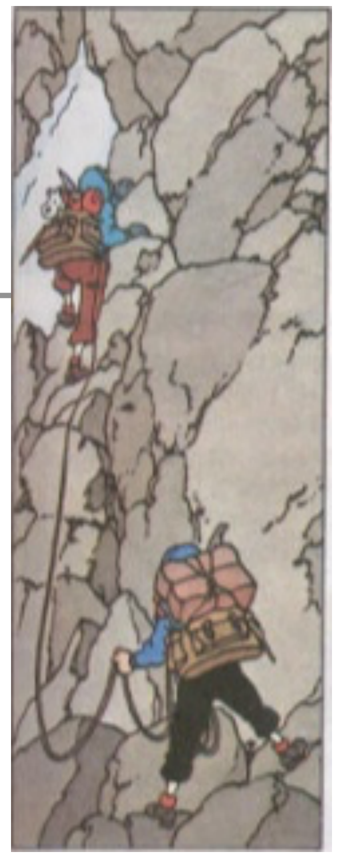
- ▼ ATLAS-SUSY-2019-08-eff
  - BkgOnly.json
  - globalInfo.txt ←
  - ▼ SR_HM_High_MCT
    - dataInfo.txt
    - TChiWH.txt
  - ▼ SR_HM_Low_MCT
    - dataInfo.txt
    - TChiWH.txt
  - ▶ SR_HM_Med_MCT
  - ▶ SR_LM_High_MCT
  - ▶ SR_LM_Low_MCT
  - ▶ SR_LM_Med_MCT
  - ▶ SR_MM_High_MCT
  - ▶ SR_MM_Low_MCT
  - ▶ SR_MM_Med_MCT

jsonFiles: {"BkgOnly.json" : ["SR_HM_Low_MCT", "SR_HM_Med_MCT", "SR_HM_High_MCT", "SR_MM_Low_MCT", "SR_MM_Med_MCT", "SR_MM_High_MCT", "SR_LM_Low_MCT", "SR_LM_Med_MCT", "SR_LM_High_MCT"]}


```
"observations": [
   {
      "data": [
         6.0,
         5.0,
         3.0
      ],
      "name": "SRHMEM_mct2"
   },
   {
      "data": [
         4.0,
         7.0,
         2.0
      ],
      "name": "SRMMEM_mct2"
   },
   {
      "data": [
         16.0,
         11.0,
         7.0
      ],
      "name": "SRLMEM_mct2"
   }
],
```


BkgOnly.json

Comments



1. Sample names, order of signal regions

Example: SUSY-2019-08: 1lept. + H(bb), EWK

SR-LM	All m_{CT} bins	Low m_{CT}	Medium m_{CT}	High m_{CT}
Observed	34	16	11	7
Expected	27 ± 4	8.8 ± 2.8	11.3 ± 3.1	7.3 ± 1.5
$t\bar{t}$	16.2 ± 3.4	4.4 ± 2.2	7.3 ± 2.5	4.6 ± 1.2
Single top	2.7 ± 1.8	1.3 ± 1.1	$0.9^{+1.0}_{-0.9}$	0.6 ± 0.6
W+jets	5.5 ± 2.0	2.0 ± 0.9	2.4 ± 1.3	1.1 ± 0.5
Di-/Multiboson	0.67 ± 0.19	0.39 ± 0.13	$0.09^{+0.11}_{-0.09}$	0.18 ± 0.04
Others	2.23 ± 0.29	0.81 ± 0.25	0.64 ± 0.15	0.77 ± 0.12

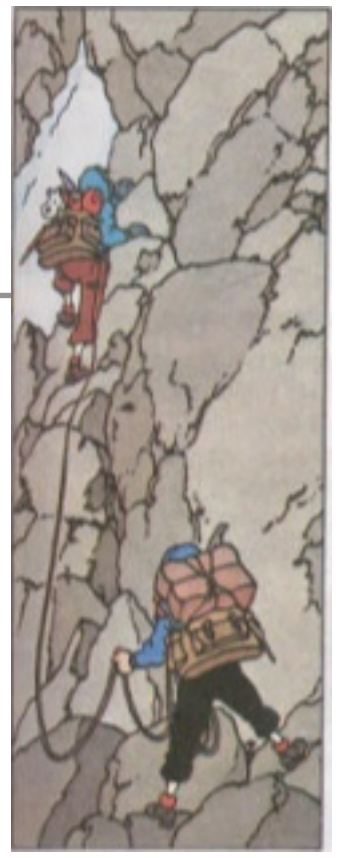
SR-MM	All m_{CT} bins	Low m_{CT}	Medium m_{CT}	High m_{CT}
Observed	13	4	7	2
Expected	8.6 ± 2.2	4.6 ± 1.7	2.6 ± 1.3	1.4 ± 0.6
$t\bar{t}$	2.7 ± 1.4	1.6 ± 0.9	0.8 ± 0.7	0.30 ± 0.24
Single top	2.7 ± 1.9	1.6 ± 1.5	$1.0^{+1.1}_{-1.0}$	$0.15^{+0.19}_{-0.15}$
W+jets	1.5 ± 0.7	0.6 ± 0.4	$0.3^{+0.4}_{-0.3}$	0.57 ± 0.26
Di-/Multiboson	0.29 ± 0.08	0.09 ± 0.04	0.065 ± 0.028	0.14 ± 0.06
Others	1.33 ± 0.27	0.69 ± 0.20	0.40 ± 0.13	0.24 ± 0.09

SR-HM	All m_{CT} bins	Low m_{CT}	Medium m_{CT}	High m_{CT}
Observed	14	6	5	3
Expected	8.1 ± 2.7	4.1 ± 1.9	2.9 ± 1.3	1.1 ± 0.5
$t\bar{t}$	1.4 ± 0.5	0.8 ± 0.4	0.36 ± 0.25	0.22 ± 0.15
Single top	$2.0^{+2.4}_{-2.0}$	$0.9^{+1.5}_{-0.9}$	0.9 ± 0.9	$0.16^{+0.26}_{-0.16}$
W+jets	3.7 ± 1.0	1.9 ± 0.8	1.4 ± 0.8	0.45 ± 0.19
Di-/Multiboson	0.21 ± 0.06	0.057 ± 0.025	0.075 ± 0.027	0.08 ± 0.04
Others	0.74 ± 0.16	0.34 ± 0.09	0.19 ± 0.08	0.21 ± 0.08

```
"observations": [
  {
    "data": [
      6.0,
      5.0,
      3.0
    ],
    "name": "SRHMEM_mct2"
  },
  {
    "data": [
      4.0,
      7.0,
      2.0
    ],
    "name": "SRMMEM_mct2"
  },
  {
    "data": [
      16.0,
      11.0,
      7.0
    ],
    "name": "SRLMEM_mct2"
  }
],
```

BkgOnly.json

Comments



1. Sample names, order of signal regions

Example: SUSY-2018-04: stau search

- ATLAS-SUSY-2018-04-eff
  - globalInfo.txt ← **jsonFiles:** {"SRcombined.json": ["SRlow", "SRhigh"]}
  - SRcombined.json
  - SRhigh
    - dataInfo.txt
    - TStauStau.txt
  - SRlow
    - dataInfo.txt
    - TStauStau.txt


```
"observations": [
....
 {
  "data": [
   10
  ],
  "name": "SR1cut_cuts"
 },
 {
  "data": [
   7
  ],
  "name": "SR2cut_cuts"
 },
 ....
],
```


SRcombined.json

SM process	Multi-jet CR-A -lowMass	Multi-jet CR-A -highMass	WCR	SR -lowMass	SR -highMass
Diboson	1.4 ± 0.6	1.9 ± 1.0	63 ± 21	1.4 ± 0.8	2.6 ± 1.4
W+jets	13 ± 4	4^{+7}_{-4}	850 ± 70	1.5 ± 0.7	2.5 ± 1.8
Top quark	2.7 ± 0.9	3.3 ± 1.6	170 ± 40	$0.04^{+0.80}_{-0.04}$	2.0 ± 0.6
Z+jets	$0.25^{+1.43}_{-0.25}$	1.5 ± 0.8	13 ± 7	$0.4^{+0.5}_{-0.4}$	$0.05^{+0.13}_{-0.05}$
Multi-jet	55 ± 10	16 ± 6	–	2.6 ± 0.7	3.1 ± 1.4
SM total	72 ± 8	27 ± 5	1099 ± 33	6.0 ± 1.7	10.2 ± 3.3
Observed	72	27	1099	10	7

To avoid confusion, please use same names (and same order of binning) in the paper and JSON files.

Comments

2. JSON files for “subsets” of signal regions

SUSY-2018-31: 2b + 2H(bb) + MET

	SRA	SRA-L	SRA-M	SRA-H	SRB
Observed events	17	12	3	2	3
Fitted SM bkg events	17.1 ± 2.8	8.4 ± 1.7	5.7 ± 0.8	3.0 ± 1.5	3.3 ± 0.9
$t\bar{t}$	10.1 ± 2.5	4.7 ± 1.5	3.7 ± 0.6	1.7 ± 1.4	2.3 ± 0.8
Z+jets	2.6 ± 0.4	1.3 ± 0.2	0.9 ± 0.2	0.4 ± 0.1	0.3 ± 0.1
Single-top	1.4 ± 0.3	0.4 ± 0.1	0.3 ± 0.1	0.6 ± 0.2	0.5 ± 0.1
$t\bar{t}+W/Z$	1.2 ± 0.3	0.7 ± 0.1	0.3 ± 0.1	0.1 ± 0.0	0.07 ± 0.02
$t\bar{t}+h$	1.1 ± 0.2	0.7 ± 0.1	0.3 ± 0.1	0.1 ± 0.0	0.13 ± 0.02
W+jets	0.4 ± 0.1	0.2 ± 0.1	0.1 ± 0.0	–	0.02 ± 0.01
Diboson	0.4 ± 0.1	0.3 ± 0.1	–	–	–

	SRC	SRC22	SRC24	SRC26	SRC28
Observed events	47	28	12	4	3
Fitted SM bkg events	37.9 ± 6.2	21.2 ± 4.1	10.6 ± 2.3	3.7 ± 0.9	2.4 ± 0.6
$t\bar{t}$	5.4 ± 2.6	3.9 ± 2.3	1.1 ± 0.6	0.3 ± 0.3	0.1 ± 0.1
Z+jets	17.6 ± 4.7	8.8 ± 2.5	6.0 ± 1.8	1.7 ± 0.7	1.1 ± 0.4
Single-top	5.0 ± 1.5	2.7 ± 1.0	1.2 ± 0.3	0.7 ± 0.2	0.4 ± 0.1
$t\bar{t}+W/Z$	4.3 ± 0.6	2.5 ± 0.4	1.1 ± 0.2	0.5 ± 0.1	0.2 ± 0.1
$t\bar{t}+h$	0.2 ± 0.0	0.2 ± 0.0	–	0.1 ± 0.0	0.0 ± 0.0
W+jets	3.5 ± 0.8	2.2 ± 0.5	0.6 ± 0.2	0.2 ± 0.1	0.4 ± 0.1
Diboson	1.8 ± 0.3	0.9 ± 0.2	0.6 ± 0.1	0.2 ± 0.0	0.1 ± 0.1

BkgOnlyA.json

```
"observations": [
  {
    "data": [
      12.0,
      3.0,
      2.0
    ],
    "name": "SR_meff"
  },
```

BkgOnlyB.json

```
"observations": [
  {
    "data": [
      3.0
    ],
    "name": "SR_cuts"
  }
],
```

BkgOnlyC.json

```
"observations": [
  {
    "data": [
      28.0,
      12.0,
      4.0,
      3.0
    ],
    "name": "SR_metsigST"
  },
```

Comments

2. JSON files for “subsets” of signal regions

SUSY-2018-31: 2b + 2H(bb) + MET

	SRA	SRA-L	SRA-M	SRA-H	SRB
Observed events	17	12	3	2	3
Fitted SM bkg events	17.1 ± 2.8	8.4 ± 1.7	5.7 ± 0.8	3.0 ± 1.5	3.3 ± 0.9
$t\bar{t}$	10.1 ± 2.5	4.7 ± 1.5	3.7 ± 0.6	1.7 ± 1.4	2.3 ± 0.8
Z+jets	2.6 ± 0.4	1.3 ± 0.2	0.9 ± 0.2	0.4 ± 0.1	0.3 ± 0.1
Single-top	1.4 ± 0.3	0.4 ± 0.1	0.3 ± 0.1	0.6 ± 0.2	0.5 ± 0.1
$t\bar{t}+W/Z$	1.2 ± 0.3	0.7 ± 0.1	0.3 ± 0.1	0.1 ± 0.0	0.07 ± 0.02
$t\bar{t}+h$	1.1 ± 0.2	0.7 ± 0.1	0.3 ± 0.1	0.1 ± 0.0	0.13 ± 0.02
W+jets	0.4 ± 0.1	0.2 ± 0.1	0.1 ± 0.0	–	0.02 ± 0.01
Diboson	0.4 ± 0.1	0.3 ± 0.1	–	–	–

	SRC	SRC22	SRC24	SRC26	SRC28
Observed events	47	28	12	4	3
Fitted SM bkg events	37.9 ± 6.2	21.2 ± 4.1	10.6 ± 2.3	3.7 ± 0.9	2.4 ± 0.6
$t\bar{t}$	5.4 ± 2.6	3.9 ± 2.3	1.1 ± 0.6	0.3 ± 0.3	0.1 ± 0.1
Z+jets	17.6 ± 4.7	8.8 ± 2.5	6.0 ± 1.8	1.7 ± 0.7	1.1 ± 0.4
Single-top	5.0 ± 1.5	2.7 ± 1.0	1.2 ± 0.3	0.7 ± 0.2	0.4 ± 0.1
$t\bar{t}+W/Z$	4.3 ± 0.6	2.5 ± 0.4	1.1 ± 0.2	0.5 ± 0.1	0.2 ± 0.1
$t\bar{t}+h$	0.2 ± 0.0	0.2 ± 0.0	–	0.1 ± 0.0	0.0 ± 0.0
W+jets	3.5 ± 0.8	2.2 ± 0.5	0.6 ± 0.2	0.2 ± 0.1	0.4 ± 0.1
Diboson	1.8 ± 0.3	0.9 ± 0.2	0.6 ± 0.1	0.2 ± 0.0	0.1 ± 0.1

BkgOnlyA.json

```
"observations": [
  {
    "data": [
      12.0,
      3.0,
      2.0
    ],
    "name": "SR_meff"
  },
```

BkgOnlyB.json

```
"observations": [
  {
    "data": [
      3.0
    ],
    "name": "SR_cuts"
  }
],
```

BkgOnlyC.json

```
"observations": [
  {
    "data": [
      28.0,
      12.0,
      4.0,
      3.0
    ],
    "name": "SR_metsigST"
  },
```

Comments

2. JSON files for “subsets” of signal regions

SUSY-2018-31: 2b + 2H(bb) + MET

	SRA	SRA-L	SRA-M	SRA-H	SRB
Observed events	17	12	3	2	3
Fitted SM bkg events	17.1 ± 2.8	8.4 ± 1.7	5.7 ± 0.8	3.0 ± 1.5	3.3 ± 0.9
$t\bar{t}$	10.1 ± 2.5	4.7 ± 1.5	3.7 ± 0.6	1.7 ± 1.4	2.3 ± 0.8
Z+jets	2.6 ± 0.4	1.3 ± 0.2	0.9 ± 0.2	0.4 ± 0.1	0.3 ± 0.1
Single-top	1.4 ± 0.3	0.4 ± 0.1	0.3 ± 0.1	0.6 ± 0.2	0.5 ± 0.1
$t\bar{t}+W/Z$	1.2 ± 0.3	0.7 ± 0.1	0.3 ± 0.1	0.1 ± 0.0	0.07 ± 0.02
$t\bar{t}+h$	1.1 ± 0.2	0.7 ± 0.1	0.3 ± 0.1	0.1 ± 0.0	0.13 ± 0.02
W+jets	0.4 ± 0.1	0.2 ± 0.1	0.1 ± 0.0	–	0.02 ± 0.01
Diboson	0.4 ± 0.1	0.3 ± 0.1	–	–	–

	SRC	SRC22	SRC24	SRC26	SRC28
Observed events	47	28	12	4	3
Fitted SM bkg events	37.9 ± 6.2	21.2 ± 4.1	10.6 ± 2.3	3.7 ± 0.9	2.4 ± 0.6
$t\bar{t}$	5.4 ± 2.6	3.9 ± 2.3	1.1 ± 0.6	0.3 ± 0.3	0.1 ± 0.1

in `globalInfo.txt` file:

jsonFiles: {'BkgOnlyA.json': ['SRA_L', 'SRA_M', 'SRA_H'], 'BkgOnlyB.json': ['SRB'], ‘BkgOnlyC.json': ['SRC_22', 'SRC_24', 'SRC_26', 'SRC_28']}

➥ Determine most sensitive set (best expected limit) and use only that one.

BkgOnlyA.json


```
"observations": [
  {
     "data": [
       12.0,
       3.0,
       2.0
     ],
     "name": "SR_meff"
  },
```


BkgOnlyB.json


```
"observations": [
  {
     "data": [
       3.0
     ],
     "name": "SR_cuts"
  }
],
```


BkgOnlyC.json


```
"observations": [
  {
     "data": [
       28.0,
       12.0,
       4.0,
       3.0
     ],
     "name": "SR_metsigST"
  },
```


Comments

3. Multiple fixed parameters

ATLAS-SUSY-2018-06 (3 lept., EWK):

Using TChiWZ efficiency maps with BkgOnly.json file, we always get observed ≃ expected CLs

README.md on HEPData

> # Known Issues
>
> These workspaces are the first time that multiple fixed parameters exist in the measurement definition. If using pyhf, take note of this issue [scikit-hep/pyhf#739] which will be resolved shortly after these likelihoods are public.

Nominally resolved in pyhf 0.5.2 (with Minuit as optimizer), but not working within the SModelS interface.

☞ https://github.com/scikit-hep/pyhf/issues/1032

Comments

4. Speed

- SModelS is supposed to be fast: O(1s) / model point
- Likelihood calculation with pyhf significantly increases runtime

Analysis | CPU time	Best SR (simplified L)	combination, pyhf
SUSY-2019-08, TChiWH	0.94 s	1 min 17.5 s
SUSY-2018-31, T6bbHH	1.08 s	1 min 07.7 s

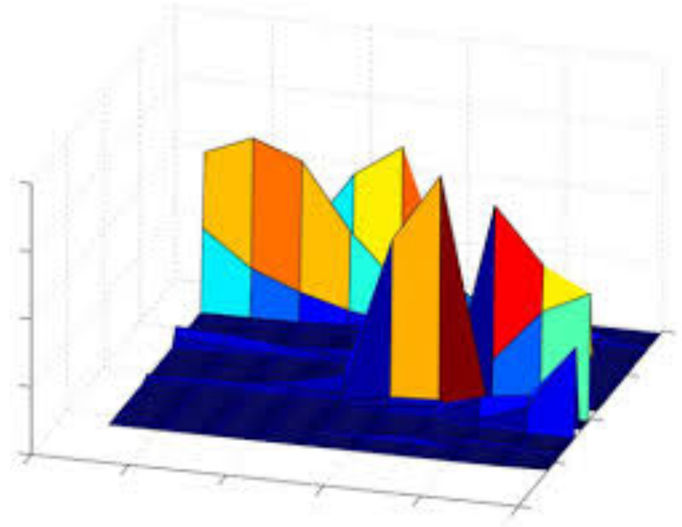
- It would be great to have a "fast" mode for pyhf,

  perhaps via a json patch where BG sources are already combined / nuisances integrated over (somewhat similar to a simplified likelihood but still accounting for asymmetries / non-Gaussian effects) ☞ https://github.com/scikit-hep/pyhf/issues/1056

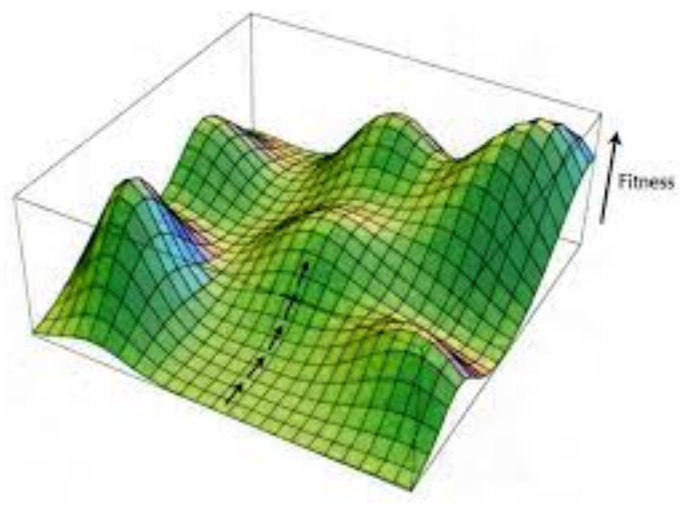
NB in addition to, not instead of, the full likelihood; e.g. to be used for large scans when a small loss in accuracy is acceptable.

Going further

Besides allowing us to better reproduce the official limits of each analysis, the full likelihoods



vs.



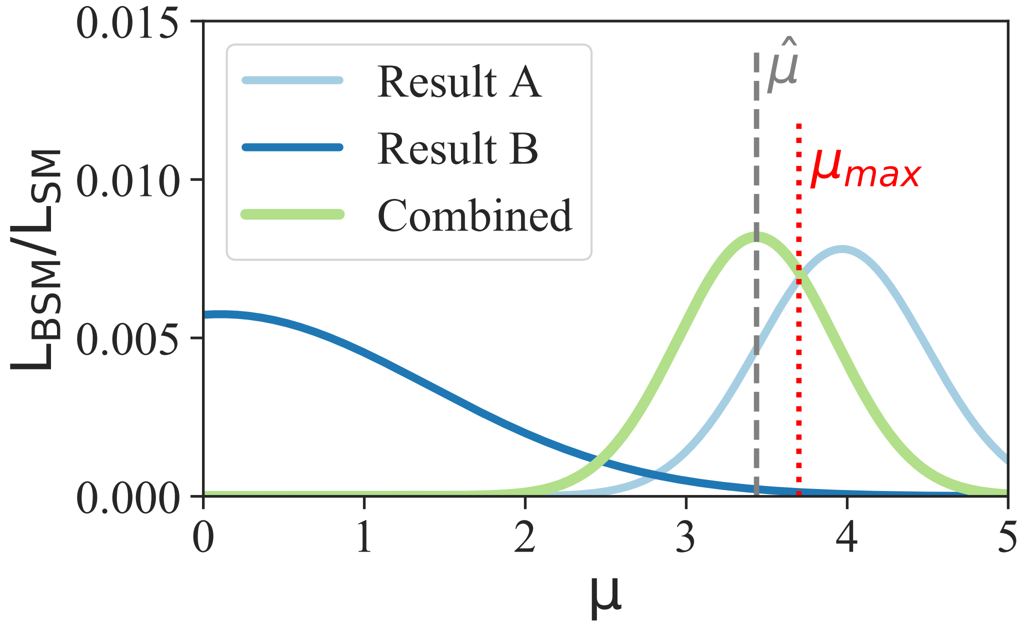
- will greatly improve global fits
- offer interesting possibilities to explore cross-analysis correlations
    - Systematic naming of nuisances?
- Both is also very useful for projects like the Protomodel Builder (cf talk by W. Waltenberger on June 4)
- Differentiability will allow for gradient-based methods in the future
- Lots to do on the pheno side, we are not yet using the full potential of full likelihoods.

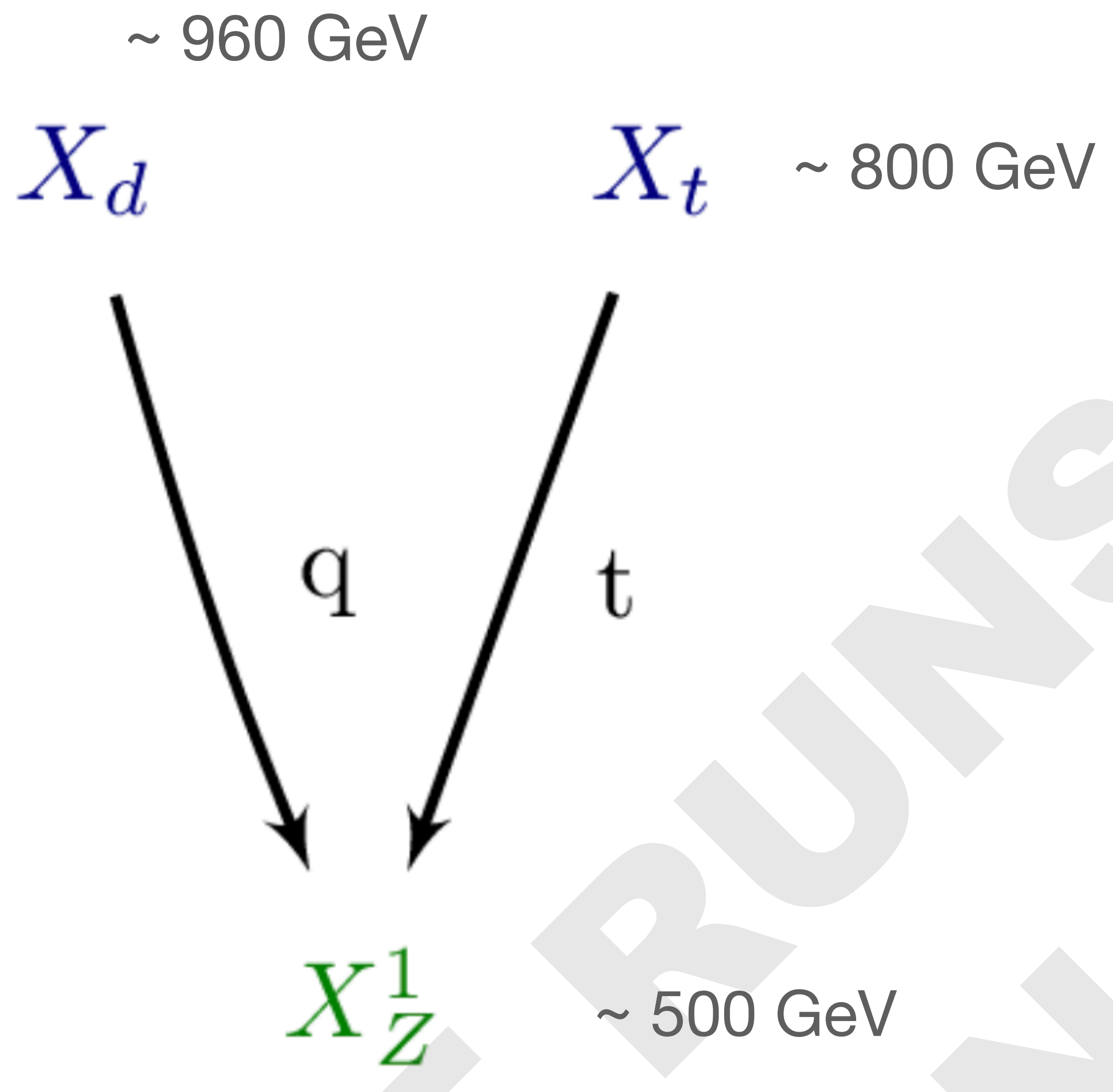
W. Waltenberger, A. Lessa, SK, in progress

Protomodel Builder

- Statistical learning algorithm to
  - identify potential dispersed signals in LHC data
  - fit "protomodels" (new particles, decay modes, signal strengths) to them
- Based on SModelS functionality and database, but could be extended much further
- Reliable likelihoods are essential for this purpose

Builder Combiner Critic





Analysis Name	Type	Dataset	Observed	Expected	Approx σ	Particles
ATLAS-SUSY-2013-02	em	SR6jtp	6	4.9 +/- 1.6	0.4 σ	X_d
ATLAS-SUSY-2013-15	em	tNboost	5	3.3 +/- 0.7	0.9 σ	X_t
ATLAS-SUSY-2016-07	em	2j_Meff_1200	611	526 +/- 31	2.2 σ	X_d
ATLAS-SUSY-2016-16	em	tN_med	50	36.3 +/- 6.6	1.5 σ	X_t
CMS-SUS-13-012	ul	-	29.4 fb	18.3 fb	1.2 σ	X_d
CMS-SUS-16-050	ul	-	106.0 fb	49.3 fb	2.3 σ	X_t

Conclusions

THANKS

for this great step forward to publish full likelihoods !

- extremely useful for long-term preservation
- opens the door for much better reinterpretation studies

I'm sure lots of interesting work is to follow

Please keep doing this systematically for all new analyses

(and please also keep providing ample simplified model A×ε maps 👍)