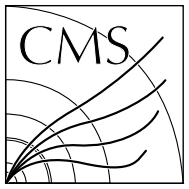
Available on CMS information server

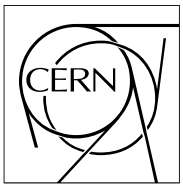
CMS DP -2018/058



The Compact Muon Solenoid Experiment

CMS Performance Note

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland



09 November 2018

Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector

The CMS Collaboration

Abstract

The identification of jets originating from b quarks (b jets) is of great importance to many physics analyses. Multivariate analysis techniques have been traditionally used in flavour tagging algorithms, however there has been a recent step change towards using deep learning algorithms due to their suitability to complex multi-classification problems. We present here the first measurement of the performance of the DeepJet algorithm using 41.9/fb of data collected from proton-proton collision at 13 TeV using the CMS detector. We derive efficiency scale factors to correct for the difference in performance on 2017 Monte Carlo simulation and data using two performance measurement techniques. We also show first studies on the dependence of the performance of DeepJet on the size of the training sample and its stability with respect to the random weight initialisation.

Performance of the DeepJet b tagging algorithm using 41.9 /fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector

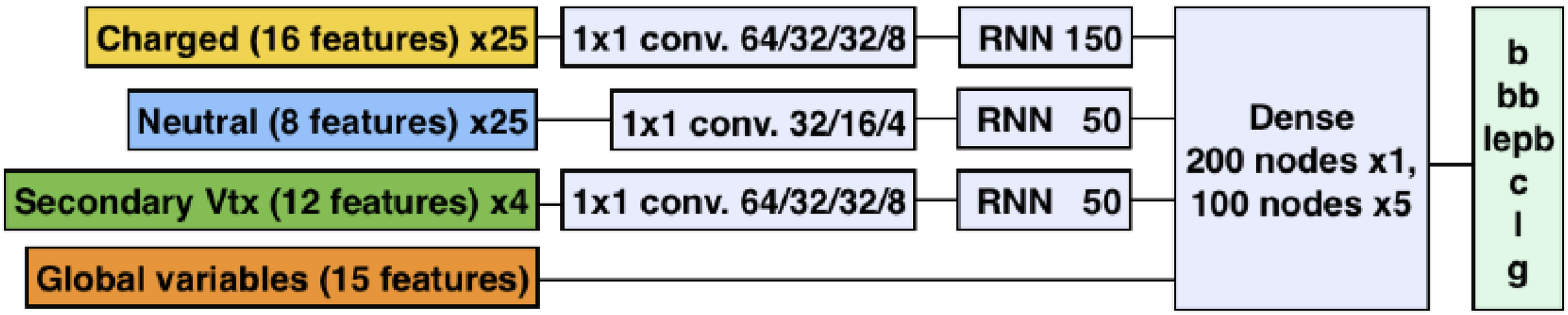
CMS Collaboration

CMS-POG-CONVENERS-BTAG@CERN.CH

Abstract

The identification of jets originating from b quarks (b jets) is of great importance to many physics analyses. Multivariate analysis techniques have been traditionally used in flavour tagging algorithms, however there has been a recent step change towards using deep learning algorithms due to their suitability to complex multi-classification problems. We present here the first measurement of the performance of the DeepJet algorithm using 41.9/fb of data collected from proton-proton collision at 13 TeV using the CMS detector. We derive efficiency scale factors to correct for the difference in performance on 2017 Monte Carlo simulation and data using two performance measurement techniques. We also show first studies on the dependence of the performance of DeepJet on the size of the training sample and its stability with respect to the random weight initialisation.

DeepJet: A deep neural network algorithm based on 16(8) properties of up to 25 charged(neutral) particle-flow jet constituents, as well as 12 properties of up to 4 secondary vertices associated with the jet. For each collection of charged and neutral particles and vertices, separate 1x1 convolutional layers are trained: 4 hidden layers with 64, 32, 32 and 8 filters respectively for charge candidates and vertices and 3 hidden layers with 32,16 and 4 filters for neutral particles. The filters act on each particle or vertex individually. The compressed and transformed output is then separately fed into 3 long short-term memories (LSTMs) with 150, 50 and 50 output nodes respectively. The outputs from the LSTMs are merged with global jet properties which are first fed through one dense layer with 200 nodes before being passed to 7 subsequent hidden layers with 100 nodes each.



* The DeepJet algorithm presented here is the same as to what has been referred to in the past as the DeepFlavour algorithm.

Working point definition

DeepJetL, DeepJetM, DeepJetT: DeepJet algorithm at the loose, medium and tight operating points, defined as the cut values in the discriminator distribution at which the rate of misidentifying a light jet as a b jet is 10%, 1% and 0.1% respectively. Each working point is derived using the performance curves of DeepJet on jets with $|\eta|<2.5$ and $p_T > 30$ GeV from QCD multijet MC. Pairs of b quarks that result from Gluon splitting are labelled as b jets.

Performance Measurement Methods

Kin Method: A method for measuring the b tagging efficiency in dilepton top pair events. Events with exactly two opposite charge leptons with p_T > 25 GeV and at least two jets with p_T > 30 GeV are selected. Final efficiencies and correction factors are derived via a template fit of a BDT discriminant that utilises kinematic variables to distinguish b jets from ISR/FSR jets.

Lifetime Tagging Secondary Vertex (LTSV) Method: A method for measuring the b tagging efficiency in multi-jet events. Jets are used if they have a p_T > 20 GeV and a muon with p_T > 5 GeV within dR < 0.4 from the jet. This method first uses the Jet Probability Tagger discriminant to evaluate the efficiency with which a secondary vertex is found. Subsequently, template fits to the distributions of the secondary vertex mass are performed to extract the final efficiencies and correction factors.

More information in CMS-BTV-16-002

Performance Measurements

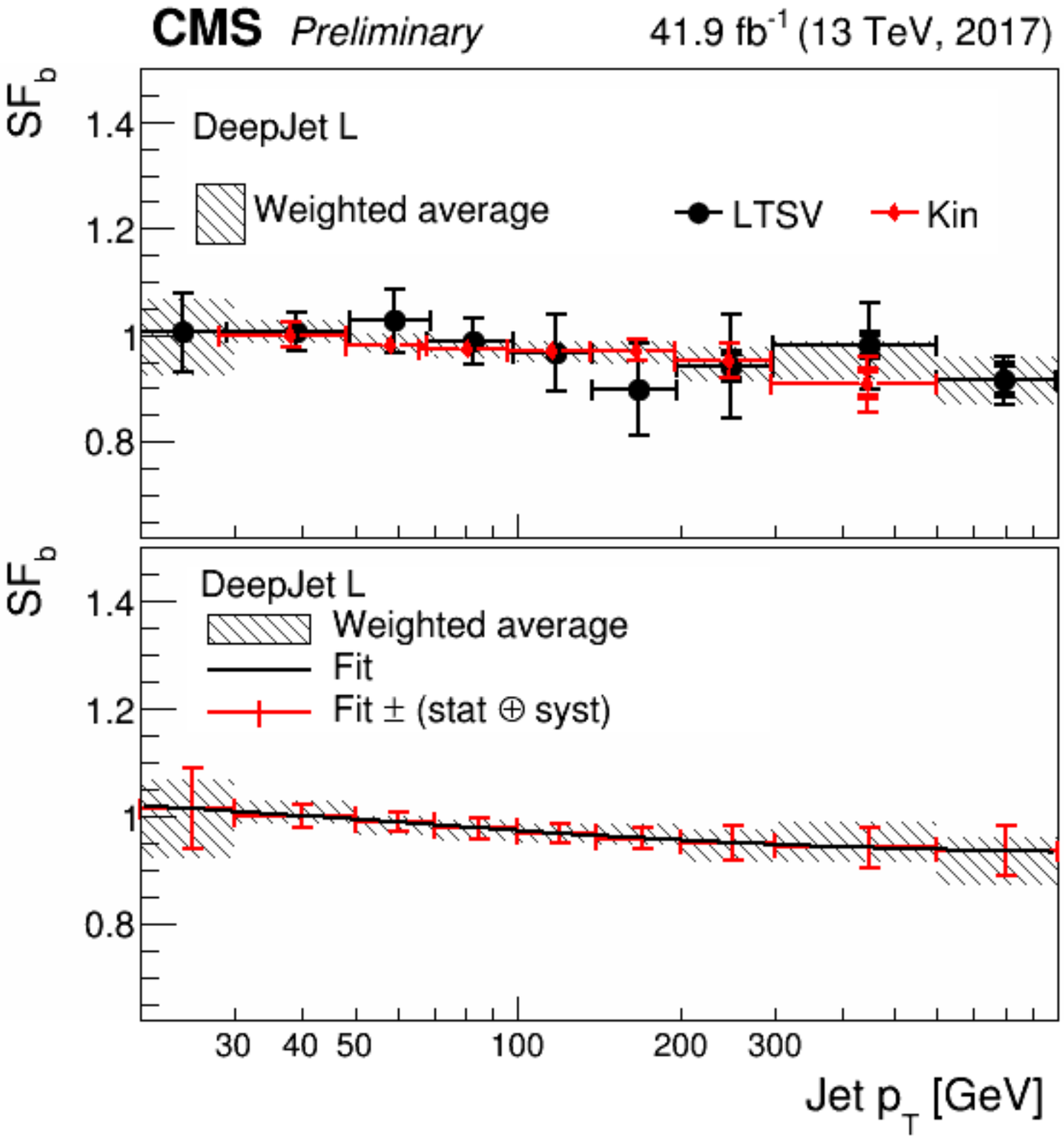


Figure 1: Data-to-simulation scale factors for b jets as a function of jet p_T for the loose working point using the DeepJet b tagging algorithm.The top plot shows the scale factors as a function of jet p_T obtained from the Kin and the LTSV methods, separately. The inner error bars represent the statistical uncertainty, whereas the outer error bars represent the total uncertainty. The scale factors obtained by combining the two measurements along with the total statistical plus systematic error are shown by the hatched area. The combination of all the methods is performed using the best linear unbiased estimator method (BLUE) method. As a result of the limited amount of data in the high jet p_T regions, statistical fluctuations can be seen to cause some discrepancy between the measured scale factors from the two methods.

The bottom panel shows the same combined scale factors with the result of a fitted function (solid curve) superimposed. The combined scale factors with the overall uncertainty are centred around the fit result. The last bin includes overflow entries.

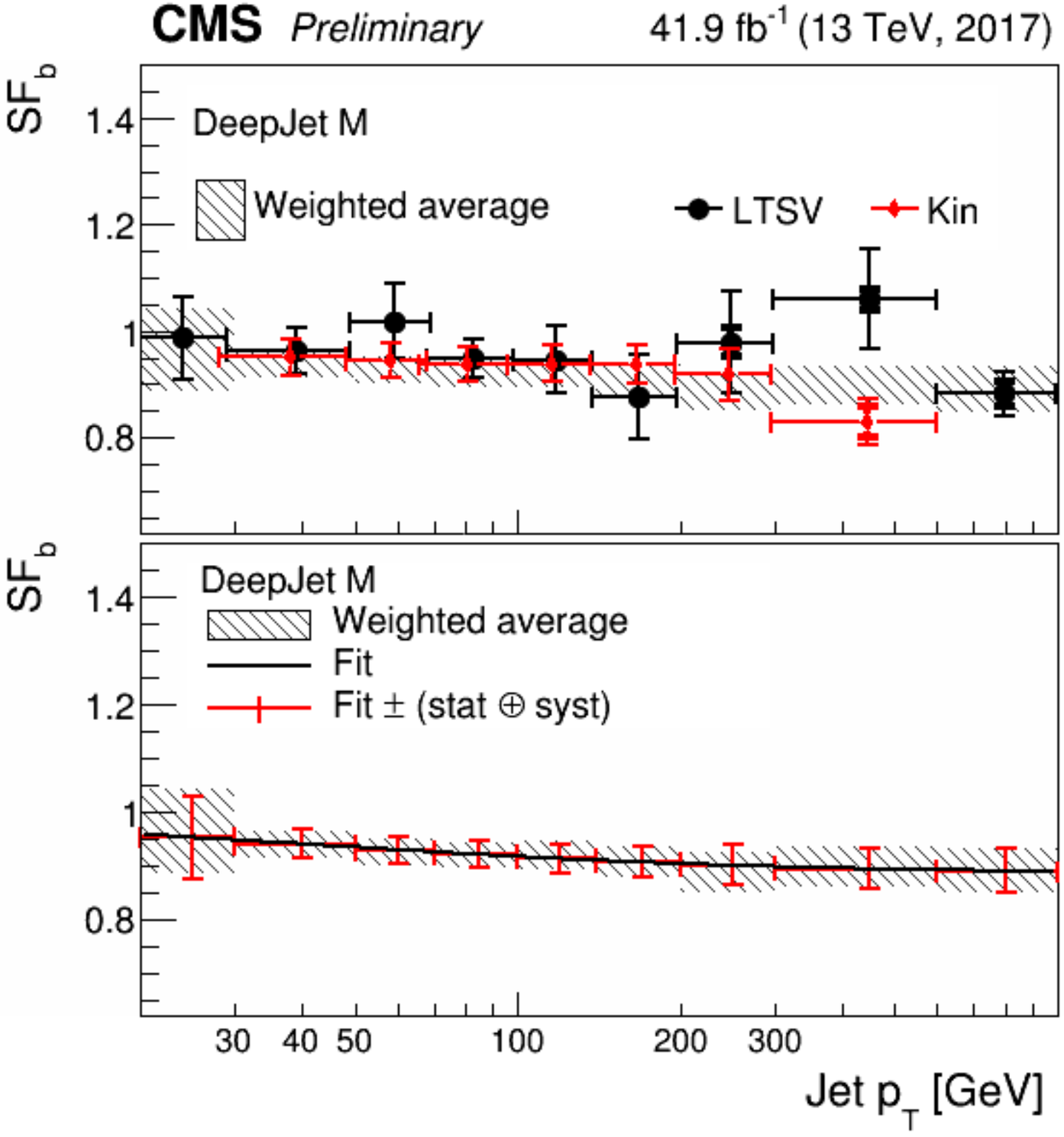


Figure 2: Data-to-simulation scale factors for b jets as a function of jet p_T for the medium working point using the DeepJet b tagging algorithm.The top plot shows the scale factors as a function of jet p_T obtained from the Kin and the LTSV methods, separately. The inner error bars represent the statistical uncertainty, whereas the outer error bars represent the total uncertainty. The scale factors obtained by combining the two measurements along with the total statistical plus systematic error are shown by the hatched area. The combination of all the methods is performed using the best linear unbiased estimator method (BLUE) method. As a result of the limited amount of data in the high jet p_T regions, statistical fluctuations can be seen to cause some discrepancy between the measured scale factors from the two methods.

The bottom panel shows the same combined scale factors with the result of a fitted function (solid curve) superimposed. The combined scale factors with the overall uncertainty are centred around the fit result. The last bin includes overflow entries.

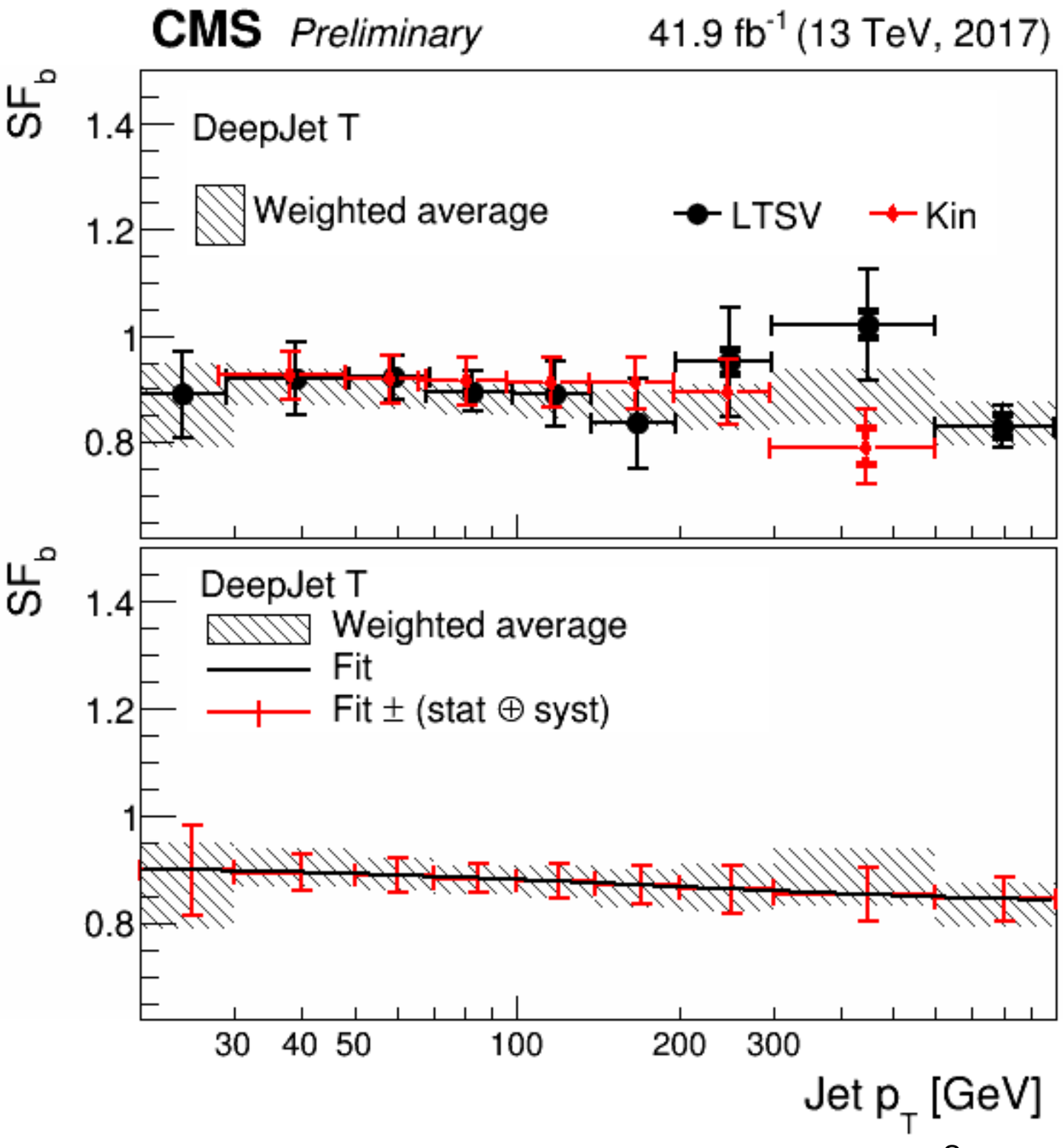


Figure 3: Data-to-simulation scale factors for b jets as a function of jet p_T for the tight working point using the DeepJet b tagging algorithm.The top plot shows the scale factors as a function of jet p_T obtained from the Kin and the LTSV methods, separately. The inner error bars represent the statistical uncertainty, whereas the outer error bars represent the total uncertainty. The scale factors obtained by combining the two measurements along with the total statistical plus systematic error are shown by the hatched area. The combination of all the methods is performed using the best linear unbiased estimator method (BLUE) method. As a result of the limited amount of data in the high jet p_T regions, statistical fluctuations can be seen to cause some discrepancy between the measured scale factors from the two methods.

The bottom panel shows the same combined scale factors with the result of a fitted function (solid curve) superimposed. The combined scale factors with the overall uncertainty are centred around the fit result. The last bin includes overflow entries.

Kin Method

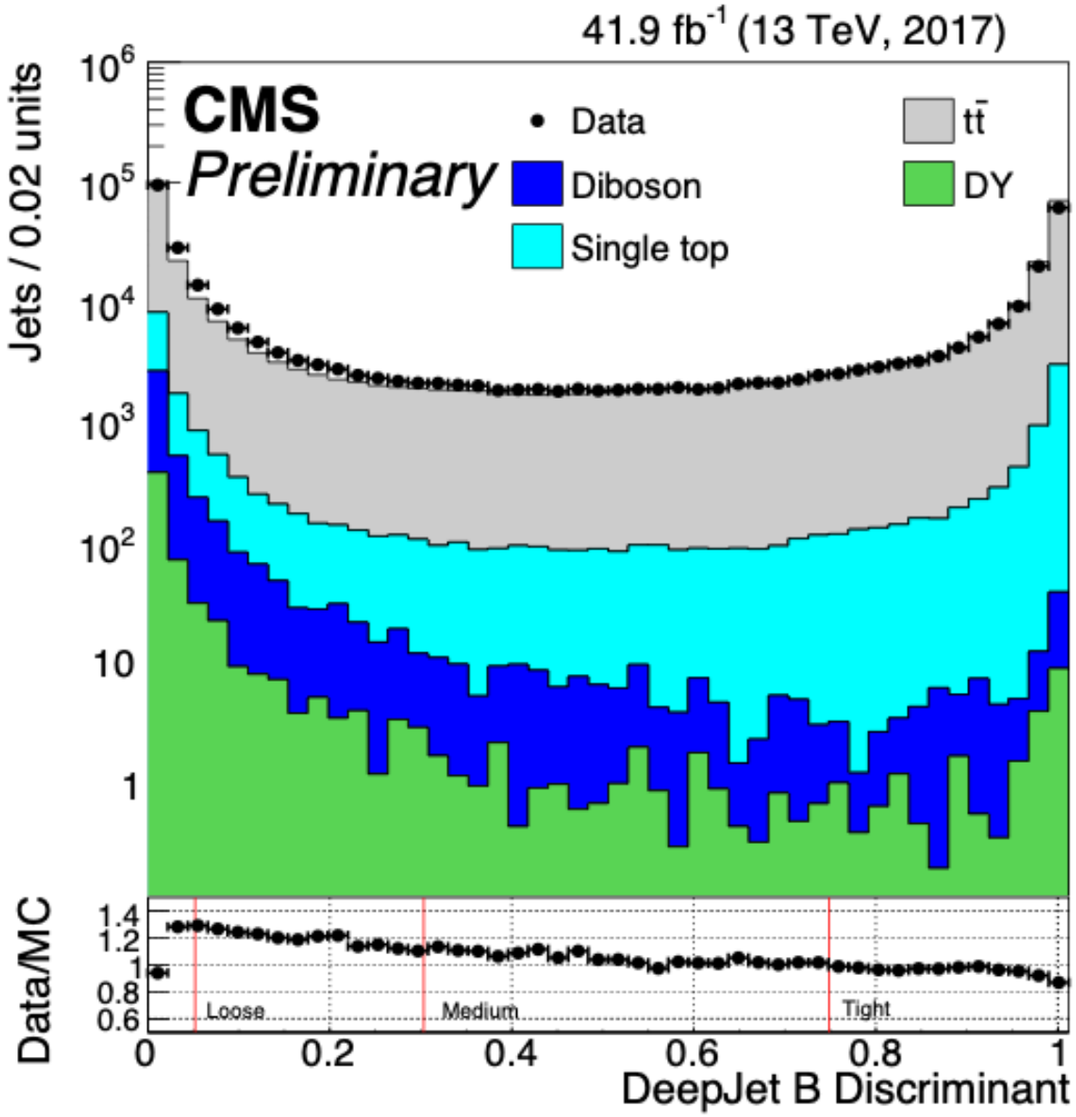


Figure 4: Comparison of the distribution of the DeepJet b tagging discriminant obtained from simulation and data after applying the top pair selection used in the Kin method. The distribution from simulated samples has been normalised to the total number of weighted events of the data distribution.

Performance

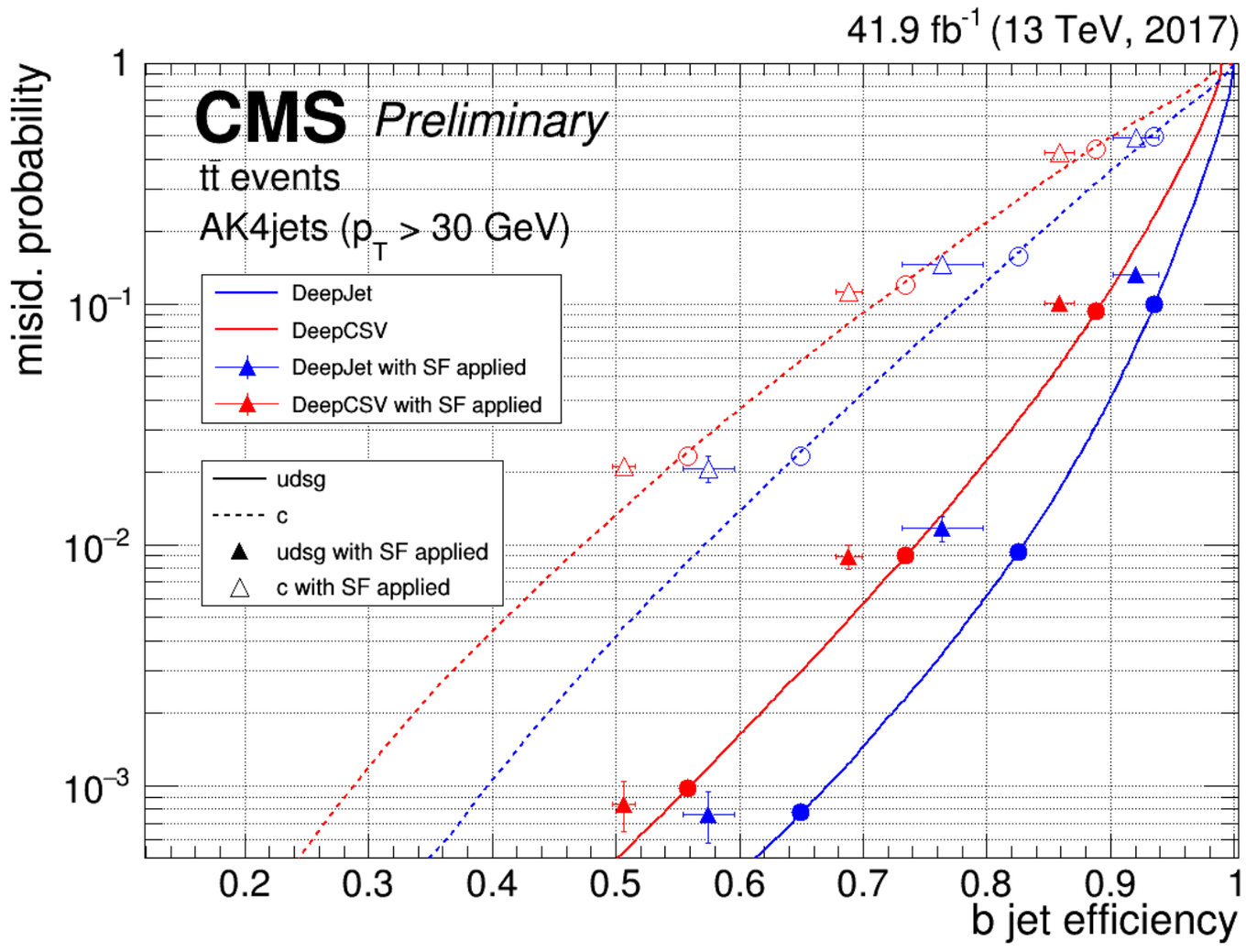


Figure 5: Performance of the DeepJet and DeepCSV algorithm. The plot shows the probability of misidentifying non-b jets as b jets with respect to the efficiency of correctly identifying b jets. The results shown were obtained using jets with $|\eta|<2.5$ and $p_T > 30$ GeV from simulated top pair events. b jets from gluon splitting yielding a pair of b hadrons are labelled as b jet. For the loose, medium and tight working points, the data-to-simulation scale factors have been applied and are represented by the cross-hairs. Circular markers represent the performance of the respective working point in simulated data.

Performance

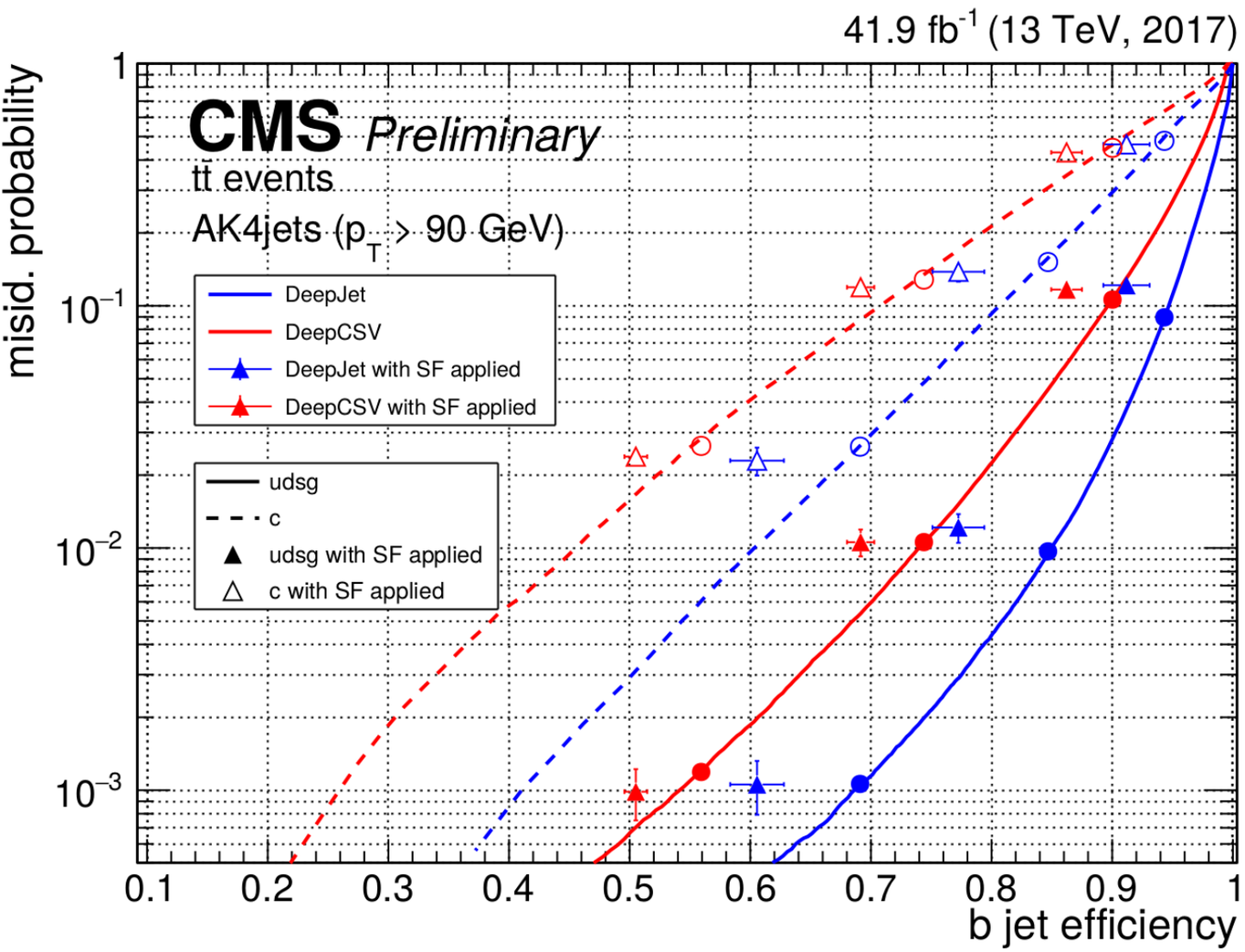


Figure 6: Performance of the DeepJet and DeepCSV algorithm. The plot shows the probability of misidentifying non-b jets as b jets with respect to the efficiency of correctly identifying b jets. The results shown were obtained using jets with $|\eta|<2.5$ and $p_T > 90$ GeV from simulated top pair events. b jets from gluon splitting yielding a pair of b hadrons are labelled as b jet. For the loose, medium and tight working points, the data-to-simulation scale factors have been applied and are represented by the cross-hairs. Circular markers represent the performance of the respective working point in simulated data.

Performance

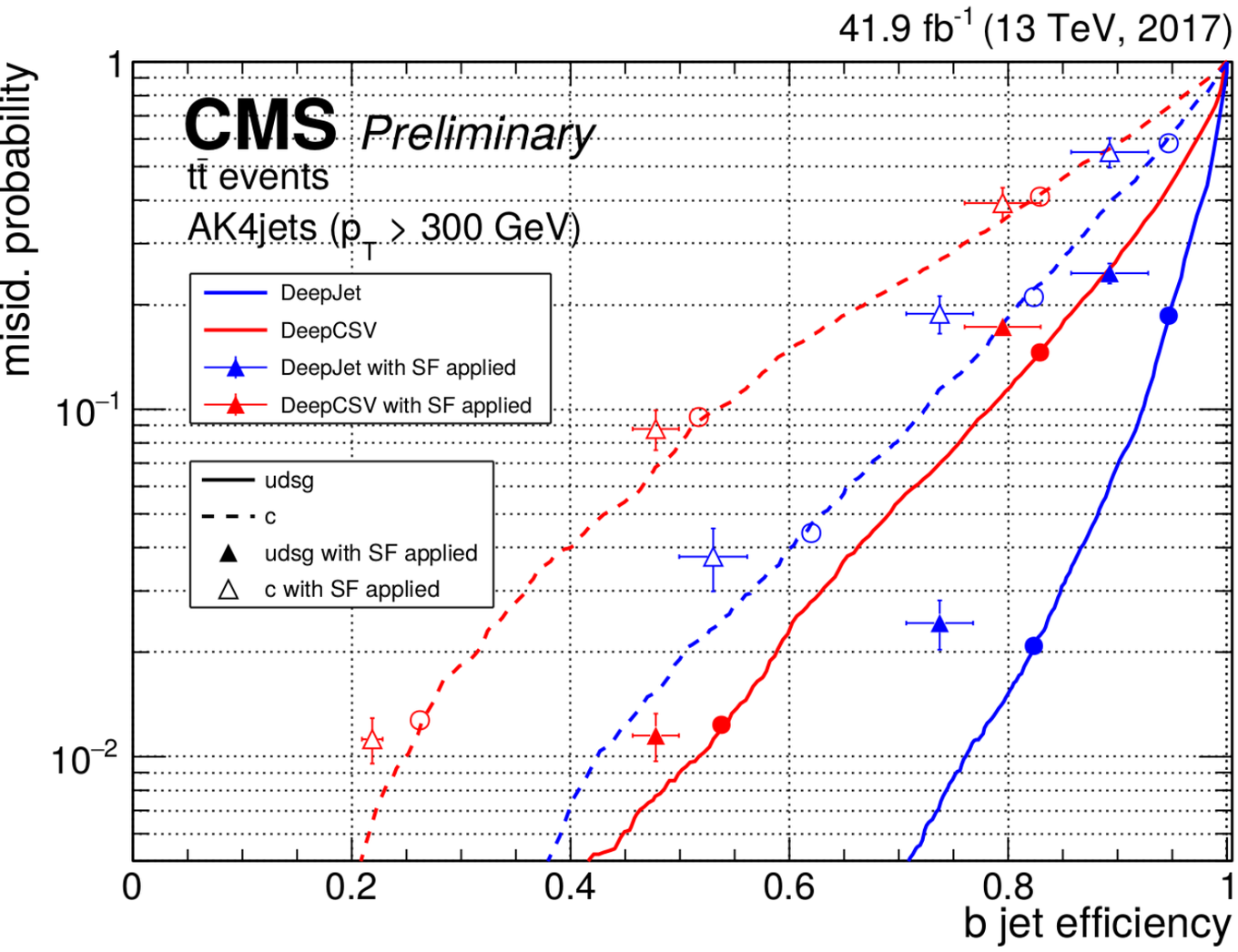


Figure 7: Performance of the DeepJet and DeepCSV algorithm. The plot shows the probability of misidentifying non-b jets as b jets with respect to the efficiency of correctly identifying b jets. The results shown were obtained using jets with $|\eta|<2.5$ and $p_T > 300$ GeV from simulated top pair events. b jets from gluon splitting yielding a pair of b hadrons are labelled as b jet. For the loose and medium working points, the data-to-simulation scale factors have been applied and are represented by the cross-hairs. Circular markers represent the performance of the respective working point in simulated data. The tight working point is not displayed due to the large statistical uncertainties of the testing sample.

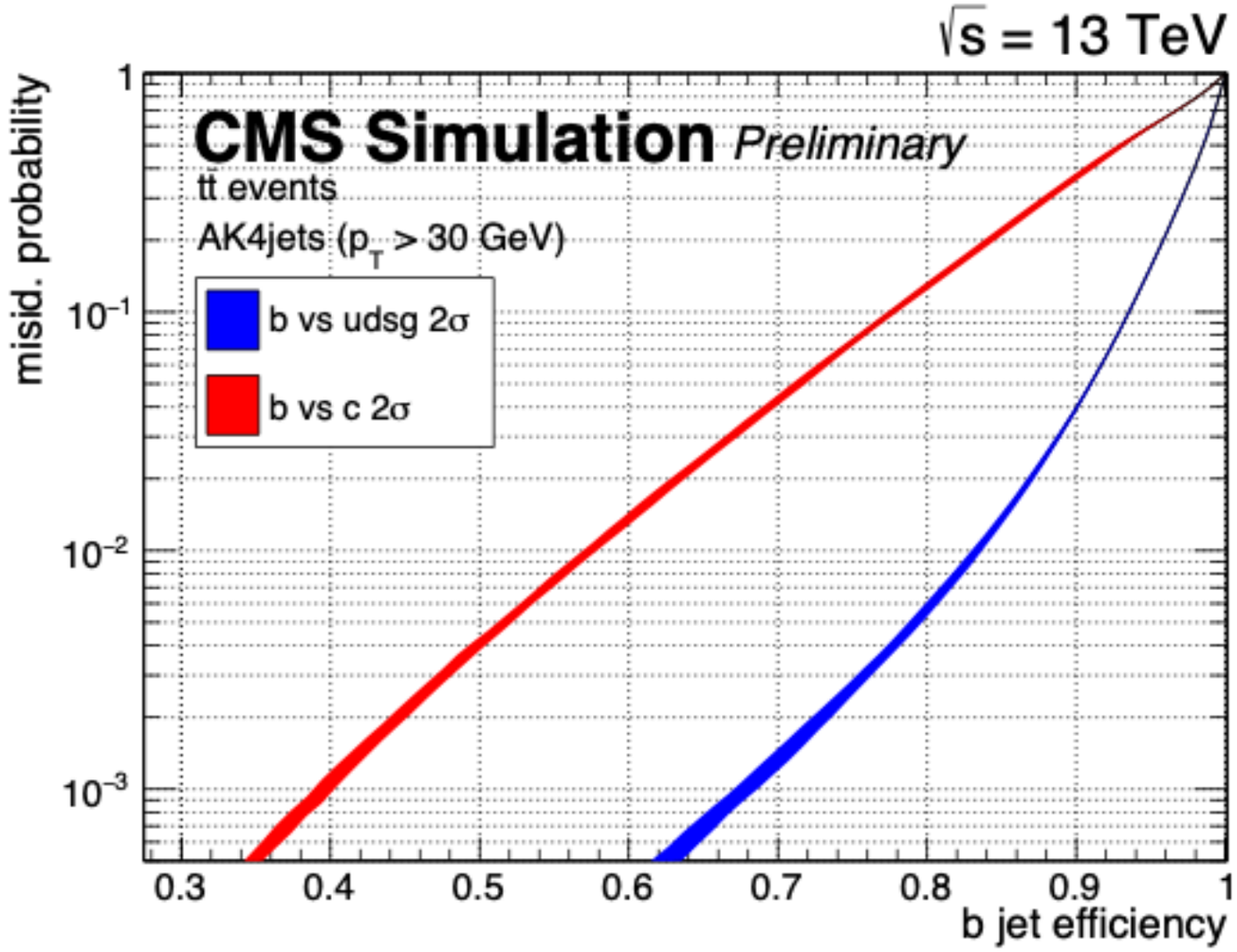


Figure 8: The plot shows the difference in performance of the DeepJet neural network when initialised using random weights. The results shown were obtained using jets with $|\eta|<2.5$ and $p_T > 30$ GeV from simulated top pair events.

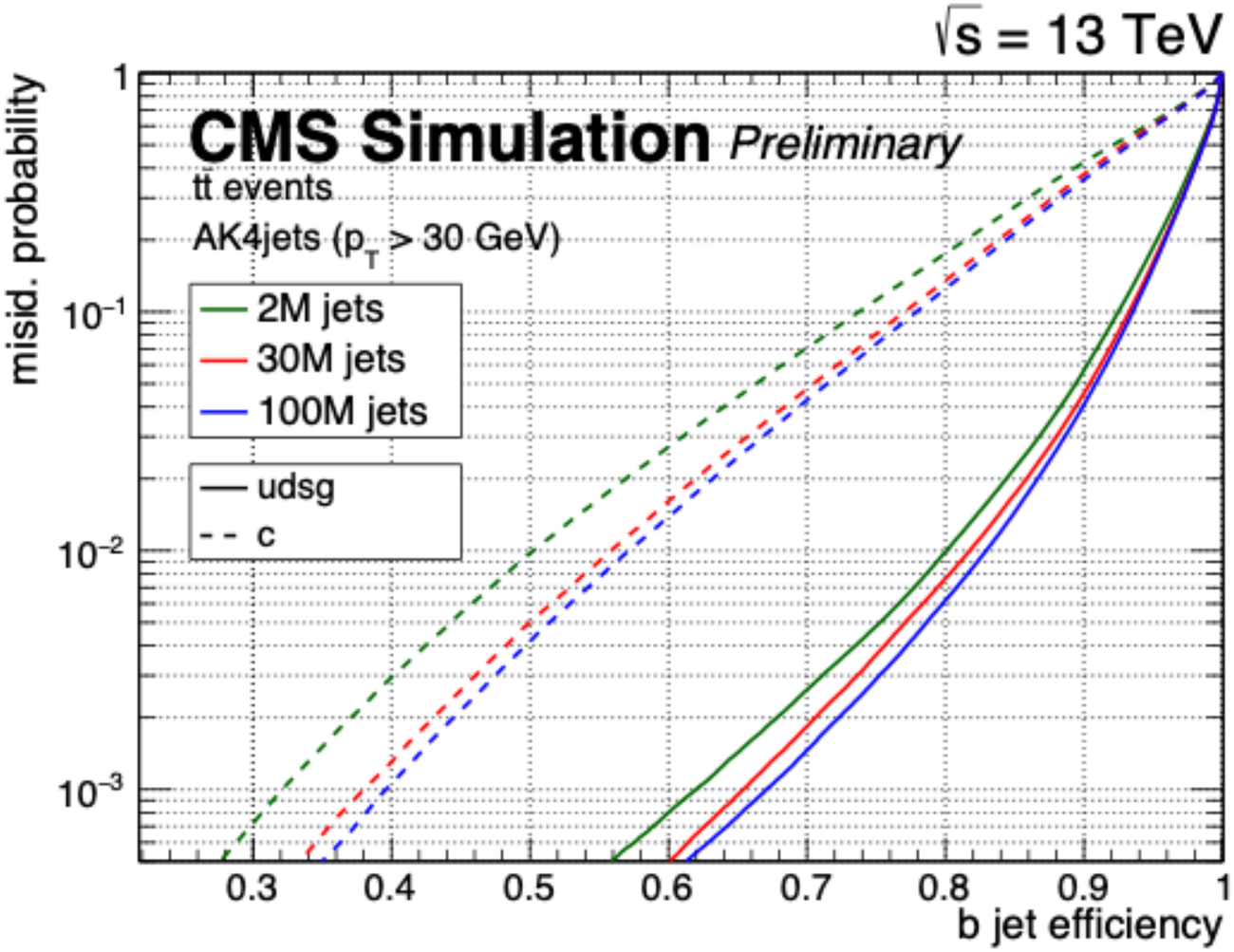


Figure 9: The plot shows the dependence of the DeepJet performance on the training set size. The results were obtained using jets with $|\eta| < 2.5$ and $p_T > 30$ GeV from simulated top quark pair events. The blue curve represents the typically available dataset in the top quark pair and QCD multijet MC production for a given year.