Machine Learning tools in the search for leptoquarks in final states with taus, b-jets and MET at the LHC

In collaboration with E. Arganda, D. Díaz, R. M. Sandá Seoane, A. Szynkman
2306.xxxxx

Instituto de Física Teórica UAM-CSIC

Universidad Autónoma de Madrid

Andres Daniel Perez
Instituto de Física Teórica UAM-CSIC

10/05/2023

Outline

Signal and background

Scalar-leptoquarks

Selection cuts and comparison with an ATLAS collaboration analysis

From Machine Learning to significances

How to link ML classifiers with standard statistical tests

Results: expected significance

95% C.L. expected exclusion contours for m(LQ) vs BR(LQ$\rightarrow q\ell$)

Signal and Backgrounds

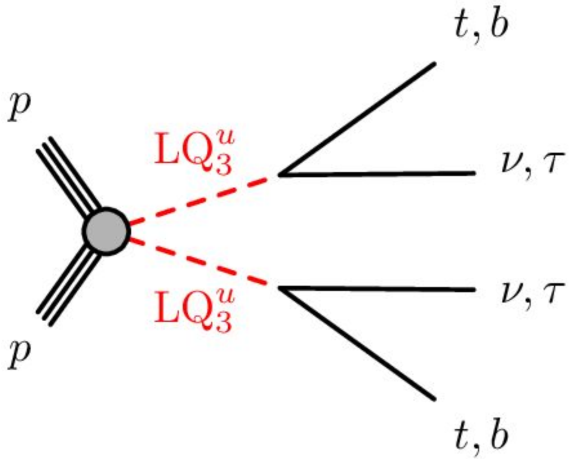
Signal: pair production of scalar-leptoquarks

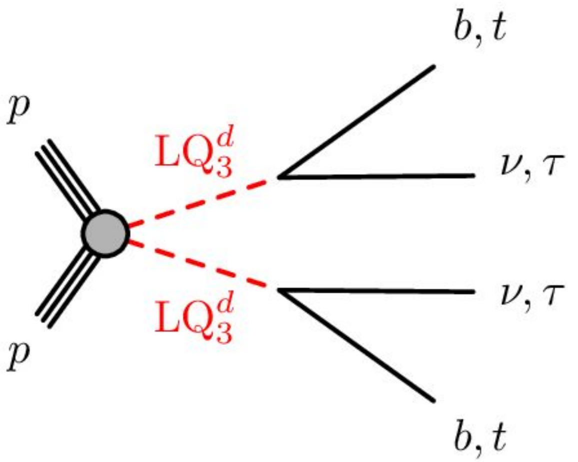
Leptoquarks are motivated by GUT (also used to explain B-anomalies)

Only decays into third-generation leptons and quarks (minimal Buchmüller–Rückl–Wyler model)

Q(LQ^u)=+⅔e

LQ^u $\rightarrow t\nu_\tau$ / $b\tau$





Q(LQ^d)=-⅔e

LQ^u $\rightarrow b\nu_\tau$ / $t\tau$

2 parameters: m(LQ$^{u/d}$), leptoquark mass and

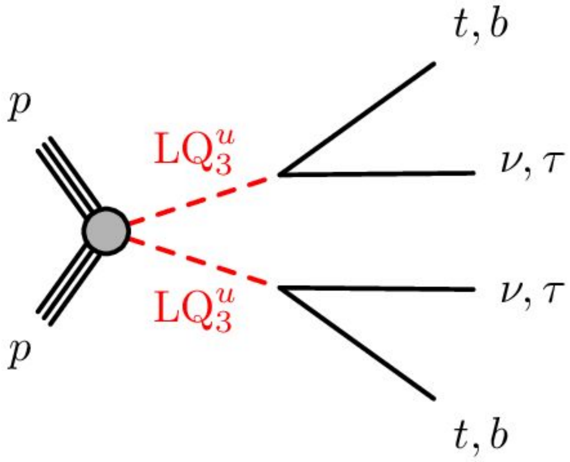
BR(LQ$^{u/d} \rightarrow q\ell$) , the branching fraction into a quark and a charged lepton.

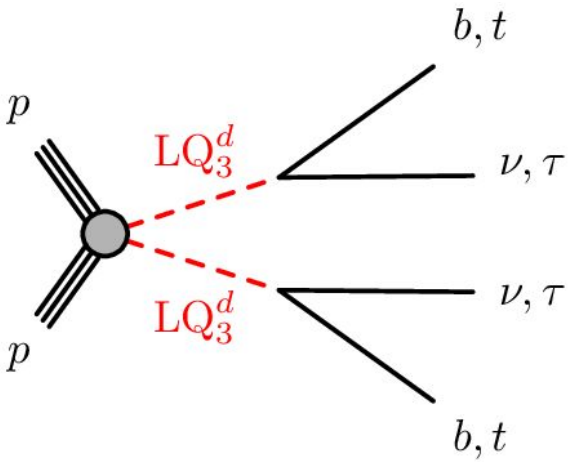
BR(LQ$^{u/d} \rightarrow q$v) = 1 - BR(LQ$^{u/d} \rightarrow q\ell$)

Signal: pair production of scalar-leptoquarks

For a $BR(\mathrm{LQ}^{u/d} \rightarrow q\ell) \sim 0.5$ most of the decays of the pair of third-generation leptoquarks yield a final state with one tau lepton, two b-jets and large MET from the tau neutrino.

Q(LQ^{u})=+⅔e
LQ^{u} $\rightarrow t\nu_\tau$ / $b\tau$





Q(LQ^{d})=-⅔e
LQ^{u} $\rightarrow b\nu_\tau$ / $t\tau$

Events with final states with:

1 hadronically decaying tau leptons (p_T>20GeV, $|\eta|$<2.5),
at least 2 b-tagged jets (p_T>20GeV, $|\eta|$<2.5),
large missing transverse momentum (MET>280 GeV)
no light leptons (e/μ)

Background: SM

ATLAS analysis, Phys. Rev. D 104, (2021) 112005, arXiv:2108.07665

	Single-tau SR (binned in $p_T(\tau)$)		
	[50, 100] GeV	[100, 200] GeV	> 200 GeV
Observed	8	6	2
Total bkg.	10.1 ±1.8	5.1 ±1.1	2.05 ±0.64
$t\bar{t}$ (2 real τ)	—	—	—
$t\bar{t}$ (1 real τ)	4.8 ±1.2	2.69 ±0.88	0.64 ±0.29
$t\bar{t}$-fake	2.83 ±0.87	0.66 ±0.17	0.185 ±0.072
Single top	$0.85^{+0.86}_{-0.85}$	0.54 ±0.54	0.57 ±0.56
W + jets	0.34 ±0.12	0.64 ±0.24	0.37 ±0.12
Z + jets	0.275 ±0.081	0.043 ±0.022	0.123 ±0.048
Multiboson	0.163 ±0.037	0.111 ±0.030	$0.030^{+0.032}_{-0.030}$
$t\bar{t} + V$	0.65 ±0.16	0.31 ±0.12	0.092 ±0.035
$t\bar{t} + H$	0.10 ±0.10	$0.060^{+0.061}_{-0.060}$	$0.028^{+0.029}_{-0.028}$
Other top	0.096 ±0.074	0.091 ±0.049	0.0120 ±0.0084

Different selection cuts

ATLAS analysis, Phys. Rev. D 104, (2021) 112005, arXiv:2108.07665

Events with final states with: 1 hadronically decaying tau leptons (p_T>20GeV, |η|<2.5),
at least 2 b-tagged jets (p_T>20GeV, |η|<2.5),
large missing transverse momentum (E_T^{miss}>280 GeV)
no light leptons (e/μ)

ATLAS signal enriched region:

$$\sum m_{\mathrm{T}}(b_{1,2}) = m_{\mathrm{T}}(b_1) + m_{\mathrm{T}}(b_2) > 700\ \mathrm{GeV}$$

$$m_{\mathrm{T}}(\tau) > 150\ \mathrm{GeV}$$

$$s_{\mathrm{T}} = p_{\mathrm{T}}(\tau) + p_{\mathrm{T}}(\mathrm{jet}_1) + p_{\mathrm{T}}(\mathrm{jet}_2) > 600\ \mathrm{GeV}$$

where $m_{\mathrm{T}}^2(\mathbf{p}_{\mathrm{T}}, \mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}) = 2\, p_{\mathrm{T}}\, E_{\mathrm{T}}^{\mathrm{miss}}\, (1 - \cos\Delta\phi(\mathbf{p}_{\mathrm{T}}, \mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}))$ is the stransverse mass

Different selection cuts

ATLAS analysis, Phys. Rev. D 104, (2021) 112005, arXiv:2108.07665

Events with final states with: 1 hadronically decaying tau leptons (p_T>20GeV, |η|<2.5),
at least 2 b-tagged jets (p_T>20GeV, |η|<2.5),
large missing transverse momentum (E_T^{miss}>280 GeV)
no light leptons (e/μ)

ATLAS signal enriched region:

$$\sum m_T(b_{1,2}) = m_T(b_1) + m_T(b_2) > 700 \text{ GeV}$$

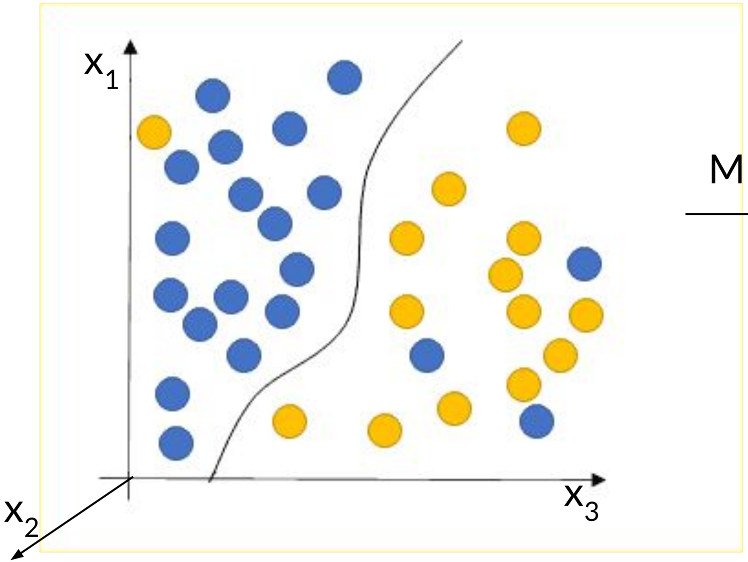
$$m_T(\tau) > 150 \text{ GeV}$$

$$s_T = p_T(\tau) + p_T(\text{jet}_1) + p_T(\text{jet}_2) > 600 \text{ GeV}$$

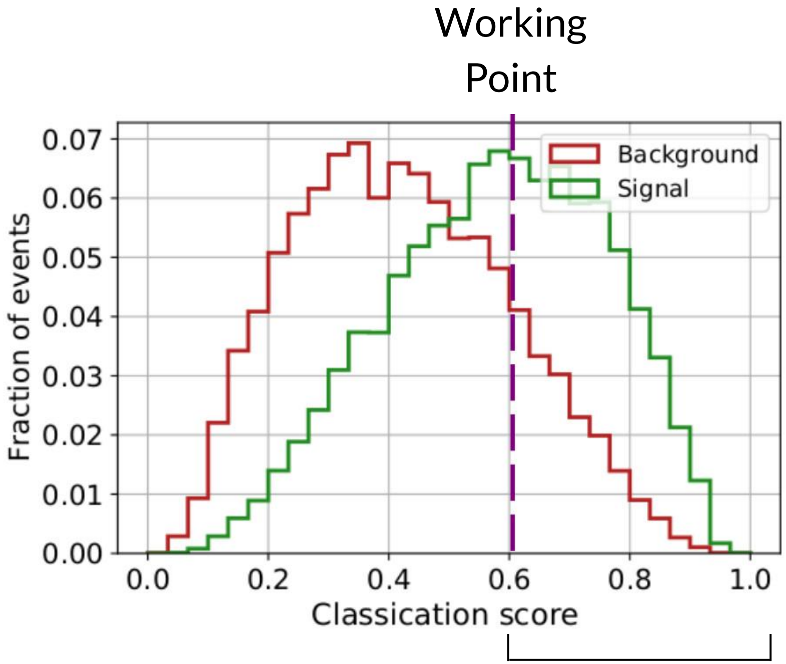
where $m_T^2(\mathbf{p}_T, \mathbf{E}_T^{miss}) = 2\, p_T\, E_T^{miss}\, (1 - \cos\Delta\phi(\mathbf{p}_T, \mathbf{E}_T^{miss}))$ is the stransverse mass

From ML to significances

Expected significance



ML →



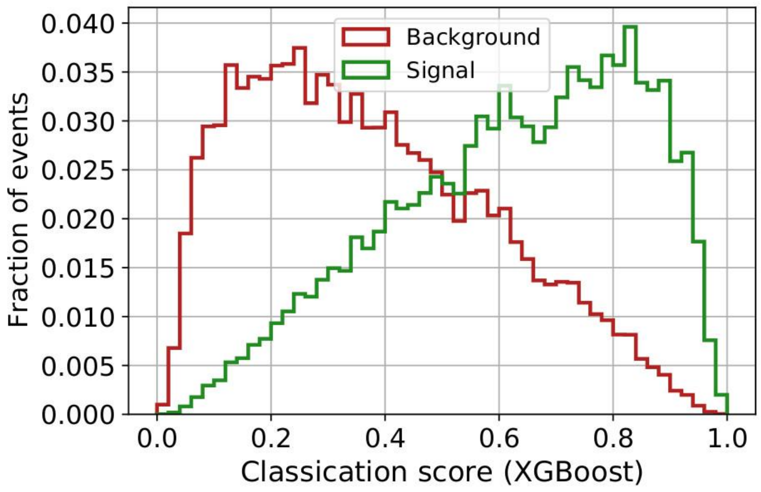
Defining a Working Point ~ design observable, define a working region

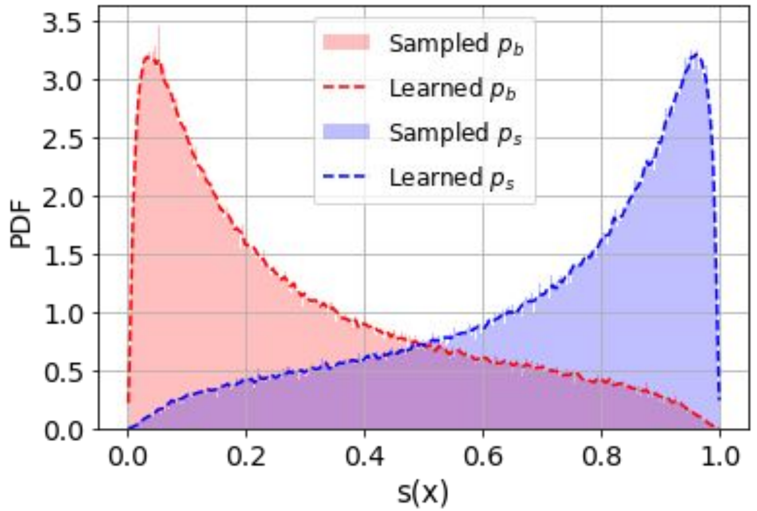
In that signal enriched region → significance ~ $\dfrac{S}{\sqrt{B}}$

(we discard events outside of it)

Expected significance

We used the ***full 1D*** ML classifier output ***o(x)*** with the standard statistical tests (without defining a working point) to compute the significance





o(x̄) Binned Likelihood method

$$\text{med}\,[Z_0^{\text{binned}}|1] = \left[2\sum_{d=1}^{D}\left((S_d + B_d)\,\text{Ln}\left(1+\frac{S_d}{B_d}\right) - S_d\right)\right]^{1/2}$$

o(x̄) Unbinned method

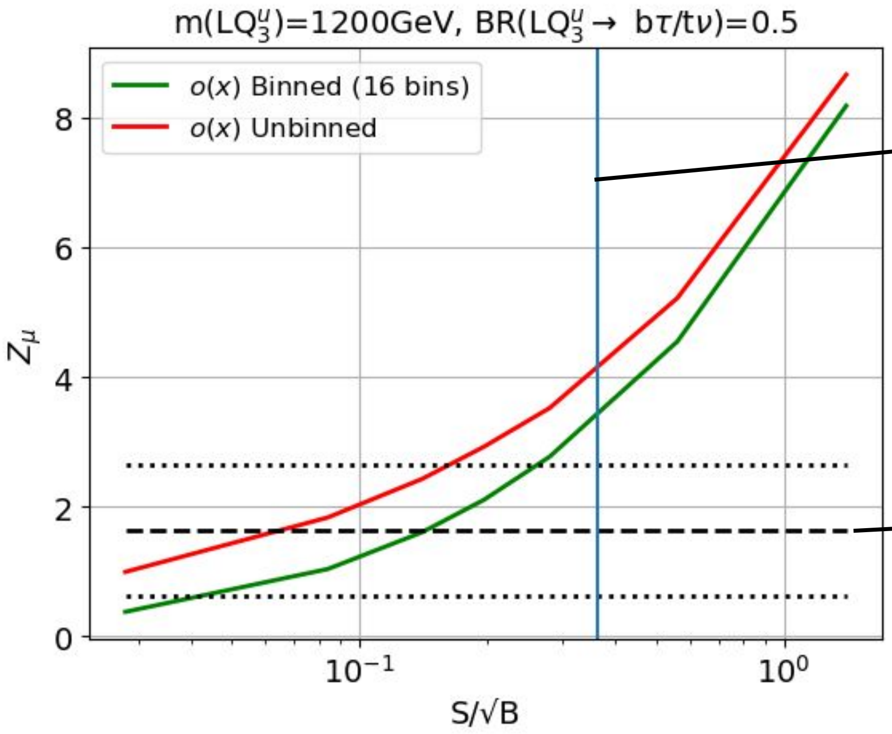
Use KDE to estimate the B and S PDFs.

↳Calculate Z building pseudo-experiments

Results

Results

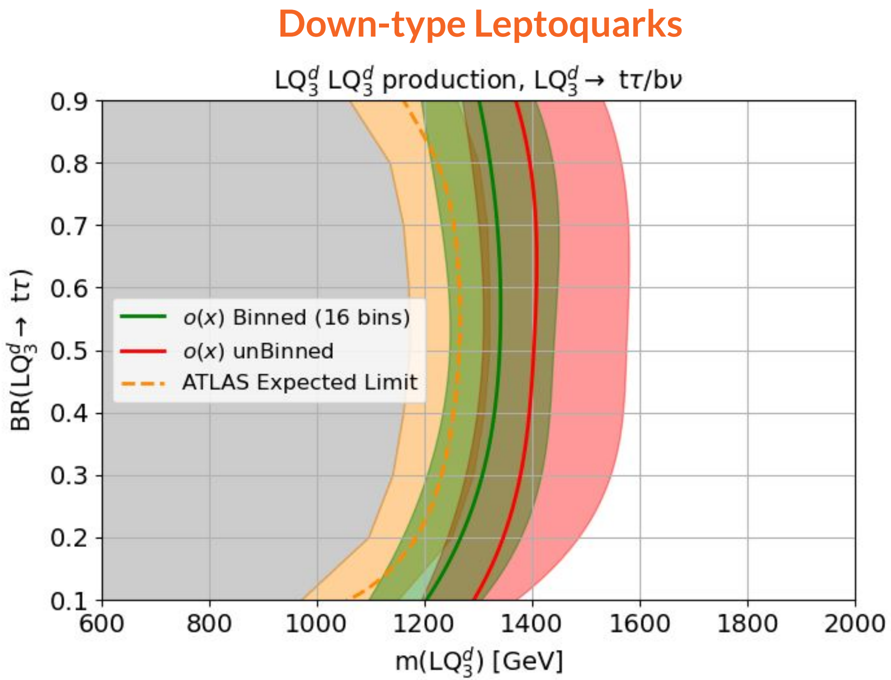
Example m(LQ^u) = 1200 GeV, and BR(LQ^u $\rightarrow q\ell$)=0.5

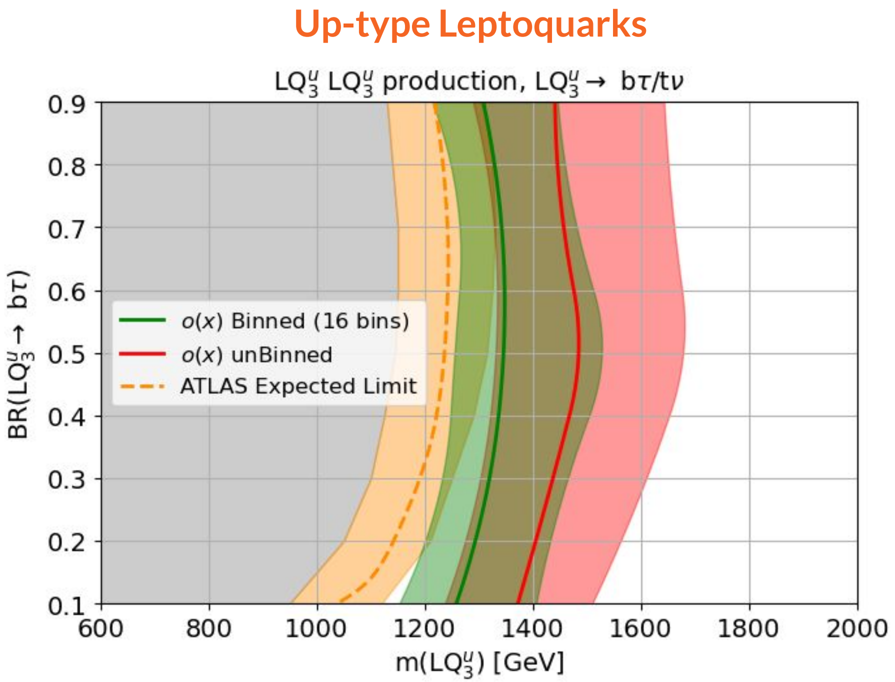


Expected signal-to-background ratio for this parameter point.

Z > 1.64 → excluded at 95%C.L.

Results





For both types of scalar leptoquarks, the expected exclusion contours extend to masses around 1.25 TeV (ATLAS), 1.35 TeV (Binned), 1.45 TeV (Unbinned) at the 95% confidence level for intermediate values of the branching fraction $BR(LQ^{u/d} \rightarrow q\ell)$

Conclusions

- Search for new phenomena in final states with hadronically decaying tau leptons, b-jets and large missing transverse momentum
  $\rightarrow$ third-generation scalar leptoquarks (motivated by GUT).

- As a ***proof of concept*** we used ML algorithms with simple selection cuts and compared them to an ATLAS analysis.

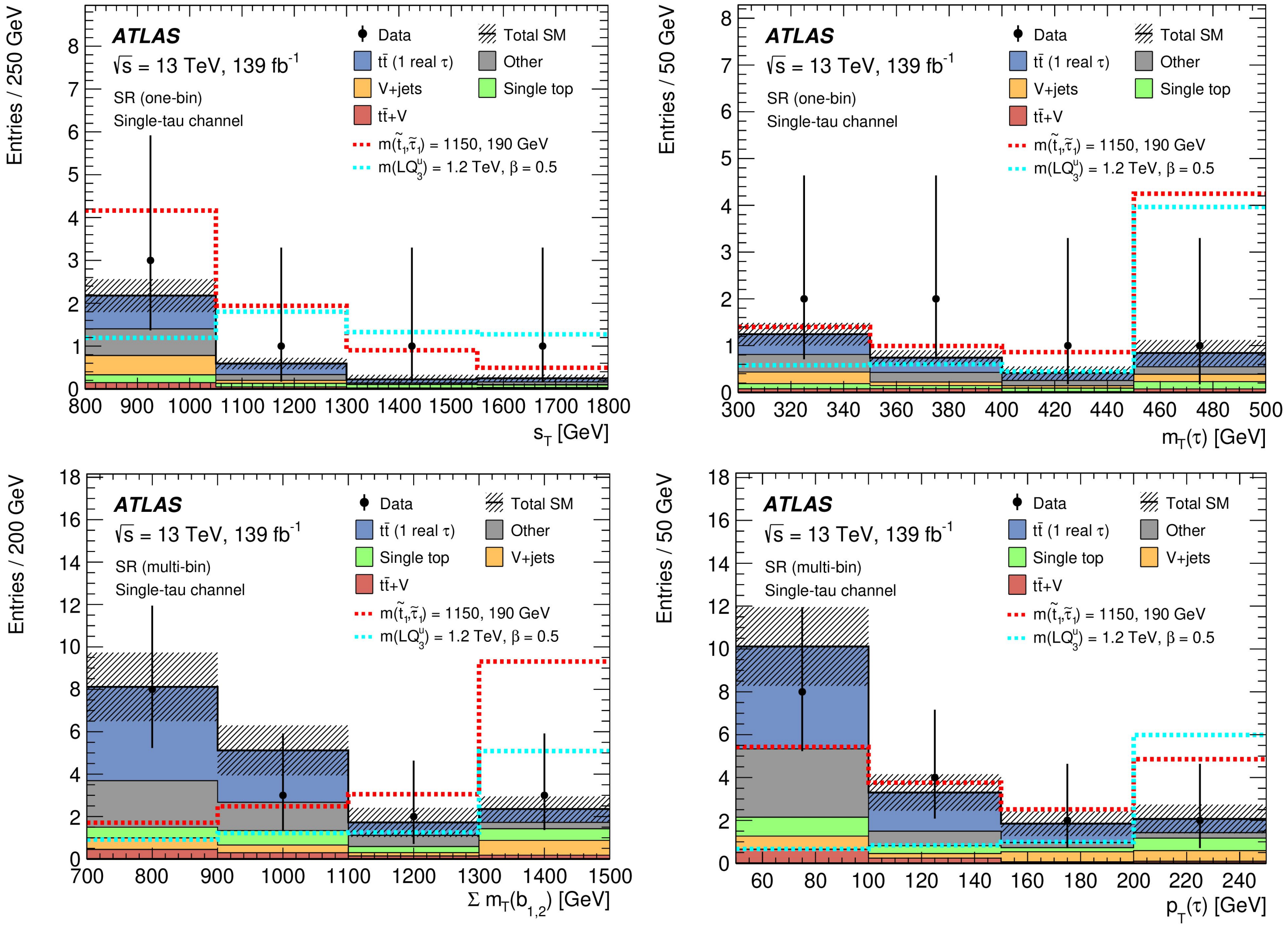
- We use the full classifier output (no working points) to estimate the significances.

- For both types of scalar leptoquarks, expected exclusion contours up to masses
  1.25 TeV (ATLAS)
  1.35 TeV (Binned)
  1.45 TeV (Unbinned)
  at the 95% confidence level for intermediate values of the branching fraction *BR*($LQ^{u/d} \rightarrow q\ell$)

Thank you!

Back up

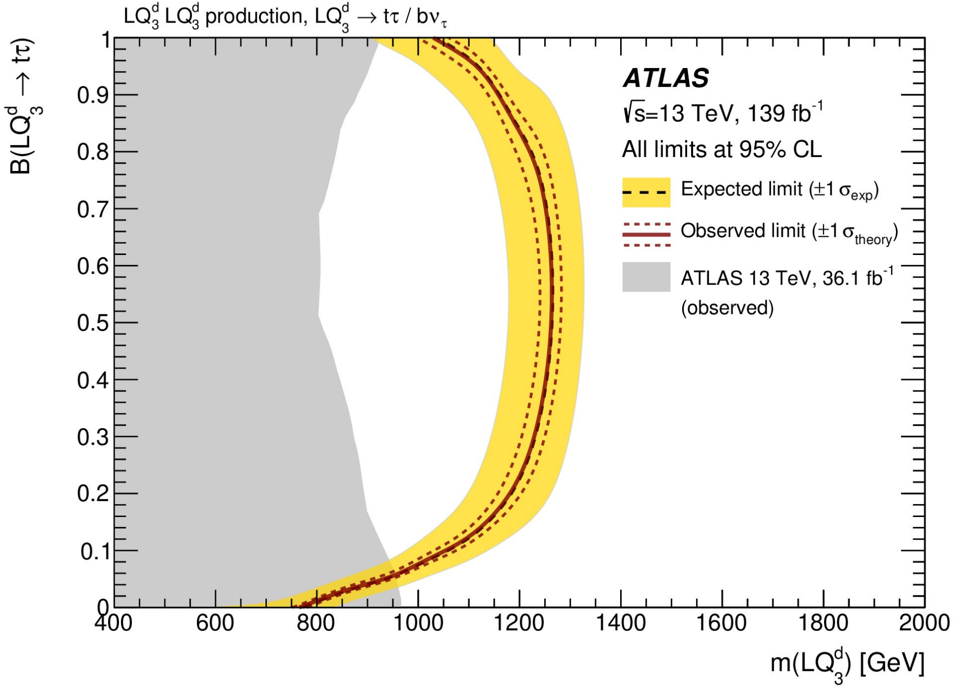
ATLAS



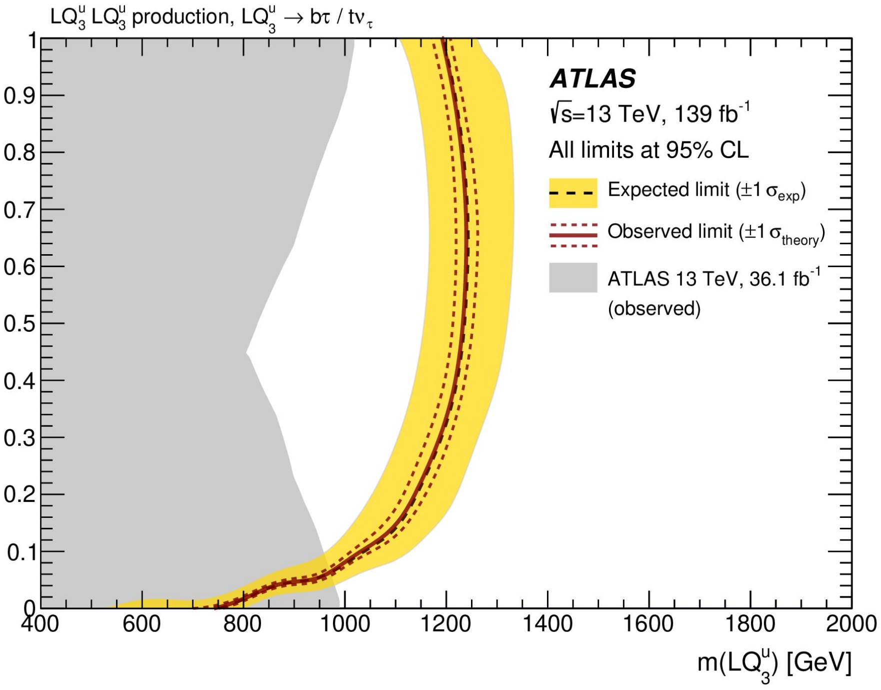
ATLAS
$\sqrt{s}$ = 13 TeV, 139 fb^{-1}
SR (one-bin)
Single-tau channel
Data
Total SM
$t\bar{t}$ (1 real τ)
Other
V+jets
Single top
$t\bar{t}$+V
m($\tilde{t}_1$,$\tilde{\tau}_1$) = 1150, 190 GeV
m(LQ_3^u) = 1.2 TeV, β = 0.5
Entries / 250 GeV
s_T [GeV]
Entries / 50 GeV
$m_T(\tau)$ [GeV]
SR (multi-bin)
Entries / 200 GeV
$\Sigma m_T(b_{1,2})$ [GeV]
$p_T(\tau)$ [GeV]

ATLAS

Down-type Leptoquarks



Up-type Leptoquarks



Leptoquark model

Buchmüller–Rückl–Wyler model, Phys. Lett. B 191 (1987) 442

We start from an effective lagrangian with the most general dimensionless, SU(3)×SU(2)×U(1) invariant couplings of scalar and vector leptoquarks satisfying baryon and lepton number conservation:

$$\mathscr{L} = \mathscr{L}_{F=2} + \mathscr{L}_{F=0}, \tag{1a}$$

$$\begin{aligned}\mathscr{L}_{F=2} = {} & (g_{1L}\bar{q}^{c}_{L} i\tau_2 \ell_L + g_{1R}\bar{u}^{c}_{R} e_R)S_1 \\ & + \tilde{g}_{1R}\bar{d}^{c}_{R} e_R \tilde{S}_1 + g_{3L}\bar{q}^{c}_{L} i\tau_2 \boldsymbol{\tau}\ell_L S_3 \\ & + (g_{2L}\bar{d}^{c}_{R}\gamma^{\mu}\ell_L + g_{2R}\bar{q}^{c}_{L}\gamma^{\mu} e_R)V_{2\mu} \\ & + \tilde{g}_{2L}\bar{u}^{c}_{R}\gamma^{\mu}\ell_L \tilde{V}_{2\mu} + \text{c.c.},\end{aligned} \tag{1b}$$

$$\begin{aligned}\mathscr{L}_{F=0} = {} & (h_{2L}\bar{u}_R \ell_L + h_{2R}\bar{q}_L i\tau_2 e_R)R_2 + \tilde{h}_{2L}\bar{d}_R \ell_L \tilde{R}_2 \\ & + (h_{1L}\bar{q}_L\gamma^{\mu}\ell_L + h_{1R}\bar{d}_R\gamma^{\mu} e_R)U_{1\mu} \\ & + \tilde{h}_{1R}\bar{u}_R\gamma^{\mu} e_R \tilde{U}_{1\mu} + h_{3L}\bar{q}_L \boldsymbol{\tau}\gamma^{\mu}\ell_L U_{3\mu} + \text{c.c.}\end{aligned} \tag{1c}$$

The Yukawa-type interaction of the leptoquarks with the quark-lepton pair are determined by two parameters:

- a common coupling strength λ (=g, h) and,
- an additional parameter β, with the coupling to a quark and a charged lepton given by $(\sqrt{\beta})\,\lambda$, and the coupling to a quark and a neutrino by $[\sqrt{(1-\beta)}]\,\lambda$.

The branching fraction *BR*(LQ$^{u/d} \to q\ell$) ≃ β, except for kinematic effects arising from the mass differences of the decay products.

λ set to 0.3 close to the electromagnetic coupling $e=\sqrt{(4\pi\alpha)}$, resulting in a LQ$^{u/d}$ width equal to ~0.2% of its mass

Leptoquark model

Buchmüller–Rückl–Wyler model, Phys. Lett. B 191 (1987) 442

Table 1
Quantum numbers of scalar and vector leptoquarks with $SU(3) \times SU(2) \times U(1)$ invariant couplings to quark–lepton pairs ($Y = Q_{em} - T_3$).

	Spin	$F = 3B + L$	$SU(3)_c$	$SU(2)_W$	$U(1)_Y$
S_1	0	-2	3^*	1	$\frac{1}{3}$
$\tilde{S}_1$	0	-2	3^*	1	$\frac{4}{3}$
S_3	0	-2	3^*	3	$\frac{1}{3}$
V_2	1	-2	3^*	2	$\frac{5}{6}$
$\tilde{V}_2$	1	-2	3^*	2	$-\frac{1}{6}$
R_2	0	0	3	2	$\frac{7}{6}$
$\tilde{R}_2$	0	0	3	2	$\frac{1}{6}$
U_1	1	0	3	1	$\frac{2}{3}$
$\tilde{U}_1$	1	0	3	1	$\frac{5}{3}$
U_3	1	0	3	3	$\frac{2}{3}$

Leptoquark model

Buchmüller–Rückl–Wyler model, Phys. Lett. B 191 (1987) 442

Table2

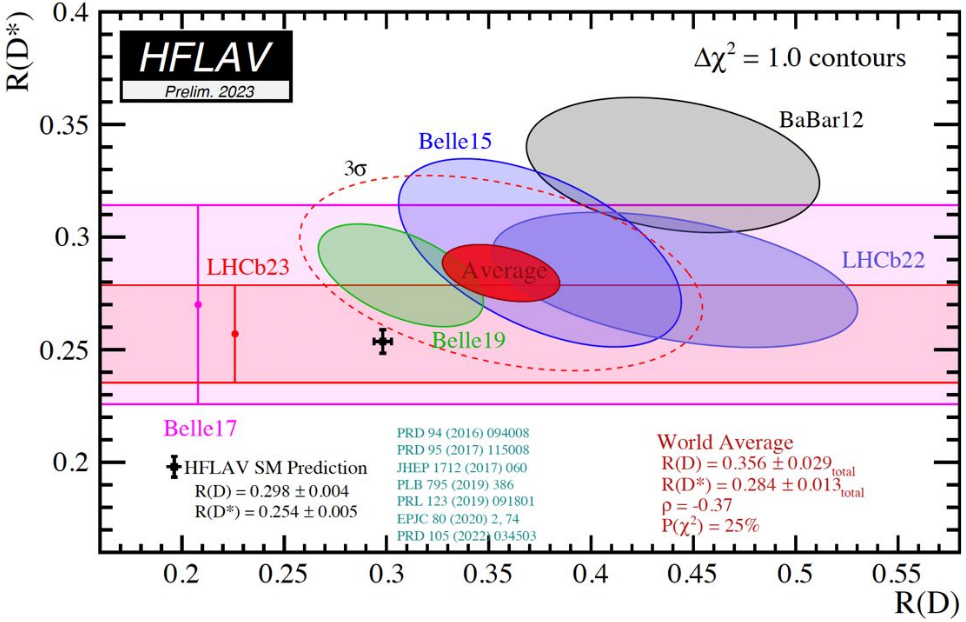
Couplings of scalar and vector leptoquarks to quark–lepton pairs. The subscripts L,R of the couplings refer to the lepton chirality.

LQ^d

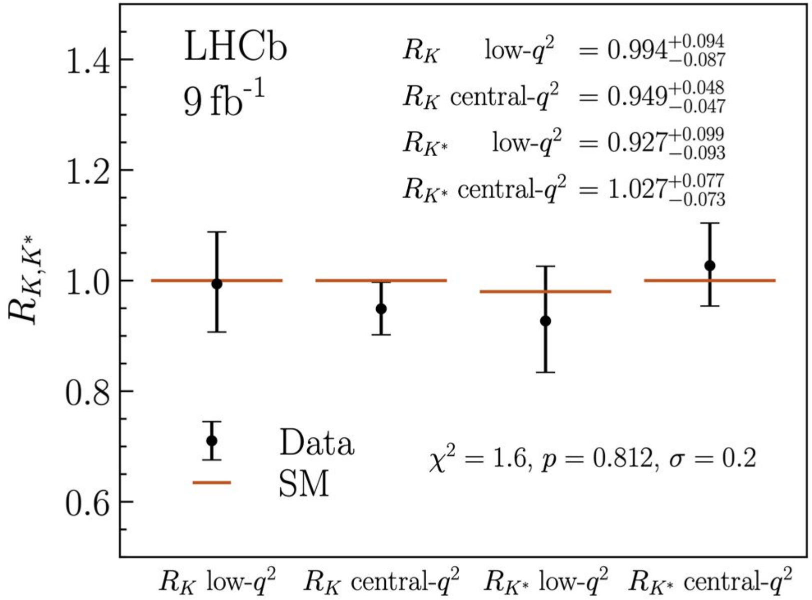
LQ^u

Channel	$F=-2$, scalars			$F=-2$, vectors	
	S_1	$\tilde{S}_1$	S_3	V_2	$\tilde{V}_2$
$e^-_{L,R}u$	$g_{1L,R}$	–	$-g_{3L}$	g_{2R}	$\tilde{g}_{2L}$
$\nu_L d$	$-g_{1L}$	–	$-g_{3L}$	g_{2L}	–
$e^-_{L,R}d$	–	$\tilde{g}_{1R}$	$-\sqrt{2}\,g_{3L}$	$g_{2L,R}$	–
$\nu_L u$	–	–	$\sqrt{2}g_{3L}$	–	$\tilde{g}_{2L}$
Channel	**$F=0$, vectors**			**$F=0$, scalars**	
	U_1	$\tilde{U}_1$	U_3	R_2	$\tilde{R}_2$
$e^-_{L,R}\bar{d}$	$h_{1L,R}$	–	$-h_{3L}$	$-h_{2R}$	$\tilde{h}_{2L}$
$\nu_L\bar{u}$	h_{1L}	–	h_{3L}	h_{2L}	–
$e^-_{L,R}\bar{u}$	–	$\tilde{h}_{1R}$	$\sqrt{2}h_{3L}$	$h_{2L,R}$	–
$\nu_L\bar{d}$	–	–	$\sqrt{2}h_{3L}$	–	$\tilde{h}_{2L}$

B-anomalies



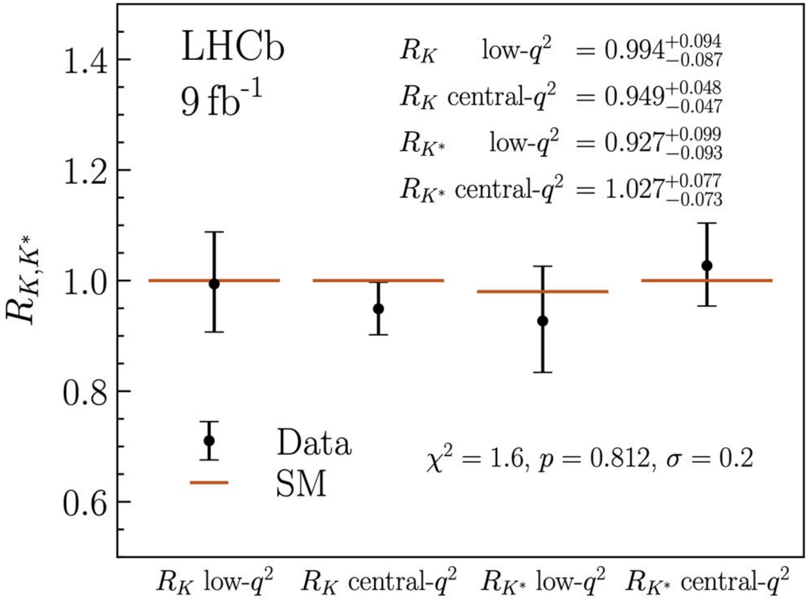
Status of the charged-current LFU ratios R(D) and R(D*).

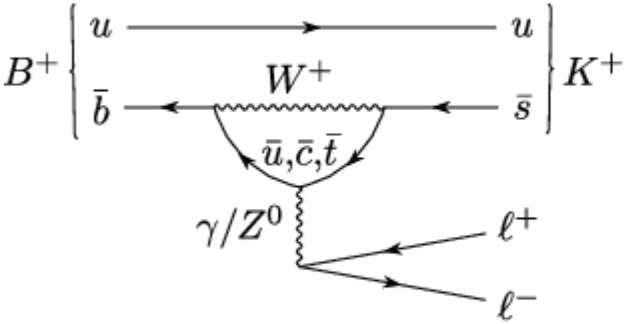


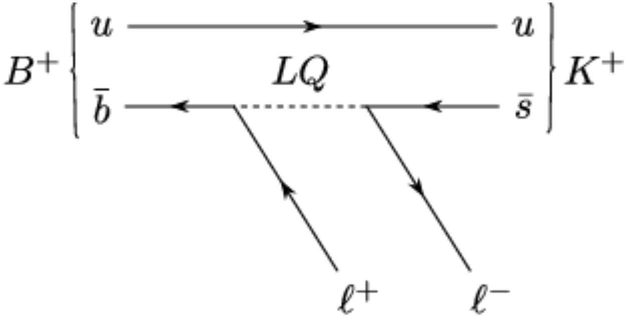
Measurements of the e/μ LFU ratios reported by LHCb in December 2022.

Fully compatible with the SM

B-anomalies



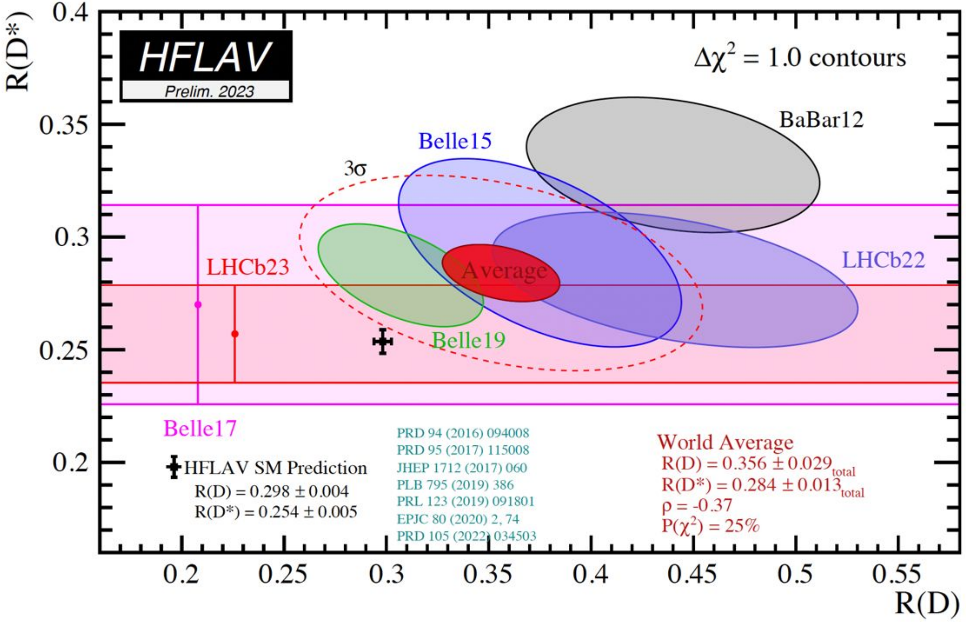




Measurements of the e/μ LFU ratios reported by LHCb in December 2022.

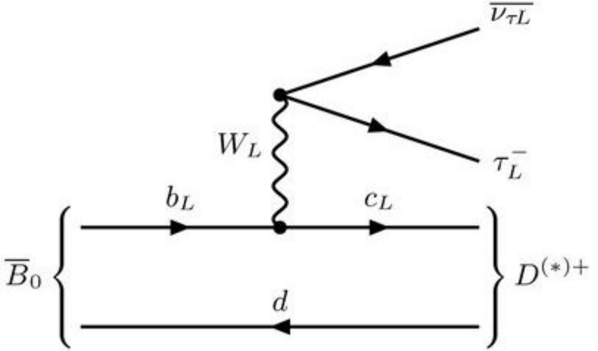
$$R_{K^{(*)}}^{[q_1^2,q_2^2]} = \frac{\mathrm{Br}'\left(B \to K^{(*)}\mu\mu\right)}{\mathrm{Br}'\left(B \to K^{(*)}ee\right)}$$

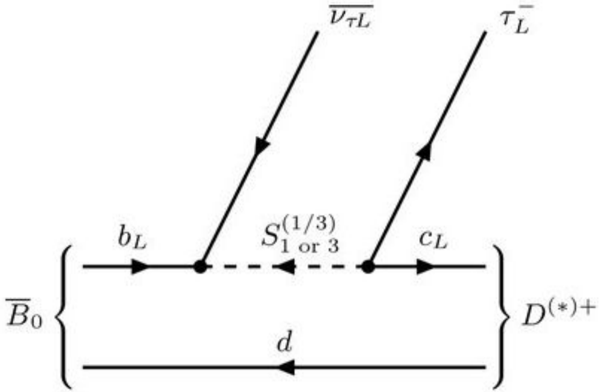
B-anomalies



Status of the charged-current LFU ratios R(D) and R(D*).

$$R_{D^{(*)}} = \left.\frac{\mathrm{Br}\left(B \to D^{(*)}\tau\bar{\nu}\right)}{\mathrm{Br}\left(B \to D^{(*)}\ell\bar{\nu}\right)}\right|_{\ell\in\{e,\mu\}}$$





Supervised Learning

Input

Labeled data $D=\{(\bar{x}_1,t_1), \dots ,(\bar{x}_n,t_n)\}$

$\{\bar{x}_i\}$: features, e.g. p_T, $\Delta\varphi_{12}$, E_t^{miss}

$\{t_i\}$: target, e.g. for classification:

1 for *signal*

0 for *background*

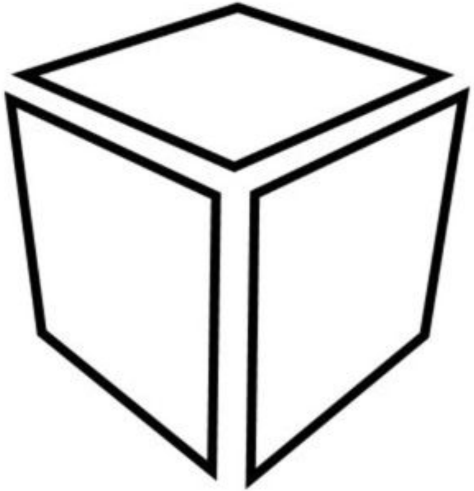
Output

The algorithm finds a mapping:

ideally $o(\bar{x}_i)=t_i$

for classification: $o(\bar{x}_i) \in [0,1]$

Train dataset $\bar{x}$



ML classifier

ML output $o(\bar{x})$

Always 1D

Supervised Learning

New data

Data sample that we do not know if it is Signal or Background

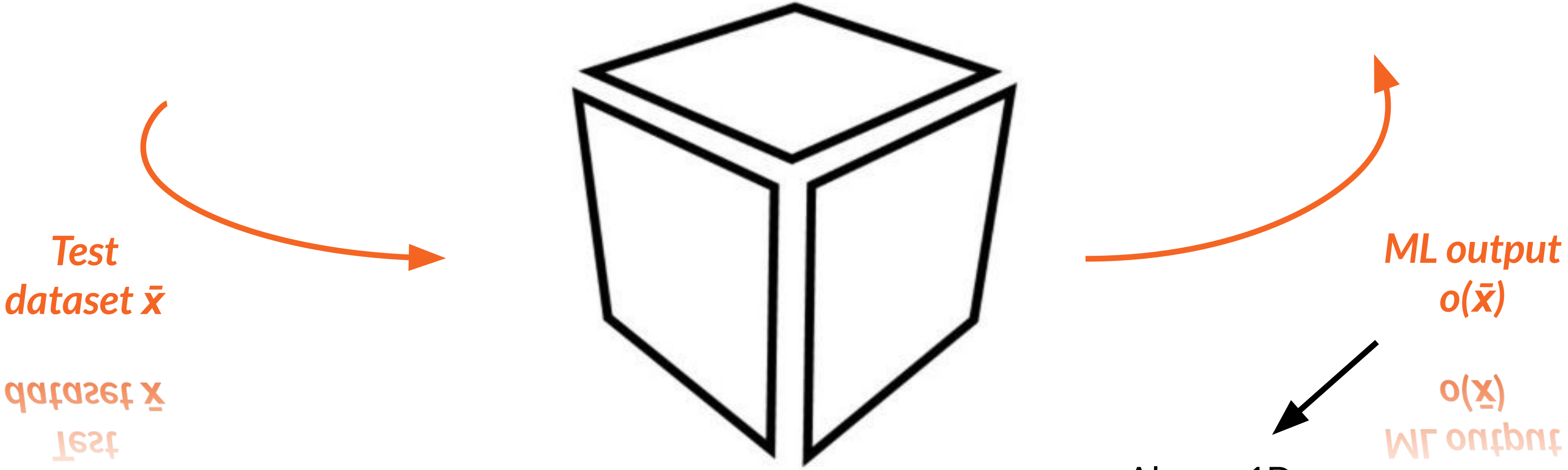
Prediction

To assign a label a threshold or working point (WP) is needed

if **o(x) < WP** label $\rightarrow$ '0' $\rightarrow$ ***B***

if **o(x) > WP** label $\rightarrow$ '1' $\rightarrow$ ***S***

S or B
label ??



Machine Learned-Likelihood (MLL)

Likelihood to define the statistical model for N independent measurements, with a set of observables x_i

$$\mathcal{L}(\mu, s, b) = p(N, \{x_i, i = 1, \ldots, N\} | \mu, s, b) = \underbrace{\mathrm{Poiss}\big(N | \mu S + B\big)}_{\substack{\sim \text{ global info} \\ \text{ensemble factor}}} \underbrace{\prod_{i=1}^{N} p(x_i | \mu, s, b)}_{\substack{\sim \text{ local info} \\ \text{event-by-event}}}$$

with:

- S the expected total signal yield
- B the expected total background yield
- $p(x | \mu, s, b) = \frac{B}{\mu S + B} p_b(x) + \frac{\mu S}{\mu S + B} p_s(x)$

$$p_s(x) = p(x|s)$$
$$p_b(x) = p(x|b)$$

- μ the signal strength defines the hypothesis we are testing for:

  background-only hypothesis → μ = 0
  background-plus-signal hypothesis → μ = 1

Machine Learned-Likelihood (MLL)

The relevant test statistic for **discovery** limits (very similar for exclusion):

discovery corresponds to studying background-only hypothesis μ = 0

$$q_0 = \begin{cases} -2\,\mathrm{Ln}\,\dfrac{\mathcal{L}(0,s,b)}{\mathcal{L}(\hat{\mu},s,b)} & \text{if } \hat{\mu} \geq 0\,, \\ 0 & \text{if } \hat{\mu} < 0\,, \end{cases}$$

using the Likelihood

$$q_0 = \begin{cases} -2\hat{\mu}S + 2\sum_{i=1}^{N} \mathrm{Ln}\left(1 + \dfrac{\hat{\mu}S}{B}\dfrac{p_s(x_i)}{p_b(x_i)}\right) & \text{if } \hat{\mu} \geq 0\,, \\ 0 & \text{if } \hat{\mu} < 0\,. \end{cases}$$

where $\hat{\mu}$ is the parameter that maximizes the likelihood

$$\sum_{i=1}^{N} \frac{p_s(x_i)}{\hat{\mu}S\,p_s(x_i)+B\,p_b(x_i)} = 1\,.$$

Machine Learned-Likelihood (MLL)

The relevant test statistic for **discovery** limits (very similar for exclusion):

discovery corresponds to studying background-only hypothesis $\mu = 0$

$$q_0 = \begin{cases} -2\,\mathrm{Ln}\,\dfrac{\mathcal{L}(0,s,b)}{\mathcal{L}(\hat{\mu},s,b)} & \text{if } \hat{\mu} \geq 0\,, \\ 0 & \text{if } \hat{\mu} < 0\,, \end{cases}$$

using the Likelihood

$$q_0 = \begin{cases} -2\hat{\mu}S + 2\sum_{i=1}^{N} \mathrm{Ln}\left(1 + \dfrac{\hat{\mu}S}{B}\dfrac{p_s(x_i)}{p_b(x_i)}\right) & \text{if } \hat{\mu} \geq 0\,, \\ 0 & \text{if } \hat{\mu} < 0\,. \end{cases}$$

where $\hat{\mu}$ is the parameter that maximizes the likelihood

$$\sum_{i=1}^{N} \frac{p_s(x_i)}{\hat{\mu}S\,p_s(x_i) + B\,p_b(x_i)} = 1\,.$$

We need

$$p_s(x) = p(x|s)$$
$$p_b(x) = p(x|b)$$

Machine Learned-Likelihood (MLL)

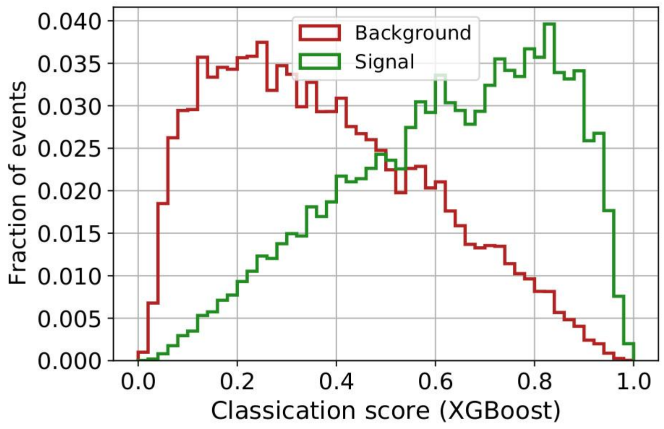
Replace the densities for the one-dimensional manifolds obtained with a machine-learning classifier.

The classification score that maximizes the binary cross-entropy approaches:

$$o(x) = \frac{p_s(x)}{p_s(x)+p_b(x)}$$

Dimensional reduction by dealing with *o(x)* instead of *x*

$$p_b(x) \to \tilde{p}_b(o(x))\,, \qquad \text{and} \qquad p_s(x) \to \tilde{p}_s(o(x))$$



where $\tilde{p}_{s,b}(o(x))$ are the distributions of *o(x)* for signal and background, obtained by evaluating the classifier on a set of pure signal or background events, respectively.

Machine Learned-Likelihood (MLL)

Then, the relevant test statistic for **discovery** limits

$$q_0 = \begin{cases} -2\hat{\mu}S + 2\sum_{i=1}^{N} \mathrm{Ln}\left(1 + \frac{\hat{\mu}S}{B}\frac{\tilde{p}_s(o(x_i))}{\tilde{p}_b(o(x_i))}\right) & \text{if } \hat{\mu} \geq 0\,, \\ 0 & \text{if } \hat{\mu} < 0\,, \end{cases}$$

with $\hat{\mu}$ the parameter that maximizes the likelihood

$$\sum_{i=1}^{N} \frac{\tilde{p}_s(o(x_i))}{\hat{\mu}S\,\tilde{p}_s(o(x_i)) + B\,\tilde{p}_b(o(x_i))} = 1$$

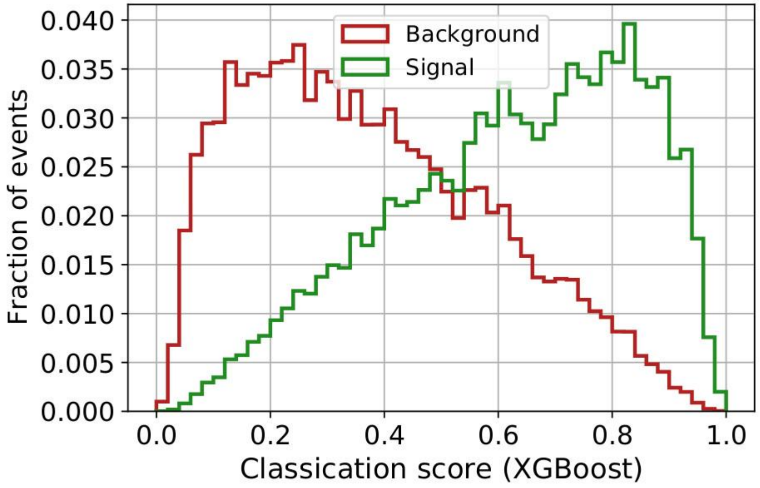
We can estimate numerically the q_0 distribution.
The median expected significance assuming signal-plus-background hypothesis (μ'=1) is

$$Z_0 \to \mathrm{med}\,[Z_0|1] = \sqrt{\mathrm{med}\,[q_0|1]}$$

Density estimation

We want to retrieve the density function from which the samples were generated

$$p_b(x) \rightarrow \tilde{p}_b(o(x))\,, \qquad \text{and} \qquad p_s(x) \rightarrow \tilde{p}_s(o(x))$$



The original space, x_i, can be high-dimensional but the classifier output ***o(x)* is always one-dimensional**

- To avoid binning, we use a non-parametric method:

Kernel Density Estimation (KDE)

Kernel Density Estimation (KDE)

$$p_{s,b}(o(x)) = \frac{1}{N}\sum_i^N \kappa_\epsilon \left[o(x) - o(x_i)\right]$$

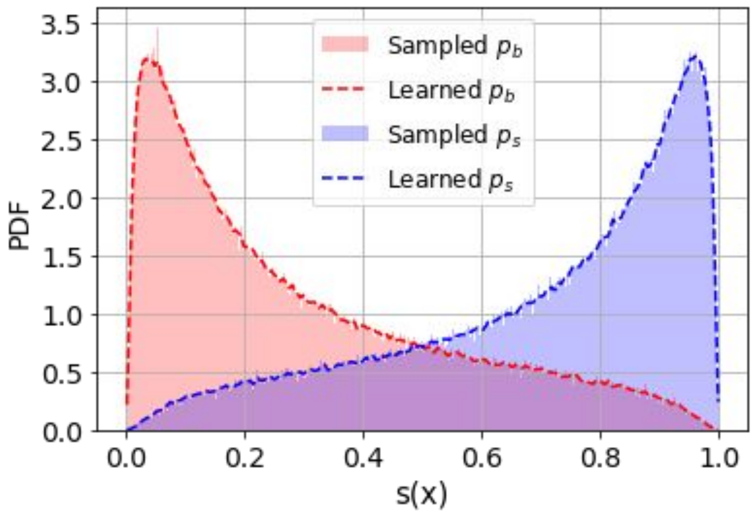
where κ_ε is a kernel function that depends on the "smoothing" scale, or bandwidth parameter ε.

We use the Epanechnikov kernel

$$\kappa_\epsilon(u) = \begin{cases} \frac{1}{\epsilon}\frac{3}{4}\left(1-(u/\epsilon)^2\right), & \text{if } |u| \le \epsilon \\ 0, & \text{otherwise} \end{cases}$$

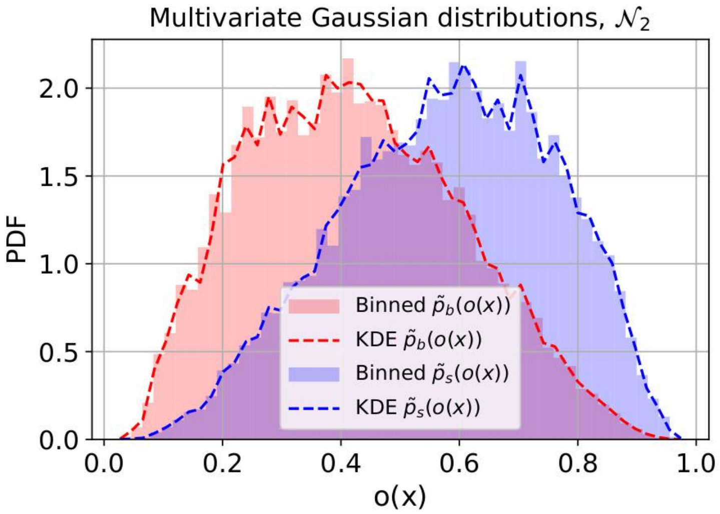
The bandwidth parameter ε is key

- if ε is too low the model may overfit
- if ε is too high the model may underfit



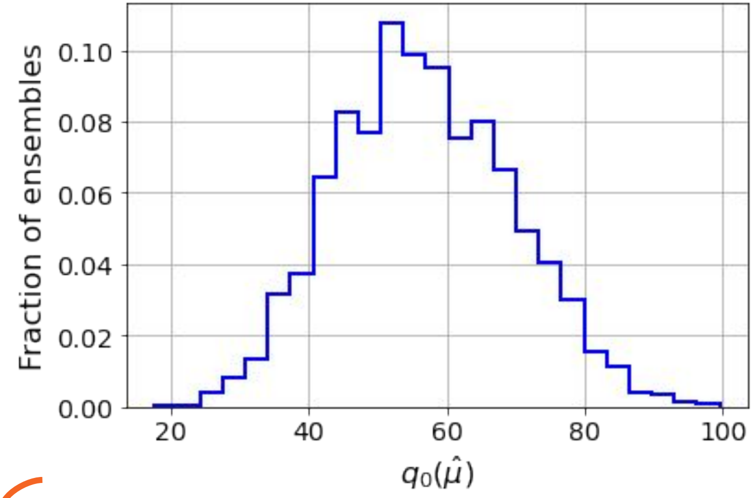
Machine Learned-Likelihood (MLL)

Train supervised per-even classifier:
XGBoost with 1M events per class



Evaluate o(x) with the test data-set
Find the distributions with KDE

Build toy ensembles of fixed B and S (each one represent a possible experimental result) and evaluate the test statistic q_0



Calculate the significance

$$Z = \sqrt{\mathrm{med}[q_0]}$$

Machine Learned-Likelihood (MLL)

First find $\hat{\mu}$ *(for each ensemble)*

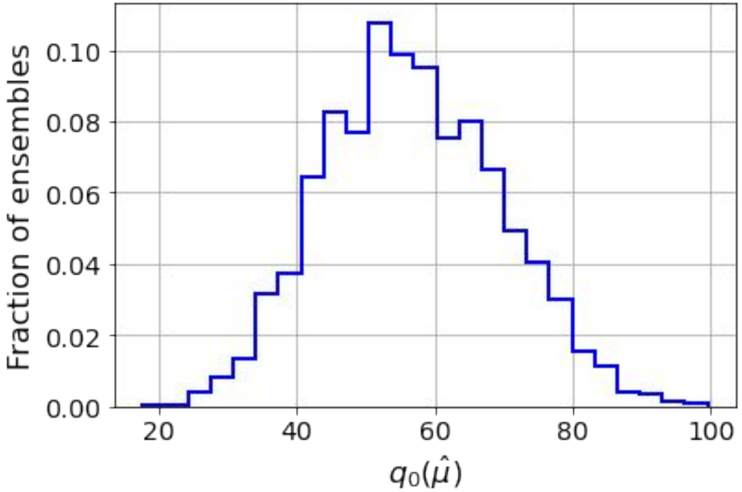
$$\sum_{i=1}^{N} \frac{\tilde{p}_s(o(x_i))}{\hat{\mu} S\, \tilde{p}_s(o(x_i)) + B\, \tilde{p}_b(o(x_i))} = 1$$

summation over the events of each ensemble (build a lot)

Build toy ensembles of fixed B and S (each one represent a possible experimental result) and evaluate the test statistic q_0

Estimate numerically the test statistic *(for each ensemble)*

$$q_0 = \begin{cases} -2\hat{\mu}S + 2\sum_{i=1}^{N} \mathrm{Ln}\left(1 + \frac{\hat{\mu}S}{B}\frac{\tilde{p}_s(o(x_i))}{\tilde{p}_b(o(x_i))}\right) & \text{if } \hat{\mu} \geq 0\,, \\ 0 & \text{if } \hat{\mu} < 0\,, \end{cases}$$



Calculate the significance

$$Z = \sqrt{\mathrm{med}[q_0]}$$

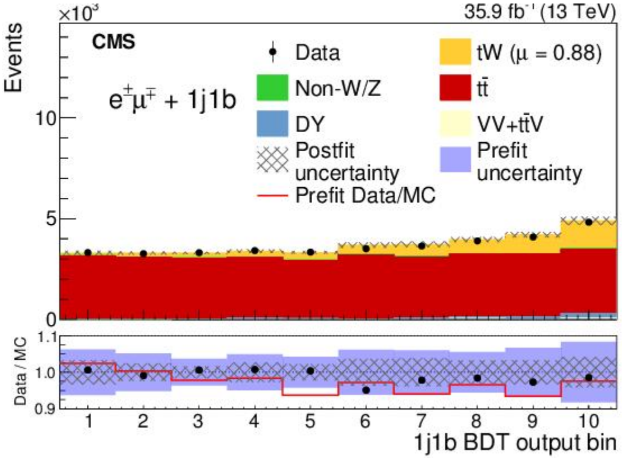
Traditional Binned-Likelihood (BL) method

The Likelihood for D bins, where in each bin d, B_d: the expected number of background events, S_d: the expected number of signal events, and N_d: the measured number of events,

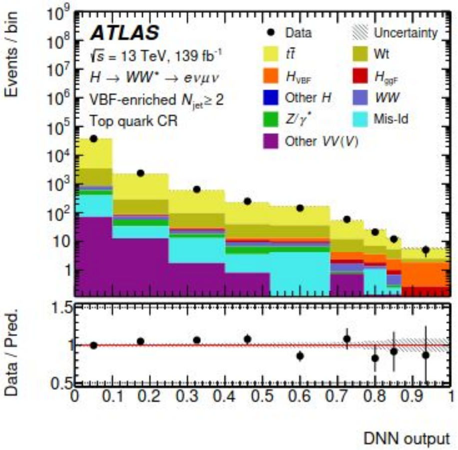
$$\mathcal{L}(\mu, s, b) = \prod_{d=1}^{D} \text{Poiss}\left(N_d | \mu S_d + B_d\right)$$

The median discovery significance

$$\text{med}\left[Z_0^{\text{binned}}|1\right] = \left[2 \sum_{d=1}^{D} \left(\left(S_d + B_d\right) \text{Ln}\left(1 + \frac{S_d}{B_d}\right) - S_d\right)\right]^{1/2} \xrightarrow[\sqrt{B} \gg 1]{S \ll B} \frac{S}{\sqrt{B}}$$



JHEP 10 (2018) 117



arXiv: 2207.00338

Machine Learned-Likelihood (MLL)

The relevant test statistic for **exclusion** limits (very similar for exclusion):

$$\tilde{q}_\mu = \begin{cases} 0 & \text{if } \hat{\mu} > \mu\,, \\ -2\,\mathrm{Ln}\,\frac{\mathcal{L}(\mu,s,b)}{\mathcal{L}(\hat{\mu},s,b)} & \text{if } 0 \leq \hat{\mu} \leq \mu\,, \\ -2\,\mathrm{Ln}\,\frac{\mathcal{L}(\mu,s,b)}{\mathcal{L}(0,s,b)} & \text{if } \hat{\mu} < 0\,, \end{cases}$$

exclusion corresponds to studying signal+background hypothesis μ = 1

using the Likelihood

$$\tilde{q}_\mu = \begin{cases} 0 & \text{if } \hat{\mu} > \mu \\ 2(\mu - \hat{\mu})S - 2\sum_{i=1}^{N} \mathrm{Ln}\left(\frac{Bp_b(x_i)+\mu Sp_s(x_i)}{Bp_b(x_i)+\hat{\mu}Sp_s(x_i)}\right) & \text{if } 0 \leq \hat{\mu} \leq \mu \\ 2\mu S - 2\sum_{i=1}^{N} \mathrm{Ln}\left(1 + \frac{\mu Sp_s(x_i)}{Bp_b(x_i)}\right) & \text{if } \hat{\mu} < 0; \end{cases}$$

where $\hat{\mu}$ is the parameter that maximizes the likelihood

$$\sum_{i=1}^{N} \frac{p_s(x_i)}{\hat{\mu}S\,p_s(x_i)+B\,p_b(x_i)} = 1\,.$$

The median expected significance assuming background-only hypothesis (μ'=0) is

$$Z_\mu \to \mathrm{med}\,[Z_\mu|0] = \sqrt{\mathrm{med}\,[\tilde{q}_\mu|0]}$$