LHC Study of Third-Generation Scalar Leptoquarks with Machine-Learned Likelihoods

In collaboration with E. Arganda, D. A. Díaz, R. M., A. D. Perez, A. Szynkman
2309.05407

Instituto de Física Teórica UAM-CSIC

Universidad Autónoma de Madrid

Rosa María Sandá Seoane
Instituto de Física Teórica UAM-CSIC

XV CPAN DAYS
October 2, 2023

Outline

Model and Final state

-Signal: Third generation scalar-leptoquarks

-Backgrounds and event selection criteria

Set up: From Machine Learning to significances

-ML classifiers and binned vs. unbinned statistical tests

Results @ 13 TeV LHC (full Run 2 dataset)

-95% C.L. expected exclusion limits for m ($LQ^{u/d}$) vs. $BR(LQ^{u/d} \to q\ell)$

Estimation of the impact of systematics (NEW!)

Prospects @ 14TeV LHC (NEW!)

Motivation:

- Third-generation scalar leptoquarks could provide an explanation to B-anomalies!
- The search of LQs at the LHC represents a very extensive program at the LHC, both in the ATLAS and CMS Collaborations.
- None of these experimental analysis, even with (binned) multivariate analysis, has been able to find significant deviations from the SM prediction.

Motivation:

- Third-generation scalar leptoquarks could provide an explanation to B-anomalies!

- The search of LQs at the LHC represents a very extensive program at the LHC, both in the ATLAS and CMS Collaborations.

- None of these experimental analysis, even with (binned) multivariate analysis, has been able to find significant deviations from the SM prediction.

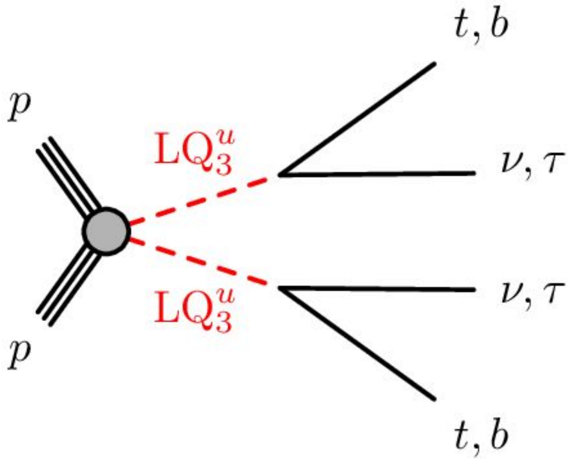
A ML unbbined approach would potentially improve results?

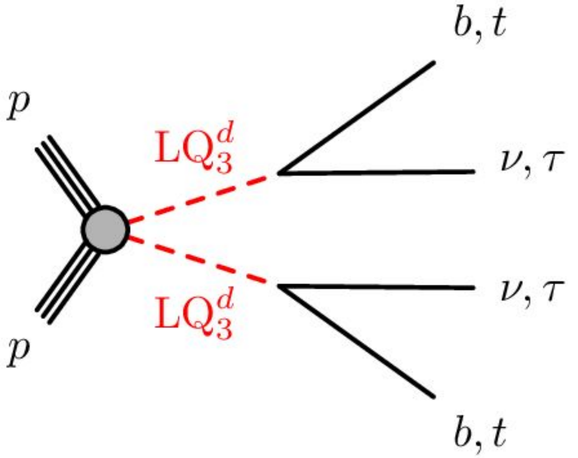
Model and Final state

Signal: pair production of scalar-leptoquarks

Leptoquarks are motivated by many extensions of the SM (used to explain $(g-2)_\mu$, B-anomalies, ...). Focus only on decays into third-generation leptons and quarks (minimal Buchmüller–Rückl–Wyler model).

Q(LQ^u)=+2/3e
LQ^u $\to t\nu_\tau$ / $b\tau$



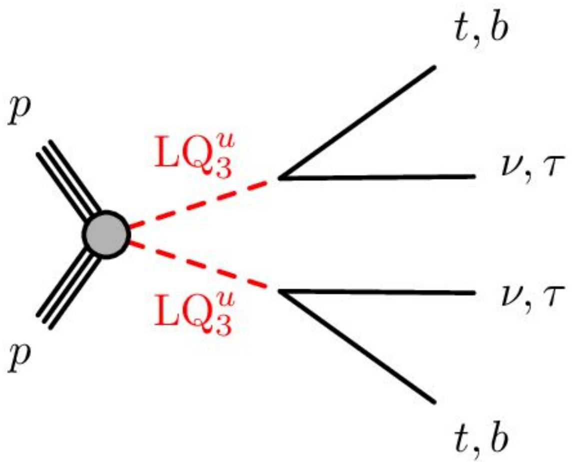


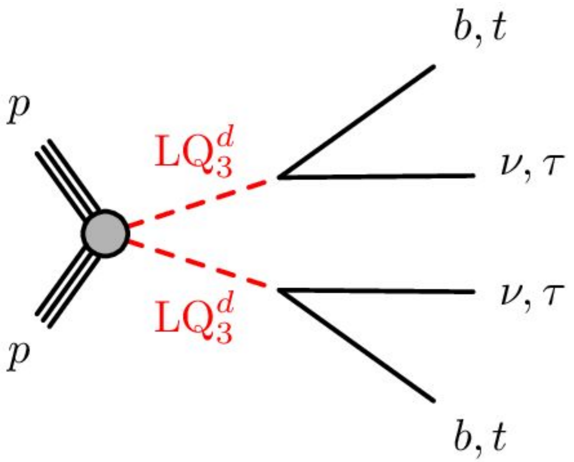
Q(LQ^d)=-1/3e
LQ^d $\to b\nu_\tau$ / $t\tau$

2 parameters: m(LQ$^{u/d}$) leptoquark mass
BR(LQ$^{u/d}$ $\to q\ell$) the branching fraction into a quark and a charged lepton
BR(LQ$^{u/d}$ $\to q\nu$) = 1 - BR(LQ$^{u/d}$ $\to q\ell$)

Signal: pair production of scalar-leptoquarks

Q(LQ^u)=+2/3e
LQ^u $\rightarrow t\nu_\tau$ / $b\tau$





Q(LQ^d)=-1/3e
LQ^d $\rightarrow b\nu_\tau$ / $t\tau$

For a *BR*(LQ$^{u/d}$ $\rightarrow q\ell$) ~ 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.

Event selection criteria

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

Single tau SR definition:

$1\ \tau_{had}$ ($p_T > 10$ GeV, $|\eta| < 2.5$)

$N_{b\text{-jets}} \geq 2$ ($p_T > 20$ GeV, $|\eta| < 2.5$)

$E_T^{miss} > 250$ GeV

Light leptons veto (e/μ)

ATLAS extra selection criteria:

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

$E_T^{miss} > 280$ GeV

$p_T(\tau_{had}) > 50$ GeV (multi-bin SR)

$m_T(\tau_{had}) > 150$ GeV (multi-bin SR)

$s_T = p_T(\tau_{had}) + p_T(j_1) + p_T(j_2) > 600$ GeV (multi-bin SR)

where $m_T^2(A) = m_T^2\,(p_T^A, E_T^{miss}) = 2\, p_T^A\, E_T^{miss}\,(1 - \cos \Delta\Phi(p_T^A, E_T^{miss}))$ is the stransverse mass

Background: SM

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

	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

“Loose” event selection criteria

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

Single tau SR definition:

$1\ \tau_{had}$ (p_T > 20 GeV, $|\eta|$ < 2.5)
$N_{b\text{-jets}} \geq 2$ (p_T > 20 GeV, $|\eta|$ < 2.5)
E_T^{miss} > 280 GeV
Light leptons veto (e/μ)

ATLAS extra selection criteria:

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

$E_T^{miss} > 280$ GeV

$p_T(\tau_{had}) > 50$ GeV (multi-bin SR)

$m_T(\tau_{had}) > 150$ GeV (multi-bin SR)

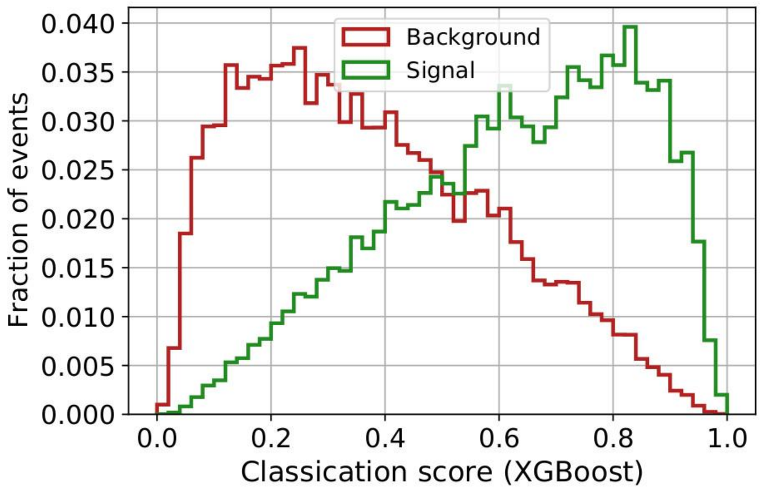
$s_T = p_T(\tau_{had}) + p_T(j_1) + p_T(j_2) > 600$ GeV (multi-bin SR)

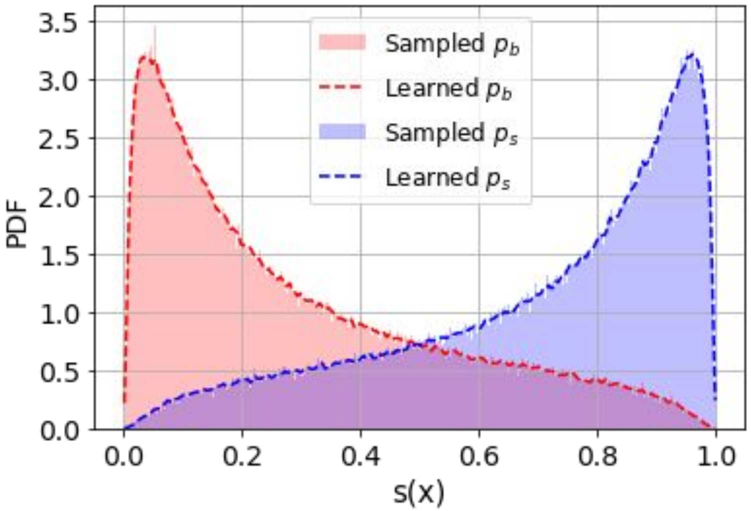
where $m_T^2(A) = m_T^2(p_T^A, E_T^{miss}) = 2\, p_T^A\, E_T^{miss}\,(1 - \cos \Delta\Phi(p_T^A, E_T^{miss}))$ is the stransverse mass

From ML to significances

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

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

o(x̄) Unbinned method

Kernel Density Estimation (KDE) to estimate the B and S PDFs.

↳Calculate Z building pseudo-experiments

Set up:

- We generate events with *MadGraph5_aMC@NLO+Pythia8+Delphes.*

- We simulated signal samples with $m(\text{LQ}^{u/d}) \in [800, 1800]$ GeV and a fixed value $BR(\text{LQ}^{u/d} \to q\ell)=0.5$, selecting BPs with a step of 200 GeV.

- We apply the “loose” event selection criteria.

- We train one supervised binary XGBoost classifier with 500k events per class (background and signal) for each BP, using simple discriminating variables (object multiplicities and low-level kinematic variables). We consider the relative weight of each background in the total background sample.

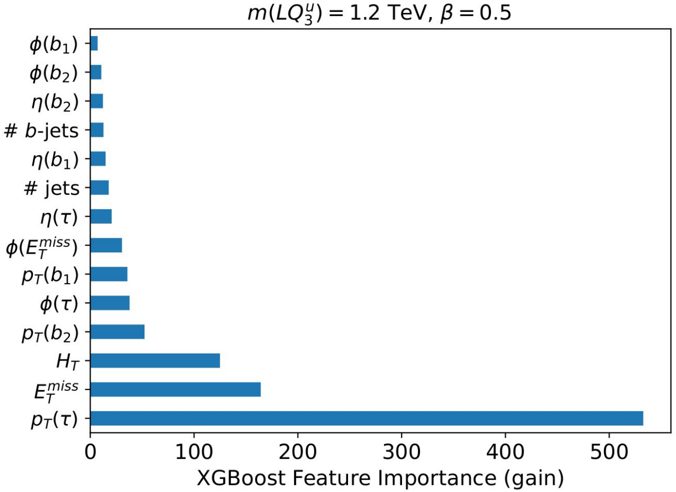
- We estimate the significance with the binned likelihood method, and for the unbinned approach, using KDEs.

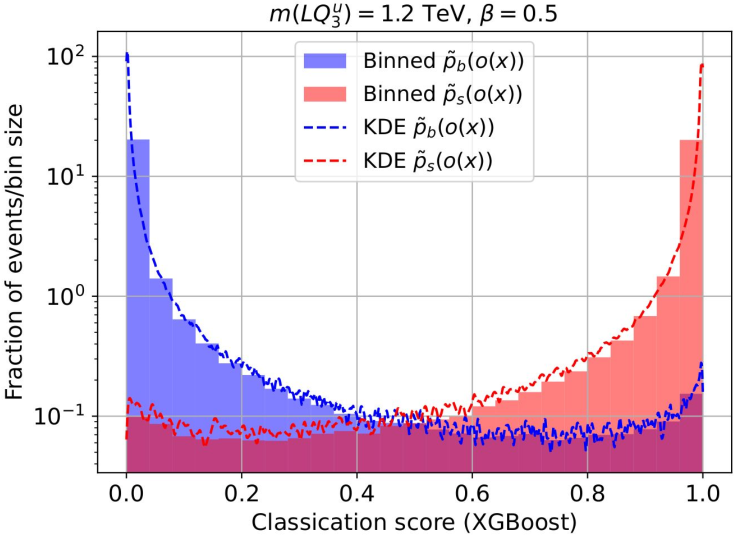
- We extend results to different branching fractions.

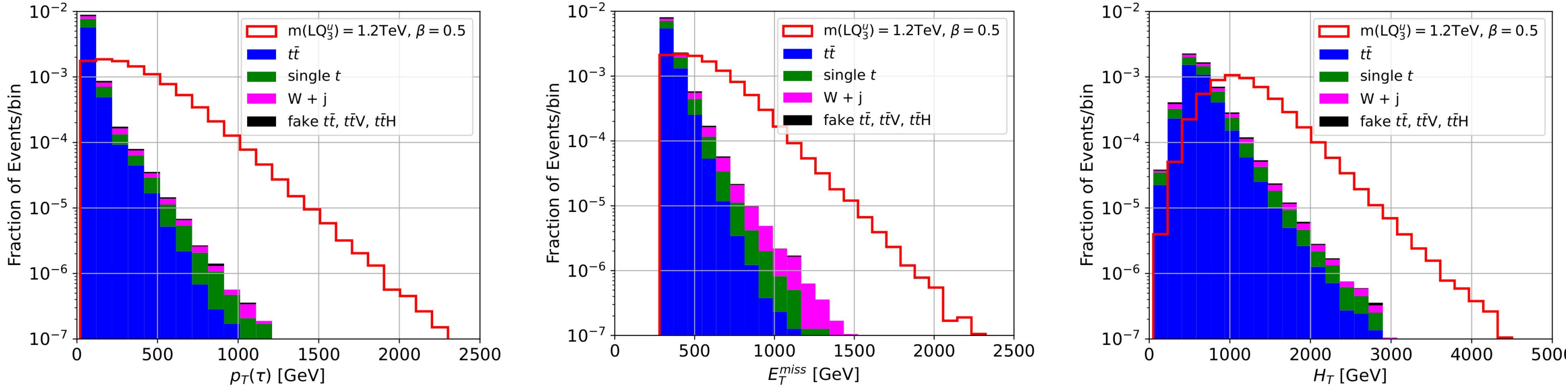
Results @ 13 TeV LHC

Results

BP with $m(\mathrm{LQ^u})$ = 1200 GeV and BR($\mathrm{LQ^u} \rightarrow q\ell$)=0.5

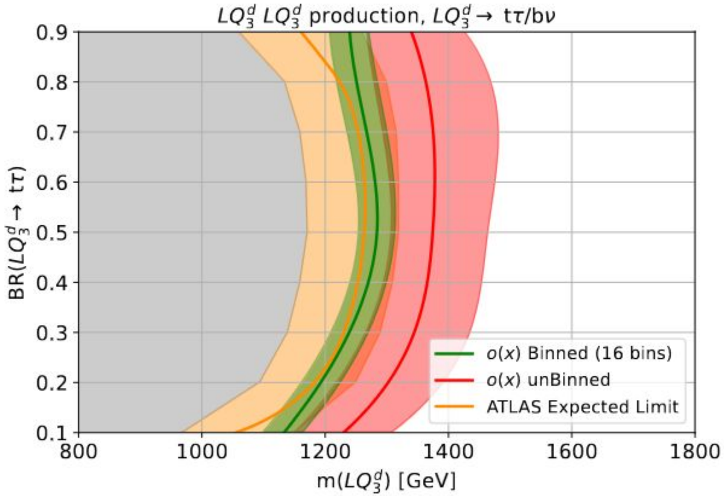




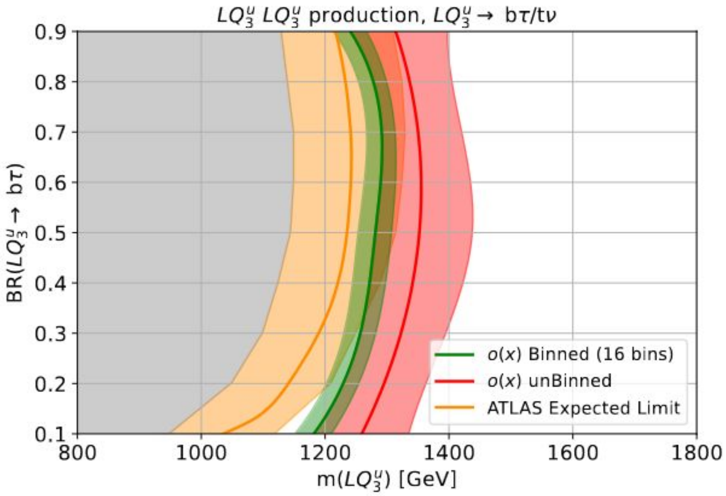


Results

Down-type Leptoquarks



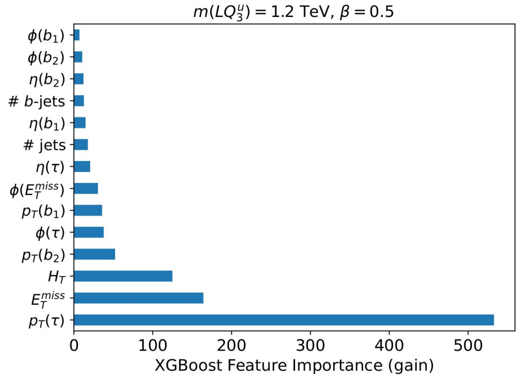
Up-type Leptoquarks



- For 139 fb^{-1} and both types of scalar LQs, the expected exclusion limits extend to ~1.3 TeV (binned ML) and ~1.35 TeV (unbinned ML) at 95% CL for intermediate values of $BR(LQ^{u/d} \to q\ell)$.
- The unbinned method provides more stringent bounds, but is computationally more expensive and has larger statistical uncertainty region.

Estimation of the impact of systematics

Idea:



- We consider uncertainties only in the most relevant features for training ($p_T(\tau)$, E_T^{miss} and H_T) and take correlations among them as not significant.

- We consider shifts of 5-10% of this variables.

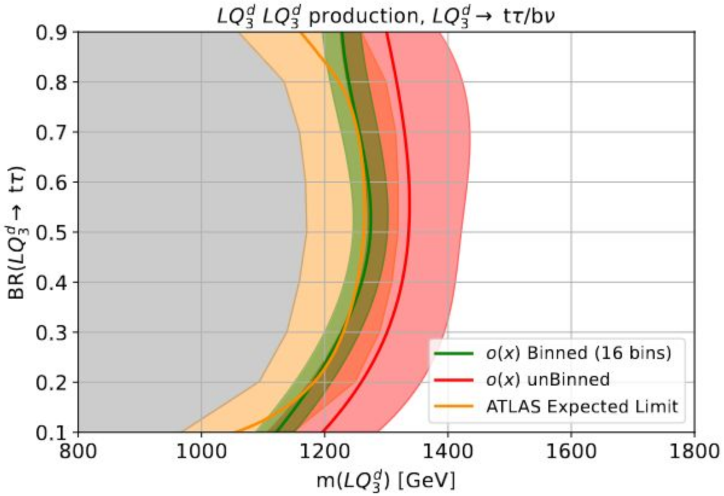
- Taking as example $p_T(\tau)$, we take the ML algorithm trained with no uncertainties, and evaluate it with two new samples with all variables unchanged but $p_T(\tau) + \Delta p_T(\tau)$, and obtained two $o^{\pm}(x)$, respectively.

- For the binned method, the uncertainty in each bin d is $\Delta o(x_d) = |o(x_d)^+ - o(x_d)^-|$ for analytical formula.

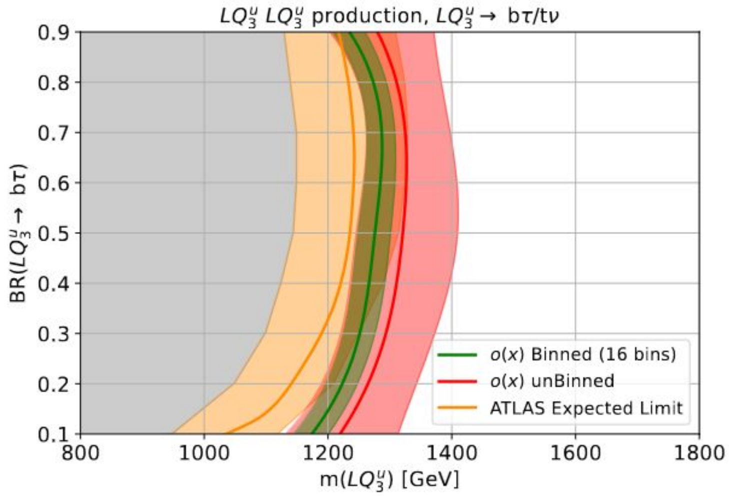
- For the unbinned method, we repeat the entire procedure with $o(x_d)^{\pm}$ and take as “modified result” the outcome with less restrictive limits.

Results

Down-type Leptoquarks



Up-type Leptoquarks

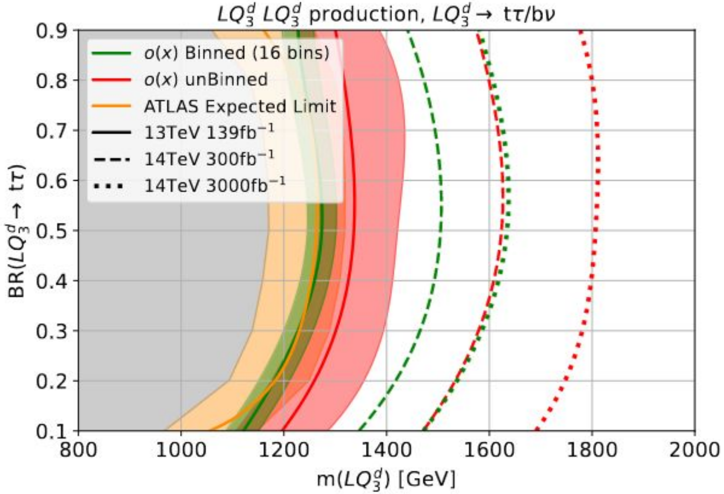


- The impact on the exclusion contours is only of a few percent, and the effect in both methods is similar $\rightarrow$ the treatment for the unbinned case provides a good numerical approximation.
- Including variations in other features does not impact significantly the results.
- Still necessary full treatment including all sources and correlations!

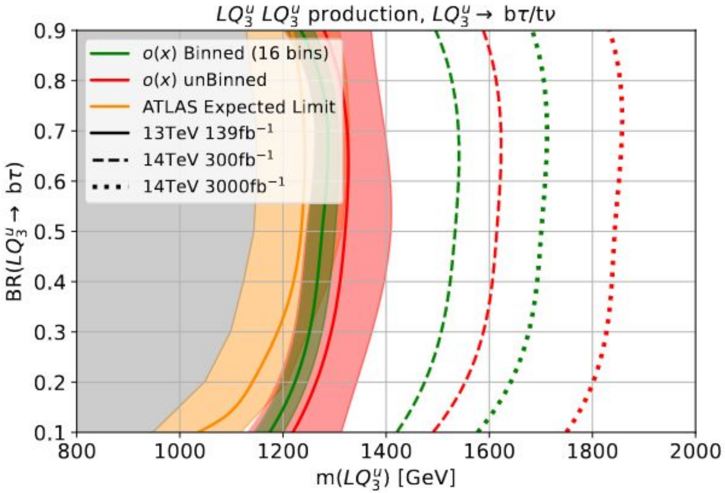
Prospects @ 14 TeV LHC

Results

Down-type Leptoquarks



Up-type Leptoquarks



- For 300 fb^{-1} and both types of scalar LQs, the expected exclusion limits extend to ~1.5 TeV (binned ML) and ~1.6 TeV (unbinned ML) at 95% CL for intermediate values of $BR(LQ^{u/d} \to q\ell)$, while for 3000 fb^{-1} extend to ~1.65 TeV and ~1.8 TeV, respectively.
- Results include the “naive” approach for the inclusion of systematic uncertainties!

Conclusions

Conclusions

Search for third-generation scalar leptoquarks in final states with one hadronically decaying tau lepton, b-jets and large missing transverse momentum. As a ***proof of concept*** we used ML algorithms with a binned and an unbinned likelihood approach and simple selection cuts.

- Tendency towards a potential improvement of the exclusion limits @13 TeV and 139 fb^{-1} reported in the ATLAS search used as reference, specially for unbinned analysis!
  - The unbinned method however is more computationally expensive and has larger statistical uncertainties.

Conclusions

- Results stable when including a naive estimation of systematic uncertainties
  - The impact is slight and similar for both the binned and unbinned approaches.
  - Still necessary a full analysis with all sources and correlations!

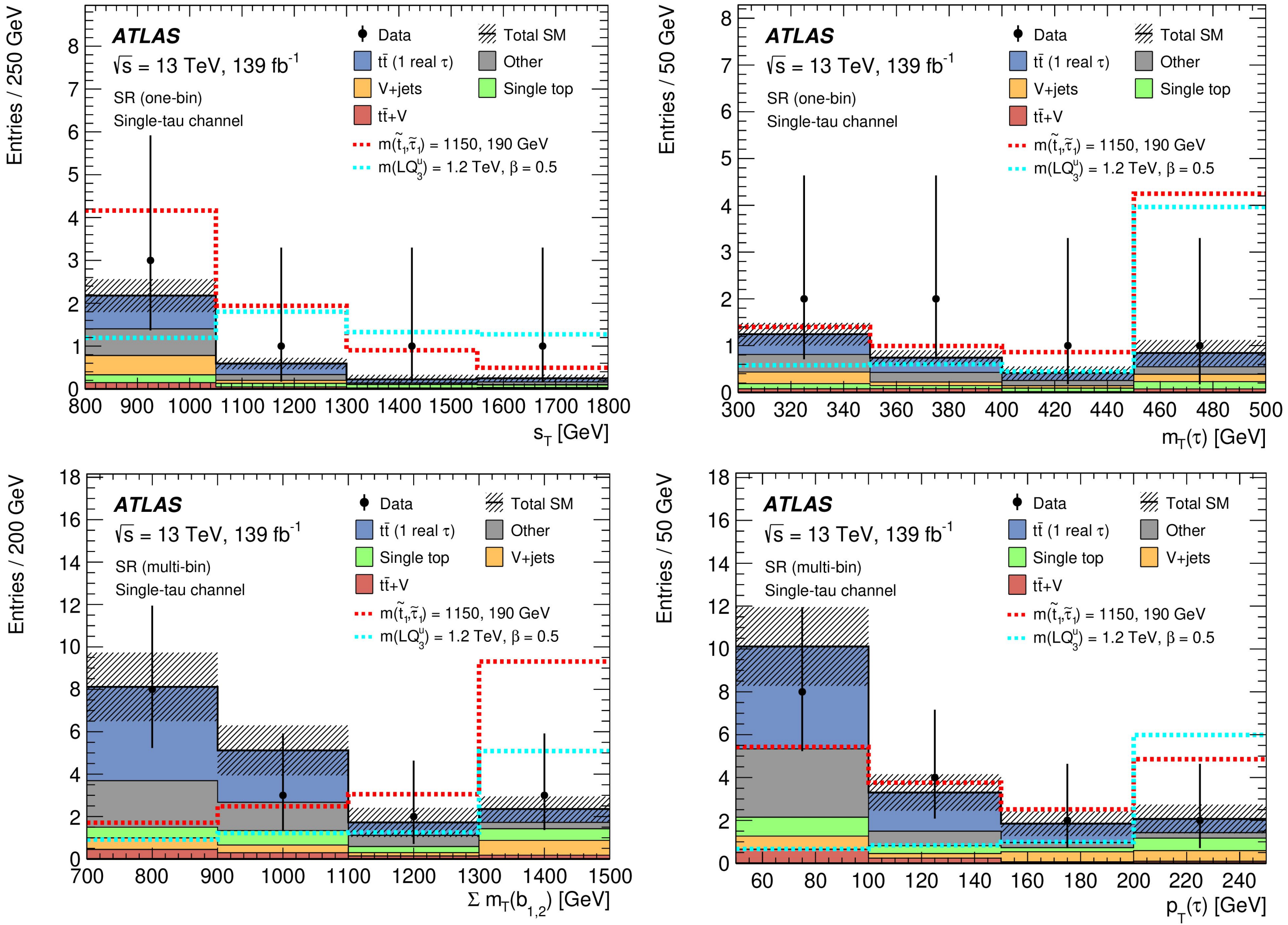
- Promising prospects for @14 TeV and both 300 and 3000 fb^{-1}
  - For both types of scalar leptoquarks, possible to exclude masses up to ~1.6 and ~1.8 TeV respectively with the unbinned approach at 95% CL for intermediate values of $BR(LQ^{u/d} \to q\ell)$.

Thank you!

(Special thanks to A. D. Perez!)

Back up

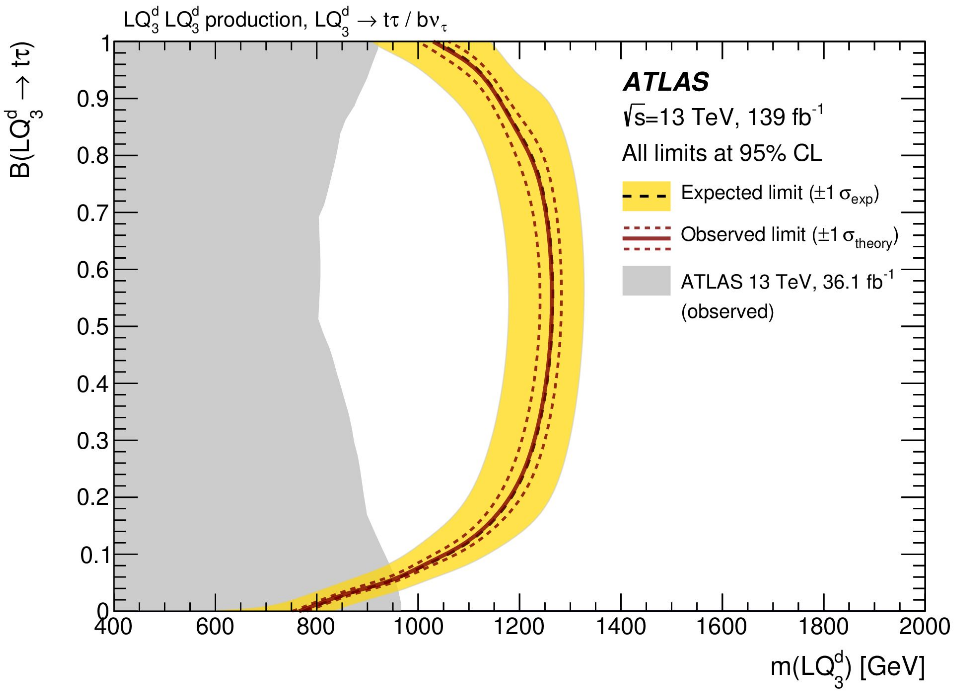
ATLAS



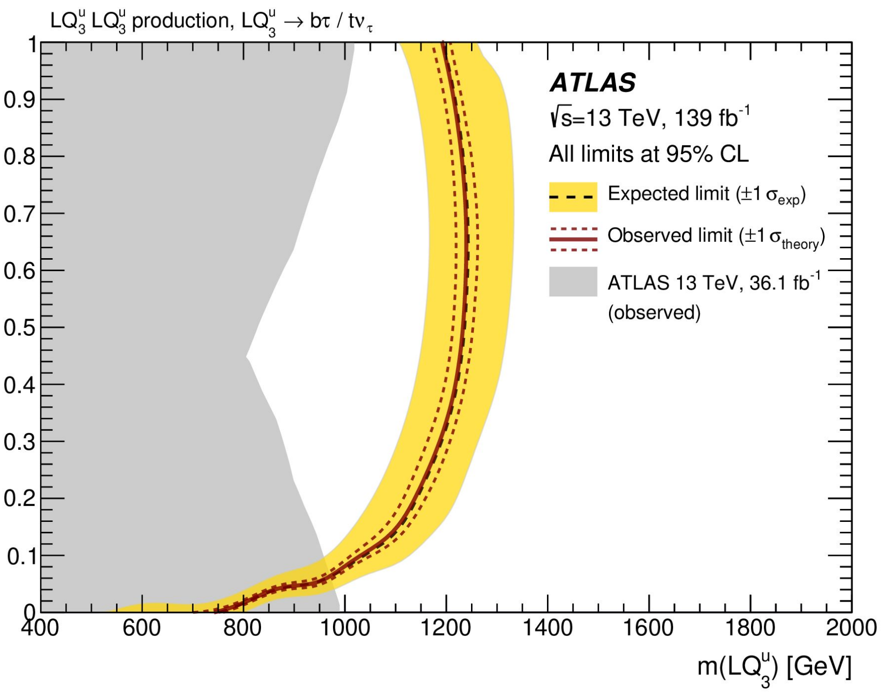
ATLAS
√s = 13 TeV, 139 fb-1
SR (one-bin)
Single-tau channel
Data
Total SM
tt̄ (1 real τ)
Other
V+jets
Single top
tt̄+V
m(t̃1,τ̃1) = 1150, 190 GeV
m(LQ3u) = 1.2 TeV, β = 0.5
Entries / 250 GeV
sT [GeV]
Entries / 50 GeV
mT(τ) [GeV]
SR (multi-bin)
Entries / 200 GeV
Σ mT(b1,2) [GeV]
pT(τ) [GeV]

ATLAS

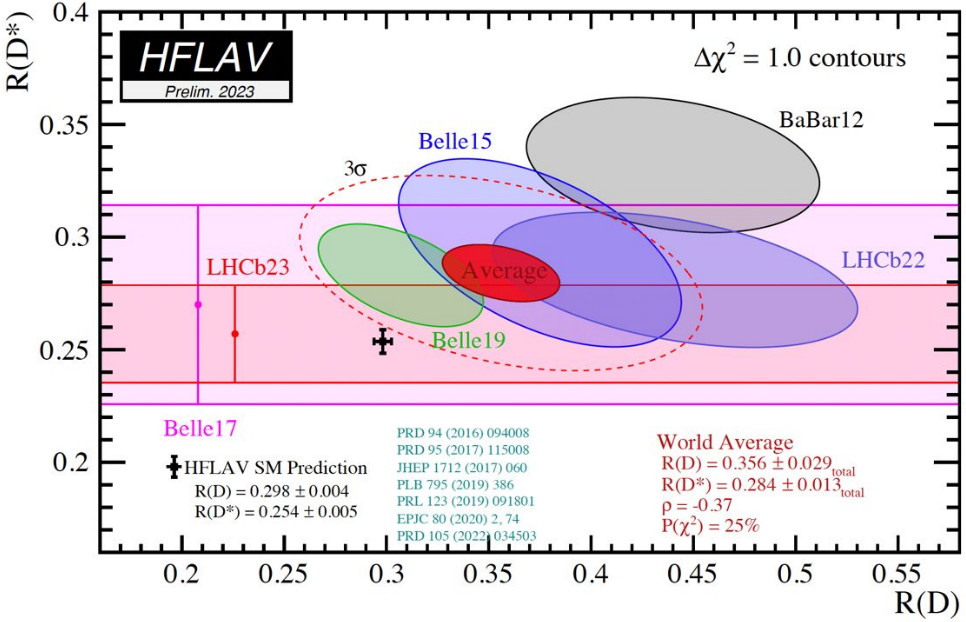
Down-type Leptoquarks



Up-type Leptoquarks

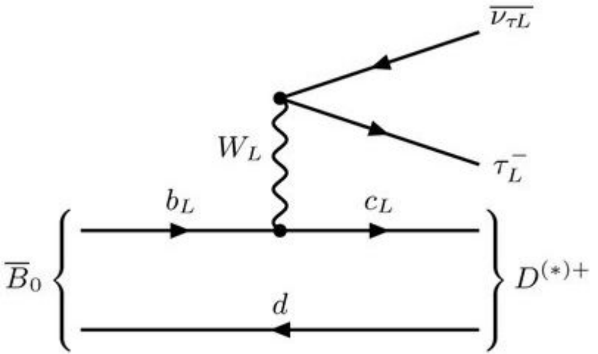


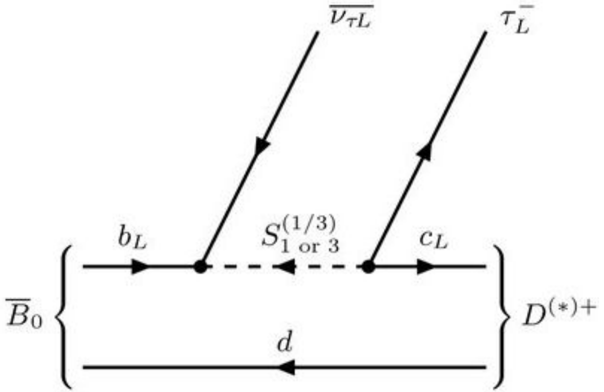
B-anomalies



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 $D=\{(\bar{x}_1,t_1), \ldots ,(\bar{x}_n,t_n)\}$

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

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

1 for *signal*

0 for *background*

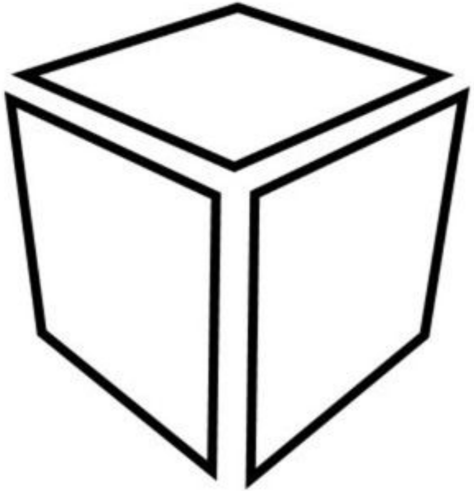
Output

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

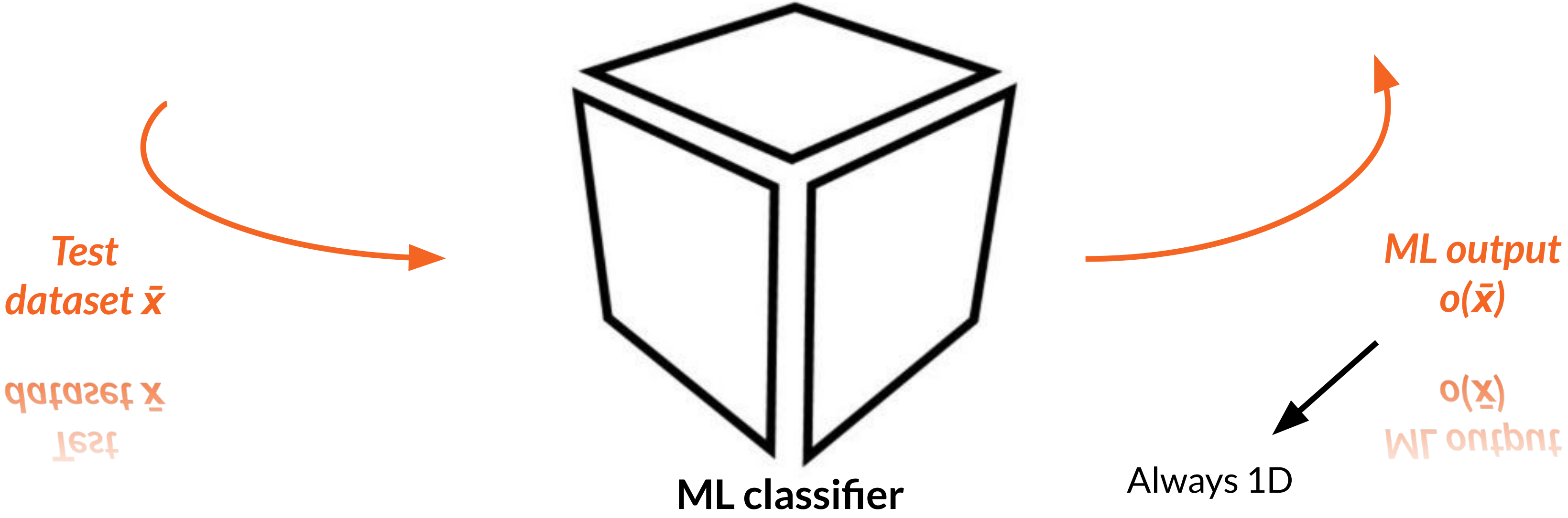
Data sample that we do not know if it is Signal or Background

Prediction

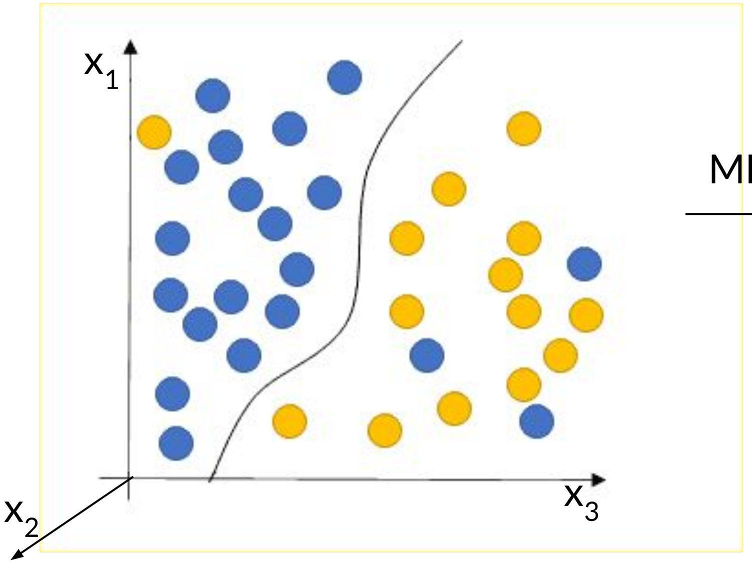
To assign a label a threshold or working point (WP) is needed

S or B
label ??

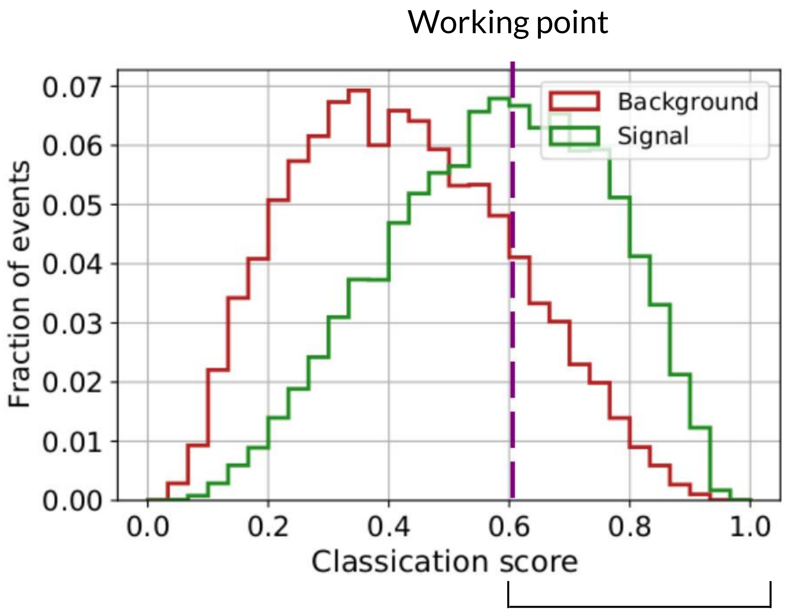
if **o(x) < WP** label → ‘0’ → ***B***
if **o(x) > WP** label → ‘1’ → ***S***



“Naive” expected significance



ML →



Defining a working point ~ Defining signal enriched region

In the defined area → Significance ~ $\frac{S}{\sqrt{B}}$
(we discard events outside of it)

Machine Learned-Likelihood (MLL)

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

$$\mathcal{L}(\mu, s, b) = p(N, \{x_i, i = 1, \ldots, N\} | \mu, s, b) = \underbrace{\mathrm{Poiss}\left(N | \mu S + B\right)}_{} \underbrace{\prod_{i=1}^{N} p(x_i | \mu, s, b)}_{}$$

~ global info
ensemble factor

~ local info
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 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

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

We need

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

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

Machine Learned-Likelihood (MLL)

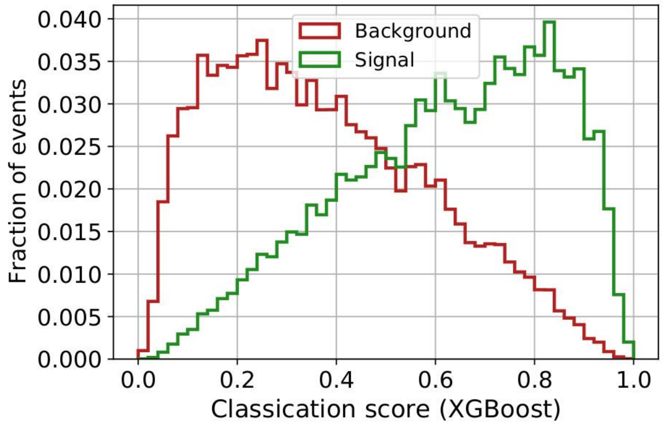
Replace the densities for the one-dimensional manifolds obtained with a machine-learning classifier.

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*

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



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.

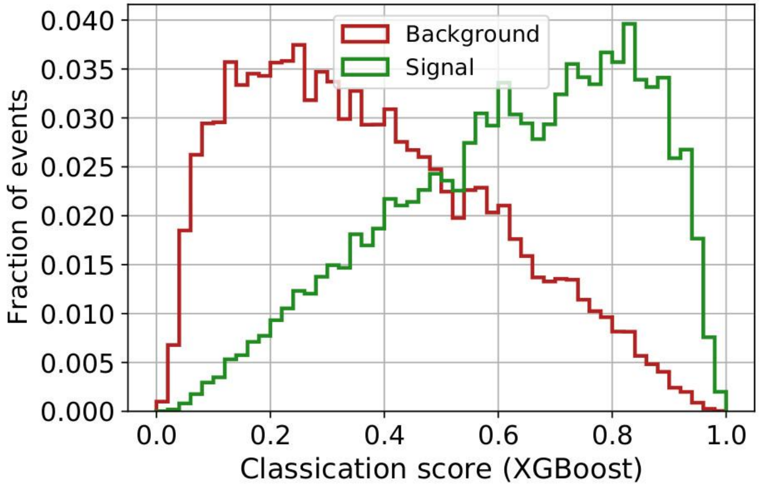
The median expected significance assuming signal-plus-background hypothesis (μ'=1) is

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

Density estimation

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

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



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)

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

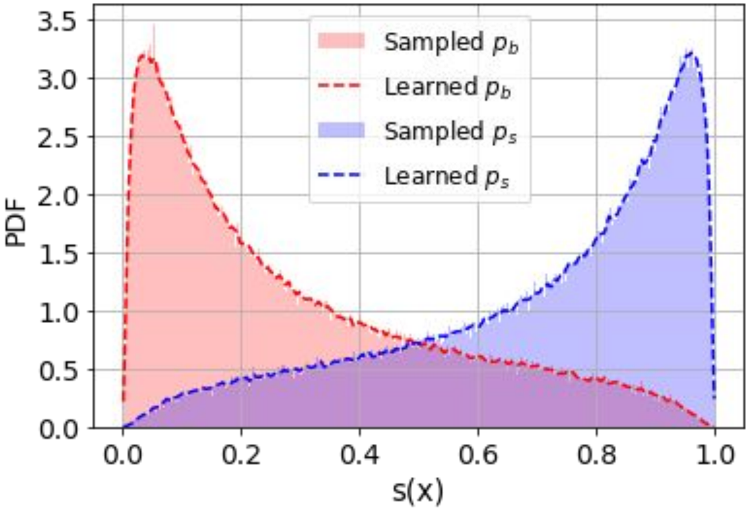
where κ_ε is a kernel function that depends on the "smoothing" scale, or bandwidth parameter ε.

We use the Epanechnikov kernel

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

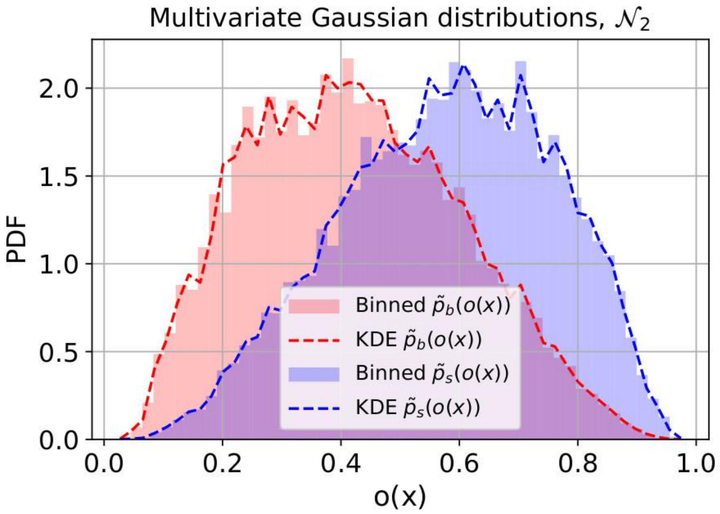
The bandwidth parameter ε is key

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



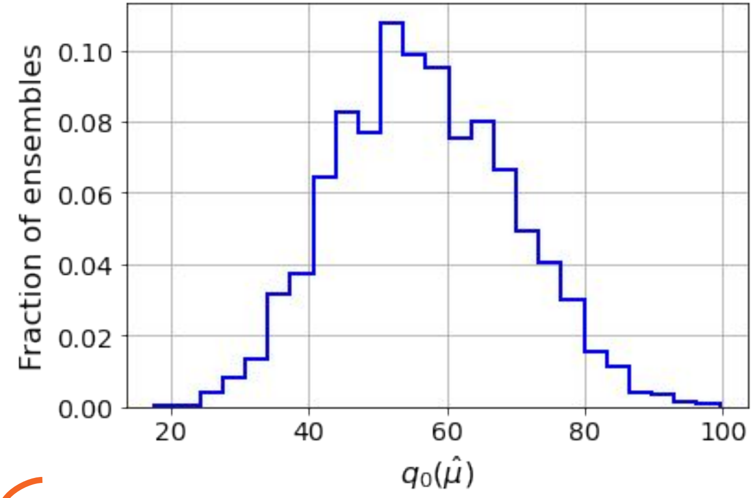
Machine Learned-Likelihood (MLL)

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

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

Machine Learned-Likelihood (MLL)

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

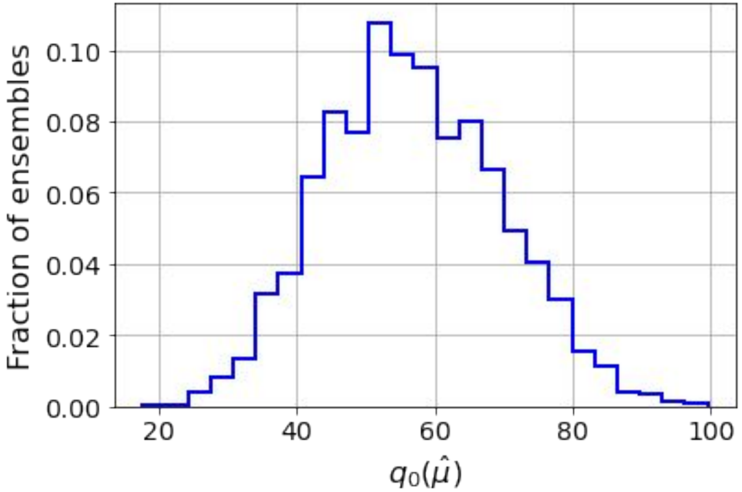
First find $\hat{\mu}$ *(for each ensemble)*

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

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

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

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

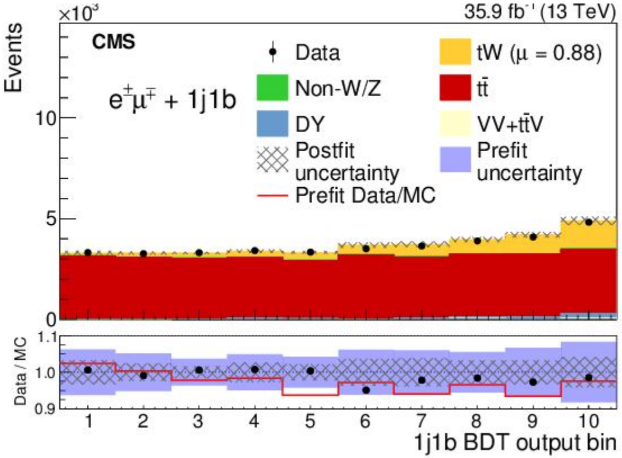
Traditional Binned-Likelihood (BL) method

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,

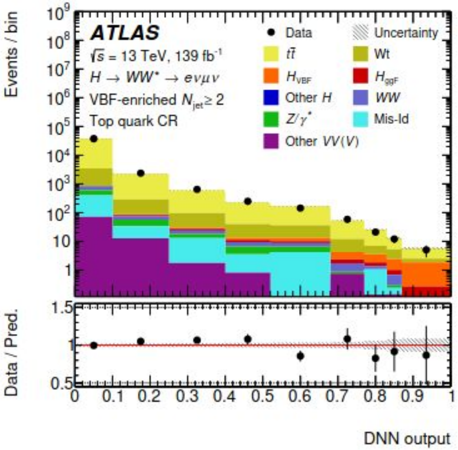
$$\mathcal{L}(\mu, s, b) = \prod_{d=1}^{D} \mathrm{Poiss}\left(N_d | \mu S_d + B_d\right)$$

The median discovery significance

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



JHEP 10 (2018) 117



arXiv: 2207.00338

Machine Learned-Likelihood (MLL)

The relevant test statistic for **exclusion** limits:

exclusion corresponds to studying signal+background hypothesis μ = 1

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

using the Likelihood

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

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

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