

$n \pi^+ \pi^- \pi^+$ BDT Study

Tom Boettcher

Indiana University

September 8, 2013

Introduction

Goals

- ▶ Use Boosted Decision Trees to separate signal and background in the $n3\pi$ channel
- ▶ Expand on work done at the July Data Analysis Workshop
 - ▶ Add discriminating variables
 - ▶ Use a larger training sample

Why $n3\pi$?

- ▶ Possible source of signal for $\pi_1(1600)$

What Are Boosted Decision Trees?

- ▶ Event classifier
- ▶ “Studies” training data to learn how to classify events

Signal sample = 250k events

Background sample = 10M events

Require $NDF > 5$ for each π^+ and π^- track for both signal and background

Require KinFit FOM ≥ 0

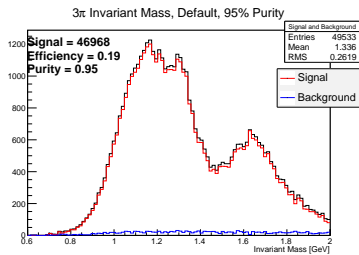
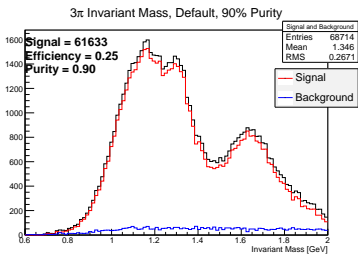
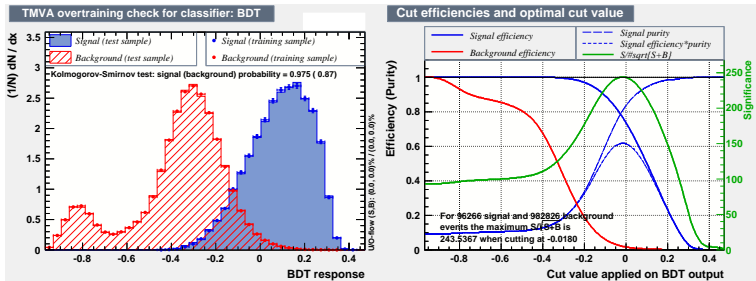
BDT Method Options:

- ▶ NTrees = 100
- ▶ nEventsMin = 100
- ▶ MaxDepth = 4
- ▶ BoostType = AdaBoost
- ▶ SeparationType = GinIndex
- ▶ nCuts = 200
- ▶ PruneMethod = NoPruning

Workshop Variables

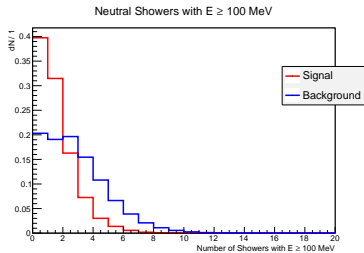
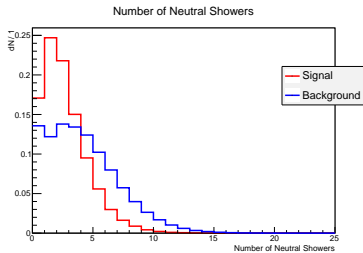
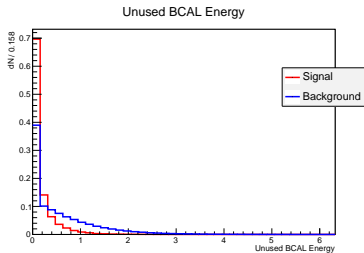
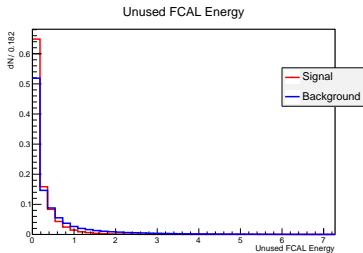
Rank	Variable	Variable Importance
1	MissingNeutron_PT	1.516e-01
2	PV_r	1.388e-01
3	FOM_KinFit	8.862e-02
4	Measured__MissingMass	8.790e-02
5	Unused__Max_Proton_FOM	6.846e-02
6	PiPlus1__Timing_FOM	6.044e-02
7	PiPlus2__Timing_FOM	5.807e-02
8	PiPlus1__NDF_Tracking	5.268e-02
9	PiMinus__NDF_Tracking	4.787e-02
10	PiPlus1__DCdEdx_FOM	4.731e-02
11	PiPlus2__NDF_Tracking	4.592e-02
12	PiPlus2__DCdEdx_FOM	3.917e-02
13	Unused__Max_KMinus_FOM	3.344e-02
14	Unused__Max_KPlus_FOM	2.777e-02
15	PiMinus__Timing_FOM	2.618e-02
16	PiMinus__DCdEdx_FOM	2.575e-02

Study using workshop variables



Same result as the July workshop

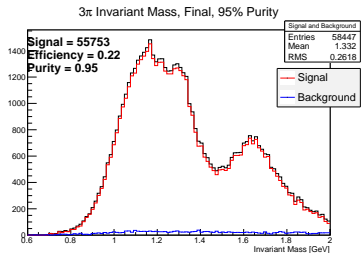
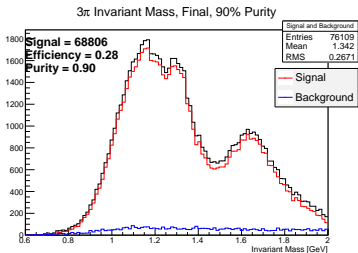
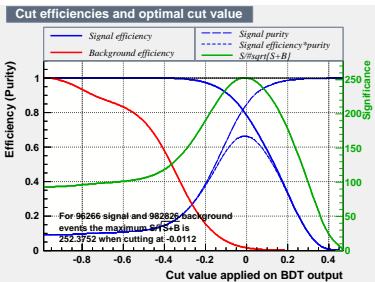
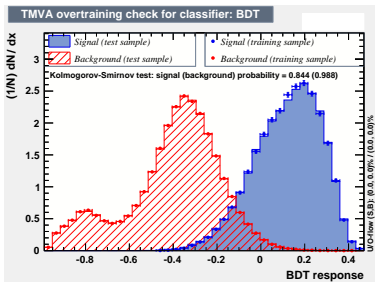
Variables Added



Full Variable Rankings

Rank	Variable	Variable Importance
1	MissingNeutron_PT	1.178e-01
2	PV_r	1.093e-01
3	Unused_BCAL	1.020e-01
4	Measured_MissingMass	8.575e-02
5	FOM_KinFit	7.115e-02
6	PiPlus1_Timing_FOM	5.654e-02
7	PiPlus1_NDF_Tracking	5.386e-02
8	Unused_FCAL	4.545e-02
9	PiPlus2_NDF_Tracking	4.378e-02
10	Unused_Max_Proton_FOM	4.216e-02
11	PiPlus2_Timing_FOM	4.106e-02
12	NeutralShowers_HighE	3.579e-02
13	PiPlus1_DCdEdx_FOM	3.554e-02
14	PiMinus_NDF_Tracking	3.418e-02
15	Unused_Max_KMinus_FOM	3.223e-02
16	PiMinus_DCdEdx_FOM	2.429e-02
17	PiPlus2_DCdEdx_FOM	2.400e-02
18	PiMinus_Timing_FOM	2.112e-02
19	Unused_Max_KPlus_FOM	1.229e-02
20	NeutralShowers	1.168e-02

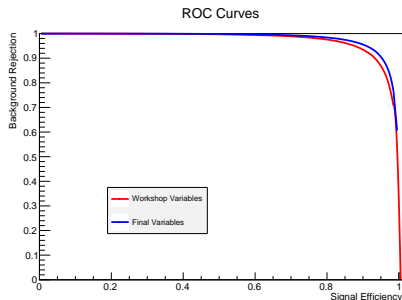
Study using added variables



Comparison

Study	Efficiency at 90% Purity	Efficiency at 95% Purity
Workshop	25%	19%
Current Study	28%	22%

Early 2011 Study 26% Efficiency at 67% Purity



Summary

3% increase in efficiency at both 90% and 95% purity since the July workshop

Larger training sample helps prevent overtraining

Further Study

- ▶ Look at remaining background events to find more potential discriminating variables
- ▶ Perform an amplitude analysis on the $n3\pi$ channel
- ▶ Repeat with other channels of interest