Progress Report on Tau Final States of TTTT

Fabio lemmi¹ Huiling Hua¹ Hongbo Liao¹ Hideki Okawa² Yu Zhang²

¹IHEP

²Fudan University

IHEP Group Meeting, 2020

IHEP Group Meeting

(I) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1))

Outline



MVA sotup and

- MVA setup and stradgy
- Correlation based selection
- Iterative Removal Method



MVA Stragdy

- Goal
 - Use the smallest set of MVA input variables to get the best performance
- Three methods we are considering
 - Correlation based selection
 - Iterative removal method
 - Combine both

BDT Study

TMVA Setup

ROOT version

- Switched to ROOT6.12/07, TMVA version 4.3.0
- New TMVA version comes with new feature

Training setup

- Signal: TTTT_TuneCP5;
- Background: all bg excerpt H and HH and minor ones
- 70% goes to training and 30% goes to testing
- Global weight and event weight same in event yield calculation
- have added more interesting variables:
 - sphericity, aplanarity
 - added some nonjets variables
- Dealing with negetive weight events :InverseBoostNegWeights(Boost With inverse boostweight), Boost With inverse boostweight(Pair events with negative and positive weights in traning sample and *annihilate* them)
- 4 boosting algorithm: BDT(A), BDTG, BDTB, BDTD (all use the default InverseBoostNegWeights)

Hyperparameters

Using the default

Step by Step Correlation Selection

- Step 1: choose the 20 most powerful variables as input to BDT
- Step 2: keep only 1 variable from pairs with correlation>90, see the performance
- Step 3: further remove variables with correlation>80, do the training
- Step 4: add more variables on the basis of step3, repeat step2 ans step3, see if we can gain any AUC(area under ROC curve).

(I) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1))

Variables Separation Power Ranking(1Tau1E)

Rank	:	Variable	:	Separation
		jets_bScore		2.686e-01
		jets_4largestBscoreSum		2.571e-01
		bjetsM_3pt		2.467e-01
		bjetsM_num		2.435e-01
				2.334e-01
				2.279e-01
				2.279e-01
		bjetsL_4pt		
10				1.968e-01
				1.941e-01
		bjetsL_invariantMass		1.833e-01
14		bjetsL_transMass		
		toptagger_HT		
				1.771e-01
				1.765e-01
18		toptagger_transMass		1.730e-01
19		bjetsM_invariantMass		
20		jets_6pt		
				1.569e-01
				1.551e-01
		jetL_7phi		
24				
				1.489e-01
				1.475e-01
				1.475e-01
28	:	bjetsL_3phi	1	1.458e-01

28 :	bjetsL 3phi	:	1.458e-01
29 :			1.437e-01
30 :			1.436e-01
31 :			1.416e-01
32 :			1.407e-01
33 :	bjetsT_3pt		1.407e-01
34 :			1.406e-01
35 :			1.394e-01
36 :			1.389e-01
37 :	bjetsT_3phi		1.388e-01
38 :	toptagger 2pt		1.338e-01
39 :	toptagger num		1.328e-01
40 :	jets 4pt		1.275e-01
41 :	nonbjetsM num		1.233e-01
42 :	bjetsL 2pt		1.181e-01
43 :	toptagger_minDeltaR_v1		1.157e-01
44 :			1.106e-01
45 :	toptagger_2eta		1.084e-01
46 :	toptagger_2phi		1.082e-01
	bjetsT_HT		1.068e-01
48 :	bjetsM_4pt		9.897e-02
49 :	bjetsM_4eta		9.849e-02
50 :	bjetsM_4phi		9.828e-02
51 :	nonbjetsT_num		9.777e-02
	bjetsT_transMass		9.641e-02
	bjetsL_minDeltaR		9.494e-02
54 :	jets_rationHT_4toRest		9.437e-02
	jets_3pt		9.341e-02
56 :	bjetsM_2pt		8.773e-02

イロト イヨト イヨト イヨト

Step1: Input Variables and Correlation

Rank	:	Variable	:	Separation
1		jets bScore		2.686e-01
		jets_4largestBscoreSum		2.571e-01
		bjetsM_3pt		2.467e-01
		bjetsM_num		2.435e-01
		bjetsL_num		2.279e-01
		bjetsL_3pt		2.279e-01
		bjetsL_4pt		1.984e-01
8		jets_number		1.968e-01
		bjetsL_HT		1.941e-01
10		jets_7pt		1.939e-01
11		bjetsL_invariantMass		1.833e-01
12		bjetsL_transMass		1.811e-01
13		toptagger_HT		1.811e-01
14		toptagger_transMass		1.730e-01
15		bjetsM_invariantMass		1.696e-01
16		jets_6pt		1.632e-01
17		bjetsM_HT		1.569e-01
18		jets_8pt		1.510e-01
19		jets_5pt		1.475e-01
20		jets_transMass		1.475e-01

Corre	elation Matrix (signal)				
	Linear correlation coefficients in % 10				
ts_transMass	11 9 17 9 16 36 21 36 72 32 58 74 24 24 48 58 61 31 67100	0			
jets_5pt	14 12 20 11 15 35 23 38 43 39 39 42 29 29 33 79 36 35 00 67				
jets_8pt	18 13 10 15 27 11 19 85 10 61 14 9 32 30 8 46 5100 35 31 80	1			
bjetsM_HT	41 39 55 46 26 48 32 5 80 6 65 79 7 7 86 25 00 5 36 61 60				
jets_6pt	15 13 17 12 16 29 21 49 32 50 32 31 32 31 25100 25 46 79 58				
invariantMass	46 45 60 54 30 47 35 9 67 8 73 64 5 5100 25 86 8 33 48				
er_transMass	6 4 4 2 8 9 9 35 10 29 5 10 97100 5 31 7 30 29 24 ⁴⁰	1			
toptagger_HT	6 4 4 2 9 9 9 3710 31 5 9100 97 5 32 7 32 29 24 20				
L_transMass	27 23 33 25 43 61 47 10 98 10 82100 9 10 64 31 79 9 42 74				
nvariantMass	31 26 35 29 53 63 54 16 86 14100 82 5 5 73 32 65 14 39 58				
jets_7pt	15 11 9 12 22 13 18 74 1100 14 10 31 29 8 50 6 61 39 32				
bjetsL_HT	30 26 36 28 47 65 51 1100 11 86 98 10 10 67 32 80 10 43 72	0			
jets_number	19 13 10 15 30 12 2100 11 74 16 10 37 35 9 49 5 85 38 36	0			
bjetsL_4pt	53 50 44 50 82 57 00 21 51 18 54 47 9 9 35 21 32 19 23 21	0			
bjetsL_3pt	42 41 52 39 60100 57 12 65 13 63 61 9 9 47 29 48 11 35 36	0			
bjetsL_num	59 51 44 56 00 60 82 30 47 22 53 43 9 8 30 16 26 27 15 16	0			
bjetsM_num	85 81 83100 56 39 50 15 28 12 29 25 2 2 54 12 46 15 11 9	U			
bjetsM_3pt	72 73100 83 44 52 44 10 36 9 35 33 4 4 60 17 55 10 20 17	0			
stBscoreSum	98 00 73 81 51 41 50 13 26 11 26 23 4 4 45 13 39 13 12 9	U			
jets_bScore	100 98 72 85 59 42 53 19 30 15 31 27 6 6 46 15 41 18 14 11	00			
	ler ler big big big big ler big ler big big ton ton big ler big ler ler ler	~ ~			
Jourge Contraction for the starting of the formation of the starting of the st					

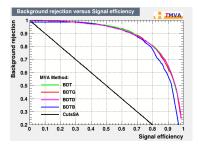
イロト イヨト イヨト イヨト

Huiling Hua (IHEP)

IHEP Group Meeting

▲ ■ ▶ ■ つへの IHEP 2020 7/46

Step 1: ROC



DataSet	MVA							
Name:	Method:	ROC-integ						
dataset	BDT	: 0.884						
dataset	BDTG	: 0.882						
dataset	BDTD	: 0.881						
dataset	BDTB	: 0.860						
dataset	CutsSA	: 0.500						

イロト イロト イヨト イヨト

Huiling Hua (IHEP)

IHEP Group Meeting

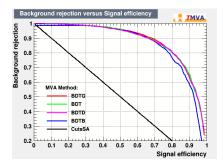
Step 2

- Remove variable with correlation >=90
- jets_bsore, jets_4largestBscoreSum; remove jets_4largestBscoreSum;
- bjetsM_3pt, bjetsM_num, remove bjetsM_num,
- bjetsL_HT, bjetsL_transMass, remove bjetsL_transMass,
- topTagger_HT, toptagger_transMass, remove topTagger_HT,

Variable		Separation
jets_bScore		2.686e-01
bjetsM_3pt		2.467e-01
bjetsL_num		2.279e-01
bjetsL_3pt		2.279e-01
bjetsL_4pt		1.984e-01
jets_number		1.968e-01
bjetsL_HT		1.941e-01
jets_7pt		1.939e-01
bjetsL_invariantMass		1.833e-01
toptagger_transMass		1.730e-01
bjetsM_invariantMass		1.696e-01
jets_6pt		1.632e-01
bjetsM_HT		1.569e-01
jets_8pt		1.510e-01
jets_5pt		1.475e-01
jets_transMass	:	1.475e-01
	<pre>: bjetsL_num : bjetsL_3pt : bjetsL_4pt : jets_number : bjetsL_HT : jets_7pt : bjetsL_invariantMass : toptagger_transMass : bjetsM_invariantMass : jets_6pt : bjetsM_HT : jets_8pt : jets_5pt</pre>	: jets_bScore : : bjetsM_3pt : : bjetsL_num : : bjetsL_3pt : : bjetsL_4pt : : jets_number : : bjetsL_HT : : jets_7pt : : bjetsL_invariantMass : : toptagger_transMass : : bjetsM_invariantMass : : bjetsM_HT : : jets_8pt : : jets_5pt : : bjets_5pt : : bjetsM_invariant : : bjetsM_HT : : jets_8pt : : jets_5pt : : bjets_5pt : : bjets_5

A (10) A (10) A (10)

Step 2: ROC



: DataSet	MVA	
: Name:	Method:	ROC-integ
: dataset	BDTG	: 0.882
: dataset	BDT	: 0.882
: dataset	BDTD	: 0.880
: dataset	BDTB	: 0.863
: dataset	CutsSA	: 0.500

イロト イロト イヨト イヨト

BDT Study

step 3

- Remove Correlation > 80
- bjetsL_num and bjetsL_4pt, remove bjetsL_4pt
- jets_number and jets_8pt, remove jets_8pt
- bjetsL_HT, bjetsM_HT(80), bjetsL_invariantMass(86), remove bjetsM_HT and bjetsL_invariantMass,
- bjetsM_invariantMass, bjetsM_HT, already remove bjetsM_HT
- bjetsM_HT and bjetsL_HT(80) and bjetsM_invariantMass(86),

Huil	ling	Hua)
Tiui	mig	ilua	(11 11-17)

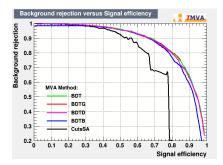
	jets_bScore bjetsM_3pt bjetsL_num bjetsL_3pt jets_number bjetsL_HT jets_7pt toptagger_transMass jets_6pt jets_5pt	1.632e-01 1.475e-01	
	jets_transMass	1.475e-01	
-	 	 	

(4) (5) (4) (5)

: Separatio

Rank : Variable

Step3: ROC



DataSet	MVA	
Name:	Method:	ROC-integ
dataset	BDT	: 0.882
dataset	BDTG	: 0.882
dataset	BDTD	: 0.878
dataset	BDTB	: 0.862
dataset	CutsSA	: 0.736

イロト イヨト イヨト イヨト

Conclusion for Correlation Removal

F	igure: s	tep1	F	igure: st	ep2	F	igure: s	tep3
DataSet Name: dataset dataset dataset dataset dataset	MVA Method: BDT BDTG BDTD BDTB CutsSA	ROC-integ : 0.884 : 0.882 : 0.881 : 0.860 : 0.500	: DataSet : Name: : dataset : dataset : dataset : dataset : dataset	MVA Method: BDTG BDT BDTD BDTB CutsSA	ROC-integ : 0.882 : 0.880 : 0.880 : 0.863 : 0.500	DataSet Name: dataset dataset dataset dataset dataset	MVA Method: BDT BDTG BDTD BDTB CutsSA	ROC-integ : 0.882 : 0.882 : 0.878 : 0.862 : 0.736

- We almost lost no AUC performance from step1 to step2 to step3
- The resulting 11 input variables are going to be our starting point for further corelation removal for iterative removal

Iterative Removal

Procedure

- Suppose we have 25 input variables for starters
- For each variable, remove it and use the left 24 as inputs to BDT, see the performance, the variable that has the least impact on the performance is ranked as the least important(25 TMVA training).
- Now we have a 24 variable set with best performance, repeat previous step (24 TMVA training)
- plot the AUC value as a function of number of input variables
- Require to be on the plateu of the curve
- Implementation
 - A bit computation expensive to implement, could take up to 325 times TMVA training for 25 variables
 - Still working on my code to implement it

Iterative Removal Implementation

- trying to implement the vSearch code to our analysis in IHEP cluster https://github.com/sitongan/vSearch
- not sure what a TMVA wrapper is. have written an email to author Sitong about it
- Sitong is not replying to my email. I am afraid I have try to do it myself
- Have some ideas in mind of how to implement it, a bit challenging but want to try them

Combined Variable Optimization

- Do correlation removal for start, keep varible list down to \approx 25 while still has very good performance
- Use the variable list from correlation removal and go through Iterative Removal.

(I) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1)) < ((1))

UL BSM Jobs Preparation

- Legacy ReReco or Legacy, aka Ultra-Legacy (UL)
- What we used before is ReReco
- for 2016, reprocessing is split into 2 parts. Analyzers need to combine them all.
 - With HIP mitigation: Run B (RAW v1, v2), C, D, E, F (most of)
 - Without HIP mitigation: F (7 runs), G, H
 - Run2016B has two sets of RAW data containing disjoint sets of runs. This is reflected in the strings "ver1" and "ver2"
 - UL2016 consists of 9 "eras": Run2016B ver1, Run2016B ver2, Run2016C, Run2016D, Run2016E, Run2016F(HIPM), Run2016F(no HIPM), Run2016G, Run2016H
 - Run2016F is split into two different reconstruction versions
 - "pre-VFP" (aka "HIPM" or "APV"): eras B-F (31 runs) are reconstructed with a reconstruction including HIP mitigation, short "HIPM".Run2016F(HIPM) has 31 runs
 - "post-VFP" (aka "no-HIPM"): eras F(7 runs)-H reconstructed with the default track reconstruction.

- The list of datasets (for example for Nanov8) should look like as follows:
 - /BTagCSV/Run2016B-ver1_HIPM_UL2016_MiniAODv1_NanoAODv2v1/NANOAOD
 - /BTagCSV/Run2016B-ver2_HIPM_UL2016_MiniAODv1_NanoAODv2v1/NANOAOD
 - /BTagCSV/Run2016C-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
 - /BTagCSV/Run2016D-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
 - /BTagCSV/Run2016E-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
 - /BTagCSV/Run2016F-HIPM_UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
 - /BTagCSV/Run2016F-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
 - /BTagCSV/Run2016G-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
 - /BTagCSV/Run2016H-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
- TTTT signal sample is still not available in DAS, maybe I should write to top MC contact asking about it

EventSelection

- MET fillters
- HLT requirements
 - HLT_PFHT450_SixJet40_BTagCSV_p056==1, HLT_PFHT400_SixJet30_DoubleBTagCSV_p056==1
- Ioose preselection
 - tausL.size()>0, jets.size()>3, bjetsL.size()>1
- Subchannel requirements

MC reweighting

- genWeight
- o prefireWeight
- PileUp reweighting

イロト イヨト イヨト イヨト

1Tau0L		
Raw entries:		
TTTT		237389
TT		46656
TTX		37788
single top		3839
тх		7785
DYJets		0
VV		8
VVV		98
OCD		305
Total BKG	=	7777.06
Weighted:		
TTTT	=	1729.23
TT		1.30564e+0
ттх		55880.3
		183.319
TX	1	7709.81
DYJets	Ξ.	0
VV	Ξ.	71/ 399
vvv	Ξ.	714.399 10.5086
OCD	Ξ.	268.521
Total BKG	Ξ	13936.3
TOLAL DAG	-	12920.2
Friends suited at		
Event yield:		0 47074
TTTT		9.47371
TT		6276.17
TTX	=	212.41
single top	=	120.653
тх		13.0152
DYJets		
VV		0.103319
VVV		0.982638
QCD		7285.38
Total BKG		13936.3

1TauIE Raw entries: TTTT TTT TTX single top TX DYJets VV VVV QCD Total BKG	= 80344 = 16515 = 10415 = 3361 = 2598 = 0 = 3 = 20 = 2 = 93.7667
Weighted: TTT TT TTX single top TX DYJets VV VVV QCD TOtal BKG	= 623.886 = 2.18092e+06 = 13611.5 = 147.115 = 2591.97 = 0 = 415.464 = 1.81758 = 2.17447 = 1178.13
Event yield: TTTT TTX single top TX DYJets VV VVV QCD Total BKG	= 3.4963 = 0 = 0.0537144 = 0.249651 = 2.35911

	1Tau1Mu		
	Raw entries:		
	TTTT	99667	
	TT	19225	
	TTX	11729	
	single top	2529	
	тх	3055	
	DYJets	0	
	VV	6	
	VVV	12	
	QCD		
	Total BKG	92.2959	
	Weighted:		
	TTTŤ	780.561	
	тт	2.46514e+06	
	ттх	15187.2	
	single top	139.593	
	тх	2971.26	
	DYJets	0	
	vv	465.544	
	VVV	0.622918	
	OCD	0	
	Total BKG	1324.63	
	Event yield:		
	TTTT	4.27635	
	TT	1236.92	
	ттх	54,2688	
	single top	18,7832	
	TX	4.01954	
	DYJets	0	
	VV	0.0786659	
	VVV	0.0812971	
	OCD	0	
•	Total BKG	1324.63	ŀ

Huiling Hua (IHEP)

IHEP 2020 21 / 46

1Tau20S		
Raw entries:		
TTTT		28250
TT		1265
TTX		4718
single top		3510
ТХ		789
DYJets		
VV		
VVV		5
QCD		0
Total BKG		28.0118
Weighted:		
TTTŤ		221.854
TT		92102.1
ттх		4935.03
single top		199.663
тх		
DYJets		0
vv		1.02674
VVV		0.650436
OCD		0
Total BKG		78.0374
Event yield:		
TTTT		1.21544
TT		58.7553
TTX		14.9337
single top		0.508183
тх		0.962345
DYJets		0
VV		0.0375234
VVV		0.0953384
QCD		0
Total BKG	=	78.0374
Huiling Hua		IHEP)

1Tau2SS		
Raw entries:		
TTTT	= 14204	
TT	= 63	
ттх	= 1321	
single top	= 520	
тх	= 385	
DYJets	= 0	
vv	= 0	
VVV	= 3	
QCD	= 0	
Total BKG	= 7.59988	
Weighted:		
TTTŤ	= 114.525	
тт	= 6095.78	
ттх	= 1501.54	
single top	= 18.4616	
тх	= 386.36	
DYJets	= 0	
vv	= 0	
VVV	= 0.421751	
OCD	= 0	
Total BKG	= 10.1519	
Event yield:		
тттт	= 0.627434	
TT	= 3.35655	
ттх	= 4.78014	
single top	= 0.046988	7
тх	= 0.498109	
DYJets	= 0	
vv	= 0	
VVV	= 0.0625488	3
QCD	= 0	
Total BKG	= 10.1519	
IHEP Grou	up Meeting	
	ap mooting	

1Tau3L	
Raw entries:	
TTTT	= 2848
TT	= 0
TTX	= 355
single top	= 26
TX	= 27
DYJets	= 0
VV	= 0
VVV	= 1
QCD	= 0
Total BKG	= 1.4098
Weighted:	
TTTT	= 23,2214
TT	= 0
TTX	= 358,464
single top	
TX	= 25.843
DYJets	= 0
VV	= 0
vvv	= 0.0168772
OCD	= 0
Total BKG	
TOCAL DRO	- 1.21000
Event yield:	
TTTT	= 0.12722
TT	= 0.12/22
TTX	= 0 = 1.02004
	= 0.00572658
TX	= 0.0294773
DYJets	= 0
VV	= 0
VVV	= 0.00242049
QCD	= 0
Total BKG	= 1.21668

IHEP 2020

22 / 46

2Tau0L	
Raw entries:	
TTTT	= 11422
TT	= 2848
TTX	= 4288
single top	= 2796
TX	= 1575
DYJets	= 0
VV	= 1
vvv	= 4
OCD	= 4
Total BKG	= 37.7907
Weighted:	
TTTT	= 92.454
TT	= 500241
TTX	= 4980.38
single top	= 158.423
тх	= 1562.55
DYJets	= 0
VV	= 1.14961
vvv	= 0.414473
QCD	= 1.95001
Total BKG	= 278.15
TOLAL DRO	= 278.15
Event yield:	
TTTT	= 0.506514
TT	= 246.776
TTX	= 16.2812
single top	= 6.77039
тх	= 2.56187
DYJets	= 0
VV	= 0.042014
VVV	= 0.034818
OCD	= 1.41789
Total BKG	= 278.15
TOTAL BRO	- 278.15

2Tau1E	
Raw entries:	
TTTT	= 2569
TT	= 180
TTX	= 1007
single top	= 459
TX	= 378
DYJets	= 0
VV	= 0
VVV	= 0
QCD	= 0
Total BKG	= 5.83133
Weighted:	
TTTT	= 21.8009
TT	= 22297.4
TTX	= 1020.17
single top	= 29.586
ТХ	= 372.931
DYJets	= 0
VV	= 0
VVV	= 0
QCD	= 0
Total BKG	= 16.2041
Event yield:	
TTTT	= 0.119437
TT	= 11.5052
TTX	= 2.98042
single top	= 0.0753027
TX	= 0.495236
DYJets	= 0
VV	= 0
VVV	= 0
QCD	= 0
Total BKG	= 16.2041

2TauIMu Raw entries: TTT TT single top TX DYJets VV VVV OCD	= 3264 = 190 = 1021 = 442 = 406 = 0 = 0 = 0 = 0
Weighted: TTTT TTX single top TX DYJets VV VVV QCD	= 5.98391 = 25.5952 = 21044 = 1128.94 = 24.0248 = 394.063 = 0 = 0 = 0 = 0 = 16.4386
Event yield: TTTT TT single top TX DYJets VV VV QCD Total BKG	= 0.530434 = 0 = 0 = 0 = 0 = 0

Huiling Hua (IHEP)

IHEP Group Meeting

IHEP 2020

23/46

-		
2Tau20S		
Raw entries:		
TTTT		449
TT		
TTX		111
single top		
ТХ		28
DYJets		0
VV		0
VVV		
QCD		0
Total BKG		0.59826
Weighted:		
TTTŤ		3.23368
TT		66.1142
TTX		110.585
single top		1.52484
TX		
DYJets	=	
VV		
VVV		0.0187778
QCD	=	0
Total BKG	=	
interest bills		
Event vield:		
	=	0.0177159
TTTT TT	=	
ттх	-	
single top	-	
TX	-	
DYJets	2	
VV	Ę.	
vvv	1	0.00269306
QCD	1	0.00269306
	1	0.484251
Total BKG	-	0.404251

2Tau2SS		
Raw entries:		
TTTT		190
TT		0
TTX		13
single top		
TX		
DYJets		
VV		
VVV		
QCD		
Total BKG		0.0644176
Weighted:		
TTTT		1.75733
TT		
TTX		11.9666
single top		
тх		2.99069
DYJets		
VV		
VVV		
QCD		
Total BKG		0.0578978
Event yield:		
TTTT		0.00962764
TT		
TTX		0.0344717 -8.60342e-06
single top		
TX		0.00315698
DYJets VV	=	
vv	-	
QCD	2	
Total BKG		0.0578978
TOLAL BKG	-	0.05/89/8

Huiling Hua (IHEP)

IHEP 2020 24 / 46

◆□ → ◆□ → ◆臣 → ◆臣 → ○臣 -

Backup



back up

Huiling Hua (IHEP)

IHEP Group Meeting

▲ য় ▶ য় ৩ ৭ ৫
IHEP 2020 25 / 46

イロト イヨト イヨト イヨト

1tau0l

TCut mycuts = "tausT_number==1 leptonsMVAT_number==0 jets_number>=8 bjetsM_num>=2" 1tau0l

イロト イポト イヨト イヨト

Variable Importance

- A ranking of the BDT input variables is derived by counting how often the variables are used to split decision tree nodes, and by weighting each split occurrence by the separation gain-squared it has achieved and by the number of events in the node
- This ranking is known to be unstable and sub-optimal, but widely used within the community.
- correlation information is not accounted for in TMVA Ranking.

< 口 > < 同 > < 回 > < 回 > < 回 > <

AUC

- AUC (area under curve), The AUC metric is the area under the signal versus background efficiency Receiver Operating Characteristic (ROC) curve, bounded by 0 and 1, where 1 is equivalent to perfect discrimination between signal and background, and 0.5 represents discrimination no better than random guessing.
- increase the number of input variables to see if we gain extra AUC
- we want to keep as small set of input variables as possible
- importance

イロン イ理 とく ヨン・

Backup

Iterative Removal

- A hillclimbing algorithm solves the search problem by always going in the direction with the highest gradient. It can be naive in the sense that it might get stuck in a local maximum instead of a global one, but it is still a valuable and intuitive method
- Among the 21 variable lists produced this way, the one that performs the best indicates which variable, if removed, has the least impact on the performance. This variable is ranked as the least important, and this particular variable list is used to generate a new batch of variable lists by again removing each variable once.
- One potential method to reduce this tendency towards local maxima is to implement a beam search of a certain width W
- We have shown that despite its tendency to get stuck in local maxima, iterative removal performs reasonably well for the problem of variable selection and much better than the standard TMVA Ranking method in most cases.

Huiling Hua (IHEP)

IHEP Group Meeting

1Tau20S		
Raw entries:		
TTTT		28250
TT		1265
TTX		4718
single top		3510
TX		789
DYJets		
VV		
VVV		5
QCD		0
Total BKG		28.0118
Weighted:		
тттт		221.854
тт		92102.1
TTX		4935.03
single top		199.663 778.825
тх	=	778.825
DYJets		0
VV	=	1.02674
VVV	=	0.650436
OCD		0
		78.0374
Event yield:		
TTTT	=	1.21544
TT		58.7553
TTX		14.9337
single top		0.508183
TX		0.962345
DYJets		0
VV		0.0375234
vvv		0.0953384
OCD		0
Total BKG		
Huiling Hua		INEP)

1Tau2SS		
Raw entries:		
TTTT		14204
TT		63
ттх		1321
single top		520
ТХ		385
DYJets		0
VV		0
VVV		3
QCD		
Total BKG		7.59988
Weighted:		
TTTT		114.525
TT		6095.78
ттх		1501.54
single top		18.4616
TX		386.36
DYJets		0
VV		0
VVV		0.421751
QCD		
Total BKG		10.1519
Event yield:		
TTTT		0.627434
TT		3.35655
TTX		4.78014
single top		
тх		0.498109
DYJets		
VV		
		0.0625488
QCD		
Total BKG	_	
	in	Monting

1Tau3L		
Raw entries:		
TTTT		2848
TT		0
TTX		355
single top		26
тх		27
DYJets		0
VV		0
VVV	=	1
QCD	=	0
	=	1.4098
.ocat bito		11.1000
Weighted:		
TTTT	_	23,2214
TT		0
ттх		358,464
single top		
TX		25.843
DYJets	1	
VV	1	
vvv		0.0168772
OCD		
Total BKG	=	1.21668
Event yield:		
TTTT		0.12722
TT		
TTX		1.02004
single top		
ТХ		0.0294773
DYJets		
VV		
VVV		0.00242049
QCD		
 Total BKG		1.21668

ing Hua (IHEP)

IHEP Group Meeting

IHEP 2020

30 / 46

2Tau0L		
Raw entries:		
TTTT		11422
TT		2848
TTX		4288
single top	=	2796
TX	=	1575
DYJets	=	0
VV	=	1
VVV	=	4
OCD	=	2
Total BKG		37.7907
		37.7307
Weighted:		
TTTT	_	92.454
TT		500241
		4980.38
TTX		
single top		158.423
TX		1562.55
DYJets		
VV		1.14961
VVV		0.414473
QCD		1.95001
Total BKG		278.15
Event yield:		
TTTT		0.506514
TT		246.776
TTX		16.2812
single top		6.77039
TX	=	2.56187
DYJets	=	0
VV		0.042014
vvv		0.0348188
OCD		1.41789
Total BKG	5	278.15
TOTAL BRO		2/0.15

2Tau1E	
Raw entries:	
TTTT	= 2569
TT	= 180
ттх	= 1007
single top	= 459
TX	= 378
DYJets	= 0
VV	= 0
VVV	= 0
OCD	= 0
Total BKG	= 5,83133
Weighted:	
TTTT	= 21.8009
TT	= 22297.4
TTX	= 1020.17
single top	= 29.586
TX	= 372.931
DYJets	= 0
VV	= 0
VVV	= 0
QCD	= 0
Total BKG	= 16.2041
Event yield:	
TTTT	= 0.119437
TT	= 11.5052
TTX	= 2.98042
single top	= 0.0753027
TX	= 0.495236
DYJets	= 0
VV	= 0
VVV	= 0
QCD	= 0
Total BKG	= 16.2041

TX DYJets VV VVV QCD	= 3264 = 190 = 1021 = 442 = 406 = 0 = 0 = 0 = 0 = 5,98391
Weighted: TTTT TTX single top TX DYJets VV VV VVV QCD	= 25.5952 = 21044 = 1128.94
Event yield: TTTT TTX single top TX DYJets VV VV VV VV QCD Total BKG	= 0.140225 = 11.0659 = 3.52228 = 0.0611482 = 0.530434 = 0 = 0 = 0 = 0 = 10.4386

Huiling Hua (IHEP)

IHEP Group Meeting

IHEP 2020 3

31 / 46

· -	
2Tau20S	
Raw entries:	
TTTT	= 449
TT	
TTX	= 111
single top	= 8
ТХ	= 28
DYJets	= 0
VV	= 0
VVV	= 1
QCD	= 0
Total BKG	= 0.59826
Weighted:	
тттт	= 3.23368
TT	= 66.1142
ттх	= 110.585
single top	= 1.52484
тх	= 30.1004
DYJets	= 0
VV	= 0
VVV	= 0.0187778
QCD	= 0
Total BKG	= 0.484251
Event vield:	
TTTT	= 0.0177159
тт	= 0.042666
ттх	= 0.31452
single top	= 0.00388105
тх	= 0.031774
DYJets	= 0
vv	= 0
VVV	= 0.00269306
OCD	= 0
Total BKG	= 0.484251

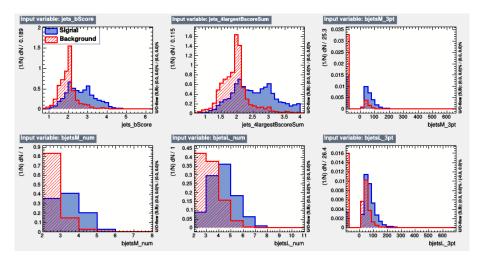
2Tau2SS		
Raw entries:		
TTTT		190
TT		0
TTX		13
single top		
тх		4
DYJets		0
VV		0
VVV		ō
QCD		0
Total BKG	=	0.0644176
Weighted:		
TTTT	=	1.75733
TT	=	
TTX		11.9666
single top		
TX		2,99069
DYJets	=	
VV	-	
vvv	=	
OCD	=	
Total BKG		0.0578978
Totat bito		0.03/03/0
Event yield:		
TTTT	=	0.00962764
TT	1	
ттх		0.0344717
single top		-8.60342e-06
TX		0.00315698
DYJets	-	
VV	2	0
vvv	1	
QCD	1	
Total BKG		0.0578978
TOLAL DAG	-	0.05/69/6

イロト イロト イヨト イヨト

Huiling Hua (IHEP)

▲ য় ▶ য় ৩ ৭ ৫
IHEP 2020 32 / 46

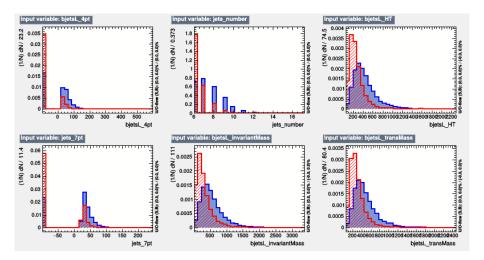
Step1: Input Varible Distribution



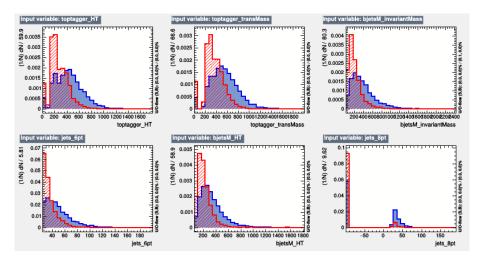
Huiling Hua (IHEP)

▶ < ≣ ▶ ≡ ∽ २० IHEP 2020 33 / 46

Step1:Input Varible Distribution

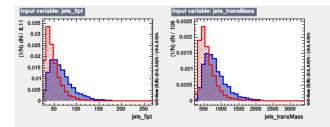


Step1:Input Varible Distribution



イロト イポト イヨト イヨト

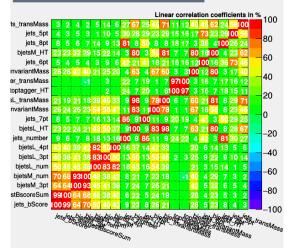
Step1:Input Varible Distribution



イロト イヨト イヨト イヨト

Step1: Correlation

Correlation Matrix (background)



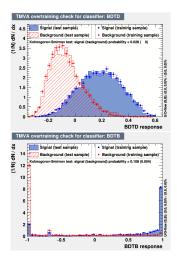
Huiling Hua (IHEP)

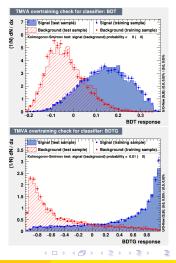
IHEP 2020 37 / 46

-

A (10) F (10)

Step 1: OverTrainining



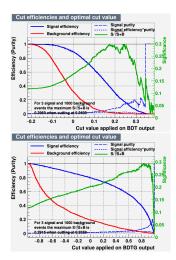


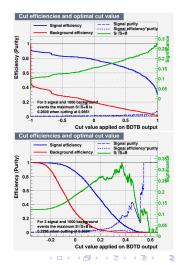
Huiling Hua (IHEP)

IHEP Group Meeting

IHEP 2020 38 / 46

Step1: Significance



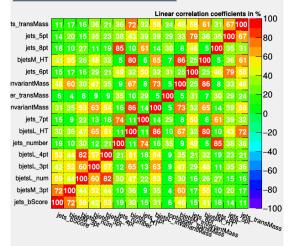


Huiling Hua (IHEP)

IHEP Group Meeting

Step 2: Correlation

Correlation Matrix (signal)



Huiling Hua (IHEP)

A (1) > A (2) > A

1 1 1 1 1 1 1

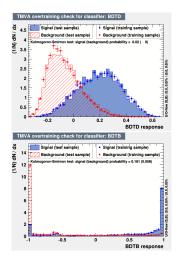
Step 2: Correlation

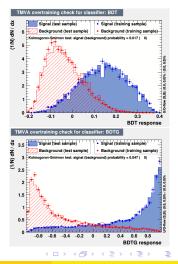
Corre	Correlation Matrix (background)																	
Linear correlation coefficients in % ts_transMass 3 4 5 14 6 27 67 25 46 12 40 46 62 24 56 100 100																		
jets_5pt	4			10		30		29	23	16	17	73		26	100	56		80
jets_8pt	8				13	81		50	8	17		36		100	26	24		
bjetsM_HT	23	32			14	3	80	3	58	7	80	16	100	4		62		60
jets_6pt	5	5				42	21	41	18	16	12	100	16	36	73	46		40
invariantMass	28	42	21		20				67	3	100	12	80		17	40		40
er_transMass						22		19	1	100		16				12	-	20
nvariantMass	26		54		47	11	83	11	100		67				23	46		0
jets_7pt	8				14	86		100		19						25		0
bjetsL_HT	23				37	9	100	9	83		63	21	80			67	_	-20
jets_number	9		18		16	100		86		22		42		81	30	27		
bjetsL_4pt	43		82	50	100				47		20		14			6		-40
bjetsL_3pt	40	41	83	100	50				58		25		22			14	_	-60
bjetsL_num	50	45	100	83	82		43		54		21					5		
bjetsM_3pt	64	100	45		39				26		42					4		-80
jets_bScore	100	64	50	40	43	9	23	8	26		28	5	23	8	4	3		_100
	-100 Jets Yestigetigetigetiget fieldera Gester Stellera Stellera Stellera Stellera Stellera Stellera Stellera Jets Yestigetigetigetigetigetigetigetigetigetige																	

Huiling Hua (IHEP)

<ロ> <四> <四> <四> <四> <四</p>

Step 2: OverTrainining



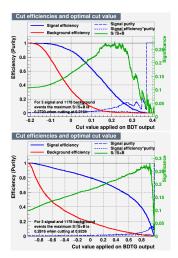


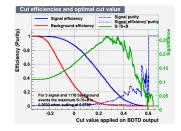
Huiling Hua (IHEP)

IHEP Group Meeting

IHEP 2020 42 / 46

Step 2: Significance





Huiling Hua (IHEP)

IHEP Group Meeting

IHEP 2020 43 / 46

(日)

Step3: Correlation

Linear correlation coefficients in % 100													
ts_transMass		17	16	36	36	72	32	24		67	100	100	
jets_5pt	14	20		35		43			79	100	67	-80	
jets_6pt	15	17		29	49		50		100	79	58	-60	
er_transMass	6							100	31	29	24	-40	
jets_7pt	15		22		74		100	29		39	32	-20	
bjetsL_HT	30	36	47		11	100	11			43	72	0	
jets_number	19				100	11	74	35			36	-20	
bjetsL_3pt	42	52		100	12	65					36	-40	
bjetsL_num	59	44	100	60	30	47	22				16	-60	
bjetsM_3pt	72	100	44	52		36			17	20	17		
jets_bScore	100	72	59	42	19	30	15	6		14	11		

Correlation Matrix (signal)

Huiling Hua (IHEP)

<ロ> <四> <四> <四> <四> <四</p>

Step3: Correlation

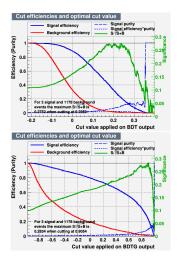
Corre	Correlation Matrix (background)											
Linear correlation coefficients in %												
ts_transMass	3	4	5	14	27	67	25	12	46	56	100	100
jets_5pt	4	5			30	28	29		73	100	56	-80
jets_6pt	5	5			42		41		100	73	46	-60
er_transMass					22	7	19	100	16		12	-40
jets_7pt	8				86		100	19	41		25	-20
bjetsL_HT	23	24	43	50	9	100	9		21		67	-0
jets_number	9				100	9	86	22	42		27	-20
bjetsL_3pt	40	41	83	100	13	50					14	-40
bjetsL_num	50	45	100	83	18	43					5	-60
bjetsM_3pt	64	100	45								4	
jets_bScore	100	64	50	40	9	23	8		5		3	
jets_bStore_Spit_num_Spit_number_HT_forlagge_Spit_spite_transMass												

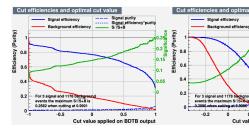
Operation Matrix (hopkersond)

Huiling Hua (IHEP)

<ロ> <四> <四> <四> <四> <四</p>

Step3: Significance





A D F A B F A B F A B F

Huiling Hua (IHEP)

IHEP Group Meeting

৸ ≣ ► ≣ ৩৭৫ IHEP 2020 46/46