Unoffical notes: Search for boosted top+Higgs resonance in the all hadronic decay mode

Lucas Corcodilos

April 28, 2021

1 Analysis strategy and selection

Dijet search for boosted $X \to tH$. The benchmark, X, is a VLQ T' produced in association with a bottom quark. The interaction with an associated top quark is currently ignored since the simulation samples are inconsistent with the UL versions available for all other simulation. The associated quark will have much lower transverse momentum than the T' decay products and so the affect on the analysis is expected to be small (ie. quark will be along beamline).

1.1 Selection

Signal region (SR)	QCD control for SR	Validation region (VR)	QCD control for VR
Two AK8 jets separated by $\Delta\phi > \pi/2$ $p_T > 350$ GeV for each jet $\lvert\eta\rvert < 2.4$ for each jet $m_{\text{jet}} > 50$ for each jet			
At least one jet top tagged		No jet top tagged	
Higgs tag pass	Higgs tag fail	Higgs tag pass	Higgs tag fail

Table 1: Analysis selection for four selection regions. A "top tag" refers to a selection of a tagger score at least 0.632 and $105 < m_{\text{jet}} < 210$ GeV, where the tagger could be DeepAK8 or ParticleNet (see Section 2). A "Higgs tag" refers to a tagger score selection of at least 0.9, where the tagger could be DeepAK8 or ParticleNet.

2 Initial studies on jet tagging

Two neural network based taggers are considered for jet tagging - DeepAK8 and ParticleNet. Several aspects of these taggers can affect this analysis and are considered when choosing between the taggers.

First and most importantly, the tagger must be decorrelated from the jet mass so as not to sculpt the jet mass distribution. The two dimensional background estimation method uses the jet mass as one of the axis in which to measure data. If the taggers sculpt the jet mass so as to create a signal-like peak, the background will be harder to discriminant from potential signal.

Additionally, the data-driven estimate of the QCD multijet background relies on a smooth ratio of the distributions of those events passing and failing the tagger. If the "pass" is sculpted and the "fail" is not, we cannot hope to use our transfer function method without introducing method for potential signal bias.

Second, the tagger should be good at distinguishing between top and Higgs jets so the tagger can be used to easily identify which side of the event is which physics object. In particular, the analysis would benefit from a tagger that correctly identifes that real Higgs jets are not top jets.

Finally, the tagger should be efficient at removing background without requiring a MC-to-data scale factor that is large or has large uncertainties that could dominant the total systematic uncertainty for the analysis.

With these three items in mind, we examined two variables - the difference in mass between the two jets and the difference in top tagging scores between the two jets - and plotted them against each other. In addition to considering the two taggers, we also considered four scenarios based on the simulation truth.

From left to right in Figs. 1 and 2:

- The first scenario does nothing with simulation truth.
- The second scenario examines when neither the top nor Higgs jet identified by the tagger were able to be matched to their respective generator particle ("Bad match").

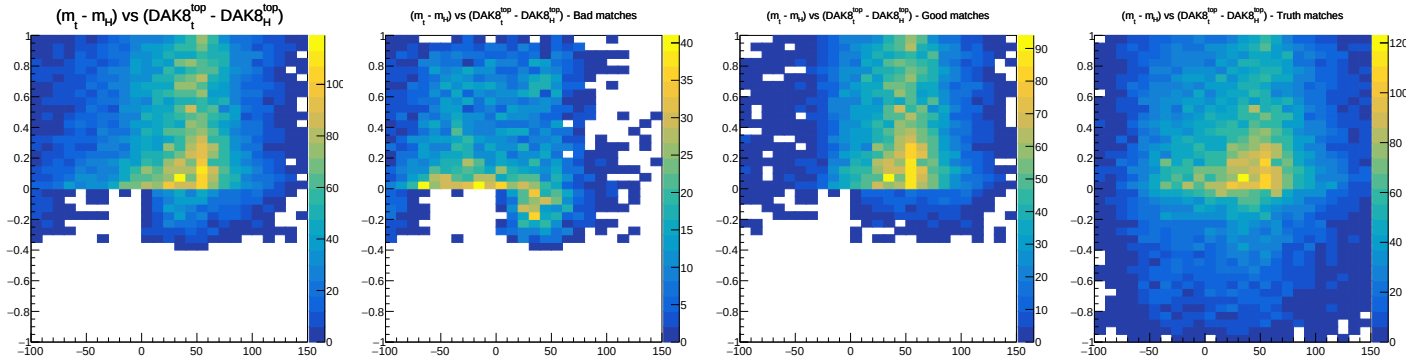

Figure 1: DeepAK8 studies

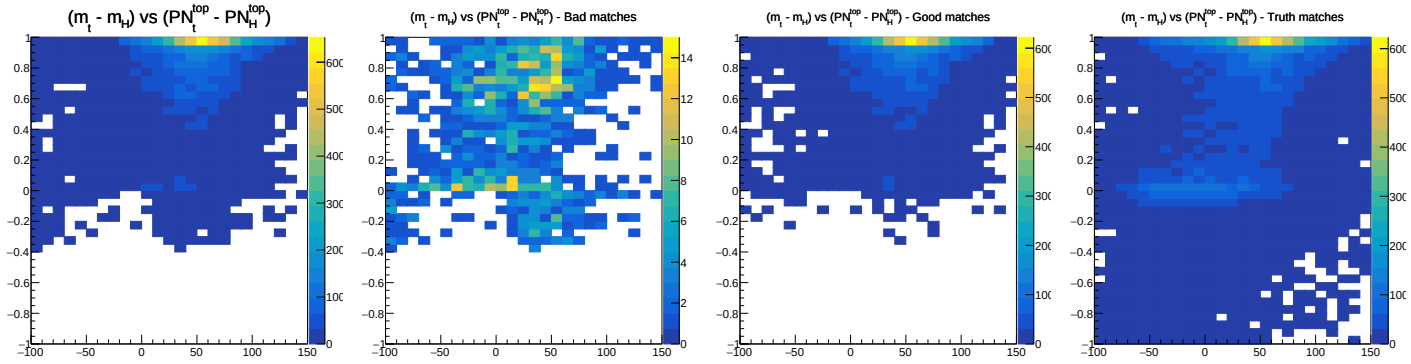

Figure 2: ParticleNet studies

- The third scenario examines when both the top and Higgs jets identified by the tagger were matched successfully to their respective generator particle ("Good match").
- The fourth scenario identifies the top and Higgs jets based on the generator particle information and then the variables plotted are calculated based on this information.

In other words, the fourth scenario flips the order of operations of the third scenario. The distributions for both taggers and the four scenarios are shown in Figs. 1 and 2 where a 2016 $T' \to tH$ signal with a T' mass of 1200 GeV is examined.

3 Kinematic distributions

We further investigate kinematic variables which may be useful in increasing the signal to background ratio in the analysis. Figures 3 and compare simulation for all of Run 2.

4 Jet tagging variables

Two variables are used to tag a canidate jet - the tagger score and the jet mass. In order to inspect these variables, a series of N-1 plots are made - one for each the top jet mass, top jet tagger score, Higgs jet mass, and Higgs jet tagger score. When one variable is plotted, the other three are cut on.

Since this means a Higgs and top cannot be identified without impacting the distributions, the top quark is assumed to be the leading jet in p_T. The MC truth shows this is true 50% of the time. The other N-1 cuts in a given plot should kill most of the mis-matched events (where the top jet is actually a Higgs and vice versa) with some fraction remaining. As can be seen in 1, that fraction is decently large when tagging a real Higgs with the DeepAK8 top tagger. This effect explains the Higgs peak in Figure 4.

Figures 4, 4, 4, and 4 compare simulation for all of Run 2.

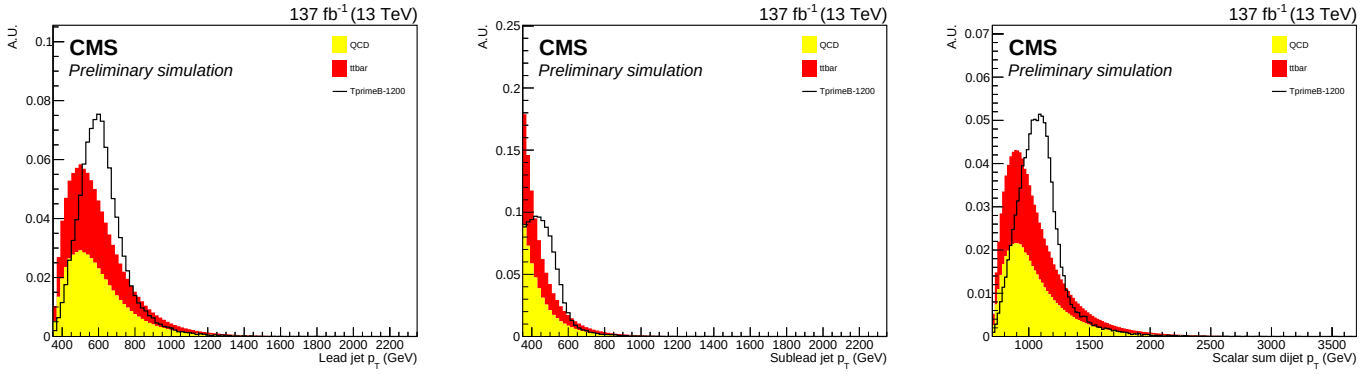

Figure 3: The p_T of the leading (left) and subleading (middle) jet in the event and the scalar sum of the two (right). The total background is represented as a stack of QCD (yellow) and ttbar (red) MC and the total is normalized to unity. The signal (black line) is separately normalized to unity.

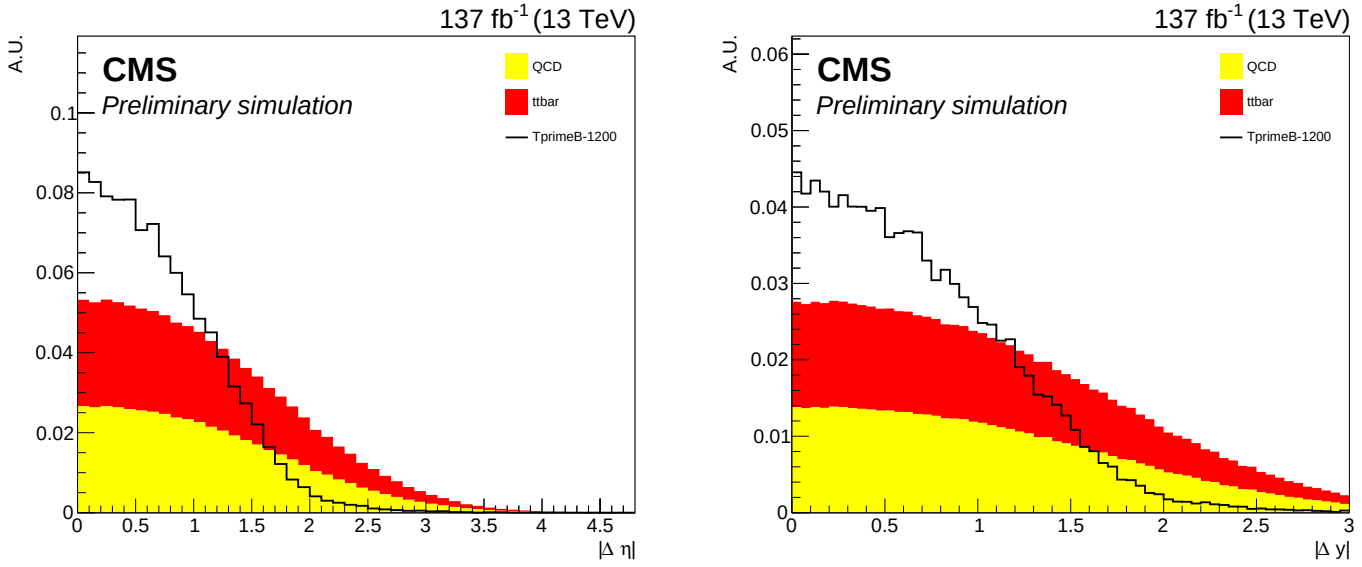

Figure 4: The $|\Delta\eta|$ between the two candidate jets in the event. The total background is represented as a stack of QCD (yellow) and ttbar (red) MC and the total is normalized to unity. The signal (black line) is separately normalized to unity.

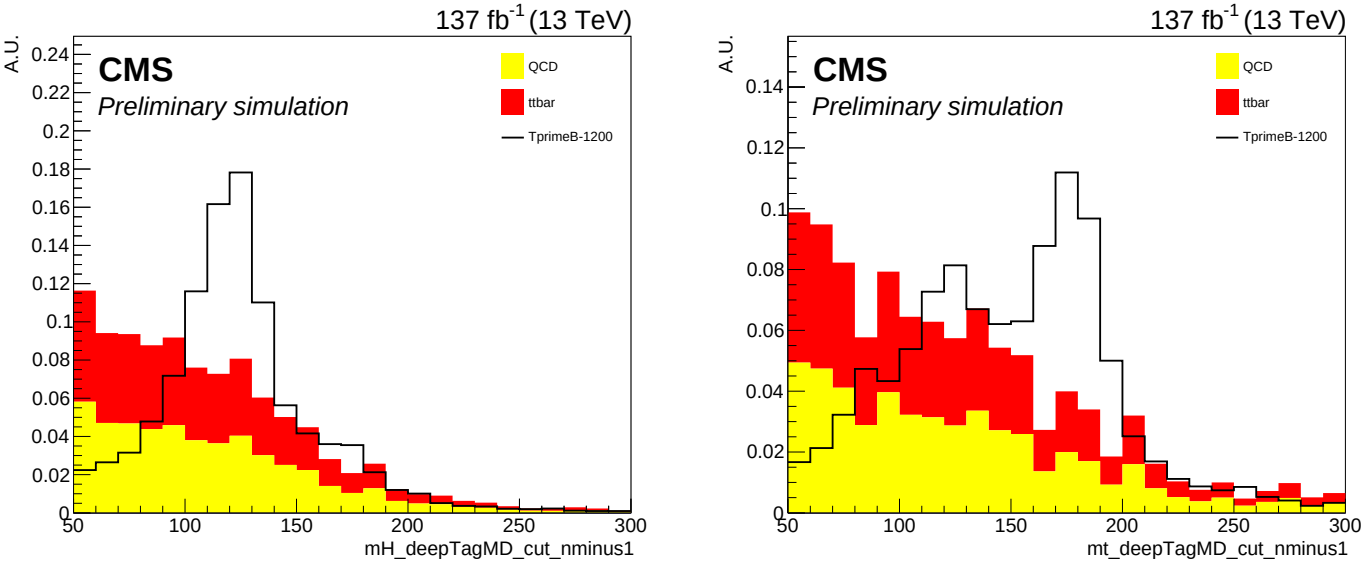

Figure 5: Shape comparison of the jet mass distributions for the Higgs (left) and top (right) jets when using the mass decorrelated DeepAK8 tagger. The total background is represented as a stack of QCD (yellow) and ttbar (red) MC and the total is normalized to unity. The signal (black line) is separately normalized to unity.

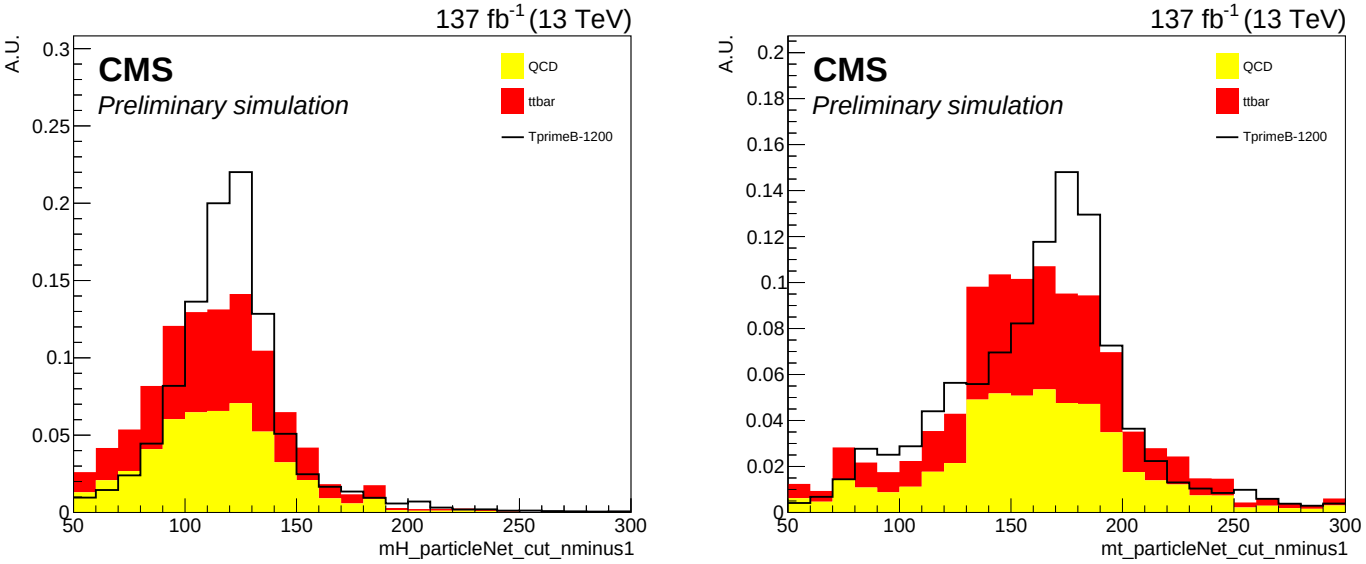

Figure 6: Shape comparison of the jet mass distributions for the Higgs (left) and top (right) jets when using the ParticleNet tagger (not mass decorrelated). The total background is represented as a stack of QCD (yellow) and ttbar (red) MC and the total is normalized to unity. The signal (black line) is separately normalized to unity.

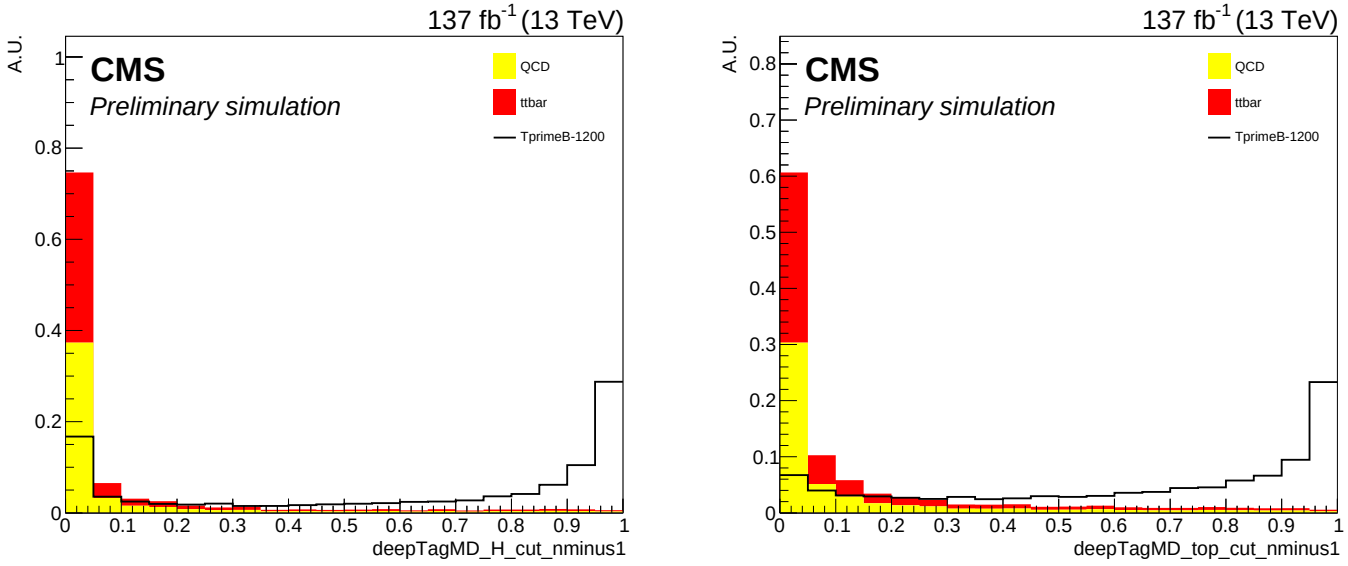

Figure 7: Shape comparison of the DeepAK8 score distributions for the Higgs (left) and top (right) jets when using the mass decorrelated DeepAK8 tagger. The total background is represented as a stack of QCD (yellow) and ttbar (red) MC and the total is normalized to unity. The signal (black line) is separately normalized to unity.

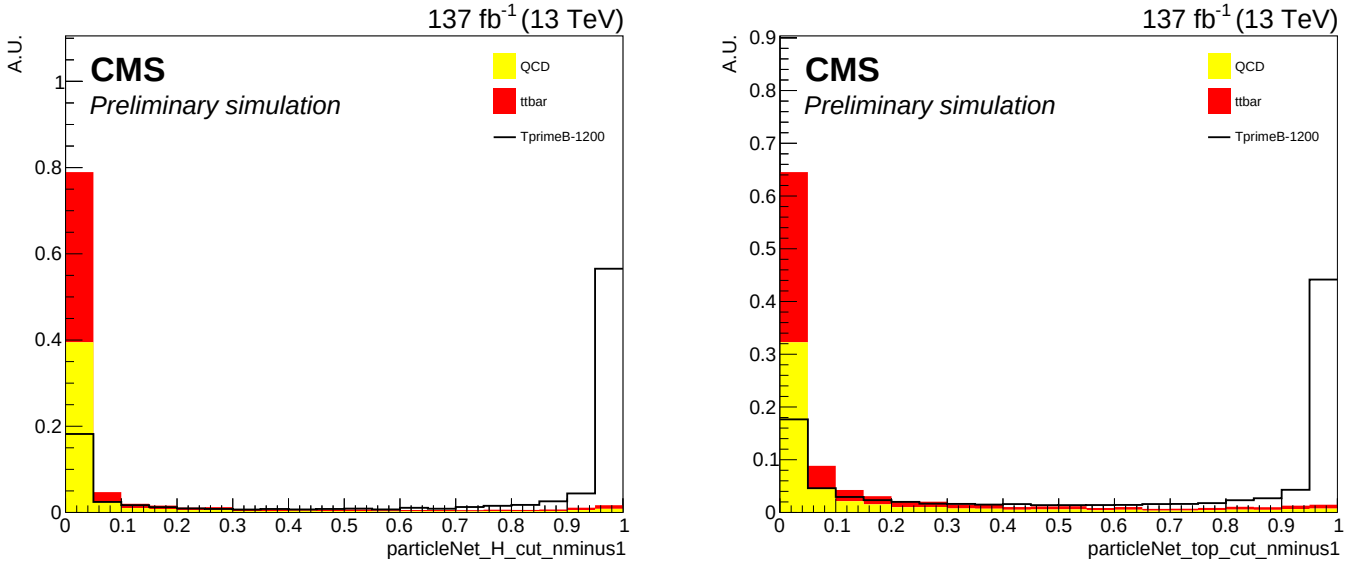

Figure 8: Shape comparison of the ParticleNet score distributions for the Higgs (left) and top (right) jets when using the mass decorrelated ParticleNet tagger. The total background is represented as a stack of QCD (yellow) and ttbar (red) MC and the total is normalized to unity. The signal (black line) is separately normalized to unity.

5 Blinded fits to data

6 Data and simulation samples

2016

Setname	DAS location
TprimeB 800 GeV	/TprimeBToTH_M-800_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
TprimeB 900 GeV	/TprimeBToTH_M-900_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
TprimeB 1000 GeV	/TprimeBToTH_M-1000_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
TprimeB 1100 GeV	/TprimeBToTH_M-1100_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
TprimeB 1200 GeV	/TprimeBToTH_M-1200_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
TprimeB 1300 GeV	/TprimeBToTH_M-1300_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
TprimeB 1400 GeV	/TprimeBToTH_M-1400_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
TprimeB 1500 GeV	/TprimeBToTH_M-1500_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
TprimeB 1600 GeV	/TprimeBToTH_M-1600_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
TprimeB 1700 GeV	/TprimeBToTH_M-1700_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
TprimeB 1800 GeV	/TprimeBToTH_M-1800_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
ttbar-allhad	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/RunIISummer19UL16NanoAODv2-106X_ mcRun2_asymptotic_v15-v1/NANOAODSIM
ttbar-semilep	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer19UL16NanoAODv2-106X_ mcRun2_asymptotic_v15-v1/NANOAODSIM
QCDHT700	/QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
QCDHT1000	/QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
QCDHT1500	/QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
QCDHT2000	/QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8/ RunIISummer19UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1/NANOAODSIM
DataB1	/JetHT/Run2016B-ver1_HIPM_UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
DataB2	/JetHT/Run2016B-ver2_HIPM_UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
DataC	/JetHT/Run2016C-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
DataD	/JetHT/Run2016D-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
DataE	/JetHT/Run2016E-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
DataF	/JetHT/Run2016F-UL2016_MiniAODv1_NanoAODv2-v2/NANOAOD
DataG	/JetHT/Run2016G-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
DataH	/JetHT/Run2016H-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD

2017

Setname	DAS location
TprimeB 800 GeV	/TprimeBToTH_M-800_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
TprimeB 900 GeV	/TprimeBToTH_M-900_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
TprimeB 1000 GeV	/TprimeBToTH_M-1000_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
TprimeB 1100 GeV	/TprimeBToTH_M-1100_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
TprimeB 1200 GeV	/TprimeBToTH_M-1200_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
TprimeB 1300 GeV	/TprimeBToTH_M-1300_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
TprimeB 1400 GeV	/TprimeBToTH_M-1400_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
TprimeB 1500 GeV	/TprimeBToTH_M-1500_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
TprimeB 1600 GeV	/TprimeBToTH_M-1600_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
TprimeB 1700 GeV	/TprimeBToTH_M-1700_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
TprimeB 1800 GeV	/TprimeBToTH_M-1800_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
ttbar-allhad	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/RunIISummer19UL17NanoAODv2-106X_ mc2017_realistic_v8-v1/NANOAODSIM
ttbar-semilep	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer19UL17NanoAODv2-106X_ mc2017_realistic_v8-v1/NANOAODSIM
QCDHT700	/QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
QCDHT1000	/QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
QCDHT1500	/QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
QCDHT2000	/QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8/ RunIISummer19UL17NanoAODv2-106X_mc2017_realistic_v8-v1/NANOAODSIM
DataB	/JetHT/Run2017B-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
DataC	/JetHT/Run2017C-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
DataD	/JetHT/Run2017D-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
DataE	/JetHT/Run2017E-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
DataF	/JetHT/Run2017F-UL2017_MiniAODv1_NanoAODv2-v2/NANOAOD

2018

Setname	DAS location
TprimeB 800 GeV	/TprimeBToTH_M-800_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
TprimeB 900 GeV	/TprimeBToTH_M-900_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
TprimeB 1000 GeV	/TprimeBToTH_M-1000_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
TprimeB 1100 GeV	/TprimeBToTH_M-1100_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
TprimeB 1200 GeV	/TprimeBToTH_M-1200_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
TprimeB 1300 GeV	/TprimeBToTH_M-1300_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
TprimeB 1400 GeV	/TprimeBToTH_M-1400_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
TprimeB 1500 GeV	/TprimeBToTH_M-1500_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
TprimeB 1600 GeV	/TprimeBToTH_M-1600_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
TprimeB 1700 GeV	/TprimeBToTH_M-1700_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
TprimeB 1800 GeV	/TprimeBToTH_M-1800_LH_TuneCP5_PSweights_13TeV-madgraph_pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
ttbar-allhad	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/RunIISummer19UL18NanoAODv2-106X_ upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
ttbar-semilep	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer19UL18NanoAODv2-106X_ upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
QCDHT700	/QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
QCDHT1000	/QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
QCDHT1500	/QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
QCDHT2000	/QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8/ RunIISummer19UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1/NANOAODSIM
DataA	/JetHT/Run2018A-UL2018_MiniAODv1_NanoAODv2-v1/NANOAOD
DataB	/JetHT/Run2018B-UL2018_MiniAODv1_NanoAODv2-v1/NANOAOD
DataC	/JetHT/Run2018C-UL2018_MiniAODv1_NanoAODv2-v1/NANOAOD
DataD	/JetHT/Run2018D-UL2018_MiniAODv1_NanoAODv2-v1/NANOAOD