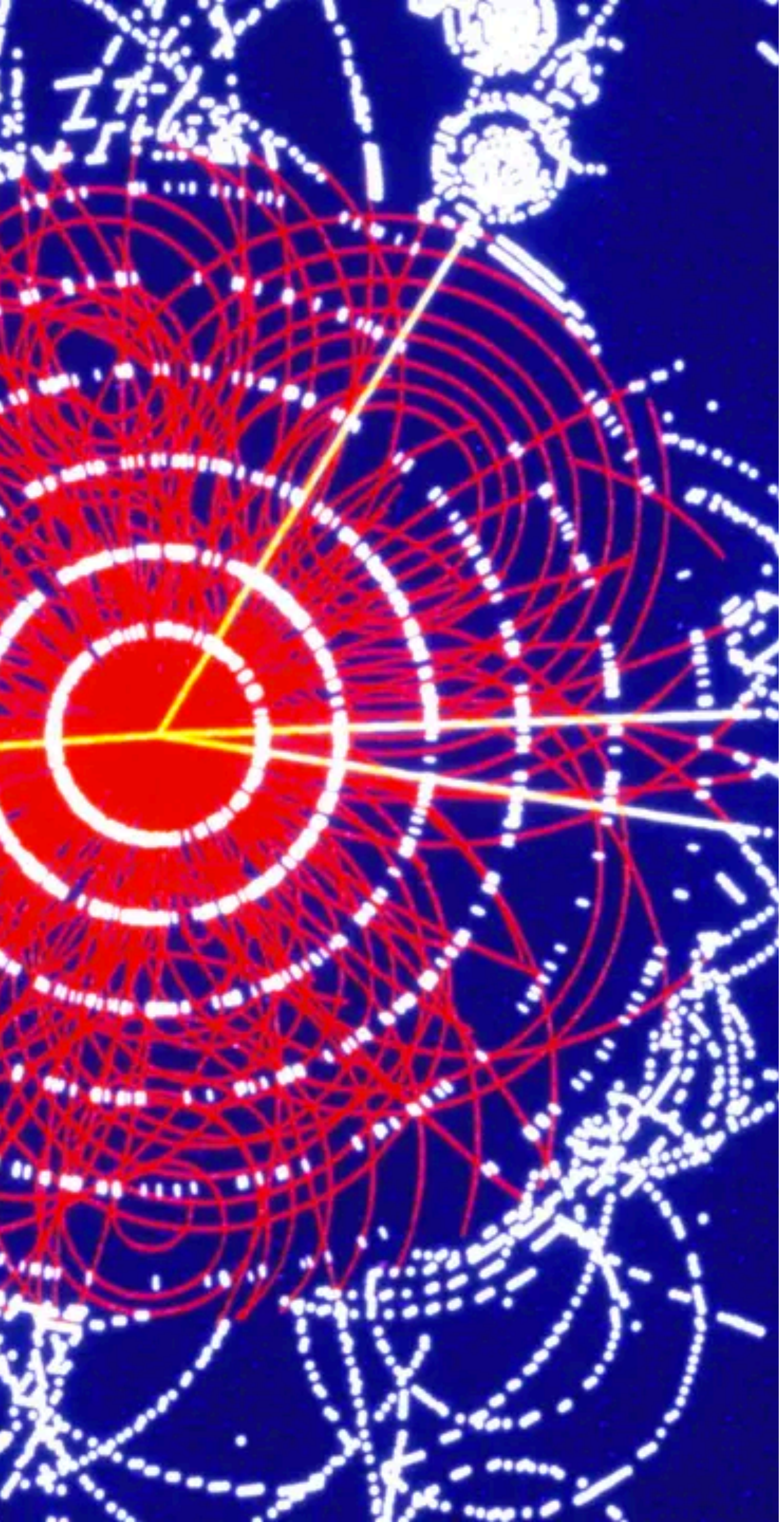


Machine-learning the likelihoods

Rafał Masełek

in collaboration with:

S. Kraml, W. Waltenberger, J. Araz, H. Reyes, A. Butter, C. Krause, J. Iturriza

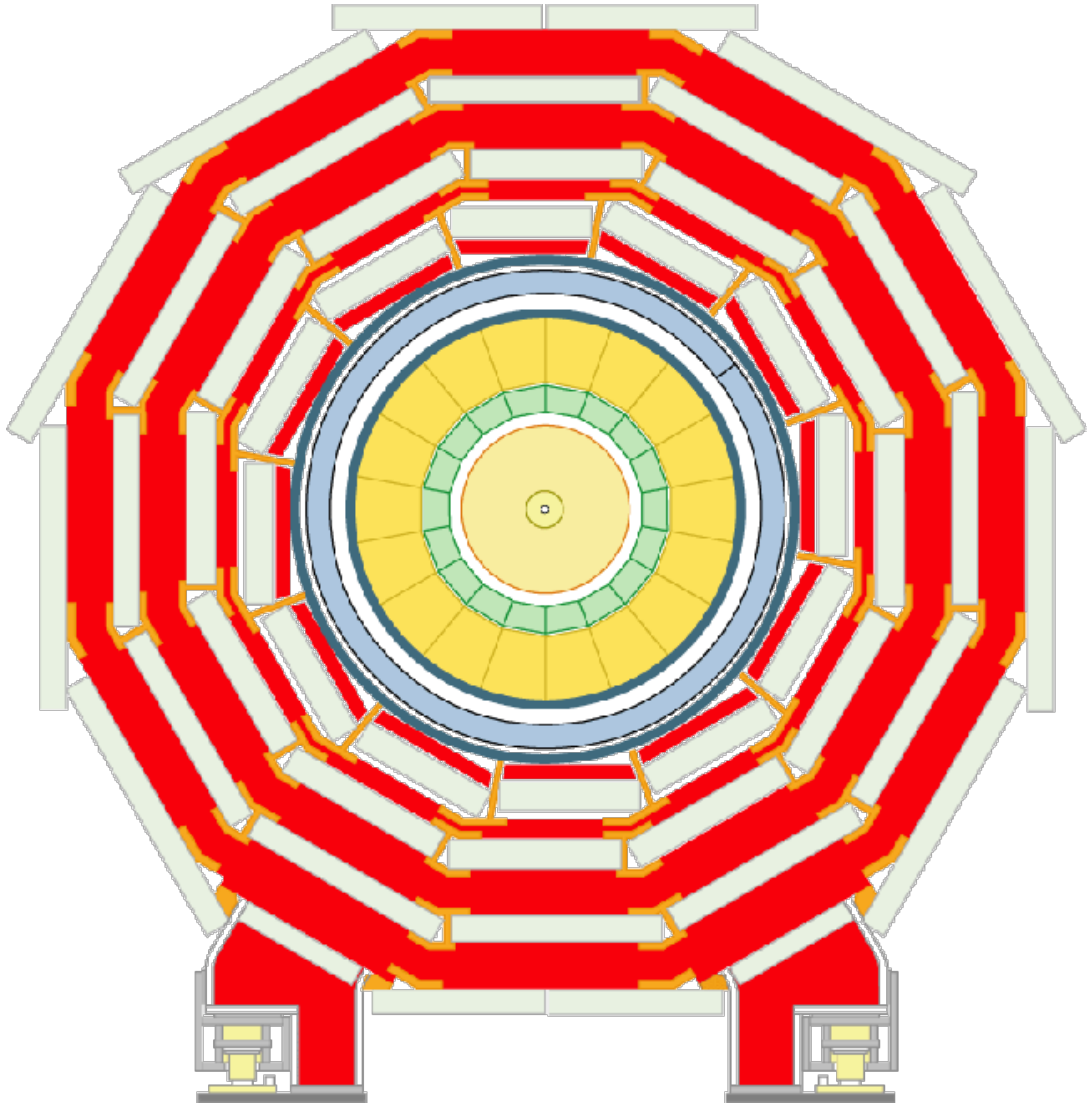
SmodelS meeting

Research supported by IN2P3 theory master project DataMATTER and Polish National Science Centre grant 2021/41/N/ST2/00972.

Reinterpretation

NEUTRINO MODELS
COMPOSITENESS
LR-SYMMETRY
LEPTOQUARK SUPERSYMMETRY
TWO HIGGS DOUBLETS
AXIONS WIMPs
EXTRA DIMENSIONS
MILICHARGED PARTICLES
DARK SECTOR

?



IMG: https://cds.cern.ch/record/1433717

Reinterpretation

Goal: Enhance and unify the statistical analysis step

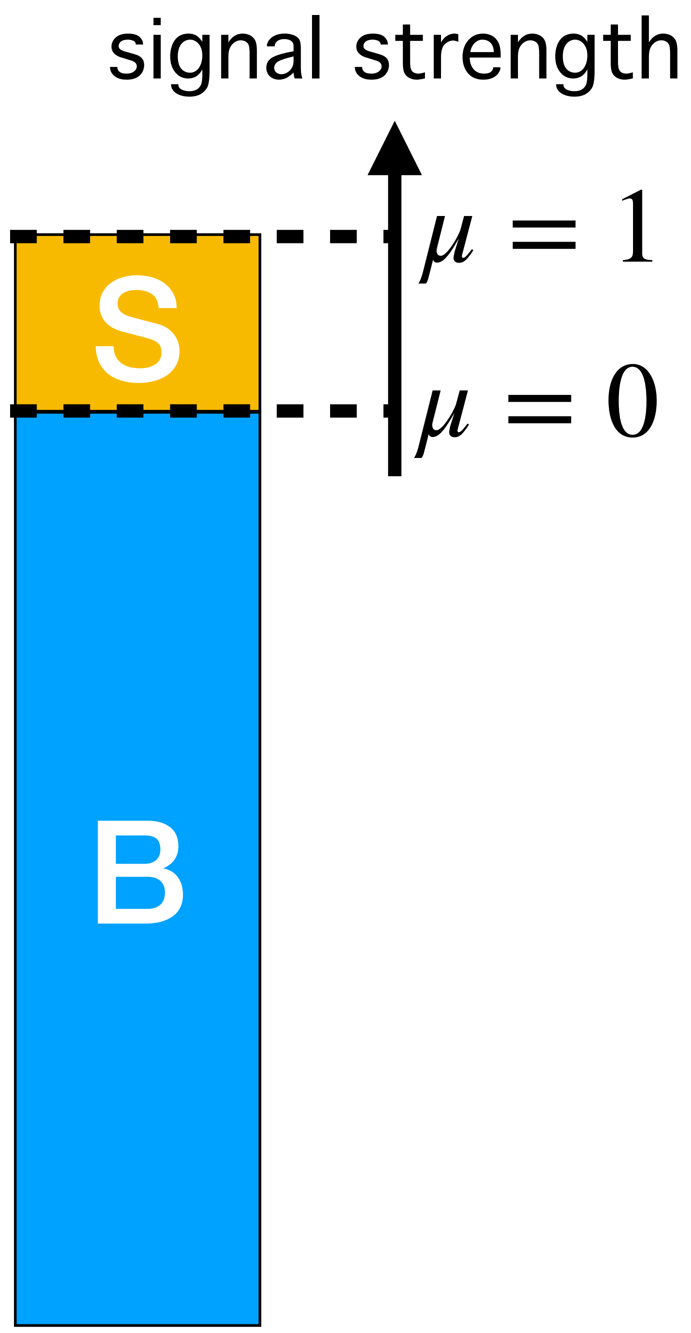
INO MODELS
COM
LR-SYMME
LEPTOQUARK SUPERS
TWO HIGGS DOUBLETS
AXIONS WIMPs
EXTRA DIMENSIONS
MILICHARGED PARTICLES
DARK SECTOR

IMG: https://cds.cern.ch/record/1433717

Likelihood template — simple

Let’s consider a simple experiment. We have a single channel with multiple bins, one signal and background contribution, and no systematics based on the discriminating variable x.

What is the probability model for obtaining n events in data where the discriminating variable for event e has value x_e?

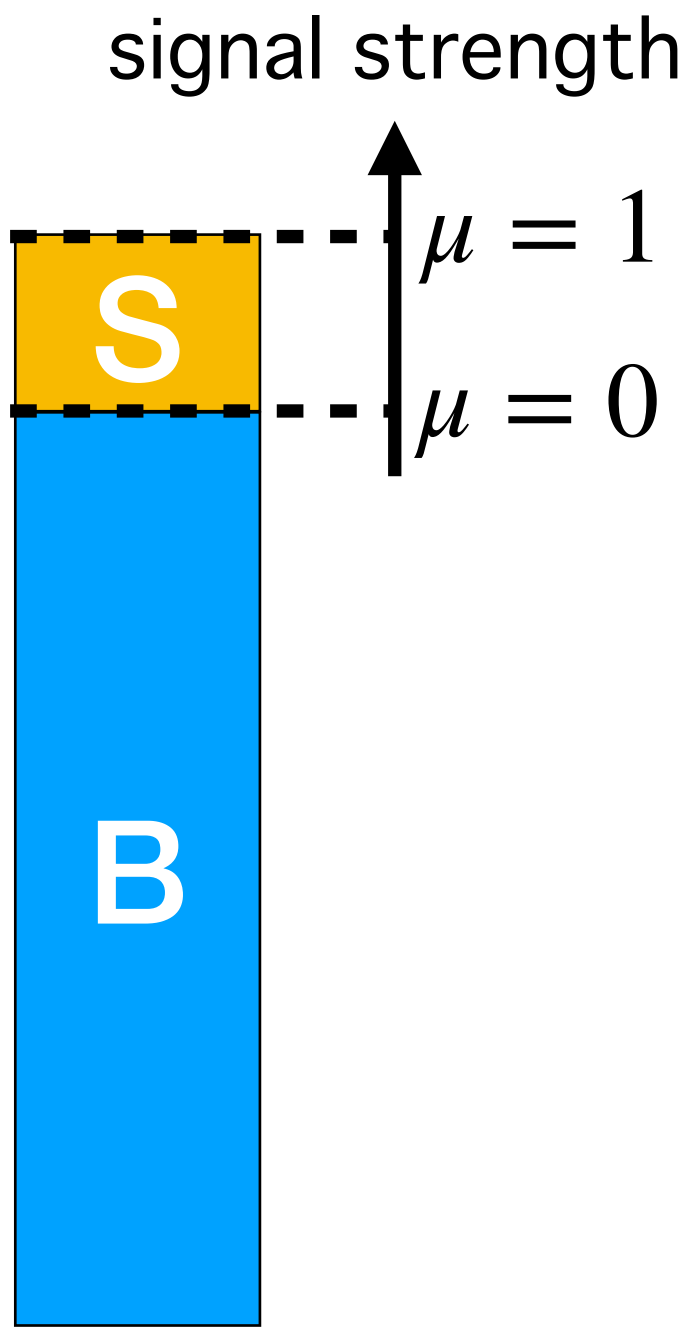


$$L(\mu) = p\left(\{x_1, \ldots x_n\} \;\; \mu\right) = \textbf{?}$$

Likelihood template — simple

Let's consider a simple experiment. We have a single channel with multiple bins, one signal and background contribution, and no systematics based on the discriminating variable x.

What is the probability model for obtaining n events in data where the discriminating variable for event e has value x_e?



$$L(\mu) = p\left(\{x_1, \ldots x_n\}\ \ \mu\right) = \mathrm{Pois}\left(n\ \ \mu S + B\right)\left[\prod_{e=1}^{n} \frac{\mu S \cdot f_{\mathrm{S}}(x_e) + B \cdot f_{\mathrm{B}}(x_e)}{\mu S + B}\right]$$

probability to observe n events given μS+B expectation

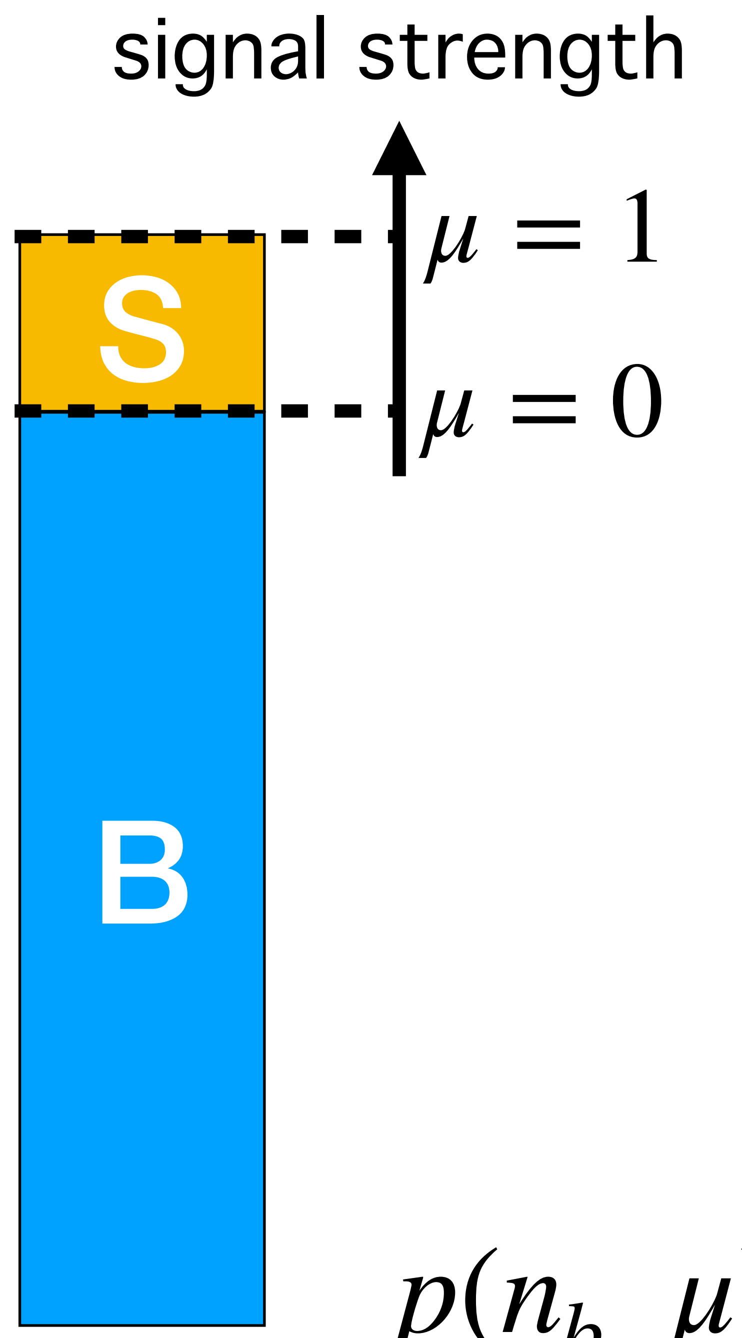
probability density of obtaining x_e based on the relative mixture of $f_{\mathrm{S}}(x)$ and $f_{\mathrm{B}}(x)$

[DOI:10.17181/CERN-OPEN-2012-016]

Likelihood template — simple

Let's consider a simple experiment. We have a single channel with multiple bins, one signal and background contribution, and no systematics based on the discriminating variable x.

What is the probability model for obtaining n events in data where the discriminating variable for event e has value x_e?



$$L(\mu) = p\left(\{x_1, \ldots x_n\} \;\; \mu\right) = \mathrm{Pois}\left(n \;\; \mu S + B\right) \left[\prod_{e=1}^{n} \frac{\mu S \cdot f_\mathrm{S}(x_e) + B \cdot f_\mathrm{B}(x_e)}{\mu S + B}\right]$$

In the binned case:

$$f_\mathrm{S}(x_e) = \frac{\nu_{b_e}^{\mathrm{sig}}}{S\Delta_{b_e}} \qquad S = \sum_b \nu_b^{\mathrm{sig}}$$

$$f_\mathrm{B}(x_e) = \frac{\nu_{b_e}^{\mathrm{bkg}}}{B\Delta_{b_e}} \qquad B = \sum_b \nu_b^{\mathrm{bkg}}$$

nominal yields

$$p(n_b \;\; \mu) = \mathrm{Pois}\left(n \;\; \mu S + B\right) \left[\prod_{b \in \mathrm{bins}} \frac{\mu \nu_b^{\mathrm{sig}} + \nu_b^{\mathrm{bkg}}}{\mu S + B}\right] = \mathcal{N}_{\mathrm{comb}} \prod_{b \in \mathrm{bins}} \mathrm{Pois}\left(n_b \;\; \mu \nu_b^{\mathrm{sig}} + \nu_b^{\mathrm{bkg}}\right)$$

counts per bin

[DOI:10.17181/CERN-OPEN-2012-016]

Likelihood template — HistFactory statistical models

We want to generalise our model to:

- combine multiple channels and correlate the parameters across the various channels
- include unconstrained scaling of the normalization of any sample
- parametrize variation in the normalization of any sample due to some systematic effect
- parameterize variations in the shape of any sample due to some systematic effect
- include bin-by-bin statistical uncertainty on the normalization of any sample
- incorporate an arbitrary contribution where each bin’s content is parametrized individually
- use the combination infrastructure to incorporate control samples for datadriven background estimation techniques
- reparametrize the model

$$L\left(n, a \;\; \mu, \theta\right) = \prod_{c}^{\text{channels}} \prod_{b}^{\text{bins}_c} \text{Pois}\left(n_{cb} \;\; \nu_{cb}(\mu, \theta)\right) \prod_{\theta} c_{\theta}\left(a_{\theta} \;\; \theta\right)$$

Likelihood template — HistFactory statistical models

We want to generalise our model to:

- combine multiple channels and correlate the parameters across the various channels
- include unconstrained scaling of the normalization of any sample
- parametrize variation in the normalization of any sample due to some systematic effect
- parameterize variations in the shape of any sample due to some systematic effect
- include bin-by-bin statistical uncertainty on the normalization of any sample
- incorporate an arbitrary contribution where each bin's content is parametrized individually
- use the combination infrastructure to incorporate control samples for datadriven background estimation techniques
- reparametrize the model

channel data

auxiliary data

$$L(n, a \mid \mu, \theta) = \underbrace{\prod_{c}^{\text{channels}} \prod_{b}^{\text{bins}_c} \text{Pois}\left(\boxed{n_{cb}} \mid \nu_{cb}(\boxed{\mu}, \boxed{\theta})\right)}_{\text{simultaneous measurement of multiple channels}} \underbrace{\prod_{\theta} c_\theta\left(\boxed{a_\theta} \mid \boxed{\theta}\right)}_{\text{constraint terms for "auxiliary measurements"}}$$

free parameters

constrained parameters

simultaneous measurement of multiple channels

constraint terms for "auxiliary measurements"

[arXiv:2211.15838]

Likelihood template — HistFactory statistical models

channel data

auxiliary data

$$L\left(n, a \mid \mu, \theta\right) = \underbrace{\prod_{c}^{\text{channels}} \prod_{b}^{\text{bins}_c} \text{Pois}\left(\boxed{n_{cb}} \mid \nu_{cb}\left(\boxed{\mu}, \boxed{\theta}\right)\right)}_{\text{simultaneous measurement of multiple channels}} \underbrace{\prod_{\theta} c_{\theta}\left(\boxed{a_{\theta}} \mid \boxed{\theta}\right)}_{\text{constraint terms for "auxiliary measurements"}}$$

free parameters

constrained parameters

simultaneous measurement of multiple channels

constraint terms for "auxiliary measurements"

$$\nu_{cb}\left(\mu, \theta\right) = \sum_{s}^{\text{samples}} \nu_{scb}\left(\mu, \theta\right) = \sum_{s}^{\text{samples}} \underbrace{\left(\prod_{\kappa} \kappa_{scb}\left(\mu, \theta\right)\right)}_{\text{multiplicative modifiers}} \left(\underbrace{\nu_{scb}^{0}\left(\mu, \theta\right)}_{\text{const. nominal rate}} + \underbrace{\sum_{\Delta} \Delta_{scb}\left(\mu, \theta\right)}_{\text{additive modifiers}}\right)$$

multiplicative modifiers **const. nominal rate** **additive modifiers**

[arXiv:2211.15838]

Likelihood template — implementation

- Full statistical models by ATLAS are available on HEPData
- They are provided as JSON files
- There are background files and signal patches
- Each patch corresponds to some signal point and contains modifiers to the background files
- There can be hundreds of modifiers
- Spey/PyHF can load and process these files


```json
"patch": [
    {
        "op": "add",
        "path": "/channels/0/samples/0",
        "value": {
            "data": [
                2.3051342964172363
            ],
            "modifiers": [
                {
                    "data": null,
                    "name": "lumi",
                    "type": "lumi"
                },
                {
                    "data": [
                        0.6571804118166927
                    ],
                    "name": "staterror_QCR1cut_cuts",
                    "type": "staterror"
                },
                {
                    "data": {
                        "hi": 1.06675,
                        "lo": 0.911403
                    },
                    "name": "PRW_DATASF",
                    "type": "normsys"
                },
```

Likelihood ratio test statistic

In the absence of the niussance parameters, the optimal test statistic (according to Neyman-Pearson lemma) is q:

$$q = -2\ln\frac{L(\mu=1)}{L(\mu=0)}$$

In the more general case, for upper limits we use:

$$\tilde{q}_\mu = \begin{cases} 0, & \mu < \hat{\mu} \\ -2\ln\dfrac{L\left(\mu, \hat{\hat{\theta}}(\mu)\right)}{L\left(\hat{\mu}, \hat{\theta}\right)}, & 0 \le \hat{\mu} \le \mu, \\ -2\ln\dfrac{L\left(\mu, \hat{\hat{\theta}}(\mu)\right)}{L\left(0, \hat{\hat{\theta}}(0)\right)}, & \hat{\mu} < 0, \end{cases}$$

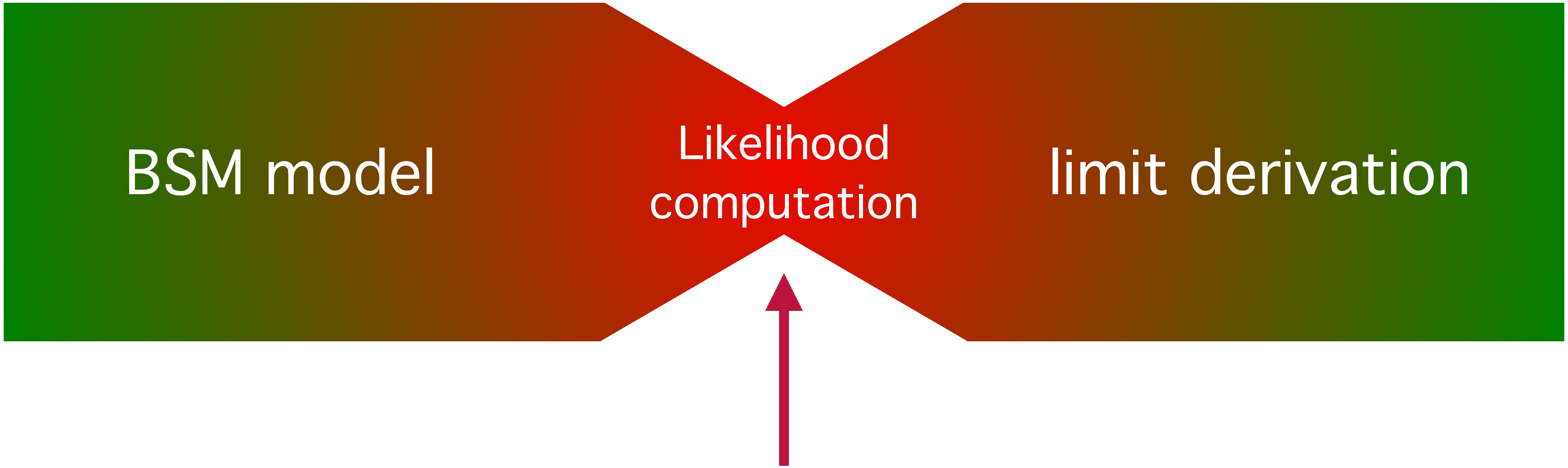
$\hat{\mu}, \hat{\theta}$ — unconditional ML estimators
$\hat{\hat{\theta}}(\mu)$ — ML estimator conditioned on μ.

$$p_{\mu,\text{obs}} = \int_{\tilde{q}_{\mu,\text{obs}}}^{\infty} f\left(\tilde{q}_\mu \ \mu'\right) d\tilde{q}_\mu$$

f — PDF of $\tilde{q}_\mu$

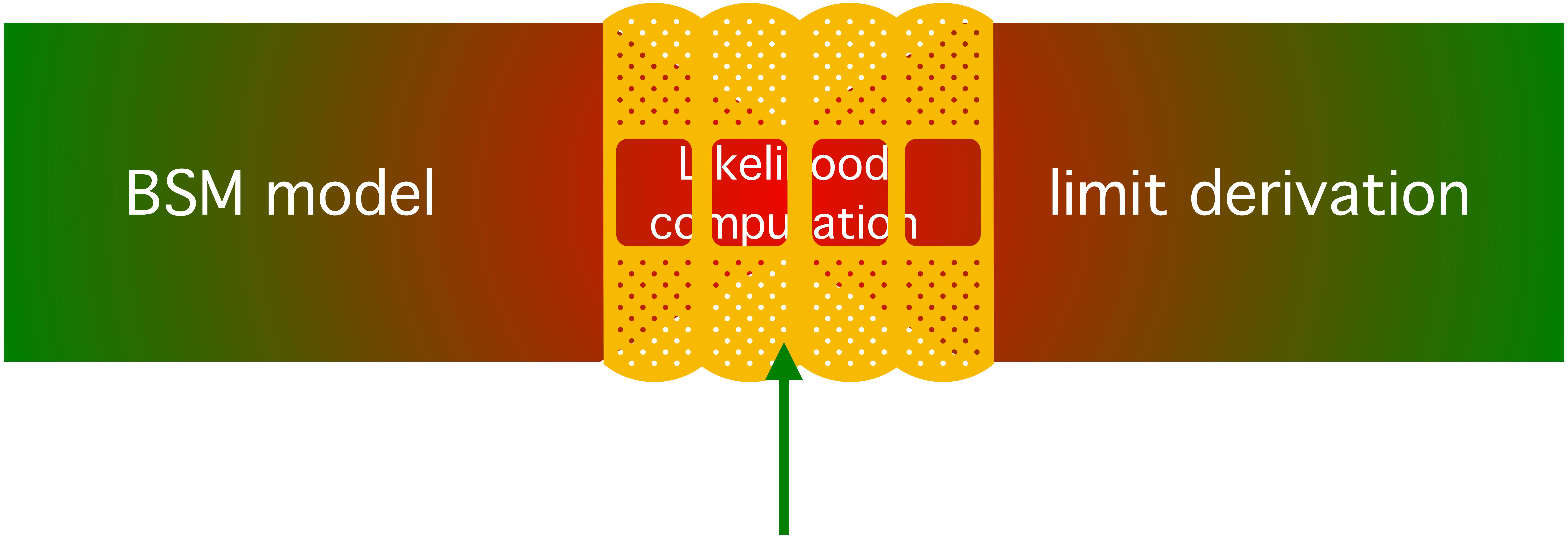
[[arXiv:1007.1727]]

Computational bottleneck



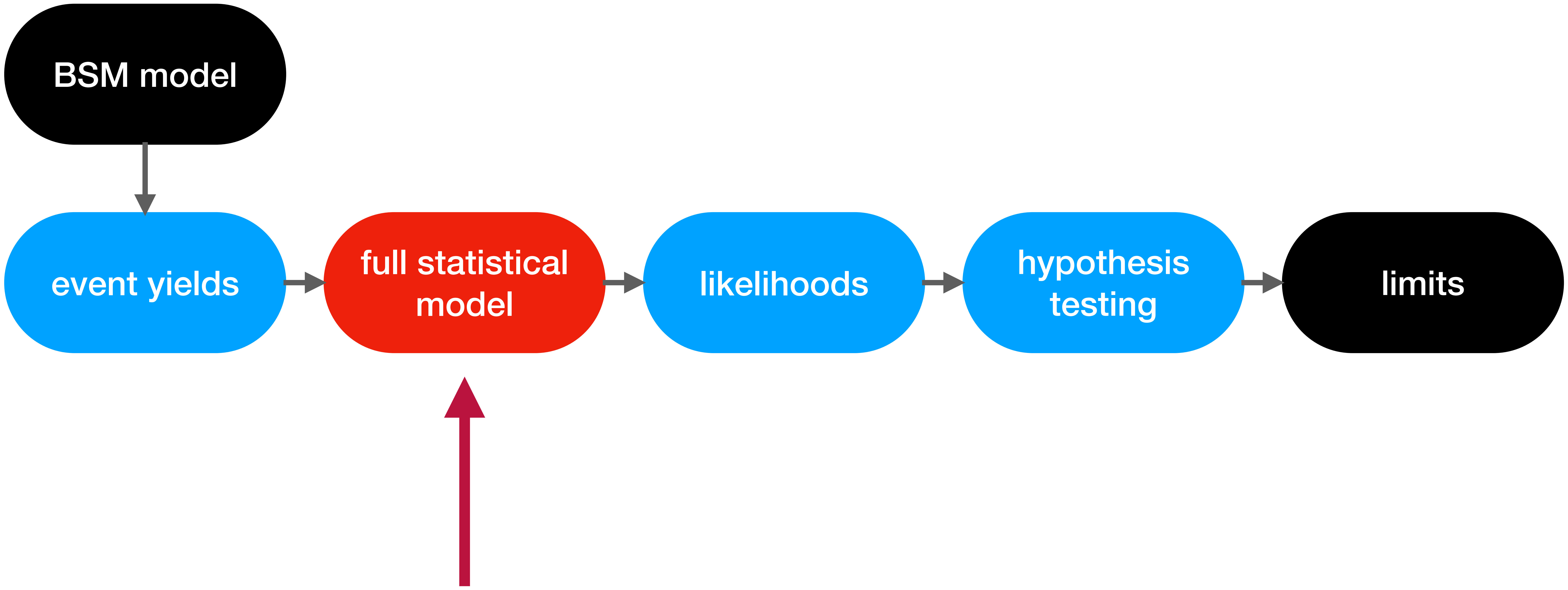
Full statistical model calculations enter here

Fixing the problem



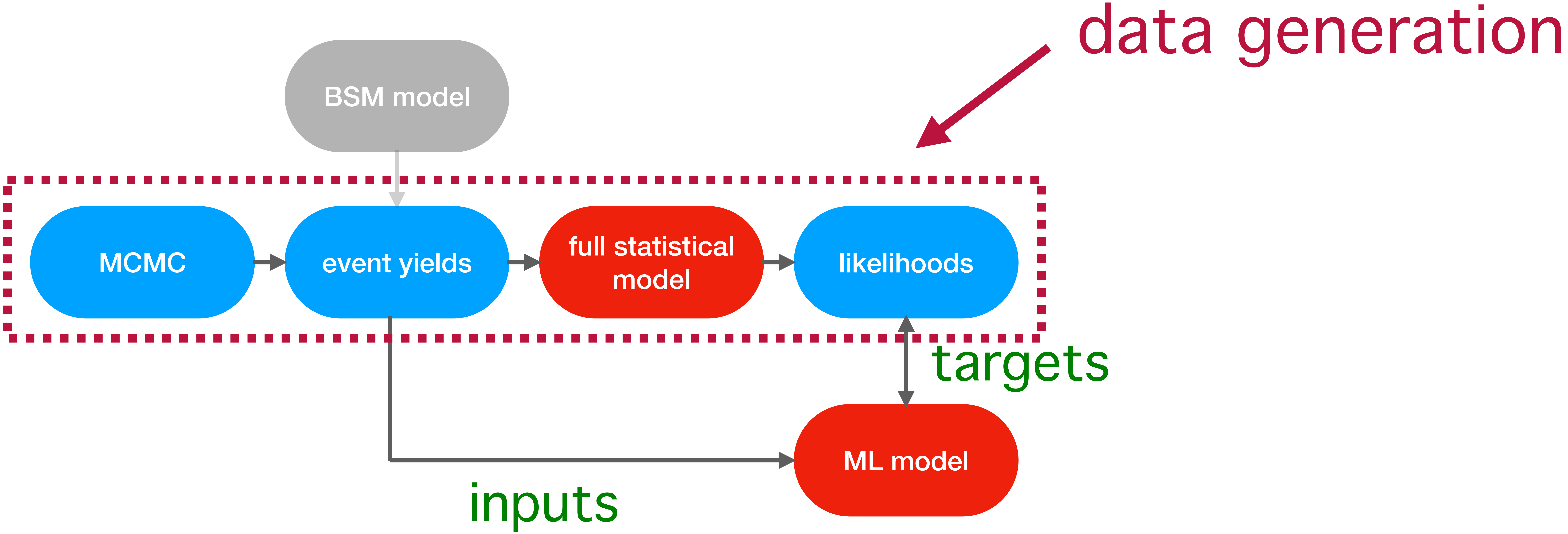
Machine Learning enters here

Old approach

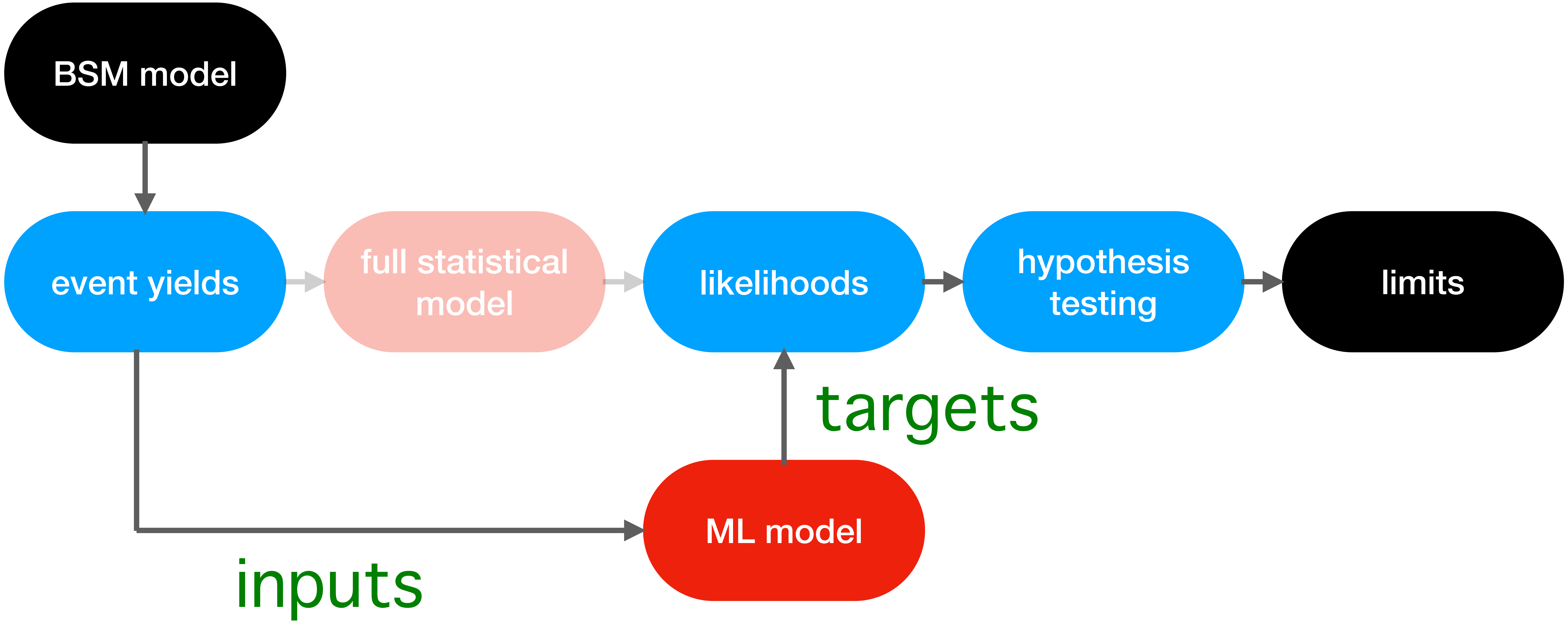


computational bottleneck

Training

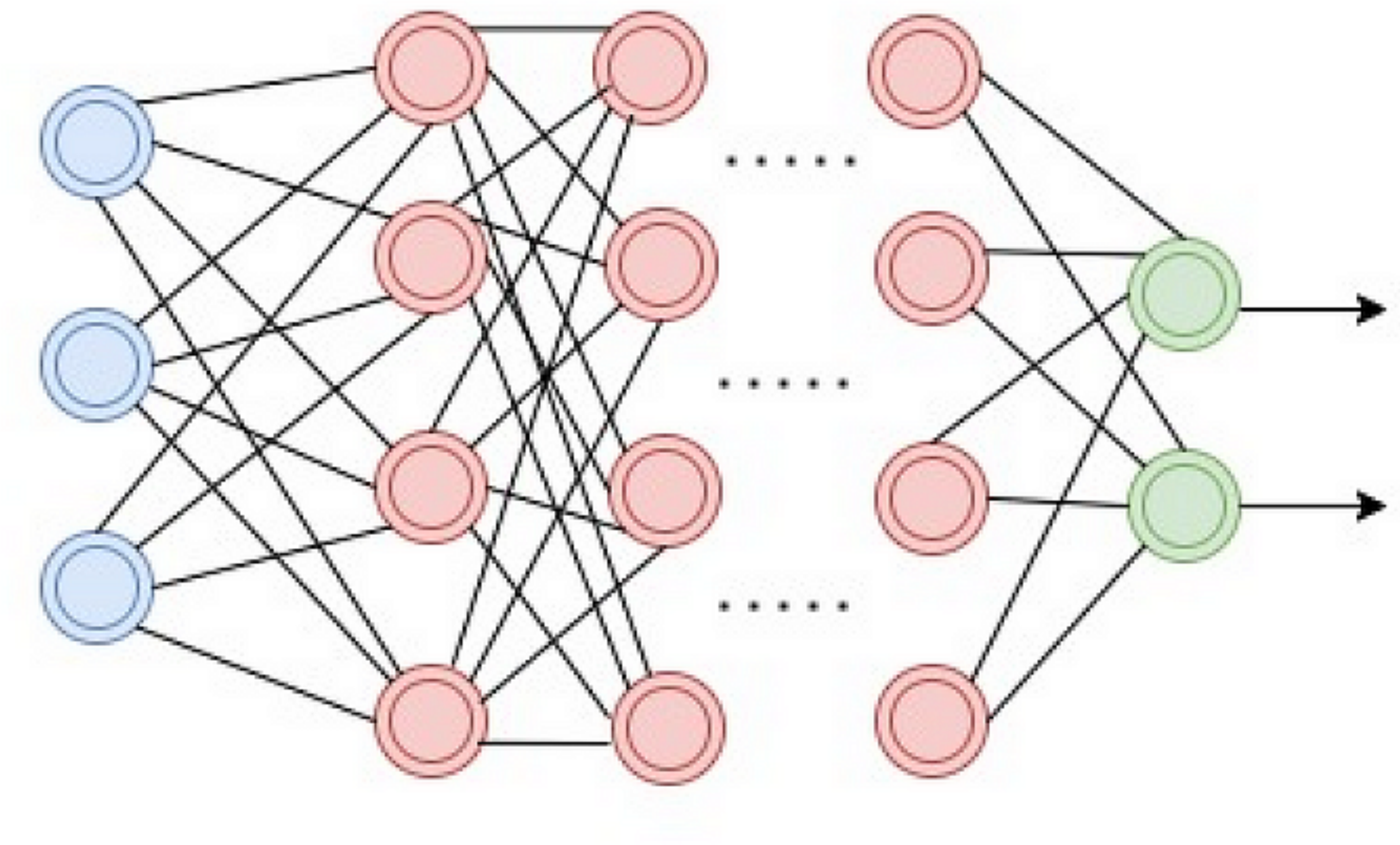

data generation
BSM model
MCMC
event yields
full statistical model
likelihoods
targets
ML model
inputs

Inference



huge speed up!

ML model



Dense 256 neurons
L2 regularization (1e-2)
batch norm
ELU

x5

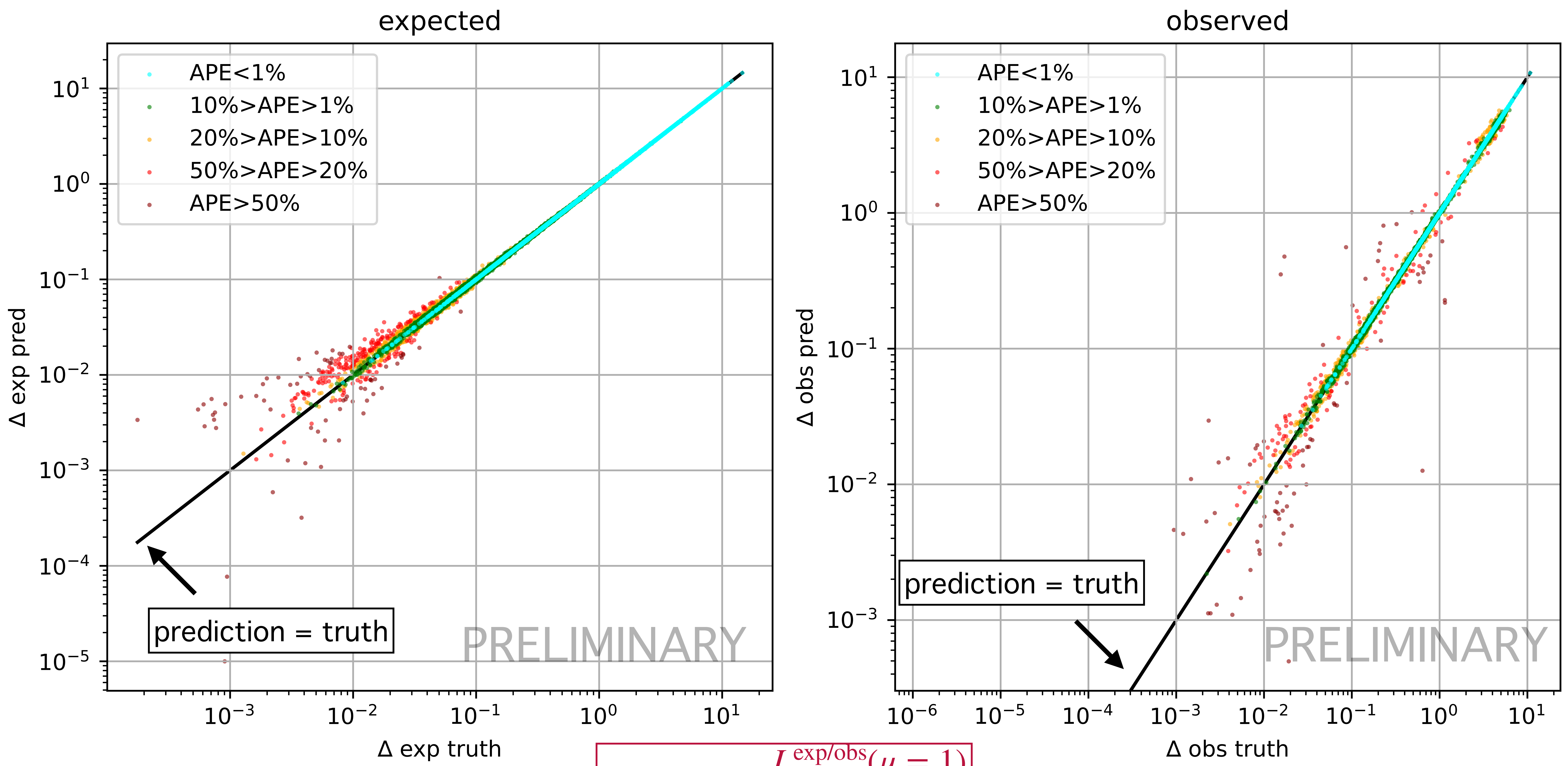
INPUTS: event yields in all bins and channels (including CRs)
OUTPUTS: negative log likelihoods (for μ=0 and μ=1), for expected and observed data
LOSS FUNCTION: MSE but others tested
OPTIMIZER: ADAM
SCHEDULER: Cosine Decay with warmup

IMG: Google Image

Preliminary results

- **ATLAS-SUSY-2018-04 [arXiv: 1911.06660]**
  - Search for direct stau production in events with two hadronic τ-leptons in √s=13 TeV pp collisions with the ATLAS detector
  - 2 signal bins, 3 control bins
- **ATLAS-CONF-2019-031 [arXiv: 1909.09226]**
  - Search for direct production of electroweakinos in final states with one lepton, missing transverse momentum and a Higgs boson decaying into two b-jets in pp collisions at √s=13 TeV with the ATLAS detector
  - 9 signal bins, 5 control bins
- **ATLAS-SUSY-2018-16 [arXiv:1911.12606]**
  - This paper presents results of searches for electroweak production of supersymmetric particles in models with compressed mass spectra
  - EWKinos: 44 signal bins, 6 control bins
  - Selectrons: 16 signal bins, 6 control bins
  - Smuons: 16 signal bins, 6 control bins
  - Sleptons: 32 signal bins, 6 control bins

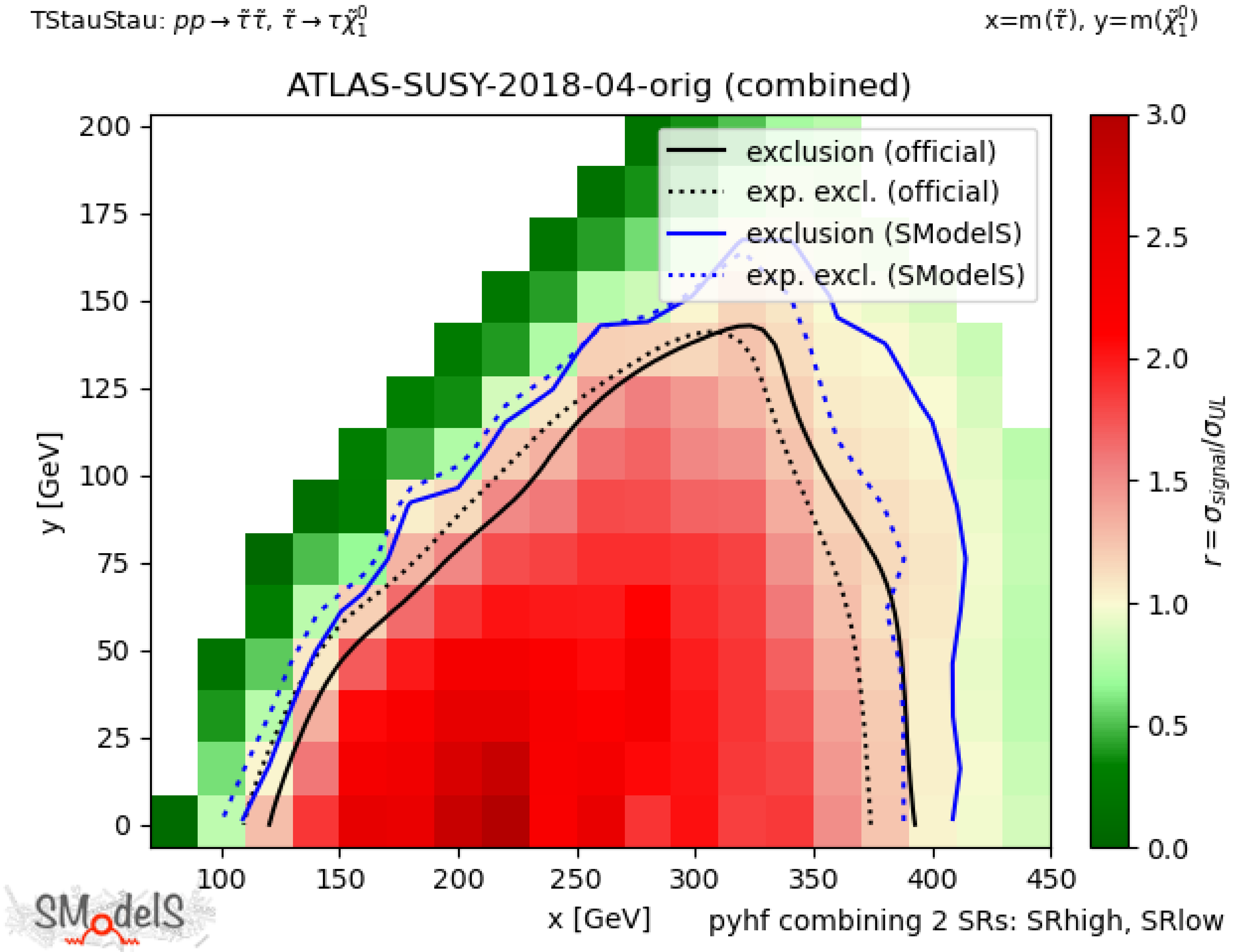
Search for direct stau production in events with two hadronic τ-leptons in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector



$$\Delta^{\text{exp/obs}} = \ln \frac{L^{\text{exp/obs}}(\mu = 1)}{L^{\text{exp/obs}}(\mu = 0)}$$

[arXiv:1911.06660]

Search for direct stau production in events with two hadronic τ-leptons in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector



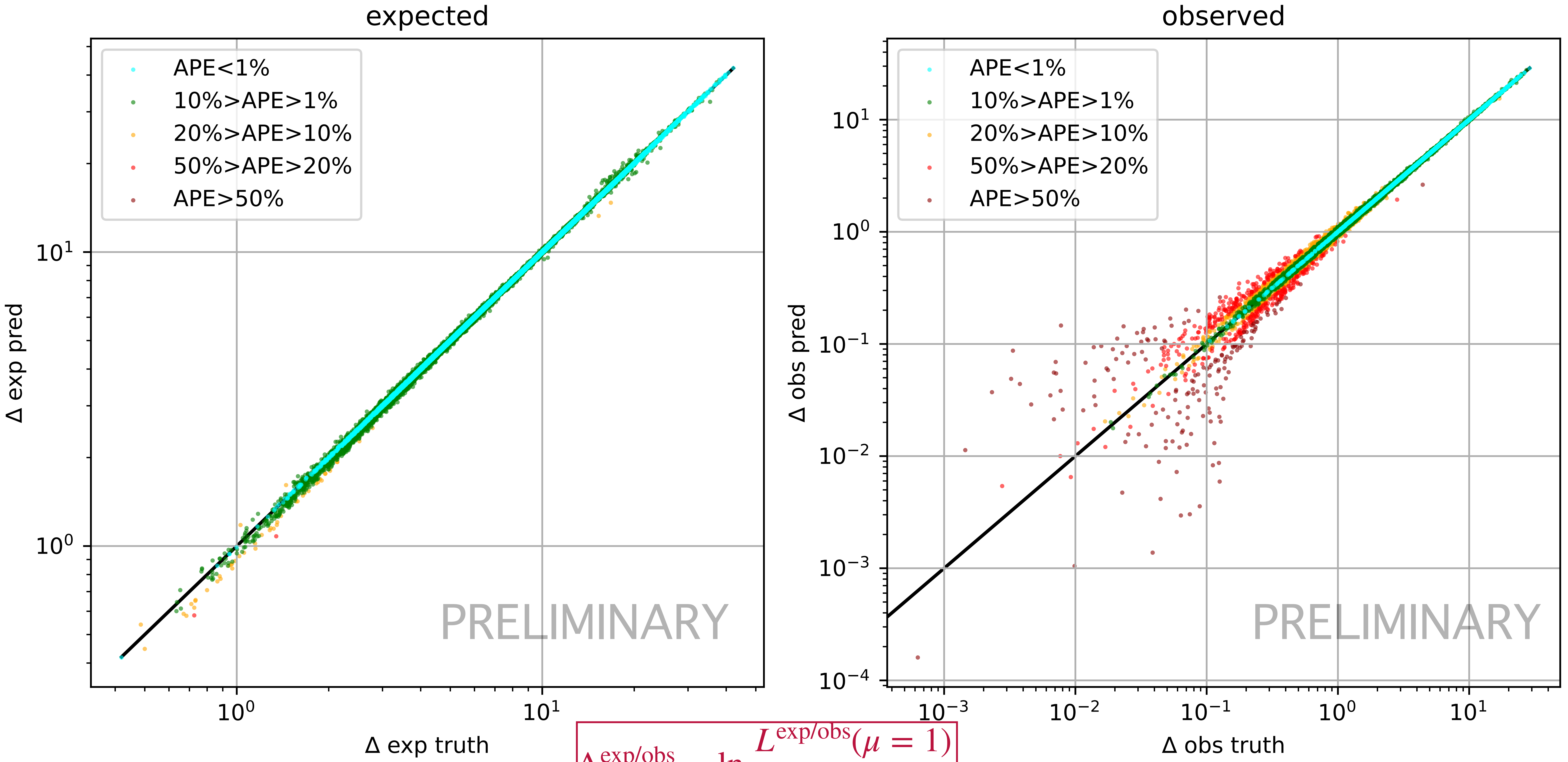
Full Likelihood Model

The reproduced limit is very off from the official one because in our fit we include CRs but apparently experimentalists don't.

->

Need for CRs removal option in the ML setup.

Search for direct production of electroweakinos in final states with one lepton, missing transverse momentum and a Higgs boson decaying into two b-jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

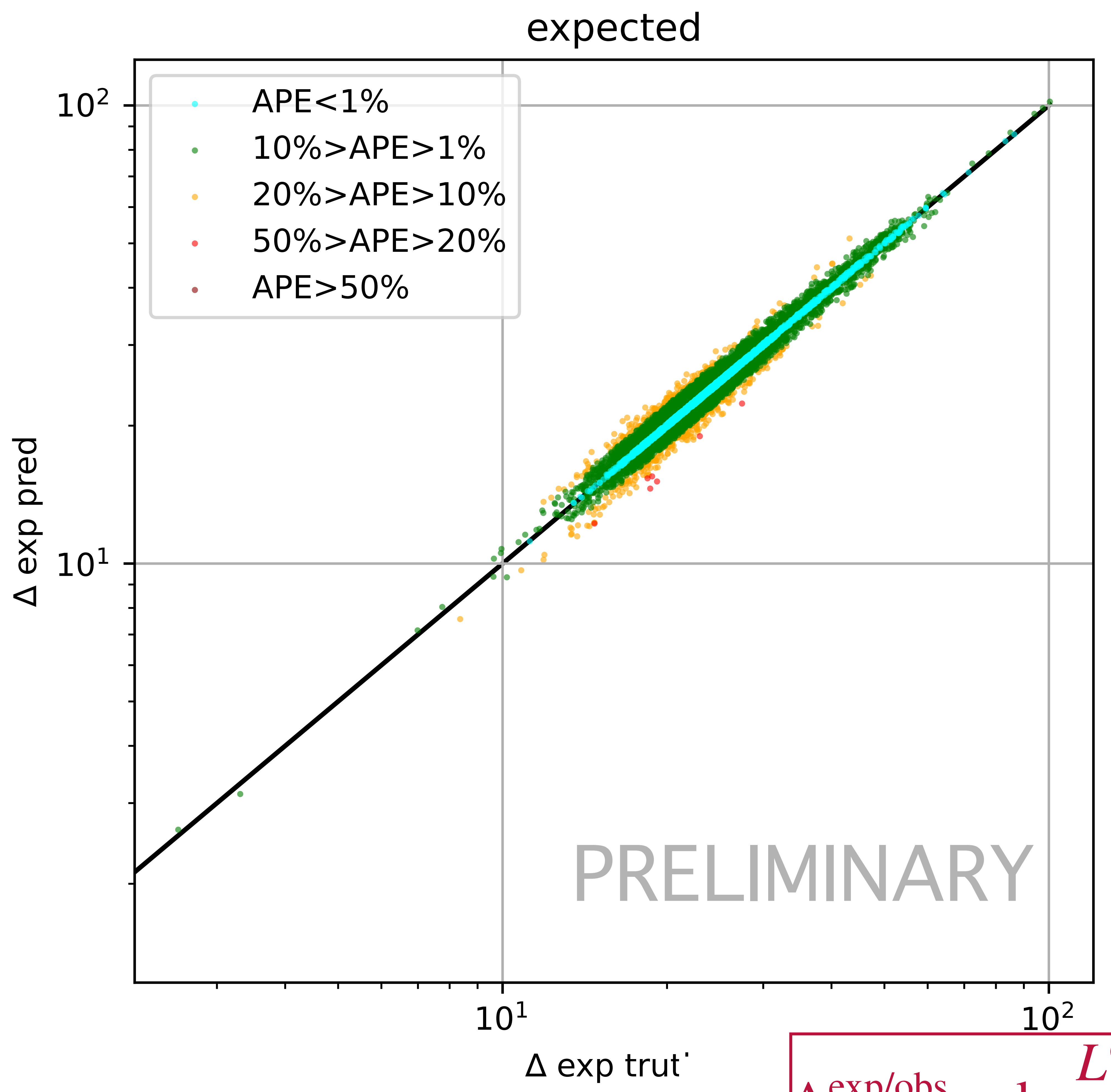


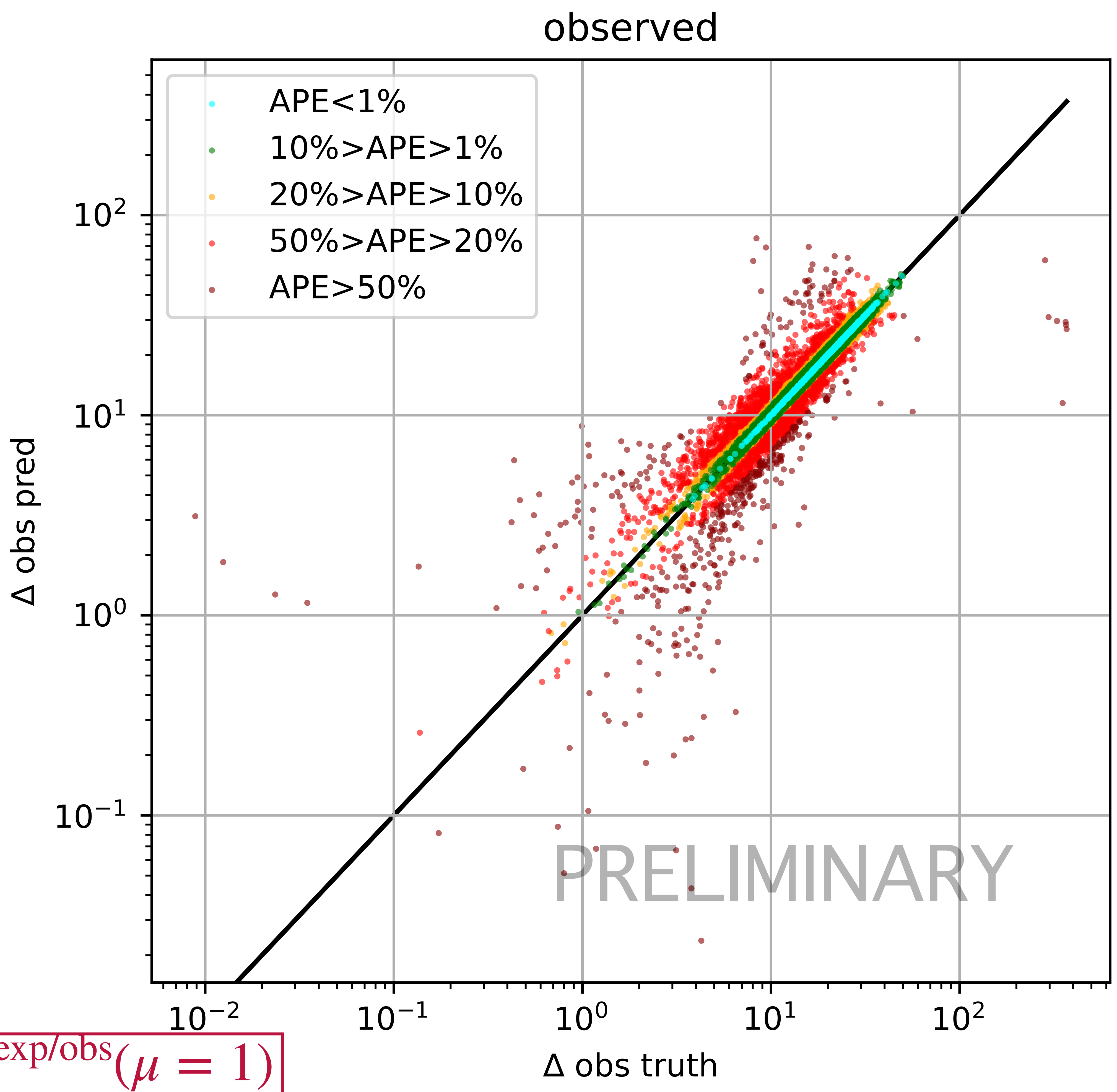
$$\Delta^{\mathrm{exp/obs}} = \ln \frac{L^{\mathrm{exp/obs}}(\mu = 1)}{L^{\mathrm{exp/obs}}(\mu = 0)}$$

[arXiv:1909.09226]

This paper presents results of searches for electroweak production of supersymmetric particles in models with compressed mass spectra. The searches use 139 fb^{-1} of $\sqrt{s}$ = 13 TeV proton–proton collision data collected by the ATLAS experiment at the Large Hadron Collider. Events with missing transverse momentum and two same-flavor, oppositely charged, low transverse momentum leptons are selected, and are further categorized by the presence of hadronic activity from initial-state radiation or a topology compatible with vector-boson fusion processes. The data are found to be consistent with predictions from the Standard Model. The results are interpreted using simplified models of R-parity-conserving supersymmetry in which the lightest supersymmetric partner is a neutralino with a mass similar to the lightest chargino, the second-to-lightest neutralino or the slepton. Lower limits on the masses of charginos in different simplified models range from 193 GeV to 240 GeV for moderate mass splittings, and extend down to mass splittings of 1.5 GeV to 2.4 GeV at the LEP chargino bounds (92.4 GeV). Similar lower limits on degenerate light-flavor sleptons extend up to masses of 251 GeV and down to mass splittings of 550 MeV. Constraints on vector-boson fusion production of electroweak SUSY states are also presented.

EWKino model





$$\boxed{\Delta^{\text{exp/obs}} = \ln \frac{L^{\text{exp/obs}}(\mu = 1)}{L^{\text{exp/obs}}(\mu = 0)}}$$

[arXiv: 1911.12606]

LPSC Grenoble

Recent updates

- Wolfgang implemented new validation
- Discrepancy between nLL without signal and nLL with signal and mu=0
- Automatic removal of CRs ->**1911.06660** (spey bug found and fixed)
- Possibility to remove arbitrary channels -> **1911.12606**
- Asimov likelihoods are now also calculated -> first scan for **1909.09226** -> problem with background estimation with spey
- Maximal likelihoods are calculated and stored in metadata -> problems with optimisation for normal likelihoods
- Main branch of the repository corresponds to current version of the project.
- Need to add more descriptive metadata and clean up the repository.

Thank you for attention!

rafal.maselek@lpsc.in2p3.fr

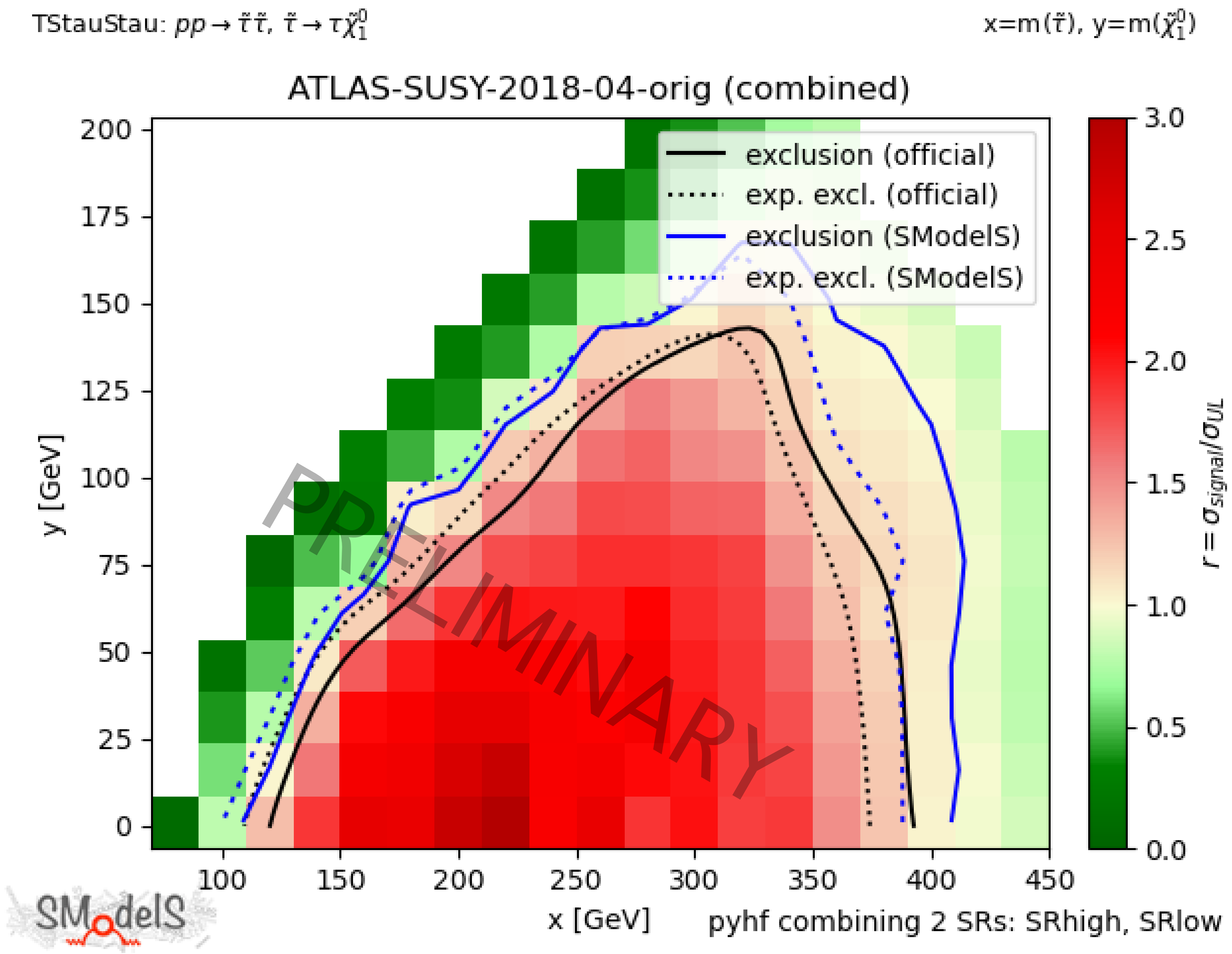
Dolina Chochołowska, Poland
photo by Piotr Kałuża

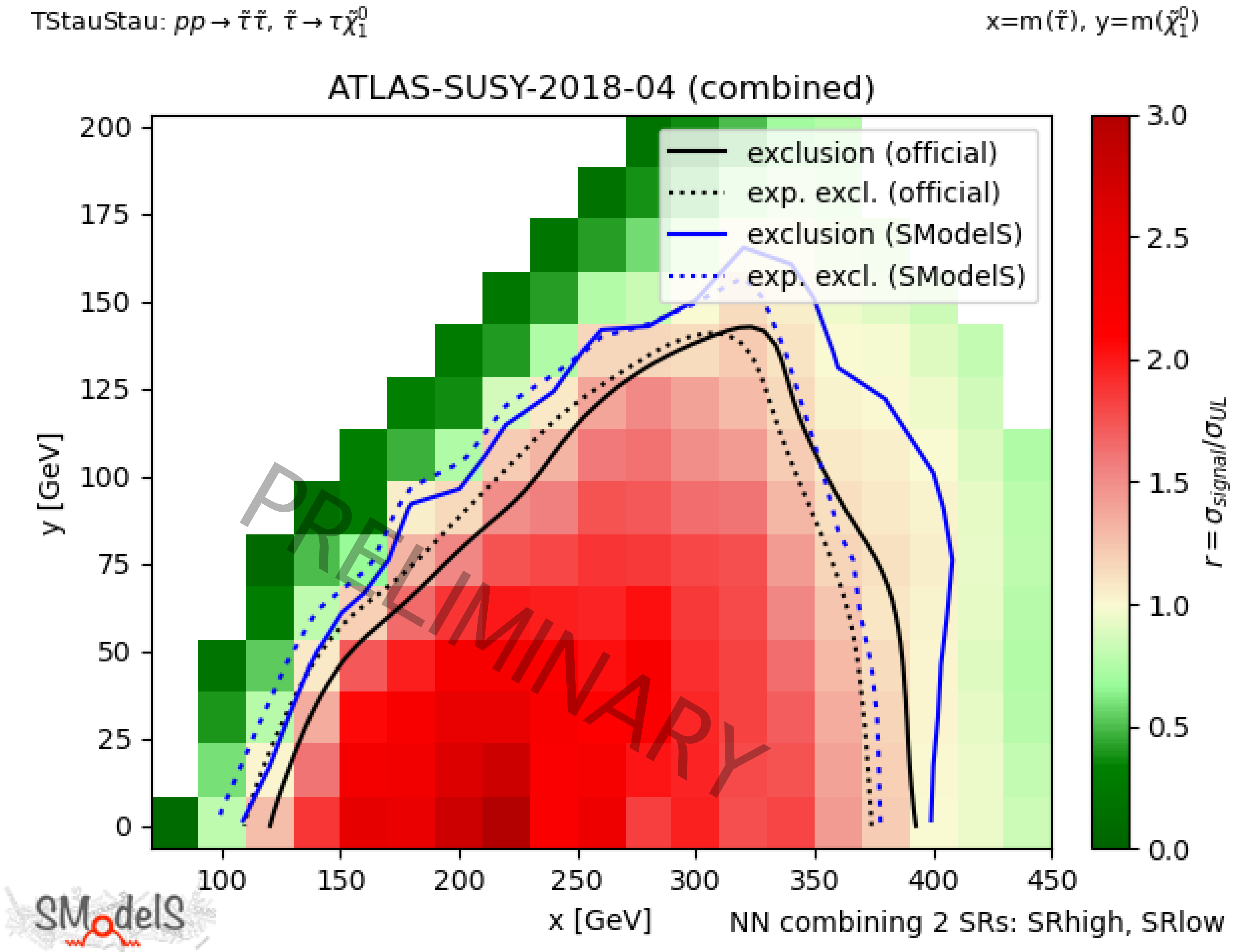
ATLAS-CONF-2019-031 [arXiv: 1909.09226]

BIN	SR	BKG	ΔBKG	OBS
SRHMEM_mct2-0	True	4.2	1.6	6.0
SRHMEM_mct2-1	True	2.9	2.2	5.0
SRHMEM_mct2-2	True	1.0	1.9	3.0
SRLMEM_mct2-0	True	11.0	7.5	16.0
SRLMEM_mct2-1	True	10.3	7.6	11.0
SRLMEM_mct2-2	True	6.9	5.1	7.0
SRMMEM_mct2-0	True	5.4	4.0	4.0
SRMMEM_mct2-1	True	2.8	3.6	7.0
SRMMEM_mct2-2	True	1.4	1.2	2.0
STCREM_cuts-0	False	178.4	68.0	155.0
TRHMEM_cuts-0	False	680.5	329.6	641.0
TRLMEM_cuts-0	False	717.5	408.6	657.0
TRMMEM_cuts-0	False	474.4	279.2	491.0
WREM_cuts-0	False	130.1	53.4	144.0

BIN	MIN	MAX
SRHMEM_mct2-0	1.6	14.1
SRHMEM_mct2-1	2.2	16.1
SRHMEM_mct2-2	0.0	12.3
SRLMEM_mct2-0	7.5	53.7
SRLMEM_mct2-1	7.6	48.8
SRLMEM_mct2-2	5.1	32.3
SRMMEM_mct2-0	4.0	24.1
SRMMEM_mct2-1	0.0	25.1
SRMMEM_mct2-2	1.2	8.0
STCREM_cuts-0	139.5	170.5
TRHMEM_cuts-0	576.9	705.1
TRLMEM_cuts-0	591.3	722.7
TRMMEM_cuts-0	441.9	540.1
WREM_cuts-0	129.6	158.4

Search for direct stau production in events with two hadronic τ−leptons in $\sqrt{s}$ = 13 TeV pp collisions with the ATLAS detector



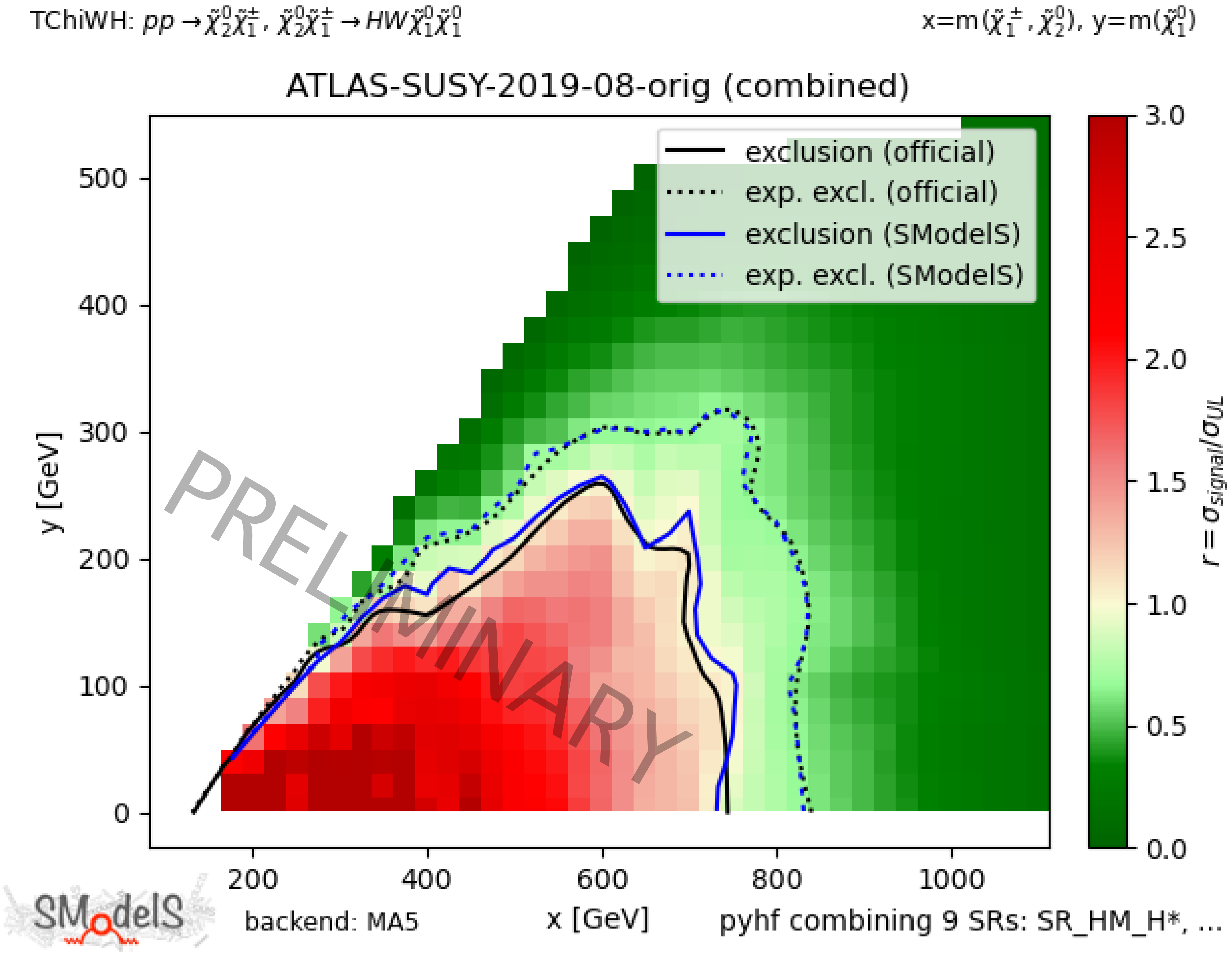


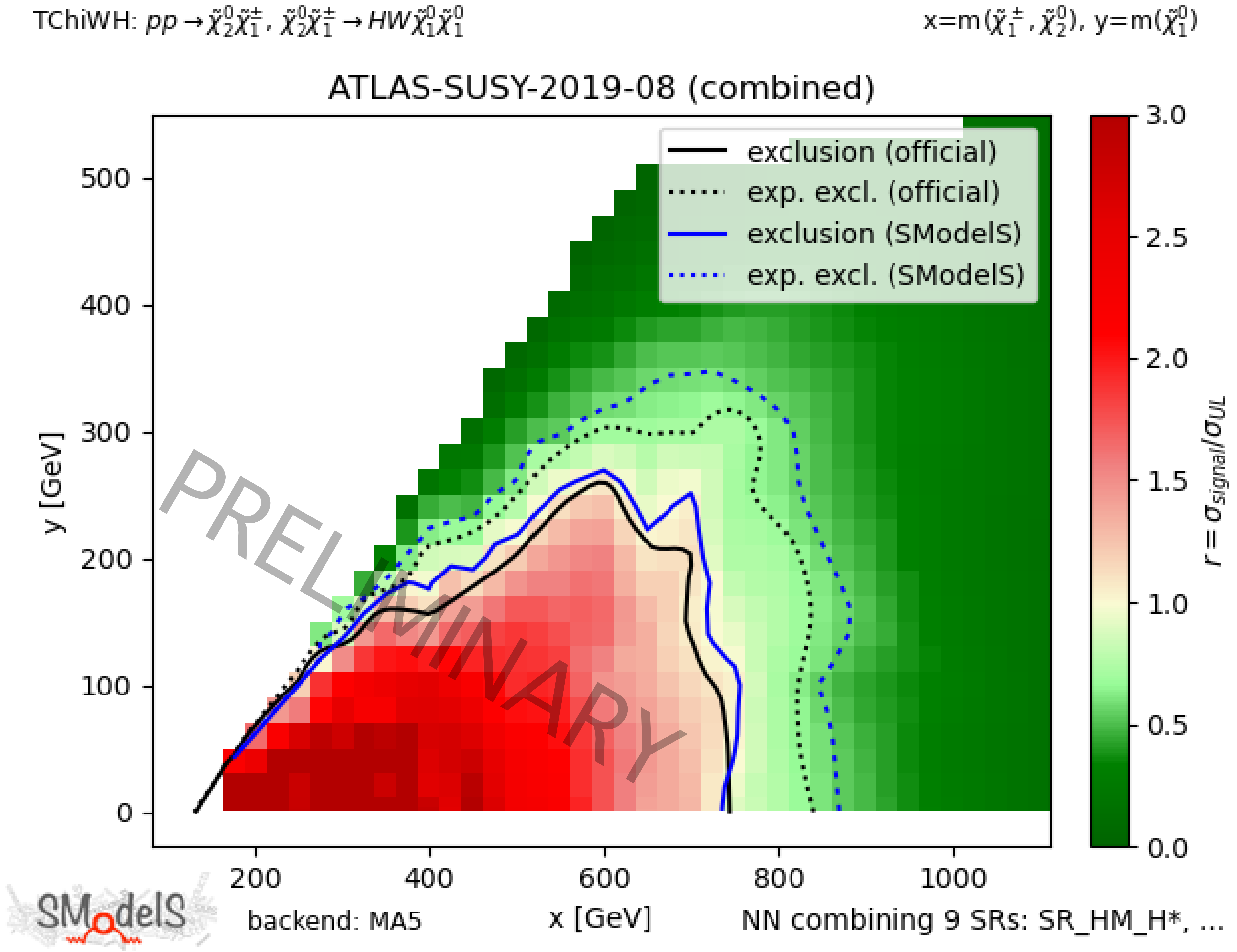
Full Likelihood Model

ML SURROGATE

[arXiv:1911.06660]

Search for direct production of electroweakinos in final states with one lepton, missing transverse momentum and a Higgs boson decaying into two b-jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector





Full Likelihood Model

ML SURROGATE

[arXiv:1909.09226]

R. Masełek, 17-12-2024