An Emerging Jets Analysis With Boosted Decision Trees

Jérémie LePage-Bourbonnais

Carleton University

Supervisor: Kevin Graham

Outline

- Emerging Jet Models
- ATLAS Run 2 Analysis Summary
- Progress Towards Run 3 Analysis with BDTs

Field	$SU(3) \times SU(2) \times U(1)$	$SU(3)_{\rm dark}$	Mass	Spin
Q_d	(1, 1, 0)	(3)	$m_d \ \mathcal{O}(\text{GeV})$	Dirac Fermion
X_d	$(3,1,rac{1}{3})$	(3)	$M_{X_d} \mathcal{O}(\text{TeV})$	Complex Scalar
Z_d	(1,1,0)	(1)	$M_{Z_d} \mathcal{O}(\text{TeV})$	Vector Boson



Dark QCD

 $\mathcal{L} \supset \bar{Q}_{d_i} (\not\!\!\!D - m_{d_i}) Q_{d_i} + (D_\mu X_d) (D^\mu X_d)^{\dagger} - M_{X_d}^2 X_d X_d^{\dagger} - \frac{1}{4} G_d^{\mu\nu} G_{\mu\nu,d} + \mathcal{L}_\kappa + \mathcal{L}_{\rm SM}$

P. Schwaller, D. Stolarski, A. Weiler - arxiv:1502.05409

Emerging Jets Models



- Dark jets "emerge" as dark mesons decay to SM particles
- Look for "four jets" and many displaced vertices



Two emerging jets shown

P. Schwaller, D. Stolarski, A. Weiler - <u>arxiv:1502.05409</u>

Emerging Jets Models



- Dark jets "emerge" as dark mesons decay to SM particles
- Look for "four jets" and many displaced vertices



Signal MC Event

P. Schwaller, D. Stolarski, A. Weiler - <u>arxiv:1502.05409</u>

Simulated Signals/Backgrounds

• 90 Model points generated uses modified Pythia 8 HiddenValley modules

	Model A	Model B	Model C	Model D	Model E	
Λ_d [GeV]	10	4	20	40	1.6	
m_{ρ_d} [GeV]	20	8	40	80	3.2	
m_{π_d} [GeV]	5	2	10	20	0.8	
m_{X_d} [GeV]	1400 1000 600					
$c\tau_{\pi_d}$ [mm] 300, 150, 75, 20, 5, 2 300, 150, 75, 5, 2, 1 300, 150, 20, 2, 1, 0					0, 150, 20, 2, 1, 0.5	





P. Schwaller, D. Stolarski, A. Weiler - arxiv:1502.05409

- Uses a SU(3) single flavoured Pythia model for the dark QCD sector
- Backgrounds assumed to be completely dominated by multi-jet
- Events pass through GEANT4 ATLAS detector reconstruction
- Background MC is JZ4W slice: 410 < pT < 790 GeV, other slices are currently processing

Run 2 Analysis

- Model independent approach
- Pre-Selection cuts
 - Jet multiplicity >= 4
 - 4 Leading jets each have $pT \ge 120 \text{ GeV}$
 - 4 Leading Jets each have $|\eta| \le 2.5$
 - Scalar sum of 4-Jet pT >= 1000 GeV
- ABCD implementation for data driven
 background estimate





ABCD Background Estimation Method

- Data-driven background estimation method
- Cut positions may change with new MC background slices
- For un-correlated ABCD plane variables:

$$\frac{N_C^{bkg}}{N_D^{bkg}} \underbrace{\begin{array}{c} N_A^{bkg} \\ N_B^{bkg} \end{array}}_{N_B^{bkg}}$$

 Can solve for background estimate in region A





D. Rocha / G. Gonella

BDT Motivation

- Model independent ABCD background estimation approach doesn't necessarily maximize separation between signal and background
- See what happens if machine learning approach is implemented to classify events
- Use additional weakly separating variables

Tree Training Flow

- Dark Jets are truth level jet objects composed solely of dark particles.
- Look for reconstructed jets within an angular separation of $\Delta R < 0.1$, "match" with the highest pT reconstructed jet found

<u>ROOT TMVA BDT Implementation</u> – Default settings (e.g. 800 trees, max height 3)

- Select Training/Testing Data
 - Apply pre-selection cuts on available data
 - Randomly select fraction of events to train on
- Train Jet Level Trees
 - Two cases:
 - Train on all **dark-matched signal jets**, all background jets
 - Train on **4 leading dark-matched signal jets**, all background jets
- Train Event Level Trees
 - Use information from jet level trees + event level variables
 - Two cases:
 - Without including HT (for ABCD comparison)
 - Including HT
- Apply to 15 signal points independently

Jet-Level Input Variable Distributions Model: A 1400 GeV 20 mm width = $\frac{1}{p_T^{jet}} \sum_i p_T^i \Delta R_i$ TMVA Input Variable: Jet Eta TMVA Input Variable: Jet Width TMVA Input Variable: Number of Jet Tracks 0.08 Signal Signal Signal Normalized Scale Normalized Scale Normalized Scale 0.14 0.08 0.07 Background Background Background 0.12 0.07 0.06 0.06 0.1 0.05 0.05 0.08 0.04 0.04 0.06 TMVA Input Variable: Jet Mass 0.03 0.03 0.04 Signal 0.02 Normalized Scale 0.25 0.02 Background 0.02 0.01 0.01 0.2 0.25 150 200 250 -2 2 0.05 0.1 0.15 0.2 50 100 Ω 0 Jet Eta Jet Width Number of Jet Tracks 0.15 Trained on all dark jets 0. 0.05 TMVA Input Variable: Jet EM Fraction TMVA Input Variable: Jet CentroidR TMVA Input Variable: Number of Jet Secondary Vertices Signal Signal Signal Normalized Scale Normalized Scale Normalized Scale 0 0.09 0.14 150 200 250 Background Background ٥ 50 100 Background 0.08 Mass [GeV] 0.12 0.07 0.8 0.1 0.06 0.08 0.6 0.05 0.04 0.06 0.4 0.03 E 0.04 0.02 0.2 0.02 0.01E n 0 0.2 0.4 0.6 0.8 200025003000350040004500500055006000 2 3 4 5 6 7 8 9 10 Õ Number of Jet Secondary Vertices CentroidR [mm]

Jet EM Fraction

Jet Level Trees

Trained on all dark jets

TMVA response for classifier: BDT

Trained 4 leading dark jets

TMVA response for classifier: BDT



Event level BDT input variables Output from jet-level BDT



Model: A 1400 GeV 20 mm

Model: A 1400 GeV 20 mm

Event Level Trained on all dark jets Trained 4 leading dark jets Trees TMVA response for classifier: BDT TMVA response for classifier: BDT Normalised Scale 80.0 80.0 Normalised Scale 80.0 80.0 Full Signal (test sample) Signal (training sample) Full Signal (test sample) Signal (training sample) Background (test sample) Background (training sample) Background (test sample) Background (training sample) Including event HT 0.06 0.06 0.04 0.04 0.02 0.02 0 0.2 -0.2 0.2 0.4 -0.2 0.4 -0.4-0.4 0 n BDT response **BDT** response TMVA response for classifier: BDT TMVA response for classifier: BDT Normalised Scale 80.0 900 Full Signal (test sample) Normalised Scale 80.0 90.0 Signal (training sample) Full Signal (test sample) Signal (training sample) Background (test sample) Background (training sample) Background (test sample) Background (training sample) Not Including event HT 0.04 0.04 0.02 0.02 0 0.2 0.4 -0.4-0.20 -0.2 0.2 0.4 -0.40

BDT response

2 0.4 JL BDT response

ABCD Comparisons



ABCD Summary Table

Trained on all dark matched jets HT not included in training

16

	Model	Signal Cross-section [fb]	Region A Signal Yield	Region A Weighted Background Yield	Signal Efficiency	Background Efficiency
	A 1400 GeV 20 mm	1.1	39.2+/-0.3	6662.0+/-652.3	0.948	0.015
/ 	A 1000 GeV 150 mm	15	582.0+/-3.3	14929.1+/-1339.0	0.778	0.033
	A 600 GeV 2 mm	430	4390.6+/-23.8	28838.1+/-1546.3	0.609	0.064
	B 1400 GeV 20 mm	1.1	32.6+/-0.2	9864.3+/-944.9	0.932	0.022
	B 1000 GeV 5 mm	15	864.3+/-4.3	13612.5+/-1315.7	0.876	0.030
	B 600 GeV 300 mm	430	815.0+/-13.2	22376.4+/-1547.9	0.454	0.050
	C 1400 GeV 75 mm	1.1	47.7+/-0.3	7332.5+/-908.5	0.954	0.016
	C 1000 GeV 5 mm	15	796.4+/-3.3	7399.6+/-1035.0	0.896	0.016
	C 600 GeV 2 mm	430	3344.2+/-27.3	18704.7+/-1417.6	0.692	0.042
	D 1400 GeV 2 mm	1.1	58.6+/-0.3	6529.5+/-667.2	0.967	0.015
	D 1000 GeV 1 mm	15	675.8+/-4.3	15762.7+/-1262.1	0.899	0.035
	D 600 GeV 300 mm	430	1979.6+/-21.9	17023.7+/-1328.8	0.739	0.038
	E 1400 GeV 75 mm	1.1	16.8+/-0.2	21367.4+/-1541.7	0.844	0.047
	E 1000 GeV 150 mm	15	306.4+/-2.7	23450.8+/-1479.9	0.602	0.051
	E 600 GeV 0.5 mm	430	5780.7+/-42.4	44117.9+/-2099.5	0.525	0.098

Cuts: HT: 1375 BDT: 0.0

Cuts on BDT Scores

Trained on all dark matched jets HT included in training

Cut: BDT > 0.2

Model	Cross-Section [fb]	Signal Events Above Cut	Background Events Above Cut	Signal Efficiency	Background Efficiency
A 1400 GeV 20 mm	1.1	31.8+/-0.2	629.9+/-420.8	0.770	0.001
A 1000 GeV 150 mm	15	419.5+/-2.8	343.9+/-136.2	0.560	0.001
A 600 GeV 2 mm	430	1772.7+/-15.1	659.4+/-209.3	0.246	0.001
B 1400 GeV 20 mm	1.1	23.0+/-0.2	151.5+/-58.4	0.658	0.000
B 1000 GeV 5 mm	15	507.0+/-3.3	193.3+/-71.7	0.514	0.000
B 600 GeV 300 mm	430	744.3+/-12.6	2874.2+/-1146.3	0.414	0.006
C 1400 GeV 75 mm	1.1	37.9+/-0.2	162.1+/-124.8	0.757	0.000
C 1000 GeV 5 mm	15	573.3+/-2.8	38.1+/-21.0	0.645	0.000
C 600 GeV 2 mm	430	2177.9+/-22.0	971.8+/-332.6	0.451	0.002
D 1400 GeV 2 mm	1.1	47.0+/-0.2	80.3+/-33.4	0.775	0.000
D 1000 GeV 1 mm	15	370.5+/-3.2	168.0+/-52.0	0.493	0.000
D 600 GeV 300 mm	430	1307.2+/-17.8	215.8+/-139.2	0.488	0.000
E 1400 GeV 75 mm	1.1	8.5+/-0.1	251.0+/-97.6	0.425	0.001
E 1000 GeV 150 mm	15	197.2+/-2.2	2340.4+/-781.2	0.387	0.005
E 600 GeV 0.5 mm	430	1648.5+/-22.6	4462.5+/-897.0	0.150	0.010

Conclusions

- Hidden valley models with emerging jets provide an interesting and unique signature that can be searched for at the LHC
- Run 2 model independent analysis progressing
- Progress towards a Run 3 sensitivity study started
 - Use of ML tools to optimize sensitivity looks promising
 - Could require alternate background estimate technique
 - Will necessitate careful evaluation of MC systematics
- Next Steps:
 - Evaluate additional classification variables
 - Explore alternate ML methods
 - Evaluate effects of MC systematics
 - Consider alternate background estimate strategies
- Lots of scope for expanding model space

Acknowledgements



Emerging Jets Members: J. Beacham, B. Death, G. Gonella, K. Graham, A. Haas, J. Heilman, A. Kilgallon, M. Naseri, I. Ramirez-Berend, D. Rocha, D. Strom, C. Treado

Thank you to NSERC for support under USRA program

Extra Slides



Decay Structure

- Dark mediators decay to 2 SM down quarks and 2 dark quarks
- Dark quarks undergo QCD-like showering and hadronization
- Dark mesons decay to SM particles
- This process forms multiple displaced vertices and sub-jets



P. Schwaller, D. Stolarski, A. Weiler - arxiv:1502.05409

Boosted Decision Trees

- Decision Trees Classify events through a series of yes/no questions
- Boosting Train many trees, ensemble scores events through weighted sum
- + Default ROOT TMVA BDT settings 800 trees, maximum height of 3



Jet-Level Input Variable Correlations

Model: A 1400 GeV 20 mm



Correlation Matrix (signal)



Correlation Matrix (background)

Event-Level Input Variable Correlations

Model: A 1400 GeV 20 mm

Linear correlation coefficients in % 100 event_njet 100 80 60 BDT[3] 8 -2 100 -2 40 BDT[2] -3 -3 11 100 20 BDT[1] -8 100 0 -20 -8 -2 BDT[0] 24 100 -40 SecVtxCount 24 100 -60 -80 event HT 100 15 11 -100event_nSecVtxCount BDT[2] BDT[1] BDT[3] event_njet event_HT

Correlation Matrix (signal)

Correlation Matrix (background)





Jet-Level Input Variable Distributions Model: B 1000 GeV 5 mm width = $\frac{1}{p_T^{jet}} \sum_i p_T^i \Delta R_i$ TMVA Input Variable: Jet Eta TMVA Input Variable: Jet Width TMVA Input Variable: Number of Jet Tracks 0.08 Signal Signal Signal Normalized Scale Normalized Scale Normalized Scale 0.14 0.08 Background Background Background 0.07 0.07 0.12 0.06 0.06 0.1 0.05 0.05 0.08 0.04 0.04 0.06 TMVA Input Variable: Jet Mass 0.03 0.03 0.04 Signal 0.02 Normalized Scale 0.02 0.2 Background 0.02 0.01 0.01 0.18 0.16 0.25 150 200 250 -2 2 0.05 0.1 0.15 0.2 50 100 Ω n 0.14 Jet Eta Jet Width Number of Jet Tracks 0.12 0.1 Trained on all dark jets 0.08 0.06 TMVA Input Variable: Jet EM Fraction 0.04 TMVA Input Variable: Jet CentroidR TMVA Input Variable: Number of Jet Secondary Vertices 0.02 Signal Signal Signal Scale Normalized Scale Normalized Scale 0 40 60 80 100 120 140 160 180 200 220 Background ٥ 20 Background Background 0.08 0.12 Mass [GeV] Vormalized 0.07 0.1 0.8 0.06 0.08 0.05 0.6 0.04 0.06 0.4 0.03 0.04 0.02 0.2 0.02 0.01 0 0 2000 2500 3000 3500 4000 4500 5000 5500 0.2 0.4 0.6 0.8 2 3 7 8 9 Õ 5 6

CentroidR [mm]

Jet EM Fraction

Number of Jet Secondary Vertices

Jet-Level Input Variable Correlations

Model: B 1000 GeV 5 mm



Correlation Matrix (signal)

Correlation Matrix (background)



Jet Level Trees

Trained on all dark jets

TMVA response for classifier: BDT

Trained 4 leading dark jets

TMVA response for classifier: BDT



Event level BDT input variables Output from jet-level BDT



Model: B 1000 GeV 5 mm

Event-Level Input Variable Correlations

Model: B 1000 GeV 5 mm



Correlation Matrix (signal)

Correlation Matrix (background)

100

80

60

40

20

n

-20

-40

-60

-80

-100

100

5

3

2

10

event_njet

Model: B 1000 GeV 5 mm

Trained 4 leading dark jets

31



Event Level



Jet-Level Input Variable Distributions Model: D 600 GeV 300 mm width $= \frac{1}{p_T^{jet}} \sum_i p_T^i \Delta R_i$ TMVA Input Variable: Jet Eta TMVA Input Variable: Jet Width TMVA Input Variable: Number of Jet Tracks Signal Signal 0.18 Signal 0.08 Normalized Scale Normalized Scale Normalized Scale 0.08 Background Background Background 0.16 0.07 0.07 0.14 -0.06 E 0.06 0.12 0.05 0.05 0.1 0.04 0.04 0.08 TMVA Input Variable: Jet Mass 0.03 0.03 0.06 Signal 0.02 Normalized Scale 0.02 0.04 0.25 H Background 0.01 0.01 0.02 0 0.2 150 200 250 -2 2 0.05 0.1 0.15 0.2 0.25 50 100 Ω 0 Jet Eta Jet Width Number of Jet Tracks 0.15 Trained on all dark jets 0.1 TMVA Input Variable: Jet CentroidR TMVA Input Variable: Jet EM Fraction 0.05 TMVA Input Variable: Number of Jet Secondary Vertices Signal Signal Signal Normalized Scale Normalized Scale Normalized Scale 0 0.12 60 80 100 120 140 160 180 Background ٥ 20 40 Background Background 0.08 Mass [GeV] 0.1 0.07 0.8 0.06 0.08 0.05 0.6 0.06 0.04 0.4 0.03 0.04 0.02E 0.2 0.02 0.01 Ω 200025003000350040004500500055006000 0.2 0.4 0.6 0.8 2 3 5 6 7 8 0 4

Number of Jet Secondary Vertices

CentroidR [mm]

Jet EM Fraction

Jet-Level Input Variable Correlations

Model: D 600 GeV 300 mm



Correlation Matrix (signal)



Correlation Matrix (background)

Jet Level Trees

Trained on all dark jets

TMVA response for classifier: BDT

Trained 4 leading dark jets

TMVA response for classifier: BDT



Event level BDT input variables Output from jet-level BDT



Model: D 600 GeV 300 mm

Event-Level Input Variable Correlations

Model: D 600 GeV 300 mm

Linear correlation coefficients in % 100 event_njet 61 100 80 60 BDT[3] -3 2 3 100 40 BDT[2] 5 2 2 100 20 BDT[1] -3 3 2 3 5 100 0 -20 BDT[0] 6 5 2 2 100 -40 SecVtxCount 6 21 100 -60 -80 event HT 100 61 21 -100event_nSecVtxCount BDT[2] BDT[1] BDT[3] event_njet event_HT

Correlation Matrix (signal)

Correlation Matrix (background)



Model: D 600 GeV 300 mm

Trained on all dark jets

Event Level

Trained 4 leading dark jets





