BDT for JpsiX⁰ modes

Feb. 27th 2019, Annecy/Edinburgh meeting, M. Chefdeville

- During last 2 weeks:
 - Finished processing of Run I DiMuon data \rightarrow Ntuples
 - One week off, resumed work this Monday
 - Started to develop a BDT for neutral modes
- BDT variables
 - Kinematics & vertex (some previously used in Fisher)
 - Calo variables (CL, cone), and also VtxIsol
 - Signal photons more isolated than bkg (jet-like)
 - \rightarrow potentially strong <code>discriminating power</code>
 - Pile-up dependent
 - \rightarrow will tend to select cleaner photons (i.e. <u>better resolution</u>, less combinatorics)
 - \rightarrow not well reproduced by MC (calibration?)
- Calibration strategy: compare efficiency when training data sideband (bkg) VS <u>sPlot-data or MC-truth</u> (signal)

Status

- Use abundant JpsiK*+[K+pi0]
- Compare sWeighted data variables to MC variables
- Assign systematics to using BDT trained on MC-truth

$B^+ \rightarrow J/psi K^{*+}[K^+pi0]$, Run II

- Selections
 - Ntupling: CL(g1,g2)>0.05, PT(pi0)>0.8 GeV/c, PROBNNk(K+)>0.1, Δm(K*+)=150 MeV/c², Δm(pi0)=30 MeV/c², Δm(Jpsi)=100 MeV/c² Fisher(B+)>-1.1, DIRA>0.9995, IP<0.2, IPCHI2<20, VTXCHI2/NDOF<10
 - Offline: PT(pi0)>1.5 GeV/c, aligned with Jpsipi0 (will be used for BR meast.)



BDT: MC-truth VS data sideband

- Samples
 - Signal: MC JpsiK*+ 2015-16 Up-Dw, Sim09e, Pythia8, s28r1 (4M evt)
 - \rightarrow 50k evt after selections
 - Background: Run II data sideband (Bplus_MM>6650)

Cut tuned such as to have similar stat as signal

1.1.1	Rank :	Variable	:	Variable Importance
	1 : 2 : 3 : 4 : 5 : 6 : 7 : 8 : 9 : 10 : 11 : 12 : 13 : 14 : 15 : 16 :	Bplus_PT Bplus_DIRA min_MIPCHI2DV pi0_PT Bplus_IP max_MIPCHI2DV Bplus_VTXCHI2 Jpsi_DIRA log(Bplus_SmallestDeltaChi2OneTrack) min_ISNotH pi0_0.40_IT max_ISNotE pi0_0.40_nc_mult-2+pi0_0.40_cc_mult Jpsi_IP min_ISNotE max_ISNotE max_ISNotE max_ISNotE		8.746e-02 8.154e-02 8.023e-02 7.177e-02 6.034e-02 5.983e-02 5.951e-02 5.765e-02 5.732e-02 5.340e-02 5.339e-02 5.170e-02 4.950e-02 4.660e-02 4.469e-02 4.325e-02
:	17 :	Bplus_NumVtxWithinChi2WindowOneTrack	:	4.183e-02



BDT variables

- Correlations
 - pi0 transverse isolation & Nobjects in cone
 - Photon PID (especially for true pi0)
 - IPCHI2 of muons (especially for true Jpsi)
- Will consider removing some variables (performance VS Nvar)

									al)	gn	si	K (9	rix	at	M	on	tic	ela	Corre
1%100	Linear correlation coefficients in % 100																		
100	0	100	-46	8	-7		1	1		49	43		3	-42	47	-13	22	9	ni2OneTrack)
80	6	-46	100	-7	7			-1		-24	-21	-10	-3	19	-24	4		-4	lowOneTrack
	3	8	-7	100	-82	4						-2			3		-5	9	pi0_0.40_IT
60	7	-7	7	-82	100	-6	-7		-6		2		-6	-3	-2	-2	11	9	0.40_cc_mult
10				4	-6	100	49	24	9				-1	2		5	-4	-7	max_IsNotE
-40	1	1			-7		100	24	15					-3		-2	-4	-4	min_IsNotE
-20	1	1			-5	24	24	100	48	2				-4		-3	-4	-1	max_IsNotH
					-6		15	48	100	1				-2	1	-2	-4	-1	min_IsNotH
0	9	49	-24					2	1	100	73		6	-60	83	-3	2	6	_MIPCHI2DV
	3	43	-21		2					73	100	1	3	-53	72		7	9	_MIPCHI2DV
-20			-10	-2							1	100	4			-2			us_VTXCHI2
40	3	3	-3		-6	-1				6	3	4	100	54	14	4	-18	-5	Bplus_IP
	2	-42	19		-3	2	-3	-4	-2	-60	-53		54	100	-55	39	-25	-7	Bplus_DIRA
-60	7 –	47	-24	3	-2		1		1	83	72	1	14	-55	100	22	-12		Jpsi_IP
	3	-13	4		-2		-2	-3	-2	-3		-2	4	39	22	100	-39	-7	Jpsi_DIRA
-80	2	22	-9	-5	11	-4	-4	-4	-4	2	7	1	-18	-25	-12	-39	100	44	Bplus_PT
10(9	9	-4	9	9	-7	-4	-1	-1	6	9		-5	-7		-7	44	100	pi0_PT
-100 304	g(Bn		Bo	2.80	axpi	n.M	axm.	in Te	axm	in Di		ou Br	HI BA	si ^B h	Silp	HUE	OB	Þi	
Unit Stx SU	thun mult	ic t	E n	Vot	HP II	SV∂į SV	R	Y E	EA	V73	A	UIR	- 0	"RA	MA	, o			
Z+pio_0																			

					-														
Linear correlation coefficients in %																			
ni2OneTrack)		11	-5	9	-2	13	-5	8	10	1	2	1	5	-10	8	-39	100		100
lowOneTrack	-9	-13	2	-18		-4	-42	-12	-14	-2	-2	-2			-7	100	-39	_	80
pi0_0.40_IT	5		-3	2		4	-2	10		8	7			-80	100	-7	8		
0.40_cc_mult	8		2			-2				-8	-7		-8	100	-80		-10	-	60
max_IsNotE	1	1	2	-3	4	-4	-4			24	40	55	100	-8			5		10
min_lsNotE	3				2			3		33	36	100		-9	8	-2	1		40
max_IsNotH	6		3		2	-2		3		57	100	36	40	-7	7	-2	2	_	20
min_lsNotH	6		2		2	-1		2	3	100	57	33	24	-8	8	-2	1		
_MIPCHI2DV	2		25	46	-1	24	7	35	100	3	4			-9	9	-14	10	-	0
_MIPCHI2DV	1	13		14	-6	14	4	100	35	2	3	3		-9	10	-12	8		
us_VTXCHI2	14	6		23	-12	-9	100	4	7				-4	5	-2	-42	-5		-20
Bplus_IP	-1	-2	-18	43	35	100	-9	14	24	-1	-2		-4	-2	4	-4	13		-40
Bplus_DIRA	11	11	17	-5	100	35	-12	-6		2	2	2	4				-2		
Jpsi_IP	-1	-22	47	100	-5	43	23	14	46				-3		2	-18	9	-	-60
Jpsi_DIRA	6	-23	100	47	17	-18			25	2	3		2	2	-3	2	-5		
Bplus_PT	30	100	-23	-22	11	-2		13								-13	11		-80
pi0_PT	100	30	6	-1	11	-1	14	1	2	6	6	3	1	8	5	-9			_100
	Pi	B	HUS	Silp	si ^B A	y Br	DIL BA		in Mi	axm	n.De	axm/	n Ma	ax ^{Di}	2, 819	BA	ulgo	Bn	100
		- 1		pp	"RA		UIR	A	0-TX	EA	J2	R	SV SV	Noti	Vot	E n		NUM Ult- 2	Stx SPIF
																		. <	+Pi0_0.

Correlation Matrix (background)

BDT performance



- Now, to compare simulated VS real performance, let's fit the B+ mass. Then:
 - Determine sWeighted BDT distribution, or
 - Cut & fit & count

Fit model

- Signal S1: double-CB with tails fixed on 2016 $\ensuremath{\mathsf{MC}}$
- S2: 1 signal photon + 1 random photon
 - \rightarrow Gaussian with expo. tails fixed on same MC, S1/(S1+S2) ratio = 0.833, only σ & yield floating
- B1: combinatorial bkg: exponential
- B2: part. reco'ed B⁰ \rightarrow JpsiK^{*}(pi): HistoPdf using RapidSim mass shape (13 TeV, PT cuts)
- B3: part. reco'ed B₊ \rightarrow Psi(2S)[Jpsi (pipi)]K^{*}: HistoPdf using RapidSim mass shape
- B4: peaking $B^+ \rightarrow JpsiK^+$ combined with random pi0. Use fullsim sample Sim09e, s28r1, 2016 (4M) Ignored for now (yield under investigation)







RapidSim, apparté

- FullSim, no mass constrain:
 - Signal σ = 41.5 MeV/c²
- RapidSim, after tuning calo constants on other decays:
 - Signal σ = 39.05 MeV/c²
- \rightarrow Peak width compatible (but no pile-up tails in RapidSim)



Fit result

- Some parameters:
 - $N_s = 93.8 \times 10^3$ signal B⁺ (good enough for BDT training)
 - Signal σ = 33.0 MeV/c² VS 28.9 MeV/c² in full-sim MC
- Part. reco'ed RapidSim B^o shape to be improved (Argus⊗Gauss)



Fit result

- Some parameters:
 - $N_s = 93.8 \times 10^3$ signal B⁺ (good enough for BDT training)
 - Signal σ = 33.0 MeV/c² VS 28.9 MeV/c² in full-sim MC
- Part. reco'ed RapidSim B^o shape to be improved (Argus⊗Gauss)



Next

1) Fit B⁺ mass @ various BDT cuts

- Compare efficiency: MC VS real data
- Quantify discrepancy (can we leave with it?)
- 2) Alternative: train BDT on sWeighted data signal
 - sPlot B+ mass
 - Use sWeights in TMVA

3) Apply BDT to other modes

- If 1) \rightarrow train on signal MC
- If 2) \rightarrow find a way to export JpsiK* BDT