

## Extending the Bump Hunt with Machine Learning

#### Based on: Phys. Rev. Lett. 121, 241803 (2018) [1805.02664] Jack Collins, Kiel Howe, Ben Nachman





## Is there new physics at the LHC?

#### ATLAS SUSY Searches\* - 95% CL Lower Limits

**ATLAS** Preliminary  $\sqrt{s} = 13 \text{ TeV}$ 

March 2019

	Model	Signature $\int \mathcal{L} dt$ [fl	Mass limit			Reference
Inclusive Searches	$ ilde{q} ilde{q}, ilde{q} ightarrow  ilde{\chi}_1^0$	$\begin{array}{ccc} 0 \ e, \mu &  ext{ 2-6 jets } E_T^{ ext{miss}} &  ext{ 36.1 } \\  ext{mono-jet } &  ext{ 1-3 jets } & E_T^{ ext{miss}} &  ext{ 36.1 } \end{array}$		0.9 1.55 0.71	$m(\bar{\chi}_1^0) < 100 \text{ GeV} \ m(\bar{q}) - m(\bar{\chi}_1^0) = 5 \text{ GeV}$	1712.02332 1711.03301
	$\tilde{g}\tilde{g},\tilde{g}\!\rightarrow\!q\bar{q}\tilde{\chi}^{0}_{1}$	0 <i>e</i> , $\mu$ 2-6 jets $E_T^{\text{miss}}$ 36.1	ğ ğ	2.0 Forbidden 0.95-1.6	$m(\tilde{\chi}_{1}^{0}) < 200 \text{ GeV}$ $m(\tilde{\chi}_{1}^{0}) = 900 \text{ GeV}$	1712.02332 1712.02332
	$\tilde{g}\tilde{g}, \tilde{g} \rightarrow q\bar{q}(\ell\ell)\tilde{\chi}_1^0$	$\begin{array}{cccc} 3 \ e, \mu & \ 4 \ { m jets} & \ 36.1 \\ e e, \mu \mu & \ 2 \ { m jets} & \ E_T^{ m miss} & \ 36.1 \end{array}$	Ĩ Ĩ	1.85 1.2	$m(\tilde{\chi}_{1}^{0}) < 800 \text{ GeV}$ $m(\tilde{g})-m(\tilde{\chi}_{1}^{0}) = 50 \text{ GeV}$	1706.03731 1805.11381
	$\tilde{g}\tilde{g}, \tilde{g} \rightarrow qqWZ\tilde{\chi}_1^0$	$\begin{array}{cccc} 0 \; e, \mu & & \mbox{7-11 jets} & E_T^{\rm miss} & \mbox{36.1} \\ \mbox{3} \; e, \mu & & \mbox{4 jets} & & \mbox{36.1} \end{array}$	Ĩġ Ĩġ	1.8 0.98	${f m}( ilde{\chi}_1^0)$ <400 GeV ${f m}( ilde{g})$ - ${f m}( ilde{\chi}_1^0)$ =200 GeV	1708.02794 1706.03731
	$\tilde{g}\tilde{g}, \tilde{g} \rightarrow t t \tilde{\chi}_1^0$		$\begin{array}{c} \tilde{g} \\ \tilde{g} \\ \tilde{g} \end{array}$		m(ℓ̃))<200 GeV m(ĝ)-m(ℓ̃1)=300 GeV	ATLAS-CONF-2018-041 1706.03731
3 <sup>rd</sup> gen. squarks direct production	$\tilde{b}_1 \tilde{b}_1, \tilde{b}_1 {\rightarrow} b \tilde{\chi}_1^0 / t \tilde{\chi}_1^{\pm}$	New Ph	vsics is			1708.0920 <mark>6, 1711.03301</mark> 170 <b>3.09266</b> 1706.03731
	$\tilde{b}_1 \tilde{b}_1,  \tilde{b}_1 \rightarrow b \tilde{\chi}_2^0 \rightarrow b h \tilde{\chi}_1^0$	$0 e, \mu$ $6 b E_T^{\text{miss}}$ 139	δ1         0.23-0.48			SUS /-2018-31 SUS /-2018-31
	$\tilde{t}_1 \tilde{t}_1, \tilde{t}_1 \rightarrow W b \tilde{\chi}_1^0 \text{ or } t \tilde{\chi}_1^0$ $\tilde{t}_1 \tilde{t}_1, \text{ Well-Tempered LSP}$ $\tilde{t}_1 \tilde{t}_1, \tilde{t}_1 \rightarrow \tilde{\tau}_1 b \nu, \tilde{\tau}_1 \rightarrow \tau \tilde{G}$	0-2 e, $\mu$ 0-2 jets $1^2 b E_T^{\text{miss}}$ to Multip $1_{\tau+1}^{2} e_{\mu,\tau}^{2} = 2$ jets $\pi b E_T^{\text{miss}} = 36.1$	o heavy			1506.08616, 17C9.04183, 1711.11520 1709.04183, 1711.11520 18C3.10178
	$\tilde{t}_1 \tilde{t}_1, \tilde{t}_1 \rightarrow c \tilde{\chi}_1^0 / \tilde{c} \tilde{c}, \tilde{c} \rightarrow c \tilde{\chi}_1^0$	2. to	o rare			1805.01649 1805.01649 1711.03301
	$\tilde{t}_2 \tilde{t}_2, \tilde{t}_2 \rightarrow \tilde{t}_1 + h$	1-2 $e, \mu$ 4 $b$ $E_T^{\text{miss}}$ 36.1	Ĩ2			170 <b>6.03986</b>
EW direct	$ ilde{\chi}_1^{\pm}  ilde{\chi}_2^0$ via $WZ$	$\begin{array}{ccc} 2-3 \ e, \mu & E_T^{\text{miss}} & 56.4 \\ ee, \mu\mu & \geq 1 & E_T^{\text{miss}} & 56.1 \end{array}$	t covered b	ov existing	$m(\tilde{\chi}_{1}^{0})=0$ $m(\tilde{\chi}_{1}^{\pm})-m(\tilde{\chi}_{1}^{0})=10 \text{ GeV}$	1403.529 <mark>4, 1806.02293</mark> 1712.08119
	$\tilde{\chi}_1^{\pm} \tilde{\chi}_1^{\mp}$ via <i>WW</i> $\tilde{\chi}_1^{\pm} \tilde{\chi}_2^0$ via <i>Wh</i> $\tilde{\chi}^{\pm} \tilde{\chi}^{\mp}$ via $\tilde{\ell}$	$\begin{array}{cccc} 2  e, \mu & E_T^{\text{miss}} & 139 \\ 0 -1  e, \mu & 2  b & E_T^{\text{miss}} & 36.1 \\ \end{array}$	〒 0.42 〒/虎	0.68	$m(\bar{\chi}_1^0)=0$ $m(\bar{\chi}_1^0)=0$	ATLAS-CONF-2019-008 1812.09432
	$\tilde{\chi}_1^{\pm} \tilde{\chi}_1^{\mp} / \tilde{\chi}_2^0, \tilde{\chi}_1^{\pm} \to \tilde{\tau}_1 \nu(\tau \tilde{\nu}), \tilde{\chi}_2^0 \to \tilde{\tau}_1 \tau(\nu \tilde{\nu})$	program	ime of ded	icated sea	arcnes	1703.07875
	$\tilde{\ell}_{L,R}\tilde{\ell}_{L,R},  \tilde{\ell} {\rightarrow} \ell \tilde{\chi}_1^0$					ATLAS-CONF-2019-008
	$\tilde{H}\tilde{H},\tilde{H}{ ightarrow}h\tilde{G}/Z\tilde{G}$	$\begin{array}{llllllllllllllllllllllllllllllllllll$	<u>й</u> 0.13-0.23 <u>Й</u> 0.3		$BR(\tilde{\mathcal{X}}_{1}^{0} \to h\tilde{G})=1$ BR( $\mathcal{X}_{1}^{0} \to ZG$ )=1	1806.04030 1804.03602
Long-lived particles	Direct $\tilde{\chi}_1^+ \tilde{\chi}_1^-$ prod., long-lived $\tilde{\chi}_1^\pm$	Disapp. trk 1 jet $E_T^{\text{miss}}$ 36.1	$ \tilde{\chi}_{1}^{\pm} = 0.46 $ $ \tilde{\chi}_{1}^{\pm} = 0.15 $		Pure Wino Pure Higgsino	1712.02118 ATL-PHYS-PUB-2017-019
	Stable $\tilde{g}$ R-hadron Metastable $\tilde{g}$ R-hadron, $\tilde{g}  ightarrow qq \tilde{\chi}_1^0$	Multiple 36.1 Multiple 36.1	ĝ	2.0 2.05 2.	4 m $(\tilde{\chi}_1^0)$ =100 GeV	1902.01636,1808.04095 1710.04901,1808.04095
RPV	$ \begin{array}{l} LFV pp \rightarrow \tilde{\mathbf{v}}_{\tau} + X, \tilde{\mathbf{v}}_{\tau} \rightarrow e\mu/e\tau/\mu\tau \\ \tilde{\chi}_{1}^{+} \tilde{\chi}_{1}^{\tau} / \tilde{\chi}_{2}^{0} \rightarrow WW/Z\ell\ell\ell\ell\nu\nu \\ \tilde{g}\tilde{g}, \tilde{g} \rightarrow qq\tilde{\chi}_{1}^{0}, \tilde{\chi}_{1}^{0} \rightarrow qqq \end{array} $	$e\mu, e\tau, \mu\tau$ 3.2 4 $e, \mu$ 0 jets $E_T^{miss}$ 36.1 4-5 large- $R$ jets 36.1 Multiple 36.1	$ \begin{split} \bar{\tilde{v}}_{\tau} & \\ \bar{\tilde{X}}_{1}^{\pm}/\bar{X}_{2}^{0}  [\lambda_{t33} \neq 0, \lambda_{12k} \neq 0] \\ \bar{\tilde{g}}  [m(\tilde{x}_{1}^{0})=200 \text{ GeV}, 1100 \text{ GeV}] \\ \bar{\tilde{g}}  [\lambda'_{112}=2e\cdot4, 2e\cdot5] \end{split} $	1.9 0.82 1.33 1.3 1.9 1.05 2.0	$\begin{array}{c} \lambda_{311}'=\!0.11,\lambda_{132/133/233}\!=\!0.07\\ m(\tilde{\xi}_1^0)\!=\!100~{\rm GeV}\\ {\rm Large}\;\lambda_{1/2}'\\ m(\tilde{\xi}_1^0)\!=\!200~{\rm GeV},{\rm bino-like} \end{array}$	1607.08079 1804.03602 1804.03568 ATLAS-CONF-2018-003
	$\begin{split} \widetilde{tt}, \widetilde{t} \rightarrow \widetilde{tt}_{1}^{0}, \widetilde{\lambda}_{1}^{0} \rightarrow tbs \\ \widetilde{t}_{1}\widetilde{t}_{1}, \widetilde{t}_{1} \rightarrow bs \\ \widetilde{t}_{1}\widetilde{t}_{1}, \widetilde{t}_{1} \rightarrow d \end{split}$	Multiple         36.1           2 jets + 2 b         36.7           2 e,μ         2 b         36.1           1 μ         DV         136	$\begin{array}{c c} \tilde{g} & [\lambda'_{323} = 2e\cdot4, 1e\cdot2] & 0 \\ \hline \tilde{t}_1 & [qq, bs] & 0.42 \\ \hline \tilde{t}_1 & [1e\cdot10 < \lambda'_{23k} < 1e\cdot8, 3e\cdot10 < \lambda'_{23k} < 3e\cdot9] \end{array}$	.55 1.05 0.61 0.4-1.45 1.0 1.6	$\begin{split} \mathbf{m}(\bar{k}_1^0) &= 200 \; \mathrm{GeV}, \; \mathrm{bino-like} \\ & \mathrm{BR}(\bar{t}_1 {\rightarrow} be/b\mu) {>} 20\% \\ & \mathrm{BR}(\bar{t}_1 {\rightarrow} q\mu) {=} 100\%, \; \mathrm{cos}\theta_i {=} 1 \end{split}$	ATLAS-CONF-2018-003 1710.07171 1710.05544 ATLAS-CONF-2019-006
*Only	a selection of the available mas	ss limits on new states or	10 <sup>-1</sup>	<u></u> 1	Mass scale [TeV]	

<sup>•</sup>Only a selection of the available mass limits on new states or phenomena is shown. Many of the limits are based on simplified models, c.f. refs. for the assumptions made.

## **LHC Searches**





A->22  $A \rightarrow WW \rightarrow 4q$  $A \rightarrow EE \rightarrow bb 1992$ 569962  $A \rightarrow 2\phi_1 \rightarrow 4\phi_2 \rightarrow \dots \rightarrow N_q$ A -> Dark shower

#### Yesterday from Yvonne





#### 1. It is very simple (bump hunt)



 It is very simple (bump hunt)
 It is very complex (jet substructure)



#### O(100)-dimensional phase space:

- Particle 3-momenta Vertices
- Particle ID (non-isolated leptons, photons)





With 150/fb, exclusion on 3 TeV dijet resonance is 5000 events



More Model Specific

## Dark Matter Detour

# An Example Target (Dark Matter version)



## An Example Target (Dark Matter version)



# An Example Target (Dark Matter version)



Weak Supervision References:

[1708.02949] E. M. Metodiev, B. Nachman, J. Thaler
 [1702.00414] L. M. Dery, B. Nachman, F. Rubbo, A Schwartzman
 [1801.10158] P. T. Komiske, E. M. Metodiev, B. Nachman, M. D. Schwartz
 [1706.09451] T. Cohen, M. Freytsis, B. Ostdiek



Weak Supervision[1708.02949] E. M. Metodiev, B. Nachman, J. ThalerReferences:[1702.00414] L. M. Dery, B. Nachman, F. Rubbo, A Sc[1801.10158] P. T. Komiske, E. M. Metodiev, B. Nachr

[1702.00414] L. M. Dery, B. Nachman, F. Rubbo, A Schwartzman [1801.10158] P. T. Komiske, E. M. Metodiev, B. Nachman, M. D. Schwartz [1706.09451] T. Cohen, M. Freytsis, B. Ostdiek



Weak Supervision[1708.02949] E. M. Metodiev, B. Nachman, J. ThalerReferences:[1702.00414] L. M. Dery, B. Nachman, F. Rubbo, A Schwartzman[1801.10158] P. T. Komiske, E. M. Metodiev, B. Nachman, M. D. Schwartz[1706.09451] T. Cohen, M. Freytsis, B. Ostdiek



## **Cross Validation**













## No Signal → No Bump!























# **Signal Characteristics**



## Summary

- 1) Factorize space of observables into:
  - a) One **test observable** (e.g. mJJ) in which bg is smooth and signal has a sharp feature (*doesn't need to be a bump*).
  - b) An additional space of **auxiliary observables** (either particle4-vectors or expert features).
- 2) May need to decorrelate auxiliary observables form test.
- 3) Define signal and sideband regions based on test observable
- 4) **Train NN on auxiliary observables** to discriminate sideband from signal region.
- 5) Use NN output as a selection cut to select events in a statistically independent sample.
- 6) Perform a shape-based hypothesis test on the test observable.



# ML with mixed samples

[1708.02949] E. M. Metodiev, B. Nachman, J. Thaler
[1702.00414] L. M. Dery, B. Nachman, F. Rubbo, A Schwartzman
[1801.10158] P. T. Komiske, E. M. Metodiev, B. Nachman, M. D. Schwartz
[1706.09451] T. Cohen, M. Freytsis, B. Ostdiek

2

2

p(data|1)p(data|0)



## **Machine Learning for Jets**

#### Simulation $\neq$ data



Figure taken from Ben Nachman's talk at BOOST 2018 https://indico.cern.ch/event/649482/contributions/2993322/attachments/1688082/2715256/WeakSupervi sion\_BOOST2018.pdf

## **Anomaly Detection Landscape**

#### "Anomalous Event Detection"



#### Autoencoders (weak supervision)

[1808.08992] Marco Farina, Yuichiro Nakai, David Shih [1808.08979] Theo Heimel, Gregor Kasieczka, Tilman Plehn, Antonio Aquilar-Saavedra Jennifer M. Thompson [1811.10276] Olmo Cerri, Thong Nguyen, Maurizio Pierini, Maria Spiropulu, Jean-Roch Vlimant [1903.02032] Tuhin Roy, Aravind Vijay

#### 'Model independent training sample' (fully supervised)

[1709.01087] Jack H Collins, Rashmish Mishra, Juan

(See also [1707.07084] Amit Chakraborty, Abhishek Iyer, Tuhin Roy for similar, non-ML ideas)

## **Anomaly Detection Landscape**



Data vs Data Or Simulation vs data

# Background-only training vs signal/sideband:

#### **Background-only**

Tagger performance does not depend on signal statistics.

Tagger can never learn the *specific* peculiar features of the signal, and so **cannot improve with greater signal rate**.

#### Signal / Sideband

Tagger relies on there being sufficient signal statistics for training.

Tagger can learn the *specific* peculiar features of the signal, and so **improves with greater signal rate**, and allows for **signal characterization**.



??



Stronger in limit of very high signal statistics



## **Performance Comparison**



Figure 11. Truth-label ROC curves for taggers trained using CWoLa with varying number of signal events, compared to those for a dedicated tagger trained on pure signal and background samples (solid black) and one trained to discriminate W and Z jets from QCD (dashed black). The CWoLa examples have B = 81341 in the signal region and S = (230, 352, 472, 697).

## **Nested Cross-Validation**



Figure 7. Illustration of the nested cross-validation procedure. Left: the dataset is randomly partitioned bin-by-bin into five groups. Center: for each group, an ensemble classifier is trained on the remaining groups. For each of the four possible combinations of these four groups into three training groups and one validation group, a set of invidual classifiers are trained and the one with best validation performance is selected. The ensemble classifier is formed by the average of the four selected individual classifiers. Right: Data are selected from each test group using a threshold cut from their corresponding ensemble classifier. The selected events are then merged into a single  $m_{JJ}$  histogram.

## **Toy Statistics**

$$\mathcal{L}(\mu, \theta) = \text{Poiss}(n|b + \theta + \mu)e^{-\theta^2/(2\sigma^2)}$$

$$\lambda_0 = \frac{\mathcal{L}(\mu = 0, \hat{\theta})}{\mathcal{L}(\hat{\mu}, \hat{\theta})}$$

