Machine learning approaches for parameter reweighting in MC samples of top quark production in CMS

Valentina Guglielmi$^{1,*}$

Deutsches Elektronen-Synchrotron (DESY), Notkestrasse 85, Hamburg, Germany

E-mail: valentina.guglielmi@desy.de

In particle physics, Monte Carlo (MC) event generators are needed to compare theory to the measured data. Many MC samples have to be generated to account for theoretical systematic uncertainties, at a significant computational cost. Therefore, the MC statistic becomes a limiting factor for most measurements and the significant computational cost of these programs a bottleneck in most physics analyses. In this contribution, the Deep neural network using Classification for Tuning and Reweighting (DCTR) approach is evaluated for the reweighting of two systematic uncertainties in MC simulations of top quark pair production within the CMS experiment. DCTR is a method, based on a Deep Neural Network (DNN) technique, to reweight simulations to different model parameters by using the full kinematic information in the event. This methodology avoids the need for simulating the detector response multiple times by incorporating the relevant variations in a single sample.

European Physical Society, EPS2023
20-25 August, 2023
Hamburg, Germany

$^1$For the CMS Collaboration.
$^*$Speaker

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).
1. Introduction

In particle physics, Monte Carlo (MC) event generators are essential for comparing theoretical predictions with experimental data. Generating numerous MC samples is necessary to account for theoretical uncertainties, which comes at a significant computational cost. As a result, the limitations of MC statistics hinder many measurements, and the high computational expense poses a bottleneck for physics analyses. For instance, in a recent study on top quark-antiquark pair production, the main source of uncertainty in the mass of the top quark came from the MC statistics of the samples used for systematic estimation [1]. To address this issue, reweighting the MC samples can be a solution. This approach involves generating only a sample with nominal values, and then variations are obtained by reweighting this nominal sample. By doing so, the need for simulating the detector response multiple times is eliminated, reducing the MC statistics and computational cost. In contrast to traditional reweighting, which compares distributions in specific bins at a truth level, Machine Learning (ML) reweighting offers a more flexible approach. Standard reweighting is limited by the choice of binning and the number of input dimensions. ML reweighting, using a neural network, does not have such limitations. It can utilize all event information as input, improving reweighting precision. Additionally, it allows for the simultaneous reweighting of multiple MC parameters, considering their correlations. This project introduces Deep Neural Network (DNN) using Classification for Tuning and Reweighting (DCTR) [2] and tests its performance in two scenarios [3]: discrete reweighting of the $h_{\text{damp}}$ variation in parton-level Powheg HVQ events and continuous reweighting of a b-fragmentation function parameter in particle-level Pythia8 events.

![Figure 1: The PFN architecture used in the DCTR method is depicted. It parametrizes the per-particle mapping $\Phi$ (on the left) and the function $F$ (on the right), shown for the case of a latent space of dimension $l = 8$. The latent observable is $O_a = \sum_i \Phi_a(p_T, y, \phi, m, \text{PID})$ [4].](image)

2. $h_{\text{damp}}$ discrete reweighting

The default MC sample of top pair production in CMS is generated using the Heavy Quark Process (HVQ) [5] of the event generator Powheg [6][7]. In Powheg, the resummation of the next-to-leading-order radiation is regulated by the $h_{\text{damp}}$ variable, which enters the damping parameter
Machine learning approaches for parameter reweighting in MC samples of top quark production in CMS
Valentina Guglielmi

\[ D = \frac{h_{\text{damp}}^2}{p_T^2 + h_{\text{damp}}^2} \]  

(1)

where \( p_T \) is the transverse momentum of the particle and \( h_{\text{damp}} \) a parameter defined as \( h_{\text{damp}} = h \cdot m_t \), where \( m_t \) is the mass of the quark top (\( m_t=172.5 \) GeV) and \( h \) is a real number. Since the parameter \( h_{\text{damp}} \) is not physical, an arbitrary value must be chosen in the simulation and varied to calculate the associated systematic uncertainty. Two variations (down and up) from its nominal value are considered in CMS. The nominal value of \( h_{\text{damp}} \) is set to \( 1.379 \cdot m_t \) in black and the nominal value of \( h_{\text{damp}} \) in green. The red line shows the nominal sample reweighted to the down \( h_{\text{damp}} \) variation using the DCTR method [2]. The bottom pad shows the ratio of the nominal sample to the \( h_{\text{damp}} \) down variation before and after the reweighting. The error bars represent the statistical uncertainties due to the limited statistics of the MC.

Figure 2: The normalised differential cross section as a function of the \( p_T \) (left) or \( \eta \) (right) of the \( t\bar{t} \) system at parton level for a MC simulated \( t\bar{t} \) sample generated with the POWHEG hvq program, [6] [7], for the CMS down variation of \( h_{\text{damp}} \) (0.8739 \( \cdot \) \( m_t \)) in black and the nominal value of \( h_{\text{damp}} \) (1.379 \( \cdot \) \( m_t \)) in green.
Machine learning approaches for parameter reweighting in MC samples of top quark production in CMS
Valentina Guglielmi

Figure 3: The distribution of $x_b$ (left) or $p_T^B$ of the B-hadron (right) of the $t\bar{t}$ system for a MC simulated
$t\bar{t}$ sample generated with Pythia 8 [8], for the CMS nominal variation of $r_b$ (0.855) in black and a second
value of $r_b$ (1.1) in green. The red line shows the sample with the variation of $r_b$ (0.855) in black and a second
generated with the nominal $r_b$ using the DCTR method, [2]. The bottom pad shows the ratio of the sample
with the variation of $r_b$ to the sample generated with the nominal $r_b$ before and after the reweighting. The
error bars represent the statistical uncertainties due to the limited statistics of the MC.

3. B-fragmentation continuous reweighting

Another significant source of uncertainty in top physics is the B-fragmentation. It involves the
decay of a top quark into a b quark and the subsequent formation of a B-hadron through a process
described by the Lund string model in Pythia [8]. The probability distribution for heavy quarks
during this process is given by Eq. 2:

$$f_B(z) \propto \frac{1}{z^{1+b r_b m_t^2}} (1-z)^a \exp(-b m_t^2/z)$$

(2)

where $z$ is the quark longitudinal momentum, $m_t$ the top quark mass and $a$ and $b$ free parameters
to be tuned to experimental data. For $b$ quarks, the Bowler modification is also considered,
involving the $b$ quark mass $m_b$ and the $r_b$ parameter in Pythia. To assess systematic uncertainties,
CMS recommends varying the $r_b$ parameter during MC event generation. In this project, the
DCTR technique is employed to estimate the uncertainty related to B-fragmentation, which affects
physical observables like $x_b$ and $p_T^B$ representing the energy fraction transferred from the b quark
to the B-hadron and the transverse momentum of the B-hadron, respectively. In this scenario, a
continous reweighting with 10 different values of $r_b$ parameter in Pythia in the range [0.6, 1.4] has
been performed training a single DNN model to reweight the samples generated with the different
$r_b$ values to the nominal sample with $r_b$ (0.855). For the nominal $r_b$ sample 5M events and 5k events
for each $r_b$ variation sample are passed as inputs to the DNN for a total amount of 10M events to
train. The $x_b$ variable, defined in Eq 3, is passed as the only input to the DNN:

$$x_b = \frac{2 p_B \cdot q}{m_t^2} \cdot \frac{1}{1 - m_W^2/m_t^2}$$

(3)

where $p_B$ and $q$ are the four vector of the B-hadron and of the top quark and $m_W$ and $m_t$ the W boson and top quark mass, respectively. The performance of reweighting is checked by applying the weights obtained in the training to the $x_b$ and $p_B^B$ observables of the $t\bar{t}$ system. The results for one of the 9 values of $r_b$ (1.1) trained are shown in Fig. 2. The original sample and the reweighted one agree also in this case with a method closure within 2%. Similar results are obtained for all the other 8 values of $r_b$ tested, as reported in [3].

### 4. Results and Conclusions

In these proceedings, the DCTR method is evaluated in two different scenarios: discrete reweighting of the $h_{damp}$ variation in Parton-level Powheg HVQ events and continuous reweighting of a $b$-fragmentation function parameter in particle-level Pythia events. The method is found to work very well in both scenarios, with a method closure within 2%. The method is implemented into CMS software framework for both cases. This approach has the potential to be applied to various other intriguing cases in the field of top physics and, more broadly, in other physics areas.

References


