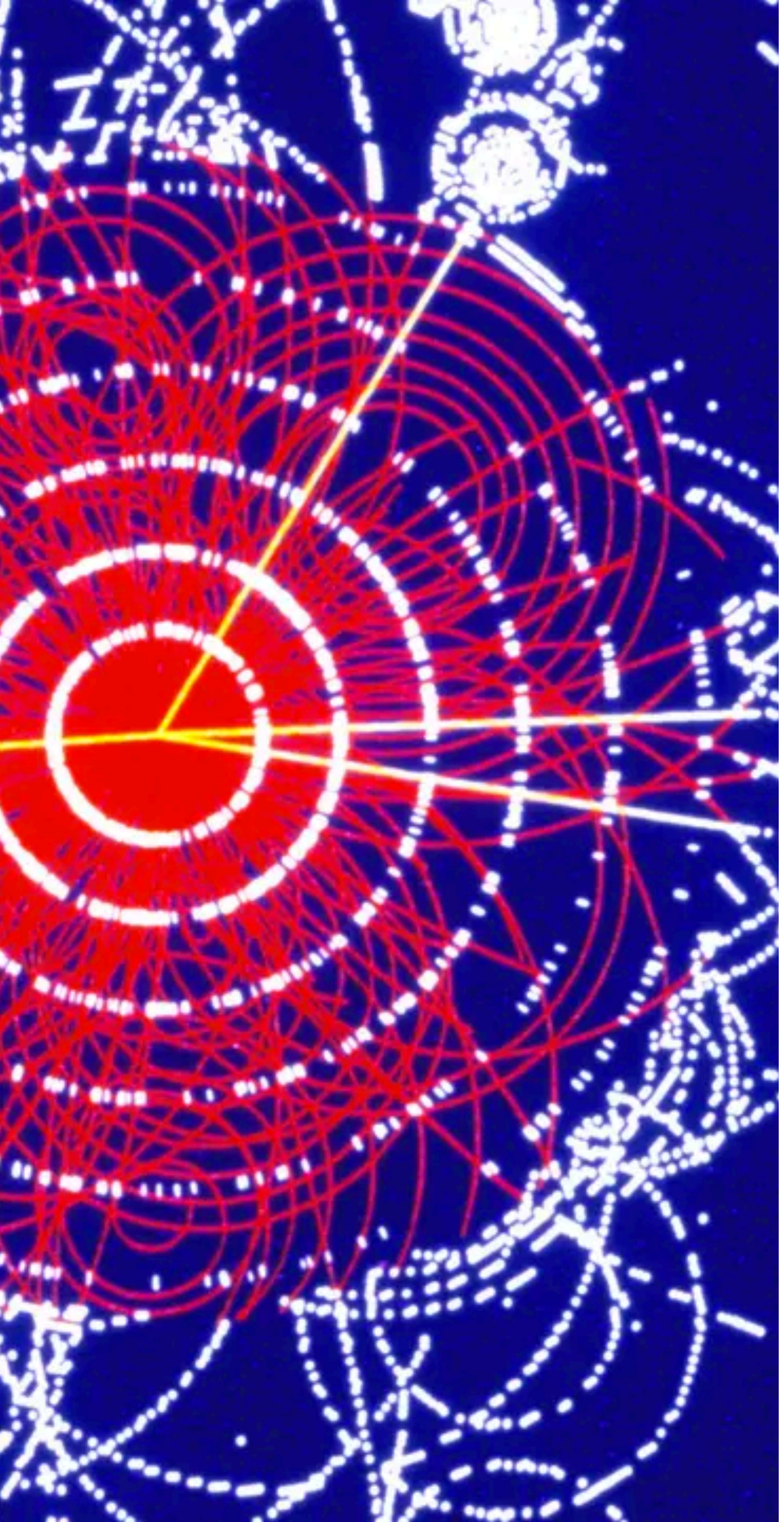


Machine-learning the likelihoods

Rafał Masełek

in collaboration with:
S. Kraml, W. Waltenberger, J. Araz, H. Reyes, A. Butter, C. Krause

ML4Jets 2024

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

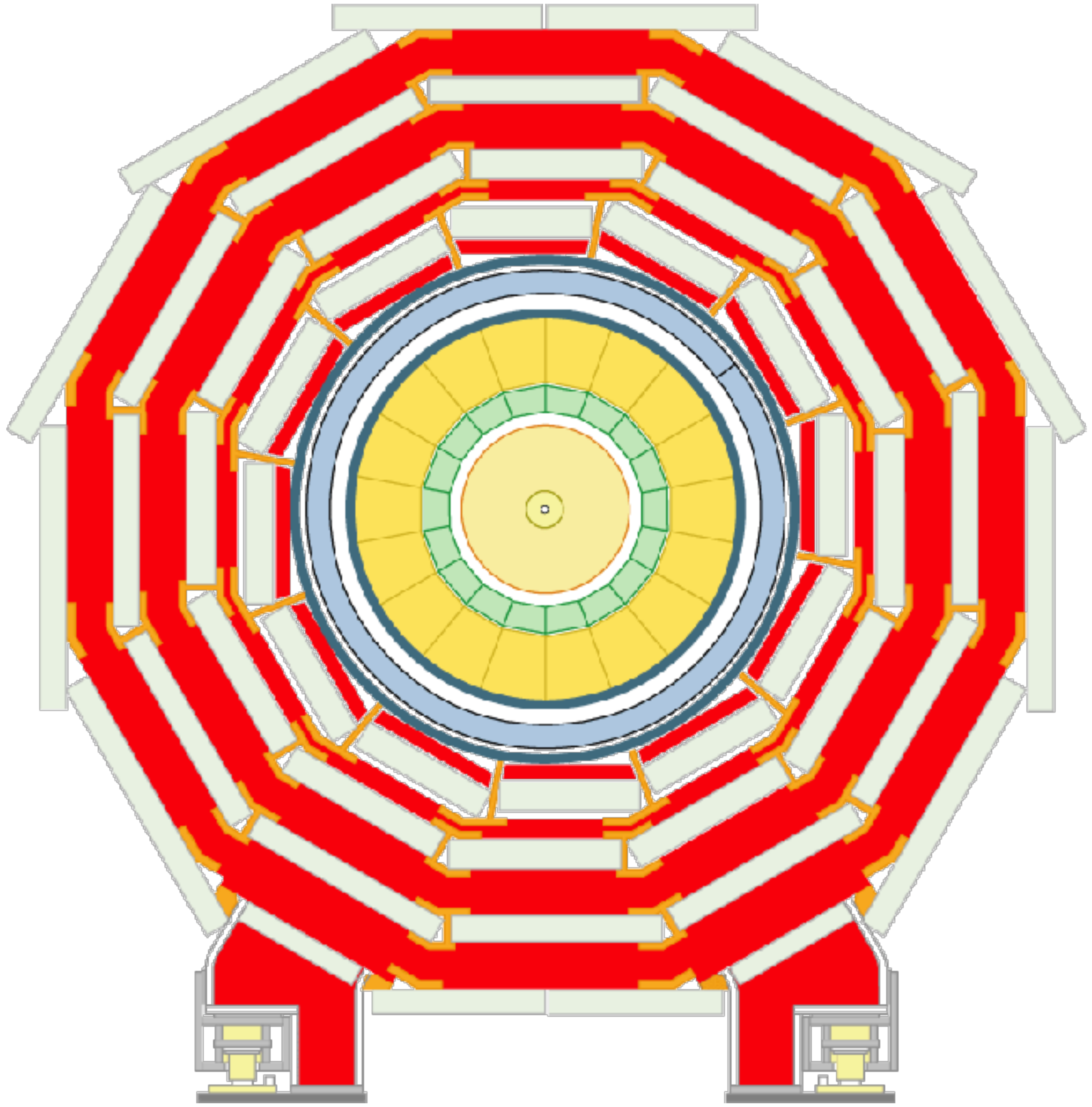
AXIONS WIMPs

EXTRA DIMENSIONS

MILICHARGED PARTICLES

DARK SECTOR

?



Reinterpretation

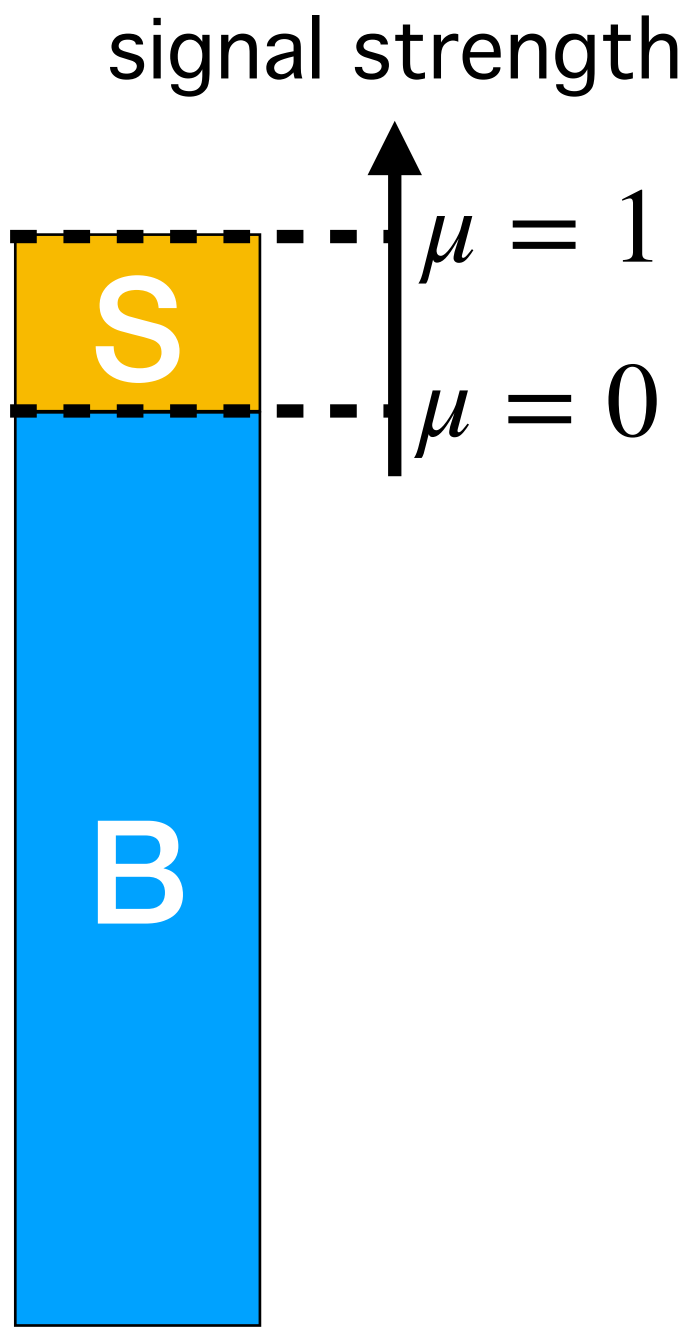
Goal: Enhance and unify the statistical analysis step

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

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?

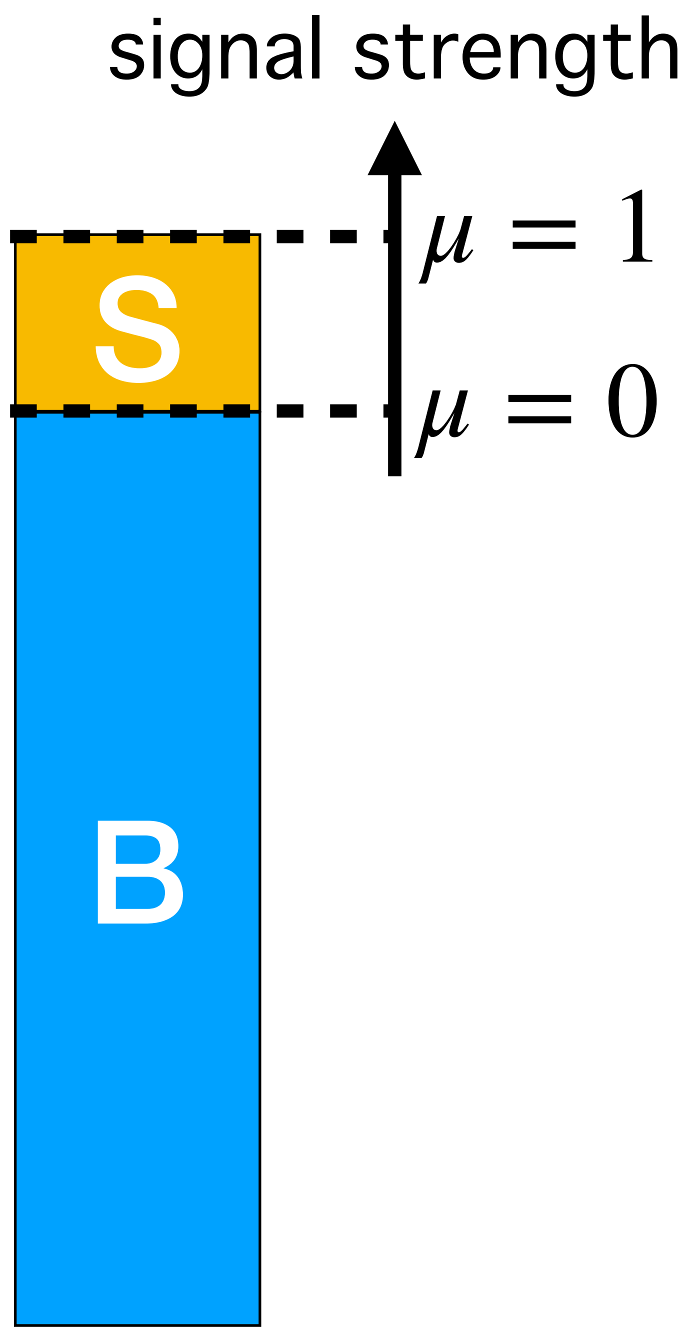


$$L(\mu) = p\left(\{x_1, \ldots x_n\} \;\; \mu\right) = \textbf{?}$$

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?



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

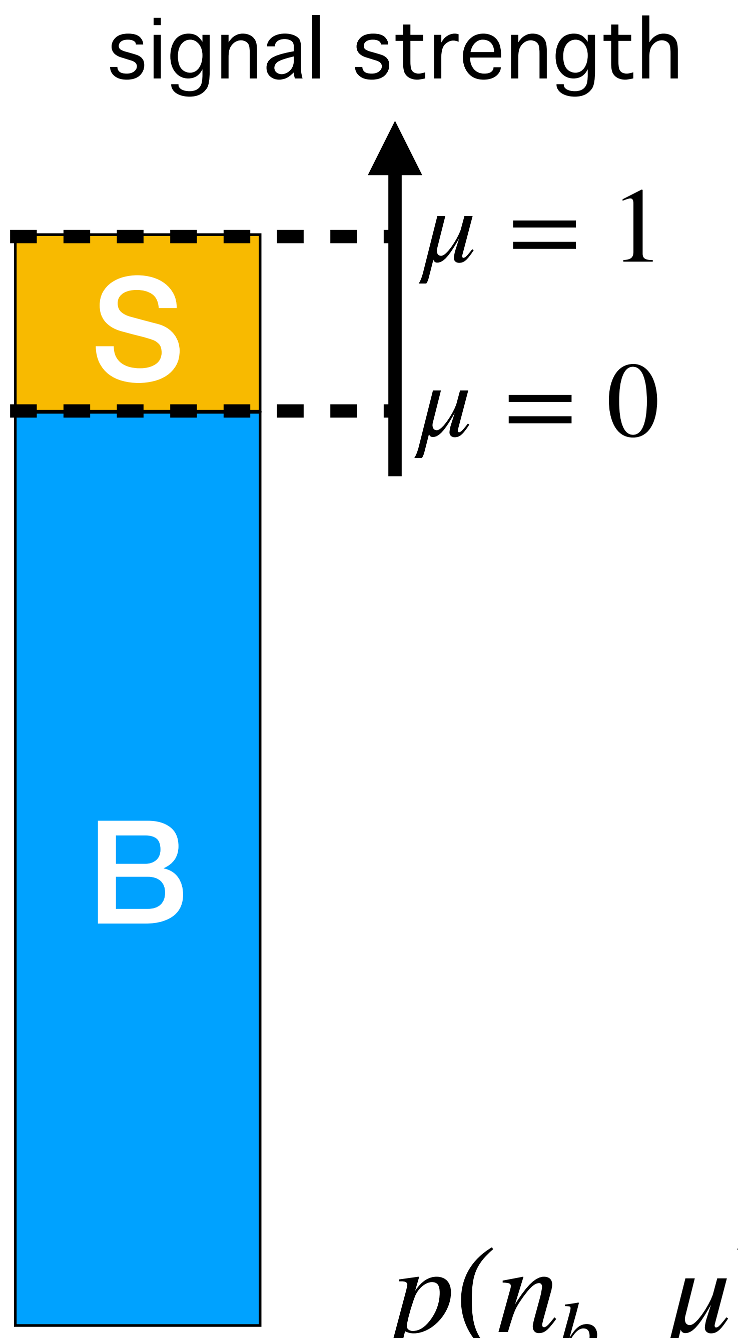
probability to observe n events given μS+B expectation

probability density of obtaining x_e based on the relative mixture of $f_S(x)$ and $f_B(x)$

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

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

auxiliary data

channel data

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

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

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

free parameters

constrained parameters

simultaneous measurement of multiple channels

constraint terms for "auxiliary measurements"

$$\nu_{cb}(\mu, \theta) = \sum_{s}^{\text{samples}} \nu_{scb}(\mu, \theta) = \sum_{s}^{\text{samples}} \left(\prod_{\kappa} \kappa_{scb}(\mu, \theta)\right)\left(\nu_{scb}^{0}(\mu, \theta) + \sum_{\Delta} \Delta_{scb}(\mu, \theta)\right)$$

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


```
"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"
                },
```


[arXiv:2211.15838]

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:

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

$\hat{\mu}, \hat{\theta}$ — unconditional ML estimators

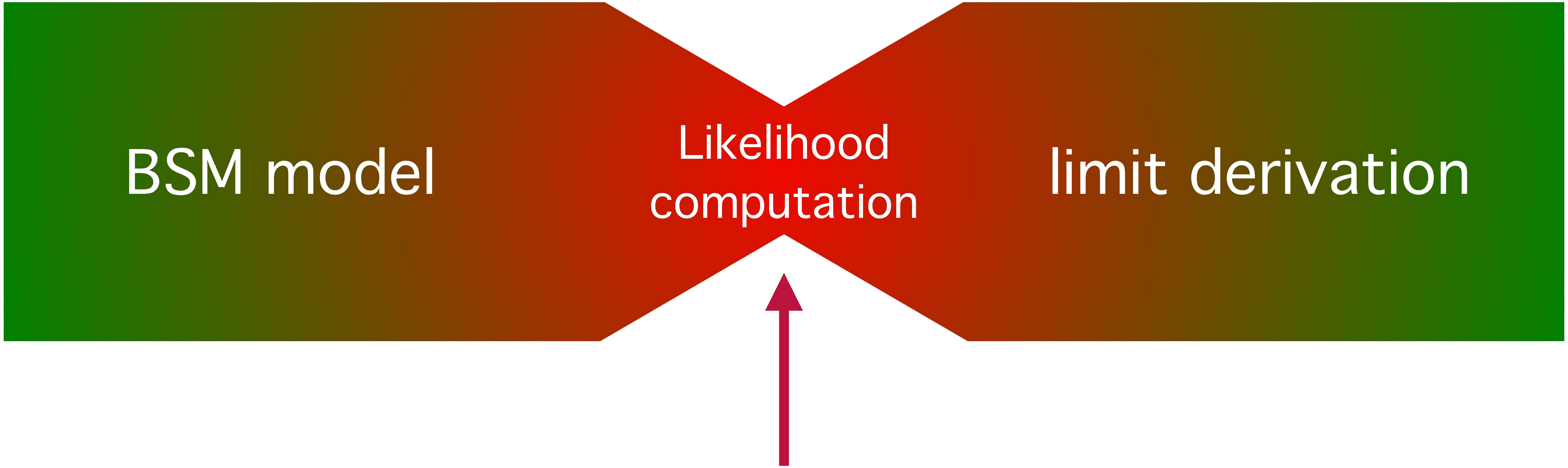
$\hat{\hat{\theta}}(\mu)$ — ML estimator conditioned on μ.

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

f — PDF of $\tilde{q}_\mu$

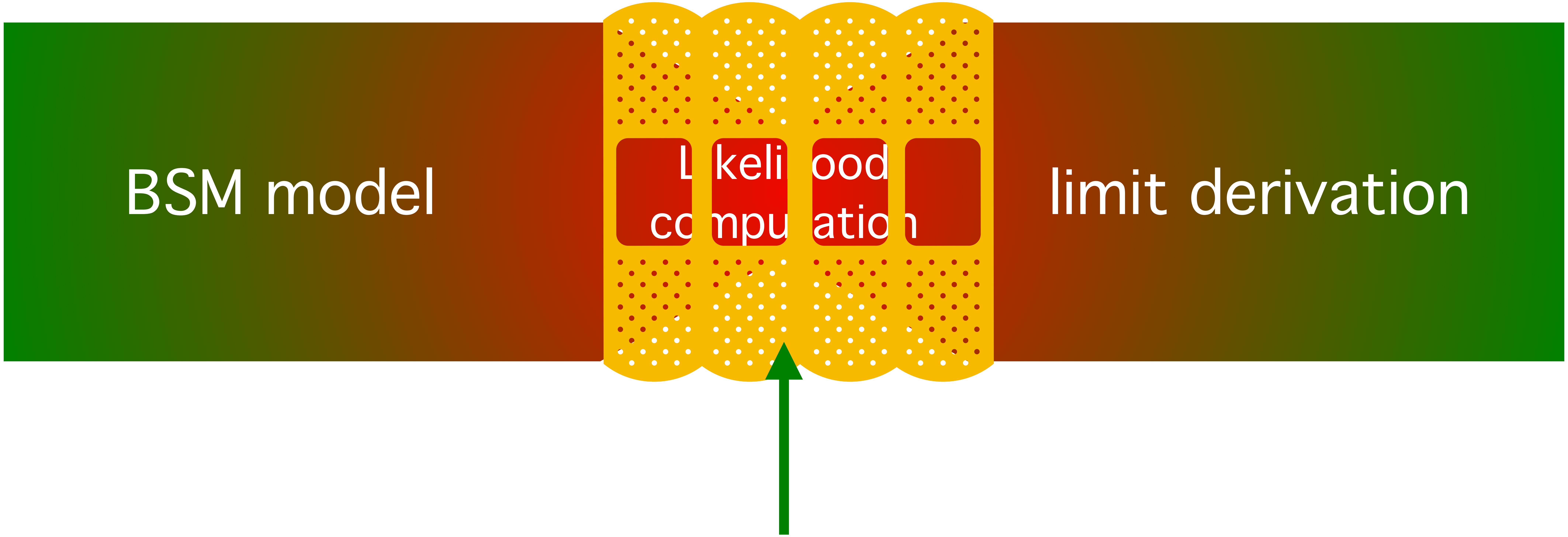
[[arXiv:1007.1727]

Computational bottleneck



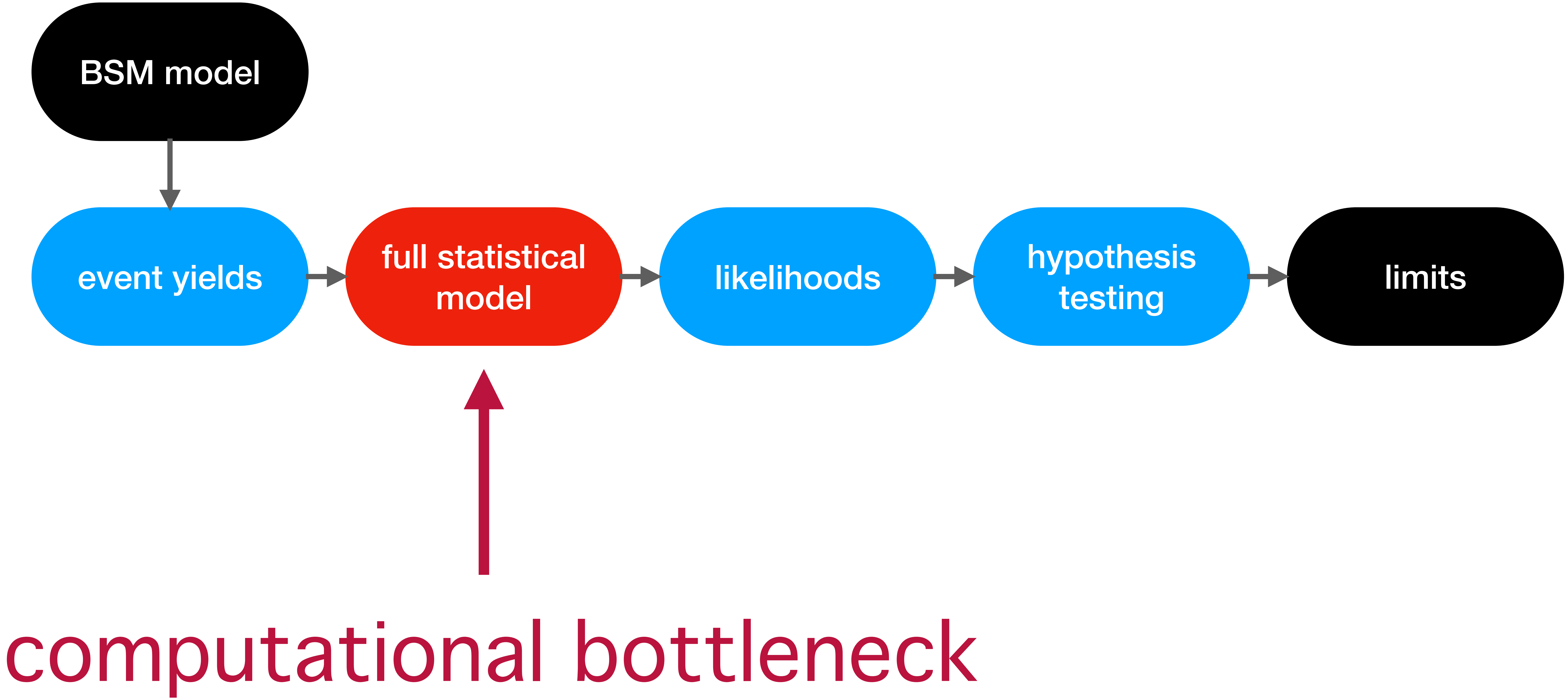
Full statistical model calculations enter here

Fixing the problem

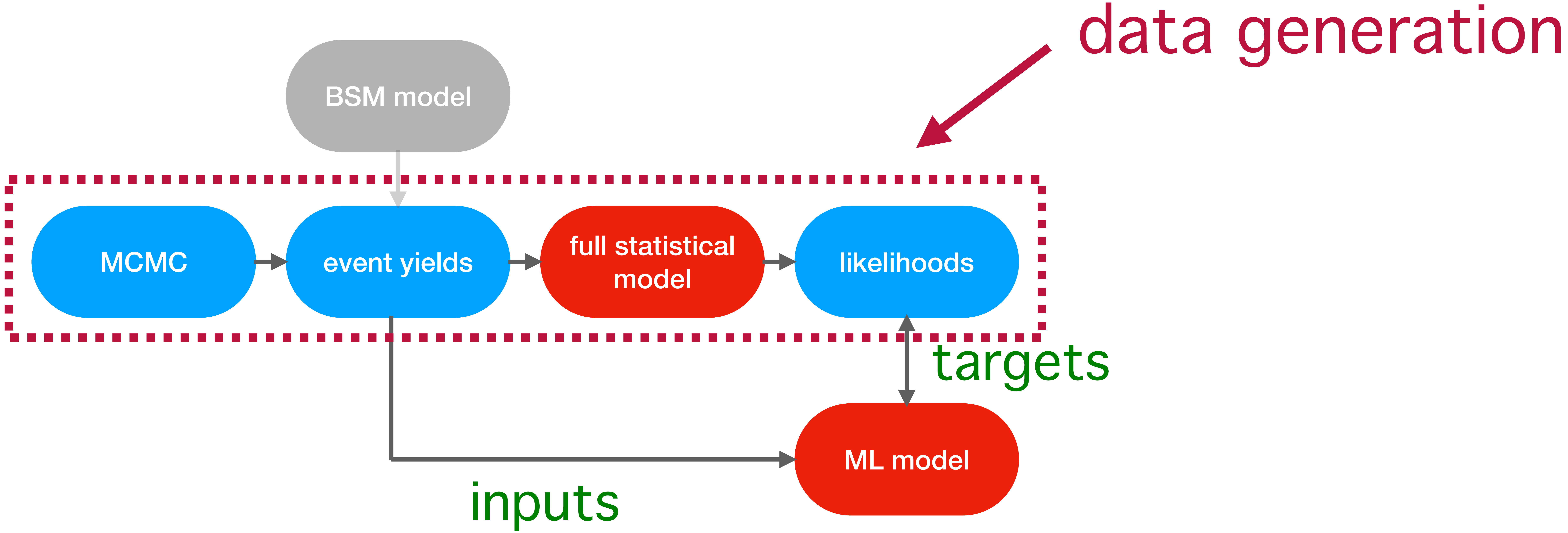


Machine Learning enters here

Old approach

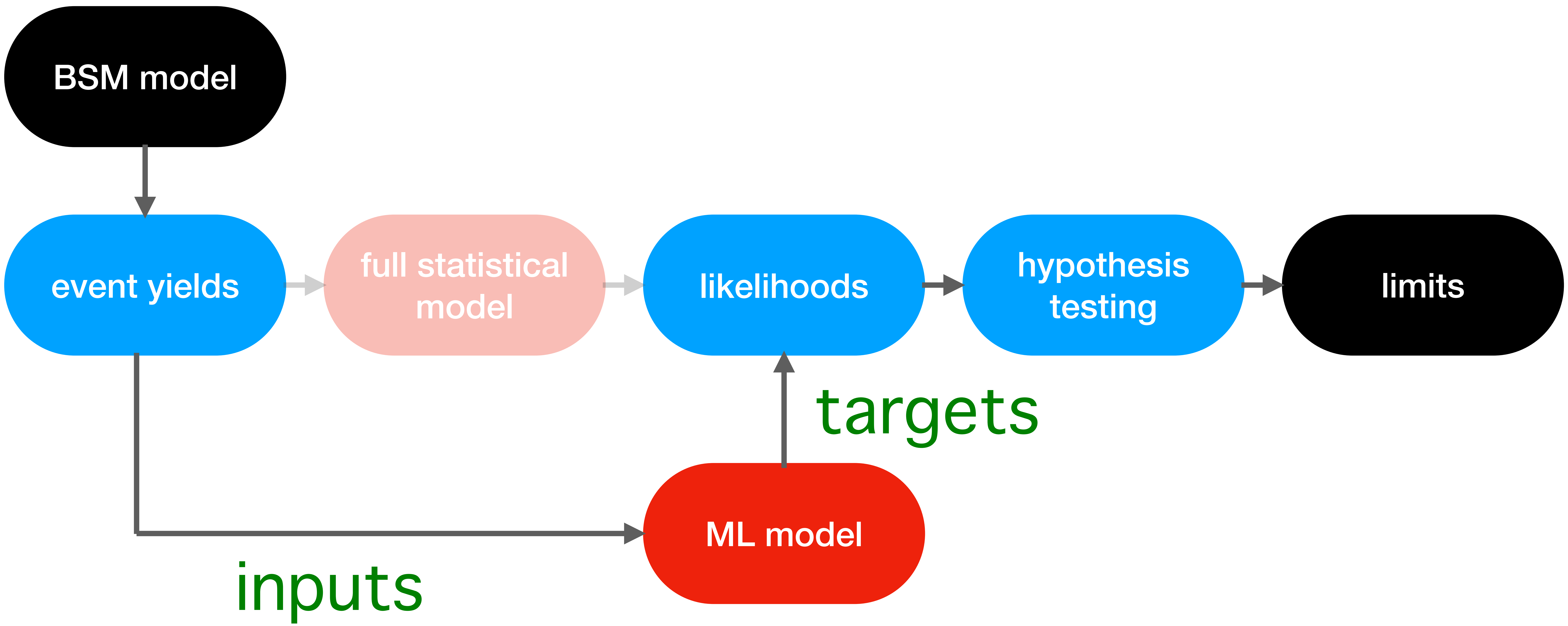


Training



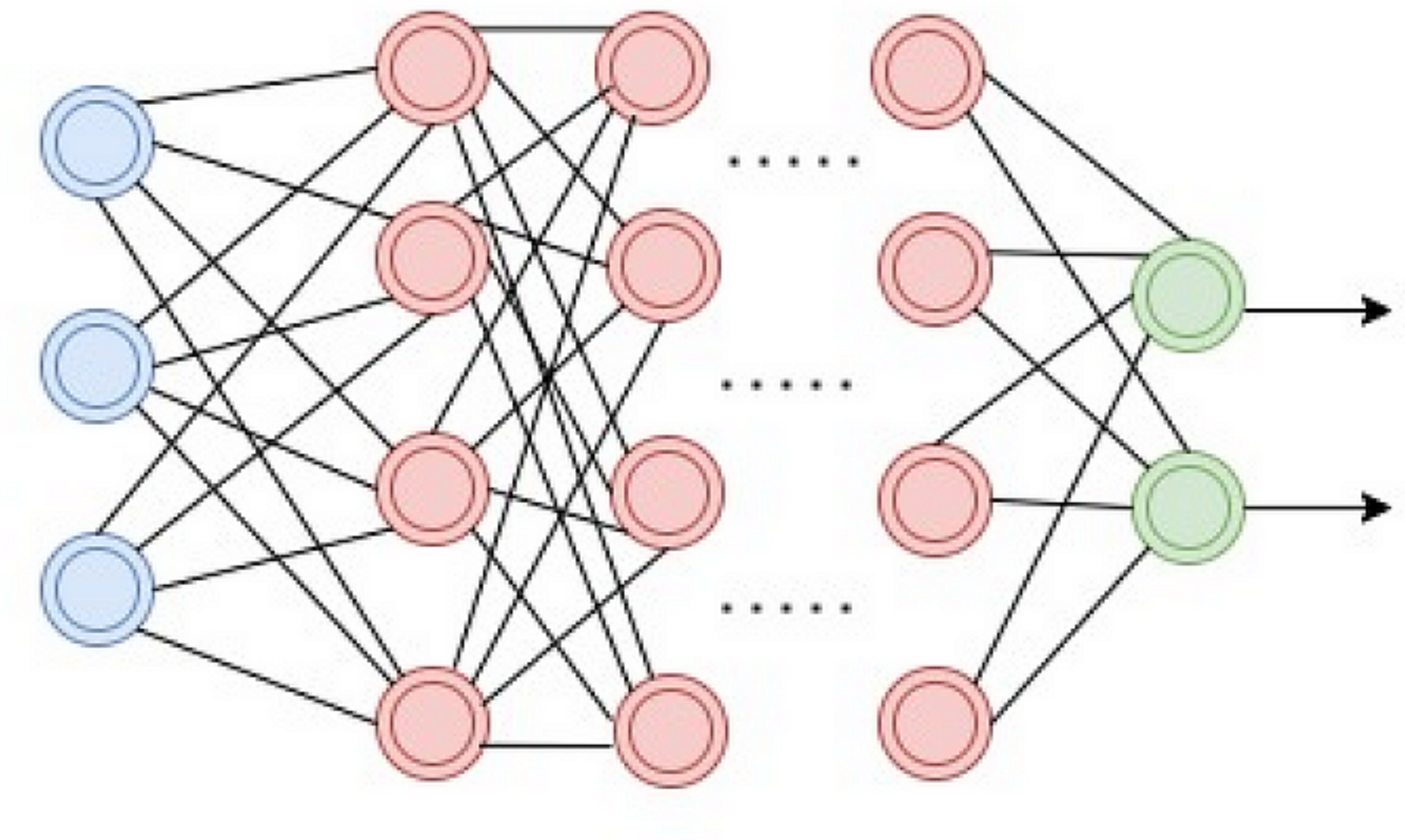
data generation
BSM model
MCMC
event yields
full statistical model
likelihoods
targets
ML model
inputs

Inference



huge speed up!

ML model



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

x5

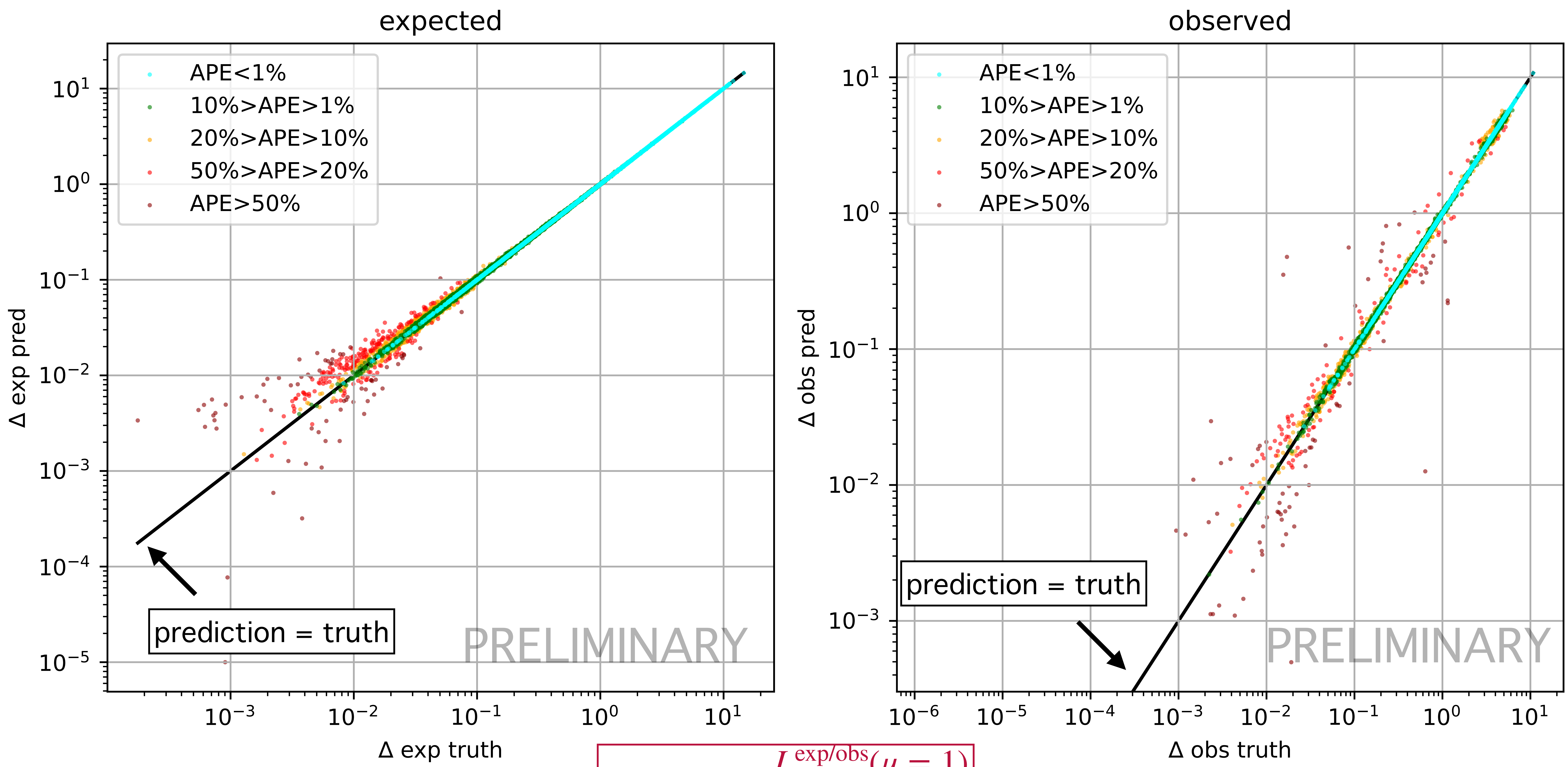
INPUTS: event yields in all bins and channels (including CRs)
OUTPUTS: negative log likelihoods (for $\mu=0$ and $\mu=1$), for expected and observed data
LOSS FUNCTION: MSE but others tested
OPTIMIZER: ADAM
SCHEDULER: Cosine Decay with warmup

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

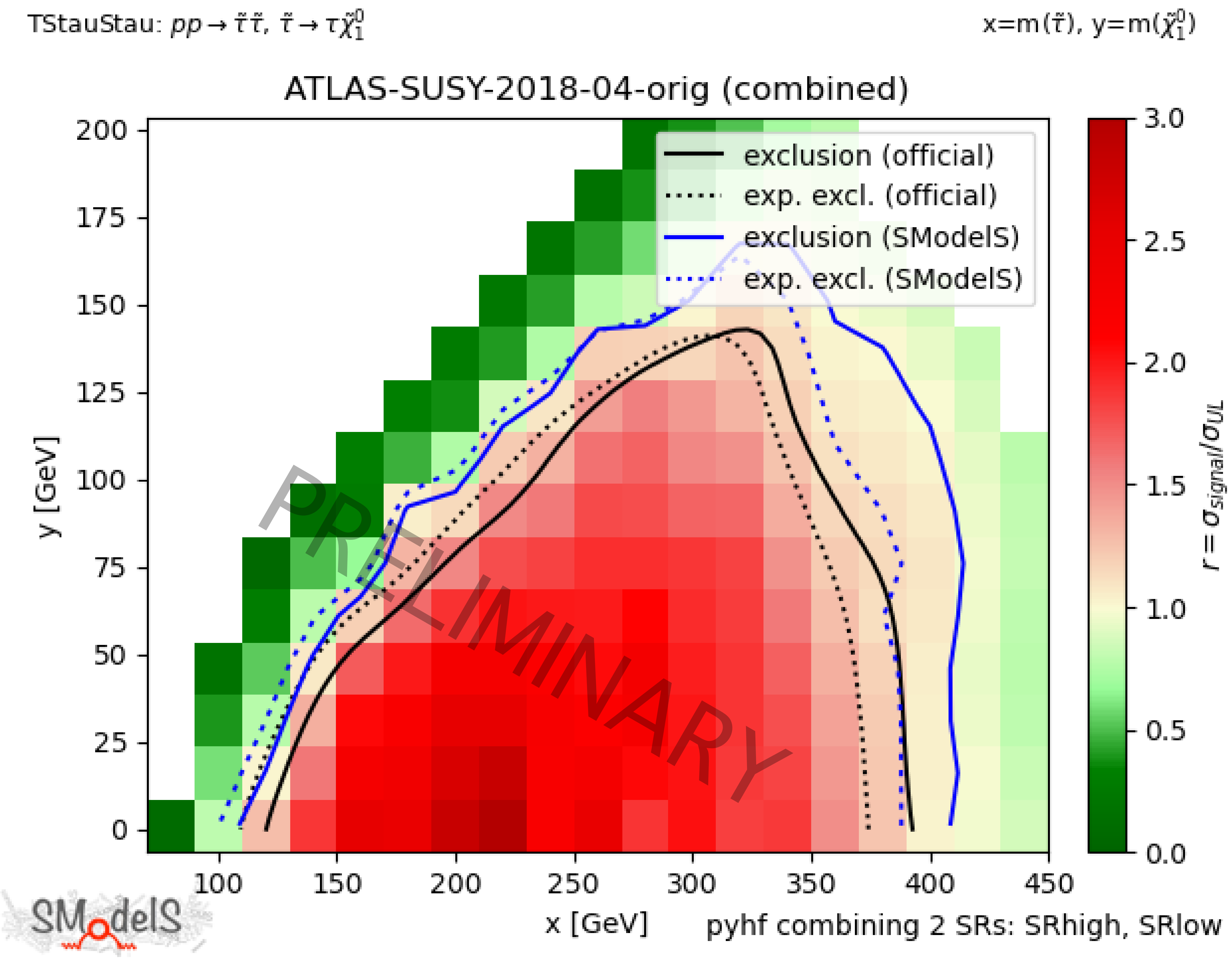
Search for direct stau production in events with two hadronic τ−leptons in $\sqrt{s}$ = 13 TeV pp collisions with the ATLAS detector



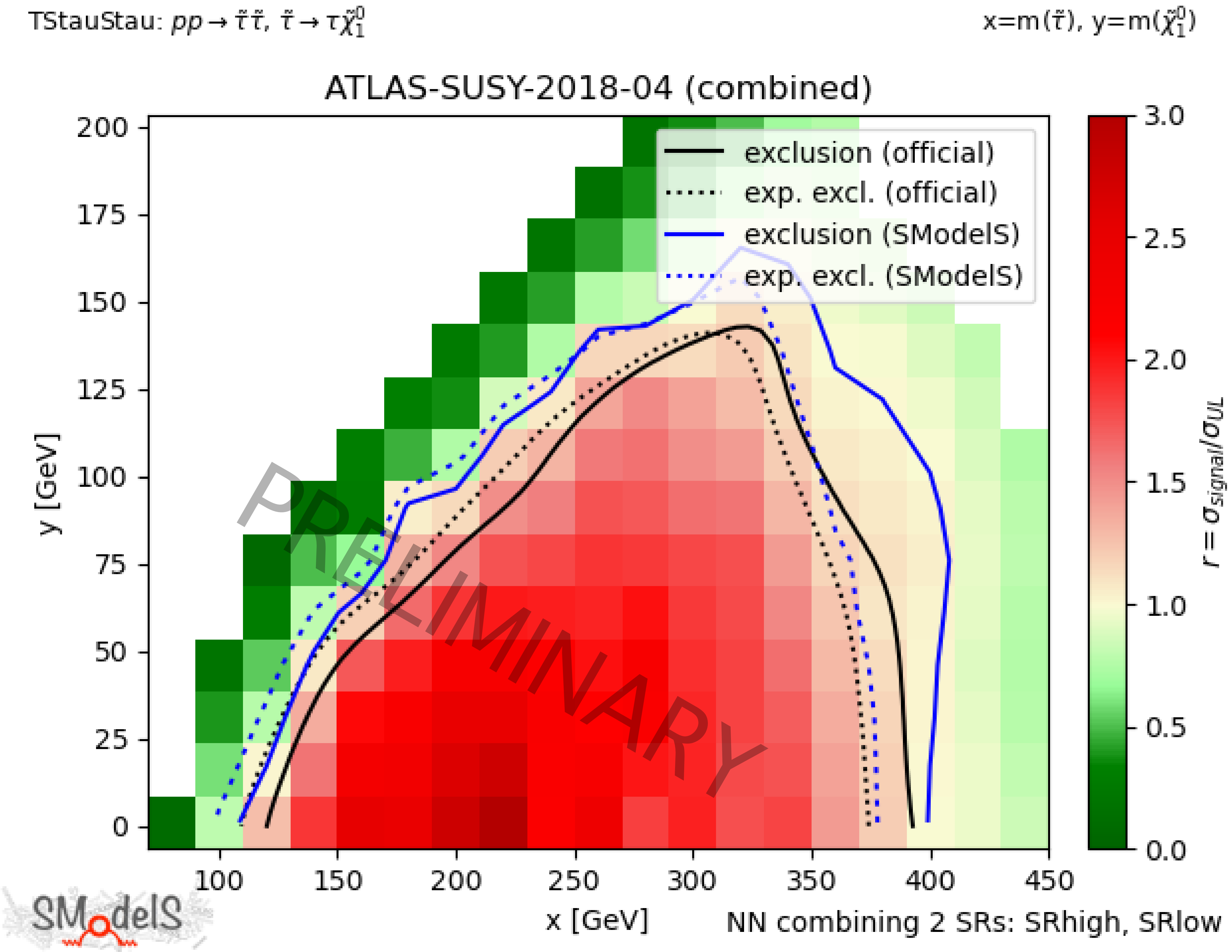
$$\Delta^{\text{exp/obs}} = \ln \frac{L^{\text{exp/obs}}(\mu = 1)}{L^{\text{exp/obs}}(\mu = 0)}$$

[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



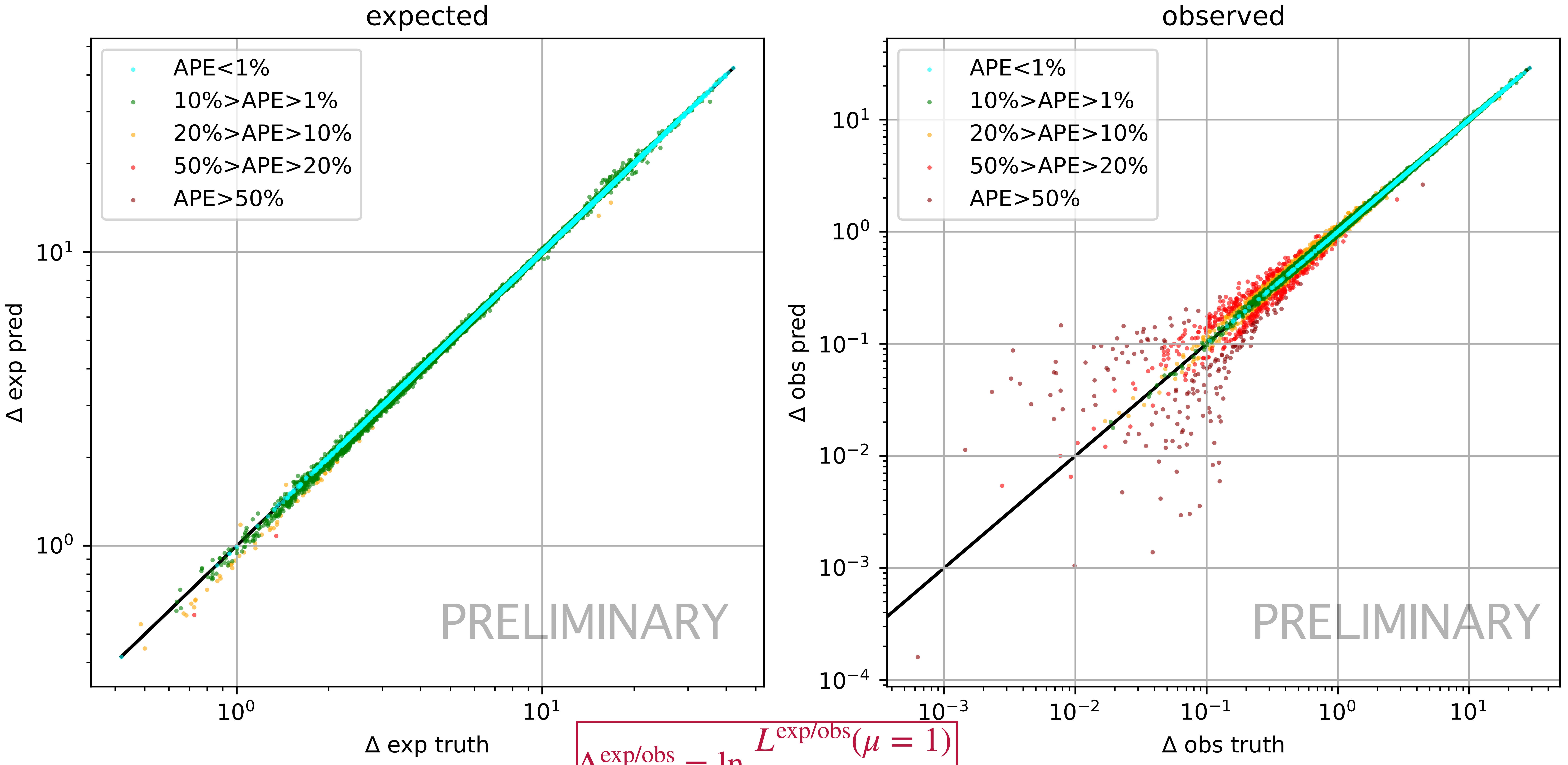
Full Likelihood Model



ML SURROGATE

[arXiv:1911.06660]

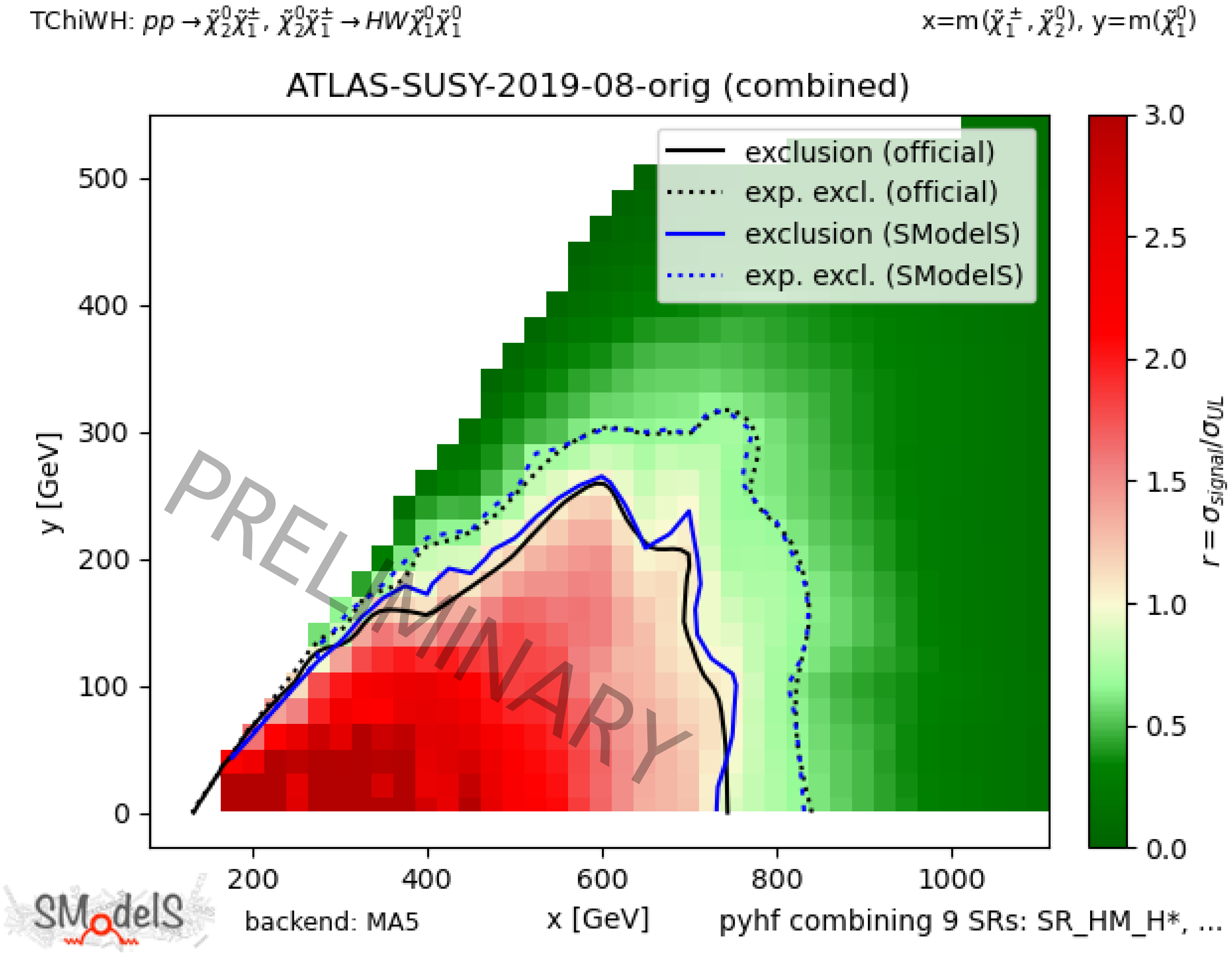
Search for direct production of electroweakinos in final states with one lepton, missing transverse momentum and a Higgs boson decaying into two b-jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector



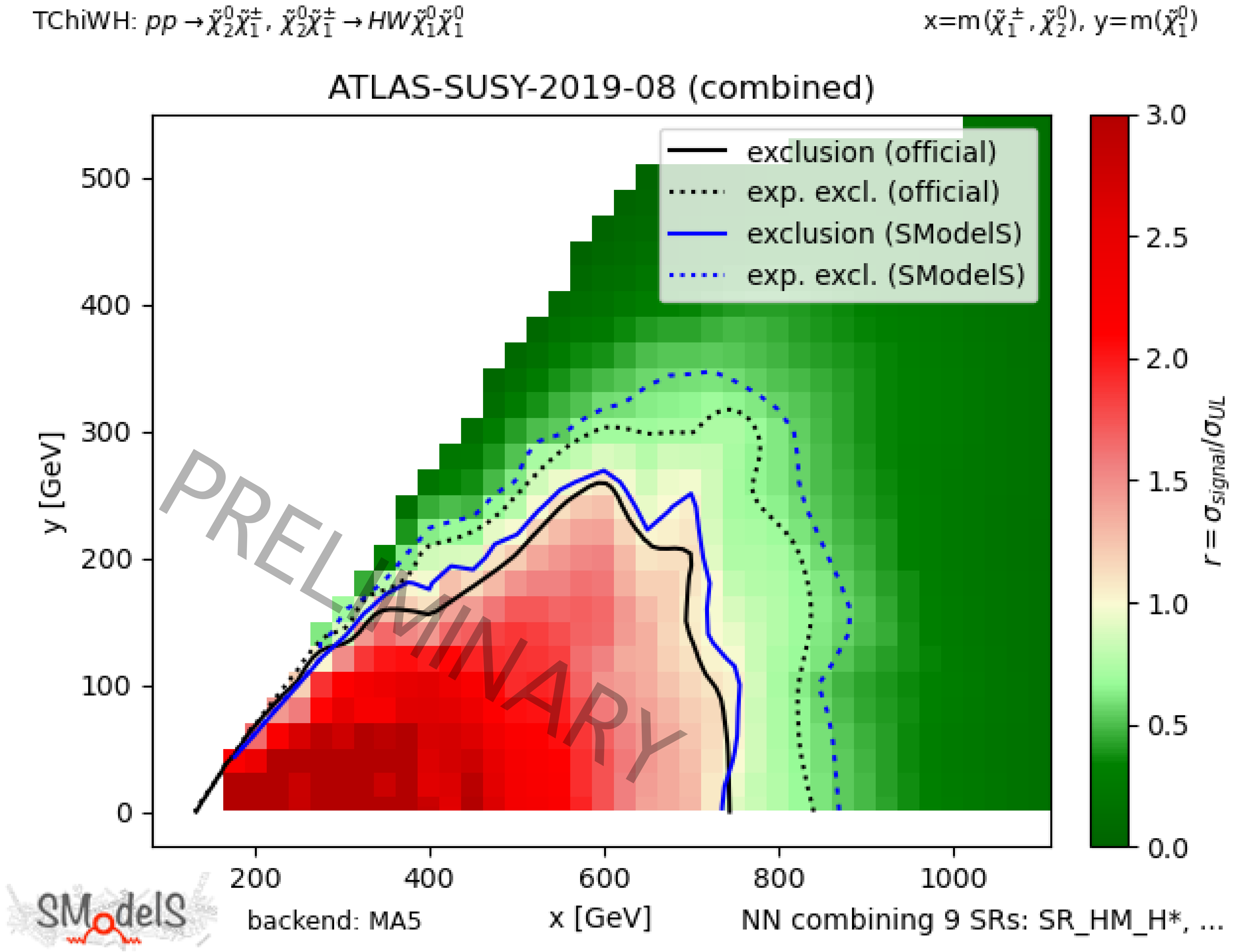
$$\Delta^{\text{exp/obs}} = \ln \frac{L^{\text{exp/obs}}(\mu = 1)}{L^{\text{exp/obs}}(\mu = 0)}$$

[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 $\sqrt{s} = 13$ TeV with the ATLAS detector



Full Likelihood Model



ML SURROGATE

[arXiv:1909.09226]

Progress summary and outlook

- **Task I — data generation**
  - MCMC sampling
  - positive and negative signal
  - fluctuations in CRs
  - parallelization
- **Task II — optimizing and training neural networks**
  - automatic hyperparameter optimization
  - training
  - exporting results to ONNX model with metadata
- **Task III — validation**
  - comparing predictions with truth values
  - reproduction of official limits with SmodelS
- **Task IV — publish models**
  - providing a complete data base with all published models
  - ensuring FAIRness
  - maintaining and keeping updated

Thank you for attention!

rafal.maselek@lpsc.in2p3.fr

Dolina Chochołowska, Poland
photo by Piotr Kałuża