Determination of missing transverse momentum using multivariate methods for a differential top pair production cross section measurement at CMS

Master Thesis

von
Philipp Nattland

vorgelegt der
Fakultät für Mathematik, Informatik und Naturwissenschaften der
RWTH Aachen

im Oktober 2022

erstellt im
I. Physikalischen Institut B

bei
Prüfer: Prof. Dr. Lutz Feld
Zweitprüfer: Prof. Dr. Alexander Schmidt

Abstract

In this thesis, a deep neural network (DNN) is developed, aiming to correct for detector effects in the reconstruction of observables, which are used in the binning of a differential top pair ($\mathrm{t\bar{t}}$) production cross section measurement. These binning variables are given by the momentum imbalance in the plane transversal to the beam axis ($p_{\mathrm{T}}^{\mathrm{miss}}$) and the azimuthal angle between $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and the lepton closest to it ($|\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}, \text{nearest } \ell)|$). The DNN is trained based on MC simulations corresponding to the data recorded at the CMS experiment in Run 2 of the LHC, with a center-of-mass energy of 13 TeV and total integrated luminosity of $138\,\mathrm{fb}^{-1}$.
The input features of the DNN are validated by performing goodness-of-fit tests of the feature distributions between data and simulation. Subsequently, the settings of the DNN are optimized in a Bayesian search. The performance of the optimized DNN is then studied. The resolution and bias of $p_{\mathrm{T}}^{\mathrm{miss}}$ and $|\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}, \text{nearest } \ell)|$ are compared for reconstruction with and without application of the DNN-based correction in simulated samples of the $\mathrm{t\bar{t}}$ process and its backgrounds. Additionally, the event migrations corresponding to the reconstruction of the $\mathrm{t\bar{t}}$ binning variables are studied for reconstruction with and without the DNN.
The studies showed that by applying the DNN-based correction, the reconstruction resolution of the binning variables could be increased significantly, resulting in less event migrations in the binning of the differential $\mathrm{t\bar{t}}$ cross section measurement.

Contents

1 Introduction

The Standard Model of particle physics (SM) [1, 2] successfully describes the fundamental particles and their interactions. Particle accelerators are crucial in validation and extension of this model. The currently most powerful of these accelerators is the Large Hadron Collider (LHC) [3], which was successfully employed in the discovery of the Higgs Boson [4, 5], the latest addition to the SM. Two of the four large experiments at the LHC are the multi-purpose detectors ATLAS (A Torodial LHC Apparatus) [6] and CMS (Compact Muon Solenoid) [7], both used to study among others proton-proton collisions.
Even though the history of the SM has been a success story, there are clear indications of its incompleteness. Shortcomings include the missing description of gravitation and dark matter, the inability to describe the unification of the fundamental forces, or a fine tuning problem with regards to the mass scale of the Higgs boson [1]. To explain these phenomena, physics beyond the Standard Model (BSM) is needed. In order to extract information on possible extensions of the SM, its predictions are tested in precision measurements, where deviations from the SM predictions could hint to BSM physics.
This thesis contributes to such a precision measurement, where the pair production cross section of the top quark is measured. The data set used in the measurement was recorded at the CMS experiment in 2016 to 2018 at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and corresponds to a combined integrated luminosity of $138\,\text{fb}^{-1}$ [8–10]. The cross section is measured differentially in two specifically chosen observables to further increase the sensitivity to BSM scenarios. As a result, the binning of the $\text{t}\bar{\text{t}}$ cross section measurement is a function of the so-called missing transverse momentum ($\vec{p}_{\text{T}}^{\,\text{miss}}$), explained in more detail in Sec. 2.4. It is derived as the sum over the transverse momenta $\vec{p}_{\text{T}}$ of all visible particles and is, due to momentum conservation, an estimator for the sum over the transverse momenta of all invisible particles in an event. In the case of the dileptonic top pair decay, the $\vec{p}_{\text{T}}^{\,\text{miss}}$ can be used to approximate the momentum of the two tree-level neutrinos ($\vec{p}_{\text{T}}^{\,\nu\nu}$), which emerge from the decay.
A precise measurement of the missing transverse momentum is therefore crucial, but an accurate reconstruction is challenging, since measurement errors of all transverse momenta are directly propagated to $\vec{p}_{\text{T}}^{\,\text{miss}}$. To correct for these detector effects and thus to improve the estimation of the true momentum of the neutrino system $\vec{p}_{\text{T}}^{\,\nu\nu}$, a correction on the missing transverse momentum is determined. This correction, derived using a deep neural network, is subject of this thesis.
The structure of the thesis can be summarized as follows. The theoretical background is introduced in Sec. 2, including an explanation of the observables of the differential $\text{t}\bar{\text{t}}$ measurement. An overview of the experimental setup is given in Sec. 3, the used data sets and the background estimation are presented in Sec. 4. In the following Sec. 5, the object definitions as well as the event selection and the systematic uncertainties are summarized. Foundations of the structure and functionality of deep neural networks (DNNs) are outlined in Sec. 6 together with a Bayesian optimization algorithm to improve the DNN performance. The development and the implementation of the DNN-based $\vec{p}_{\text{T}}^{\,\text{miss}}$ correction are summarized in Sec. 7, introducing the input features of the DNN and discussing the input feature validation process. Subsequently, in Sec. 8, the performance of the DNN, obtained from the Bayesian optimization, is evaluated. More specifically, resolution and bias of the $p_{\text{T}}^{\text{miss}}$ reconstruction are compared with and without application of the correction yielded by the DNN. This is shown for the simulated $\text{t}\bar{\text{t}}$ sample as well as simulated samples of background processes. Additionally, the impact of the DNN on the event migration in the binning of the $\text{t}\bar{\text{t}}$ cross section measurement is presented. The thesis concludes with a summary and an outlook in Sec. 9.

2 Theoretical Foundation

In this section, the unit system of this thesis is discussed and the theoretical foundations of the Standard Model of particle physics are presented. Despite the successes and robustness of the SM, there are indications of its shortcomings. These are explained in the following, before outlining a possible extension to the SM called Supersymmetry. Lastly, the methodology of the differential $t\bar{t}$ production cross section measurement is explained.

2.1 Units

In the natural unit system, the reduced Planck constant $\hbar$ and the speed of light c are set to unity:

$$\hbar = c = 1 \ . \tag{2.1}$$

This choice results in concurrent units of mass, energy and momentum, which can be given in GeV. Due to this simplicity, natural units are used in this thesis. Length and time are stated in the metric system and cross sections are given in barn ($1\,\mathrm{b} = 10^{-28}\,\mathrm{m}^2$). Charges are denoted in multiples of the elementary charge e.

2.2 Standard Model of Particle Physics

The Standard Model of particle physics [1, 2] describes three of the four known fundamental forces as interactions between elementary particles. The Lagrangians of these forces are directly derived from requiring invariance under local gauge transformation. The SM is thus based on the combination of the gauge symmetry groups

$$SU(3) \times SU(2) \times U(1) \, . \tag{2.2}$$

Here, $SU(3)$ is the gauge group of the strong interaction, while $SU(2) \times U(1)$ is associated with the electroweak interaction, combining weak and electromagnetic interactions into one unified gauge theory.
The particle content of the SM, summarized in Fig. 2.1, is split into two main groups, fermions and bosons, with fermions having half-integer spin and bosons having integer spin. Each fermion has an associated anti-particle that has the opposite sign in all charge-like quantum numbers. In this thesis, no distinction is made between particles and their anti-particles, unless otherwise stated.
Fermions are subdivided into leptons and quarks. They can be distinguished by the types of interactions they partake in. Only quarks interact by the strong interaction, while all fermions interact by the weak interaction. Leptons are grouped into electrically charged leptons with weak isospin $I_3 = -1/2$ and electrically neutral neutrinos with weak isospin $I_3 = +1/2$. The three charged leptons are ordered into generations according to their masses. To each charged lepton (electron e, muon μ, tau τ), one neutrino of the same flavor (electron neutrino ν_e, muon neutrino ν_μ, tau neutrino ν_τ) is associated. Neutrinos are (according to the SM) massless and interact only via the weak interaction. Direct measurements of neutrinos are therefore generally very challenging.
Quarks carry charges of all three interactions including the charge of the strong interaction, called color charge. Analogously to leptons, quarks are split into generations according to their masses. There are two quarks in each generation, one up-type quark with an electrical charge of $+2/3$ and weak isospin of $+1/2$, and one down-type quark with an electrical charge of $-1/3$ and weak isospin of $-1/2$. The first generation with the quarks of the lowest masses contains the up

and down quark, the second generation contains the charm and strange quark. Most relevant in this thesis is the third generation, made up by the top and bottom quark. Quarks can decay into less massive quarks via the weak interaction, the likelihood of such decays is given by the CKM matrix. Its small off-diagonal and large diagonal elements show that decays within one generation are generally much more likely than between generations, especially for the decay of the top quark. Hence, the top quark almost exclusively decays into a bottom quark, resulting in a decay as shown in Fig. 2.2.

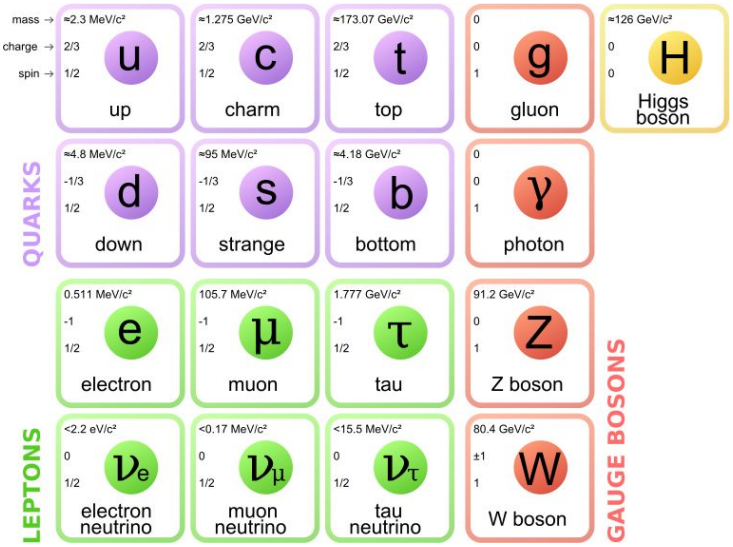


Figure 2.1: Particle content of the Standard Model [11]. Quarks (violet) and leptons (green) are fermions, sorted into generations, bosons include gauge bosons (orange) and Higgs boson (yellow). Mass, charge and spin of all particles are given.

For bosons, one distinguishes between gauge bosons with a spin of one and the Higgs boson which carries spin zero. Gauge bosons mediate the fundamental forces. The weak interaction is mediated by the massive vector bosons, Z and $W^{\pm}$, that carry electrical charge in the case of $W^{\pm}$. Photons are associated with the electromagnetic interaction, massless and electrically neutral. Gluons are also massless and mediate the strong interaction. They carry color charge and therefore self-interact. This leads to an effect called color-confinement [12]. It describes that, when pulling apart two strongly interacting particles, the gluon field between the particles forms a narrow flux tube, resulting in a constant force. Therefore, at a certain distance between the particles, formation of a new quark-antiquark pair is energetically favorable. Strongly interacting quarks thus cannot propagate freely, but from bound, color-neutral states (hadrons) of two quarks (mesons) or three quarks (baryons). Hence, when a quark leaves the interaction point (vertex) of a proton-proton collision, it hadronizes, forming a hadronic shower in the detector, called jet. The lifetime of hadrons depends on their quark constituents. Since bottom quarks cannot decay further in the same generation, the decay of b quarks is suppressed via the CKM matrix, resulting in a longer lifetime of B hadrons. This characteristic signature can be used to identify bottom quarks.
The Higgs boson is the latest addition to the Standard Model. It was postulated in the 1960s as a consequence of the Brout-Englert-Higgs mechanism to preserve the local gauge invariance of the SM while solving the problem of the SM which predicted vector bosons to be massless. The observation of the Higgs boson in 2012 at CMS and ATLAS [4, 5] was a milestone in particle physics, fulfilling one of the main goals for the LHC.

2.3 Limits of the Standard Model

Despite the successes of the SM, there are notable shortcomings. For instance, it does not describe gravity and is thus only applicable at scales below the Planck scale $\mathcal{O}(10^{19}\,\mathrm{GeV})$.
Furthermore, various astrological observations show that a large part of matter in our universe does not interact via strong or electromagnetic forces [13–16]. For this so-called dark matter, no suitable candidate is included in the SM.
The hierarchy problem of the SM is a fine-tuning problem. When calculating processes in particle physics, loop corrections have to be considered. In these corrections, the momentum of the involved particles is undetermined and must be integrated over the whole energy scale of the theory. Therefore, the energy cutoff scale Λ_{UV} of the theory appears. For scalar particles, such as the Higgs boson, this leads to corrections to the mass that are squared in this scale:

$$m_{\mathrm{H}}^2 = m_0^2 - \frac{\left|\lambda_t^2\right|}{8\pi^2}\Lambda_{\mathrm{UV}}^2 + \dots \,, \tag{2.3}$$

where m_{H} and m_0 are the corrected and tree-level Higgs boson mass, respectively, and λ_t corresponds to the coupling between top quark and Higgs boson. If the cutoff scale is at the Planck scale ($\sim 10^{19}\,\mathrm{GeV}$), the tree-level Higgs boson mass has to be chosen at a precision of $2 \cdot 19 - 2 \cdot 2 = 34$ digits to achieve the effective Higgs boson mass of $\sim 10^2\,\mathrm{GeV}$ that we observe [17].
Another shortcoming of the SM is that it does not describe the unification of all fundamental forces. Describing different phenomena in a unified way within a new, more general theory can provide new insights and help to produce a more consistent picture of nature. This was the case for the unification of the weak and the electromagnetic interaction to the electroweak interaction. The next step, a unification of electroweak and strong interaction, is not possible in the SM since it would require the gauge couplings of the SM to converge at high energies. Such an intersection point of the forces cannot be found in the SM.
The SM also predicts neutrinos to be massless. However, several appearance and disappearance measurements have shown that neutrinos can oscillate between different flavors [18–20]. These neutrino-oscillations require neutrinos to be massive, in conflict with the SM prediction.

2.3.1 Supersymmetry

One possible extension to the SM, that is able to solve some of its shortcomings, is Supersymmetry (SUSY) [21]. The particle content of the SM is extended, adding one supersymmetric partner to each SM particle. These superpartners differ in spin by 1/2, such that fermions have bosonic partners and vice versa. All other quantum numbers are shared between the partners. As none of these SUSY particles have been observed in the experiments, they must have different masses to their partners. Therefore, SUSY has to be a so-called broken symmetry.
Supersymmetry includes suitable dark matter candidates. The unification of the three fundamental forces could also be possible in SUSY, since there is an intersection of the gauge coupling constants of all forces at high energies. Additionally, if the mass difference between particles and their superpartners is not too large, it could solve the hierarchy problem [22].
So far, there has been no direct evidence for Supersymmetry. Since the lower mass range has been explored extensively, searches for SUSY especially focus on high energies, presuming large masses of the superpartners. As some of the SUSY particles, such as neutralinos, only interact weakly, a phase space of large missing transverse momentum is of particular interest. In this phase space, top pair production is one of the largest backgrounds and a precise determination of its cross section is therefore integral for many SUSY searches.

2.4 Differential $t\bar{t}$ Production Cross Section Measurement

The analysis this thesis is embedded in, measures the differential cross section of the $t\bar{t}$ production in the dileptonic decay channel. As mentioned previously, top quarks almost always decay into bottom quarks, radiating a W boson. The W boson can in turn decay hadronically into two quarks or leptonically into charged lepton and associated neutrino. In this measurement, only final states, where both W bosons decay leptonically are considered. This so-called dileptonic $t\bar{t}$ decay is illustrated in Fig. 2.2 together with a BSM process with a similar signature. In the latter, the SUSY partner of the top quark, called stop, is pair produced. The stops then decay into neutralinos, which cannot be detected directly, and top quarks, resulting in a signature similar to the signal process.

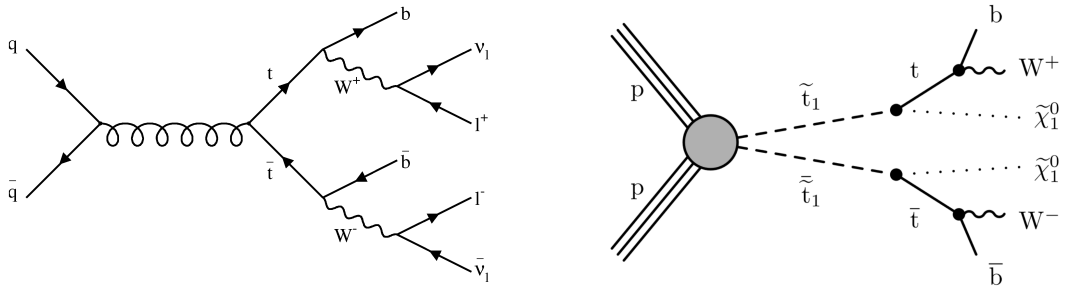


Figure 2.2: On the left a diagram of the dileptonic $t\bar{t}$ production is shown [23]. Illustrated on the right is a diagram of the pair production of stop quarks, the supersymmetric partners of the top quarks [24]. The stop quarks decay into neutralinos and top quarks.

The differential cross section measurement is performed in two steps. Firstly, data and simulation are compared in a binning of so-called detector-level observables, that are affected by the experimental setup. The distribution of recorded data is checked for deviations to the simulation that might be due to BSM processes. To guarantee good discriminative power between SM and BSM processes in this binning, the observables are chosen to exploit the different kinematics of the processes. More precisely, the additional invisible particles in BSM events, such as neutralinos, shift the momentum balance between the particles in the final states. In order to measure this imbalance, one exploits the negligible transverse momentum ($\vec{p}_{\mathrm{T}}$) of the collision partners before the collision. Due to conservation of momentum, the sum of the transverse momenta of the decay products has to vanish. Therefore, the sum over the $\vec{p}_{\mathrm{T}}$ of all visible particles must equal the $\vec{p}_{\mathrm{T}}$ sum over all invisible particles, with opposite sign. This quantity, referred to as $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, is essential in this analysis and its reconstruction is explained in more detail in section 5.1. Since $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is correlated with the neutrino momenta in the SM process and predominantly correlated with the heavier invisible BSM particles in BSM scenarios, its absolute value $|\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}| = p_{\mathrm{T}}^{\mathrm{miss}}$ is chosen as one detector-level observable, expecting generally larger $p_{\mathrm{T}}^{\mathrm{miss}}$ values in BSM events. The other detector-level observable of the $t\bar{t}$ analysis is the azimuthal angle $|\Delta\phi|$ in the transverse plane between $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and the charged lepton closest to it, denoted as $|\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}, \text{nearest } \ell)|$. This quantity tends to be small in SM events due to the kinematic correlation of the momenta of the neutrinos and their associated charged leptons. In contrast, BSM events include large additional contributions to the missing transverse momentum originating from invisible particles that are not produced together with the leptons, thus helping to distinguish SM from BSM processes.
In the second step of the differential cross section measurement, the simulated backgrounds are subtracted from the recorded data, if data and simulation are in good agreement. For compa-

rability with theoretical cross section calculations, the cross section is measured differentially in the generation-level quantities $p_{\mathrm{T}}^{\nu\nu}$ and $|\Delta\phi(p_{\mathrm{T}}^{\nu\nu},\ \text{nearest}\ \ell)|$. Therefore, the detector-level distribution of $t\bar{t}$ events has to be corrected for detector effects and for $p_{\mathrm{T}}^{\mathrm{miss}}$ resulting from for other particles, such as neutrinos from the b quark hadronization. In this so-called unfolding procedure, the mapping from generation-level to detector-level variables is estimated and inverted.
The stability of the unfolding procedure can be increased by improving the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ resolution and thus reducing the event migrations that need to be corrected. For this purpose, a deep neural network (DNN) is developed in this thesis.
The DNN is based on detector-level quantities and aims to approximate the generated missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss,\ gen}}$ in an event, based on different $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstructions (see Sec. 5.1.5) and other kinematic variables. The distributions in $p_{\mathrm{T}}^{\mathrm{miss}}$ and $|\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}, \text{nearest}\ \ell)|$, reconstructed by the DNN, can then be unfolded.
The two dimensional generator-level binning of the differential $t\bar{t}$ measurement is summarized in Table 2.1.

Table 2.1: The two dimensional binning used in the differential $t\bar{t}$ measurement.

Variable	Binning
$p_{\mathrm{T}}^{\nu\nu}$	$[0, 40, 65, 95, 125, 160, 200, \infty]$ GeV
$\lvert\Delta\phi(p_{\mathrm{T}}^{\nu\nu},\ \text{nearest}\ \ell)\rvert$	$[0, 0.64, 1.28, 3.14]$

3 Experimental Setup

The data used in this thesis was recorded by the Compact Muon Solenoid (CMS), a detector at the Large Hadron Collider (LHC), in the years 2016, 2017 and 2018. In the following, a short overview of the LHC is given, before summarizing the key features of the CMS detector and its subcomponents.

3.1 Large Hadron Collider

The Large Hadron Collider [3] is currently the largest and most powerful particle accelerator. It was built in a tunnel, 45 m to 170 m underground, close to Geneva at the main site of the European Organization of Nuclear Research (CERN). This tunnel has a circumference of 26.7 km and was built for the predecessor of the LHC, the Large Electron Position Collider (LEP) [25]. The first proton-proton collisions at the LHC were recorded in years the 2010-2012 at center-of-mass energies of $7 - 8$ TeV. For the second run, years 2015-2018, the center-of-mass energy was increased to $\sqrt{s} = 13$ TeV. Recently, the third run has started with collisions at $\sqrt{s} = 13.6$ TeV, close to the design value of 14 TeV. Heavy ions like lead ions are also collided at the LHC, at energies ranging up to 2.76 TeV per nucleon.
The layout of the LHC is illustrated in Fig. 3.1, showing the main acceleration steps and interaction points of the accelerated particles. Protons are accelerated initially in a linear accelerator (LINAC2) and pass through the Proton Synchrotron Booster into the Proton Synchrotron (PS). After further acceleration in the Super Proton Synchrotron (SPS) an energy of 450 GeV is reached and the proton beam is channeled into the LHC storage ring, both clockwise and counterclockwise. Up to 2808 bunches of 10^{11} protons can be injected into the LHC, focused and brought to collision every 25 ns at the interactions points. The interaction points are located at four of

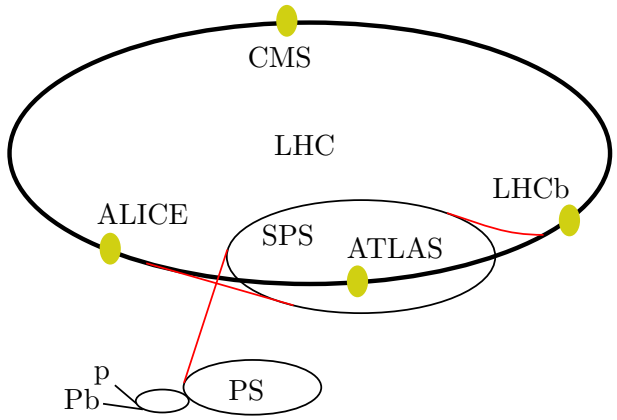


Figure 3.1: Layout of the LHC complex [26]. The four interactions points are shown in yellow and the LHC storage ring and pre accelerators are shown in black. Protons (p) and lead ions (Pb) pass through different acceleration steps before entering the Proton Synchrotron (PS) and Super Proton Synchrotron (SPS).

the eight beam pipe crossings. At each interaction point, one of the four large experiments is situated. Main purpose of the ALICE (A Large Ion Collider Experiment) [27] detector is to study the quark gluon plasma, focusing on heavy ion collisions. The LHCb (LHC beauty) [28] experiment is designed to gain insight into physics involving b quarks. ATLAS and CMS are multi-purpose detectors that are used for precision measurements of the SM and searches for BSM physics. Since both detectors have similar capabilities, but are designed differently, measuring the same phenomenon in both of them can be a valuable cross check.

The capacity of a particle storage ring like the LHC is characterized by the instantaneous luminosity $\mathcal{L}$. Integrating $\mathcal{L}$ over time, yields the integrated luminosity L_{int}. A precise measurement of this value is crucial, because it is the factor of proportionality between the cross section of a process σ and its expected number of occurrences N:

$$\mathcal{L} = \frac{n_b f n_p^2 F}{4\pi\sigma_x\sigma_y} , \qquad L_{\text{int}} = \int \mathcal{L} \cdot dt , \qquad N = L_{\text{int}} \cdot \sigma , \tag{3.1}$$

where n_b is the number of bunches, f is the revolution frequency of the bunches, n_p is the number of particles per bunch, F is a factor accounting for the beam collision geometry and σ_x and σ_y are the width and height of the beam profile, respectively. The data used in this thesis corresponds to a total integrated luminosity of $138\,\text{fb}^{-1}$ [8–10].

3.2 Compact Muon Solenoid

The Compact Muon Solenoid [7], which recorded the data used in this thesis, is one of the four large experiments at the LHC. It is operated by the CMS collaboration, formed by over 4000 people from many scientific institutes all over the world [29]. Centered around one of the interaction points, the cylindrical detector is 21.6 m long, has a diameter of 14.6 m and weighs 14000 t.
As the name suggest, one of the key features of CMS is a superconducting solenoid, that can generate a near homogeneous magnetic field of up to 3.8 T in the inner parts of the detector. Located inside of the solenoid are the silicon pixel and strip tracking systems, as well as the electromagnetic and hadronic calorimeters. Outside the solenoid the muon system is installed, embedded into the solenoid's flux return yoke. All systems are split into a part parallel to the beam pipe (barrel) and two sections orthogonal to the beam pipe (end caps), which close off the cylindric barrel at both ends. An illustration of the detector structure is shown in Fig. 3.2.
In the following, the coordinate system used in the detector is introduced and the subcomponents of the detector are presented in detail.

3.2.1 Coordinate System

The origin of the right-handed coordinate system, most commonly used in CMS analyses, is located at the nominal collision point. The x-axis points towards the center of the LHC storage ring, the y-axis points upwards, orthogonal to the beam path, which aligns with the z-axis. To benefit from the cylindrical symmetry of the detector, a polar coordinate system is used. The azimuthal angle ϕ lies in the xy-plane and starts at the x-axis, going anticlockwise. Instead of the polar angle θ, which starts at the z-axis, the pseudorapidity η can be used, defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) . \tag{3.2}$$

This is often convenient, because differences in η are invariant under Lorentz-boosts in z direction in the limit of massless particles. For the same reason, some quantities are defined as projections into the transverse plane, such as the transverse momentum $p_{\text{T}} = |\vec{p}| \cdot \sin\theta = \sqrt{p_x^2 + p_y^2}$. The combined distance ΔR in η and ϕ is given by

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} . \tag{3.3}$$

It is therefore also invariant against Lorenz boosts along the z axis.

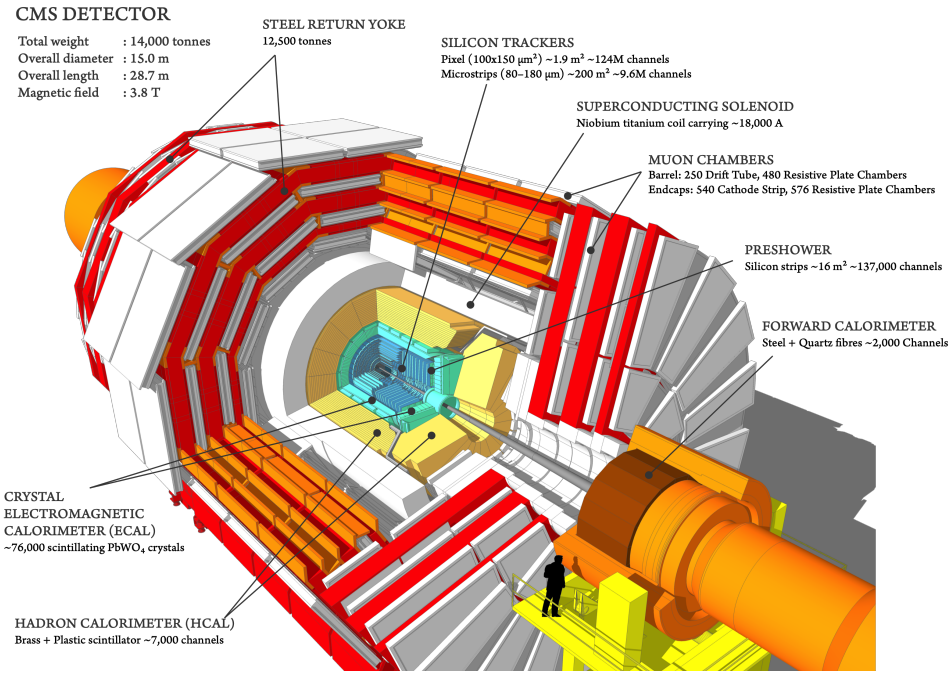


Figure 3.2: Cutaway diagram of the CMS detector, showing all subdetectors with additional technical information. The details regarding the tracking system describe its properties after the update during Run 2 of the LHC.

3.2.2 Tracking System

Closest to the interaction point, the tracking system is installed, covering a pseudorapidity range of $|\eta| < 2.5$. It consists of two parts. The inner pixel detector ranges from 2.9 cm (4.4 cm) to 16.0 cm (10.9 cm) from the beam axis in the barrel region, after (before) its upgrade between 2016 and 2017 [30]. Outside the pixel detector, lower particle flux enables the usage of silicon microstrip modules, which extend up to a radius of 1.1 m around the beam pipe.
When charged particles pass through the tracking system, they ionize their surroundings and thus produce hits in the pixels and strips of the detectors. Due to the strong magnetic field inside the tracking system, the tracks of the charged particles are bent, so that the radius and curvature of the tracks enables derivation of momentum and sign of charge of the particles. The momentum resolution peaks at $1 - 2\%$ and drops with decreasing momentum and increasing pseudorapidity. High granularity is key to be able to accurately reconstruct particle tracks and deduce the vertices they originate from. This is especially important for the identification of b quarks, since the decay of hadrons containing b quarks is suppressed (see Sec. 2.2), resulting in a measurable propagation time, thus creating secondary vertices. Additionally, the tracking system is needed to identify particles that arise from different interactions in the same bunch crossing (pileup).

3.2.3 Electromagnetic Calorimeter

Surrounding the tracking system, the electromagnetic calorimeter (ECAL) is installed. When energy is deposited as electromagnetic showers by passing particles, the lead tungstate crystals ($PbWO_4$) used as scintillators emit light. The number of photons in the scintillation light is proportional to the energy deposited and measured by photodiodes (phototriodes) in the barrel (endcap) region. The ECAL covers a range of $|\eta| < 3$ and extends up to 1.77 m from the beam

axis. This rather compact extent can be ascribed to its high density and short radiation length. The ECAL is especially important for the energy measurement of electrons and photons, as these particles deposit most of their energy in this subdetector. However, the signature of electrons and photons in the calorimeter is similar, so that for distinction between the two, the tracking system has to be used, where only electrons leave tracks.

3.2.4 Hadron Calorimeter

The hadron calorimeter (HCAL) is located between the ECAL and the superconducting solenoid, with some parts extending outside the magnetic coil. Alternating layers of brass absorbers and scintillators enable energy measurements of jets. Through the strong interaction, hadrons produce hadronic showers when crossing the brass absorbers. These showers are measured by the scintillators and the scintillation light is, after being shifted in wavelength, detected by hybrid photodiodes.
The hadronic barrel (HB), endcap (HE) and forward (HF) calorimeters are contained within the solenoid and cover η ranges of up to 1.3, 3 and 5, respectively. As the outer calorimeters (HO) are installed outside the magnetic coil, the HCAL covers almost eleven hadronic interaction lengths, and thus enables measurement of the full energy of high energetic jets and reduces punch-through into the muon system.

3.2.5 Muon System

A key feature of the CMS detecotor is the muon system. It is able to identify muons reliably, because all other detectable particles are mostly stopped before reaching the muon system, and the background is therefore relatively low.
The barrel region of the muons system ($\eta < 1.2$) consists of drift tube chambers. In the endcaps ($0.9 < |\eta| < 2.4$), cathode strip chambers are installed, due to higher particle flux and non-uniform magnetic field. For muon triggering, dedicated resistive plate chambers are installed in both barrel and endcap for $|\eta| < 1.6$.
The muon system is interspersed with the flux-return yoke of the magnetic solenoid, dividing the muon system into four layers. The magnetic field in the muon system enables an additional measurement of the momentum of muons, similar to the one performed in the tracking system.

3.2.6 Trigger System

Since bunch crossings occur at a frequency of 40 MHz at CMS, not all events can be stored and processed. Most of the events are large cross section QCD interactions, which are generally not of great interest. Therefore, a fast triggering system is needed that can decide whether an event is stored, based on reduced information of the detector readout [31, 32].
The triggering system in CMS is twofold, the first part being the hardware-based Level-1 (L1) trigger, which performs a fast, simplistic preselection. If an event is selected by the L1 trigger, it is read out by the Data Acquisition system and passed on to the High-Level Trigger (HLT). The HLT is a software trigger that can fully reconstruct events, combining information from all subdetectors. By this twofold system, the event rate can be reduced from 40 MHz to about 1 kHz.

4 Data Sets and Simulation

The fundamental concept of analyses such as the differential $t\bar{t}$ measurement, is to compare recorded data with the SM expectations. The signal process and its backgrounds are estimated with simulation.
In the following, the data sets are presented together with their corresponding triggers. Subsequently, the background estimation is introduced, outlining the Monte-Carlo (MC) simulation used in this thesis.

4.1 Data Sets and Triggers

The data samples that are used in this thesis were recorded at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ during Run 2 of the LHC in the years 2016 to 2018 by the CMS detector (see Sec. 3.2). In the first part of 2016, there was a problem with highly ionizing particles blinding the readout chips (APVs) of the strip detector [33]. This problem was solved by setting the APV's Preamplifier Feedback Voltage Bias (VFP) to zero, so that the data samples from 2016 are split into two parts, before (preVFP) and after (postVFP) adapting the VFP settings. In the following these two parts are treated separately, resulting in four periods of data taking considered this analysis: 2016 preVFP, 2016 postVFP, 2017 and 2018.
In total, the data corresponds to a combined integrated luminosity of $138\,\text{fb}^{-1}$, where the periods 2016 preVFP, 2016 postVFP, 2017 and 2018 contribute with $19.5\,\text{fb}^{-1}$, $16.8\,\text{fb}^{-1}$ [8], $41.5\,\text{fb}^{-1}$ [9] and $59.8\,\text{fb}^{-1}$ [10], respectively. As noted in Sec. 3.2, the pixel detector was updated between 2016 and 2017, which leads to different detection efficiencies and acceptances between the periods. Additionally, the number of interactions per bunch crossing and thus the pileup increased over the years [34]. This is especially important in this thesis, since it results in differing precision of the $p_{\text{T}}^{\text{miss}}$ reconstruction (explained in detail in Sec. 5.1.5) between the periods. The legacy reprocessing is applied on all data samples for more precise calibration.
Events collected by high level triggers (HLT, see Sec. 3.2.6), that target the same final states, are grouped into primary data sets (PDs). The PDs and HLTs, that are deployed in this analysis, differ between the periods of data taking and are listed in Appx. 1 and Appx. 2, respectively. The HLTs include requirements on the lepton momenta. These requirements are matched by the event selection, which is summarized in Sec. 5.2.

4.2 Simulation

The signal process $t\bar{t}$ and all background processes are simulated via Monte-Carlo (MC) simulations. The simulation is a four-step process. Firstly, an event generator simulates the hard scattering process. Secondly, the particle showering is simulated using PYTHIA [35]. This includes hadronization processes as well as initial and final state radiations. In the third step, the pileup contribution is simulated and added to the events. Lastly, GEANT4 [36] is employed to simulate the interaction of particles with the detector.
The baseline $t\bar{t}$ sample used in the differential measurement is generated at next-to-leading order (NLO) precision using the event generator POWHEG [37, 38]. An additional NLO sample generated by MADGRAPH5_AMC@NLO [39] is used in the training process of the DNN. The $t\bar{t}$ samples are normalized using next-to-next-to-leading order (NNLO) calculations [40, 41].
The relevant backgrounds in the targeted final state are single top, vector boson + jets, diboson (ZZ, WZ, WW) and $t\bar{t}X$ (X being either Z, W, or H) production, as well as backgrounds originating from $t\bar{t}$ production, such as hadronic $t\bar{t}$ decays. A list of the simulated samples and the corresponding cross sections is shown in Appx. 3.

In order to achieve comparability between simulation and recorded data, the simulated samples are rescaled with regards to the integrated luminosity L_{int} of the data, reweighting each event with:

$$w_{\text{lumi}} = \frac{\sigma \cdot L_{\text{int}}}{N_{\text{sim}}} \,, \tag{4.1}$$

where the the number of simulated events of a process is denoted as N_{sim} and the process cross section as σ. To account for differing pileup in data and simulation, an additional weight is applied to the simulation, so that the pileup distributions match. Since the p_{T} spectra of top quarks are observed to be softer in data than in simulation [42, 43], the respective events in simulation are reweighted, dependent on the top p_{T}, to match data. Furthermore, MC weights passed on by the event generator are applied. These can be negative, since they aim to compensate for overlaps between the simulations of hard scattering and parton showering in the event generation.
Further weights aim to compensate for differing efficiencies between data and MC simulation. This includes efficiencies of lepton identification, of b-tagging and of all applied HLTs. In 2016/2017, prefiring of L1 triggers leads to a trigger efficiency loss in data, which is also emulated in simulation with dedicated weights. These four additional weights are referred to as scale factor (SF) weights in the following.

5 Event Reconstruction, Event Selection and Systematic Uncertainties

To analyze collision data recorded at CMS, a dedicated algorithm is used, that aims to reconstruct particles and their kinematics from the detector readout. The particles reconstructed by this so-called Particle Flow (PF) algorithm [44], have to fulfill additional identification criteria (ID). Reconstruction and ID of the particles are exactly the same for all data taking periods.
The PF algorithm is outlined in the following, together with the definitions of objects and variables most relevant in this analysis, including the particle IDs. The event selection of this analysis is presented and lastly, the systematic uncertainties of MC and data distributions are summarized.

5.1 Object Definitions

5.1.1 Particle Flow Algorithm

The Particle Flow algorithm [44] uses the full detector readout for the reconstruction of particle objects and their kinematic properties, by linking matching information from all subdetectors. Tracks from the tracking system are associated with entries in the calorimeters and the muon system and grouped into blocks of information. When a particle type is assigned to such a block, the corresponding information is removed from the whole collection. This procedure is performed first for muons, as their identification can be done most reliably due to the muon chambers. Then electron signatures are identified and the remaining tracks are assigned to jets. Neutral particles, namely photons and neutral hadrons, are reconstructed based on information that is left in the collection.

5.1.2 Muons

Muons do not deposit much energy in the calorimeters, but information from the inner tracking and muon systems can be used for the identification. The entries in the tracking system are used to reconstruct the muon track. The same is done independently with entries in the muon system. Both tracks are then checked for compatibility with each other and, if deemed compatible, the corresponding muon candidate is denoted as "global muon".
The muon with the largest momentum (leading) is required to exceed $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, the subleading muon must exceed $p_{\mathrm{T}} > 20\,\mathrm{GeV}$. All muons have to fulfill $\eta < 2.4$. In addition, there are various identification criteria, that are posed on the muon candidates. In this thesis, the tight working point of the muon ID [45] is used. This ID includes a requirement on the number of hits in tracking and muon system to reduce the number of muons originating from in-flight decays and to guarantee a precise momentum measurement. Additionally, there has to be more than one muon chamber station that contains entries matched with the muon, to reduce entries from hadron punch-through into the muon system. The variables used in further ID requirements are explained in the following, the full ID is then summarized in Table 5.1.

$\boldsymbol{\chi^2_{track}/n_{dof}}$ measures the goodness of the fitted trajectory of the combined muon track in tracking and muon systems. It is derived via a χ^2 test and normalized with the number of degrees of freedom.

$\boldsymbol{|d_0|}$ and $\boldsymbol{|d_z|}$ are calculated as the transversal and longitudinal distance between the reconstructed track and the primary vertex. By limiting these distances, the number of particles originating from pileup can be reduced.

$\boldsymbol{r^{\mu}_{Iso}}$ describes the relative isolation of the muon candidate. It is calculated as the sum over the p_{T} of all hadrons and photons within a cone of $R < 0.4$ around the lepton, divided by the p_{T} of the lepton. In the p_{T} sum, only particles from the primary vertex (PV) should be considered. To account for the fact that neutral particles cannot be traced to the primary vertex via the tracking system and thus cannot be easily distinguished from pileup (PU) events, the pileup contribution needs to be subtracted. For this purpose, the so called δ-β correction is employed, which estimates the pileup from neutral hadrons by 0.5 times the pileup from charged hadrons. The resulting formula for the relative isolation is given by

$$r^{\mu}_{Iso} = \left(\sum_{h^{\pm} \text{ from PV}} p_{\mathrm{T}} + \max\left(0, \sum_{h^0} p_{\mathrm{T}} + \sum_{\gamma} p_{\mathrm{T}} - 0.5 \sum_{h^{\pm} \text{ from PU}} p_{\mathrm{T}}\right)\right) / p^{\mu}_{\mathrm{T}} \,, \tag{5.1}$$

where $h^{\pm}$, h^0 and γ denote charged hadrons, neutral hadrons and photons, respectively, each inside a cone of $\Delta R < 0.4$ around the muon candidate.

Table 5.1: Tight working point of the cut-based muon ID [45].

Parameter	Muon Requirement
Global Muon	true
χ^2_{track}/n_{dof}	< 10
Muon Chamber Hits	> 0
Matched Stations	> 1
$\lvert d_0 \rvert$	< 2 mm
$\lvert d_z \rvert$	< 5 mm
Pixel Hits	> 0
Track Layer Hits	> 5
r^{μ}_{Iso}	< 0.15

5.1.3 Electrons

Electrons can be traced through the tracking system and are stopped in the ECAL, depositing their energy via electromagnetic showers. While traversing the tracking system, electrons emit Bremsstrahlung, losing a significant amount of their energy due to their low mass. This is accounted for in the calculation of the energy deposition by using a special algorithm to create ECAL superclusters (SCs), that include energy deposition from Bremsstrahlung-photons. The tracks of the electrons are reconstructed using a Gaussian sum filter algorithm [46], which also considers Bremsstrahlung in its calculation. Track and super cluster are then linked by the PF algorithm.
The leading (subleading) electron is required to fulfill $p_{\mathrm{T}} > 25(20)\,\mathrm{GeV}$, as well as $\eta < 2.4$. The transition region of the ECAL between barrel and endcap is excluded by the requirement on the pseudorapidity of the ECAL super cluster $\eta_{\mathrm{SC}} \notin [1.4442, 1.5660]$. Similar to muons, electrons have to fulfill ID criteria in the tight working point [47], shown in Table 5.2. The restricted variables are explained in the following, $|d_0|$ and $|d_z|$ are defined as before.

$\boldsymbol{|\Delta\eta_{seed}|}$ and $\boldsymbol{|\Delta\phi_{In}|}$ measure the distance between the reconstructed trajectory and the linked ECAL SC in η and ϕ, respectively. They can thus be used to asses the quality of the link made by the PF algorithm.

$\boldsymbol{\sigma_{i\eta i\eta}}$ is used to differentiate electrons from hadrons by the wider shower shape of hadrons. It is derived as the energy weighted mean difference in η between the ECAL crystal with the highest energy deposition (seed) and the 5×5 crystals around it:

$$\sigma^2_{i\eta i\eta} = \frac{\sum (\eta_i - \eta_{seed})^2 \cdot w_i}{\sum w_i} \qquad i = 1, ..., 25 \ , \tag{5.2}$$

where w_i denotes the energy deposited in crystal i.

$\boldsymbol{H/E}$ is given by the energy deposited in the HCAL segment closest to the seed divided by the energy of the corresponding ECAL SC. It is expected to be much smaller for electrons than for hadrons, as electrons are mostly stopped inside the ECAL.

$\boldsymbol{|E^{-1} - P^{-1}|}$ compares the energy of the ECAL SC with the momentum derived from its associated track. If track and ECAL deposition were linked correctly by the PF algorithm, the value is expected to close to zero, since the electron mass is negligibly small.

$\boldsymbol{r^{\mathrm{e}}_{Iso}}$ serves the same purpose as r^{μ}_{Iso} in the muon ID. Main difference is the elimination of the pileup contribution, where, instead of the δ-β correction, now the so called effective area method is used. This method estimates the pileup from neutral hadrons with the product of the average energy density of each event ρ and the effective area of the cone around the lepton A_{eff}, where ρ is calculated by a k_t jet clustering algorithm [48],

$$r^{\mathrm{e}}_{Iso} = \left(\sum_{h^{\pm} \text{ from PV}} p_{\mathrm{T}} + \max\left(0, \sum_{h^0} p_{\mathrm{T}} + \sum_{\gamma} p_{\mathrm{T}} - \rho A_{\mathrm{eff}}\right)\right) / p^{\mu}_{\mathrm{T}} \ . \tag{5.3}$$

Other than for muons, each of the contributing particles is from within a cone of $\Delta R < 0.3$ around the electron candidate.

$\boldsymbol{N_{miss}}$ and **Conversion Veto** aim to suppress electrons originating from photon conversions in the tracking system. These electrons do not traverse the whole tracking system and therefore, the number of missing hits in the inner layers of the tracker N_{miss} is limited. Additionally, the conversion veto uses a conversion vertex fit probability to exclude such leptons.

Table 5.2: Tight working point of the cut-based electron ID [47]. The average energy density is denoted as ρ, the energy (pseudorapidity) of the ECAL super cluster as E_{SC} (η_{SC}) and the transverse momentum of the electron as $p^{\mathrm{e}}_{\mathrm{T}}$.

Parameter	Barrel ($\eta_{\mathrm{SC}} \leq 1.4442$)	Endcap ($\eta_{\mathrm{SC}} > 1.5660$)
$\Delta\eta_{seed}$	< 0.00255	< 0.00501
$\Delta\phi_{In}$	< 0.022	< 0.0236
$\sigma_{i\eta i\eta}$	< 0.0104	< 0.0353
H/E	$< 0.026 + 1.15/E_{\mathrm{SC}} + 0.0324\rho/E_{\mathrm{SC}}$	$< 0.0188 + 2.06/E_{\mathrm{SC}} + 0.183\rho/E_{\mathrm{SC}}$
$\lvert d_0 \rvert$	< 0.5 mm	< 1 mm
$\lvert d_z \rvert$	< 1 mm	< 2 mm
$\lvert E^{-1} - P^{-1} \rvert$	< 0.159 GeV^{-1}	< 00.0197 GeV^{-1}
r^{e}_{Iso}	$< 0.0287 + 0.506/p_{\mathrm{T}}$	$< 0.0445 + 0.963/p^{\mathrm{e}}_{\mathrm{T}}$
N_{miss}	≤ 1	≤ 1
Conversion Veto	pass	pass

5.1.4 Jets

Jets consist of various particles, produced in the hadronization (see Sec. 2.2) of quarks and gluons. The jet reconstruction therefore aims to cluster all particles associated with such a hadronization into one jet. For this clustering the anti-k_t algorithm [49] is used, with a distance parameter of 0.4.
Transverse momentum of jets is required to exceed $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. Jets with $|\eta| \geq 2.4$ and jets within a cone of $\Delta R < 0.4$ of a selected lepton are vetoed. Additionally, the ID requirements listed in Table 5.3, are exerted on the jets, which restrict the energy fractions of the jet constituents as well as the jet composition. Jet energy corrections [50] are applied on all reconstructed jets that pass these requirements. The jet energy corrections are multiplicative factors on the jet momentum and consist of an offset correction, which accounts for pileup and electronics noise contributions, a calibration factor, which targets the jet response, and residual correction factors, which aim to reduce remaining differences between data and simulation. Similar corrections are applied on the leptons, but the corrections are of smaller scale.

Table 5.3: Definition of the applied jet ID [51] in the TIGHTLEPVETO working point. Energy fractions of the jet constituents and the jet composition are restricted.

Parameter	Jet requirement
Neutral hadron fraction	< 0.90
Neutral EM fraction	< 0.90
Number of constituents	> 1
Muon fraction	< 0.8
Charged hadron fraction	> 0
Charged multiplicity	> 0
Charged EM fraction	< 0.9

Due to the hadronization of quarks and gluons, the determination of the quark flavor proves difficult. The light quarks (u, d, s) and gluons cannot be distinguished at all. Since top quarks decay to b quarks almost instantaneously, they do not form hadrons and their production is identified by final states with a b jet and a W boson per top quark. For the identification of b quarks (b-tagging), in this thesis the loose working point of the multivariate algorithm DeepJet [52] is employed. The algorithm uses characteristics of b jets as inputs, such as the position of the secondary vertex (see Sec. 2.2) and the particle contents of the jet.

5.1.5 Missing Transverse Momentum

The missing transverse momentum is calculated as the sum over the transverse momenta of all observed final-state particles:

$$p_{\mathrm{T}}^{\mathrm{miss}} = \left|\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right| = \left|-\sum_{i}^{N} \vec{p}_{\mathrm{T}_i}\right| \,. \tag{5.4}$$

Since the transverse momentum of the collision partners before the collision is negligible and momentum is conserved, $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ equals the sum of the transverse momenta of all undetected particles (see Sec. 2.4). In case of the top pair decay, it therefore approximates the $\vec{p}_{\mathrm{T}}$ sum of all neutrinos (See Fig. 2.2). However, p_{T} mismeasurements of all reconstructed particles, are directly propagated to $p_{\mathrm{T}}^{\mathrm{miss}}$. There are multiple algorithms to reconstruct $p_{\mathrm{T}}^{\mathrm{miss}}$. The algorithms used in this thesis, PF $p_{\mathrm{T}}^{\mathrm{miss}}$ [53, 54], pilup-per-particle-identification (PUPPI) $p_{\mathrm{T}}^{\mathrm{miss}}$ [55] and

Calo $p_{\mathrm{T}}^{\mathrm{miss}}$ [53, 54], denoted as $p_{\mathrm{T}}^{\mathrm{miss,\ PF}}$, $p_{\mathrm{T}}^{\mathrm{miss,\ Puppi}}$ and $p_{\mathrm{T}}^{\mathrm{miss,\ Calo}}$, respectively, are presented in the following.

$\boldsymbol{p_{\mathrm{T}}^{\mathrm{miss,\ PF}}}$ is calculated by using all PF candidates in the sum shown in Eq. 5.4. The jet energy corrections mentioned in the previous section are propagated to $p_{\mathrm{T}}^{\mathrm{miss,\ PF}}$. Additionally, a so-called xy-correction is applied, that reduces the ϕ modulation of $p_{\mathrm{T}}^{\mathrm{miss,\ PF}}$.

$\boldsymbol{p_{\mathrm{T}}^{\mathrm{miss,\ Puppi}}}$ uses the pileup-per-particle-identification (PUPPI) method [55] to mitigate the influence of pileup on the missing transverse momentum. It is derived based on Eq. 5.4 analogously to $p_{\mathrm{T}}^{\mathrm{miss,\ PF}}$, but introduces additional pileup weights for each PF candidate. These weights describe the probability of the PF candidate to originate from the PV and should therefore be close to zero for particles from pileup. In order to determine such weights, a local shape variable α is employed. For each PF candidate i, the associated shape variable α_i is defined as

$$\alpha_i = \sum_{i \neq j, \Delta R_{ij} < 0.4} \left(\frac{p_{T_j}}{\Delta R_{ij}} \right)^2 \begin{cases} \text{for } |\eta_i| < 2.5, & j \text{ are charged PF candidates from PV} \\ \text{for } |\eta_i| > 2.5, & j \text{ are all reconstruced PF candidates} \end{cases} \tag{5.5}$$

where j refers to all charged PF candidates from the PV within a cone of ΔR_{ij} around i [56]. Outside the tracker acceptance of $|\eta| < 2.5$, the particle origin cannot be traced and all PF candidates have to be used.
To determine the probability of a particle to originate from pileup, the quantity

$$\chi_i^2 = \frac{(\alpha_i - \overline{\alpha}_{\mathrm{PU}})^2}{\mathrm{RMS}_{\mathrm{PU}}^2} \tag{5.6}$$

is used, where $\overline{\alpha}_{\mathrm{PU}}$ and $\mathrm{RMS}_{\mathrm{PU}}$ denote median value and root-mean-square, respectively, of the α_i distribution of the pileup PF candidates. Since the distribution of α_i is Gaussian-like, the χ_i^2 can, as the notation suggests, be interpreted as a χ^2 distribution with number of degrees of freedom $n_{dof} = 1$. Therefore, the corresponding pileup weights of each particle w_i can be calculated using the cumulative distribution function $F_{\chi^2, n_{dof}=1}$ of the χ^2 distribution:

$$w_i = F_{\chi^2, n_{dof}=1}(\chi_i^2) \tag{5.7}$$

The quantity $p_{\mathrm{T}}^{\mathrm{miss,\ Puppi}}$ is then calculated as the $\vec{p}_{\mathrm{T}}$ sum of all PF candidates, weighted by the corresponding factors w_i. Equivalently to $p_{\mathrm{T}}^{\mathrm{miss,\ PF}}$, xy-corrections are applied to $p_{\mathrm{T}}^{\mathrm{miss,\ Puppi}}$ to improve the angular reconstruction.

$\boldsymbol{p_{\mathrm{T}}^{\mathrm{miss,\ Calo}}}$ uses calorimeter entries to derive the sum shown in Eq. 5.4. The direction of each calorimeter entry and its energy deposition are used to define pseudo particles, which are included in this sum. Entries below a noise threshold are not considered. Muons do not deposit energy significantly in the calorimeters. Their calorimetric contribution is therefore excluded and instead their momentum, reconstructed via tracking and muon system, is included. Since no xy-corrections are available for $p_{\mathrm{T}}^{\mathrm{miss,\ Calo}}$, the angular resolution is insufficient and only the absolute value of $p_{\mathrm{T}}^{\mathrm{miss,\ Calo}}$ is used in this thesis.

These three quantities are commonly used to describe $p_{\mathrm{T}}^{\mathrm{miss}}$. Due to its simplicity and robustness, $p_{\mathrm{T}}^{\mathrm{miss,\ PF}}$ is used in most CMS analyses. Since the pileup contribution has been increasing over the years [34], the PUPPI calculation of $p_{\mathrm{T}}^{\mathrm{miss}}$ is employed more frequently as it is less sensitive to pileup. Generally, $p_{\mathrm{T}}^{\mathrm{miss,\ Puppi}}$ outperforms $p_{\mathrm{T}}^{\mathrm{miss,\ PF}}$ in some phase spaces, but falls behind in

other phase spaces. The less commonly used Calo $p_{\mathrm{T}}^{\mathrm{miss}}$ utilizes a different approach to derive $p_{\mathrm{T}}^{\mathrm{miss}}$, which can be a valuable addition. Using, among others, these three $p_{\mathrm{T}}^{\mathrm{miss}}$ calculations as inputs, a deep neural network is trained in this thesis. The DNN-based $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction aims for a better description of $p_{\mathrm{T}}^{\mathrm{miss}}$ in the considered phase space of the dileptonic $\mathrm{t\bar{t}}$ decay. Other $p_{\mathrm{T}}^{\mathrm{miss}}$-related quantities used in this thesis are summarized in the following.

$\boldsymbol{p_{\mathrm{T}}^{\mathrm{miss,\,DNN}}}$ approximates $p_{\mathrm{T}}^{\mathrm{miss}}$with a deep neural network (DNN), which is explained in detail in section 6. Goal of this thesis is to improve this quantity to be able to use it in the differential $\mathrm{t\bar{t}}$ measurement.

$\boldsymbol{p_{\mathrm{T}}^{\nu\nu}}$ denotes the projection in the transverse plane of the vectorial momentum sum of the two prompt neutrinos in the dileptonic top pair decay. Since neutrinos are not detectable with the CMS detector, $p_{\mathrm{T}}^{\mathrm{miss}}$ has to be used to approximate it.

$\boldsymbol{p_{\mathrm{T}}^{\mathrm{miss,\,gen}}}$ is the amount of $p_{\mathrm{T}}^{\mathrm{miss}}$ generated before detector effects (generation-level) in an event. In case of the top pair decay, it differs from $p_{\mathrm{T}}^{\nu\nu}$ for example due to neutrinos radiated in the decays of B-mesons from the b quark hadronization.

5.1.6 Additional Variables

As inputs for the DNN, various quantities are used. The most important of these quantities are defined in the following. For all variables, the leading lepton/jet is denoted as ℓ_1/j_1, the subleading lepton/jet is denoted as ℓ_2/j_2.

$\boldsymbol{p_{\mathrm{x}}^{\mathrm{miss,\,PF}}/p_{\mathrm{y}}^{\mathrm{miss,\,PF}}}$, $\boldsymbol{p_{\mathrm{x}}^{\mathrm{miss,\,Puppi}}/p_{\mathrm{y}}^{\mathrm{miss,\,Puppi}}}$ are the projections of the respective quantities described in the previous section onto the x/y axis.

$\boldsymbol{p_{\mathrm{x}}^{i}/p_{\mathrm{y}}^{i}}$ is the x/y component of the momentum of a particle i. Instead of i one of ℓ_1, ℓ_2, j_1 or j_2 is inserted.

$\boldsymbol{p_{\mathrm{x}}^{\mathrm{all}\,j}/p_{\mathrm{y}}^{\mathrm{all}\,j}}$ are calculated as projections on the x/y axis of the vectorial sum over all transverse jet momenta.

$\boldsymbol{H_{\mathrm{T}}}$ is defined as the absolute value of the scalar sum over the p_{T} of all reconstructed jets passing the event selection (see Sec. 5.2).

$\boldsymbol{n_{\mathrm{jets}}}$ is the number of reconstructed and selected jets.

$\boldsymbol{m_{\mathrm{HT}}}$ is the invariant mass of all reconstructed and selected jets:

$$m_{HT} = \sqrt{\left(\sum_{\mathrm{all}\,j} E_i\right)^2 - \left(\sum_{\mathrm{all}\,j} \vec{p}_i\right)^2}\,. \tag{5.8}$$

$\boldsymbol{m_{\ell_1,\ell_2,\mathrm{all}\,j}}$ is the invariant mass of ℓ_1, ℓ_2, and all reconstructed and selected jets:

$$m_{\ell_1,\ell_2,\mathrm{all}\,j} = \sqrt{\left(\sum_{\ell_1,\ell_2,\mathrm{all}\,j} E_i\right)^2 - \left(\sum_{\ell_1,\ell_2,\mathrm{all}\,j} \vec{p}_i\right)^2}\,. \tag{5.9}$$

$\boldsymbol{M_{T2}}$ is a quantity based on the transverse masses of the W bosons. It is defined as

$$M_{\mathrm{T2}} = \min_{\vec{p}_{\mathrm{T1}}^{\,\mathrm{miss}} + \vec{p}_{\mathrm{T2}}^{\,\mathrm{miss}} = \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}} \left(\max \left[m_{\mathrm{T}} \left(p_{\mathrm{T}}^{\ell_1}, \vec{p}_{\mathrm{T1}}^{\,\mathrm{miss}} \right), m_{\mathrm{T}} \left(p_{\mathrm{T}}^{\ell_2}, \vec{p}_{\mathrm{T2}}^{\,\mathrm{miss}} \right) \right] \right) \tag{5.10}$$

where $\vec{p}_{\mathrm{T1}}^{\,\mathrm{miss}}$ and $\vec{p}_{\mathrm{T2}}^{\,\mathrm{miss}}$ are all possible transverse momenta such that their sum yields the missing transverse momentum [57].

5.2 Event Selection

To increase the ratio of signal events by reducing the contribution of SM backgrounds, various event selection criteria are applied on the recorded data. These criteria restrict phase space and particle composition of each event. More specifically, each event is required to have

1. at least two jets (see Sec. 5.1.4)
2. at least one jet identified as b jet
3. exactly two oppositely-charged leptons (either ee, $\mu\mu$, or eμ)
4. both leptons within $|\eta| < 2.4$
5. a leading lepton with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and a subleading lepton with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$
6. $p_{\mathrm{T}}^{\mathrm{miss,\,Puppi}} > 40\,\mathrm{GeV}$ for ee and $\mu\mu$ final states.
7. $m_{\ell\ell} < 76\,\mathrm{GeV}$ or $m_{\ell\ell} > 106\,\mathrm{GeV}$ for ee and $\mu\mu$ final states
8. leptons with an invariant mass $m_{\ell\ell} > 20\,\mathrm{GeV}$

Selection criteria 1-3 aim to reproduce the particle content of the final state of dileptonic $\mathrm{t\bar{t}}$ events. This would include two b jets, but due to a significant decrease in event statistics when requiring two b-tagged jets caused by limited identification efficiency, only one b-tagged jet is required. Requirement 4 restricts the η range of the leptons due to the coverage of the muon chambers, which reaches up to $|\eta| = 2.4$ in the endcaps. To exceed the corresponding trigger thresholds (see. Appx. 2), the lepton momenta are also restricted. Requirements 6-8 intend to reduce the Drell-Yan (DY) background specifically. The tree-level final state of this background includes only same-flavor leptons and no invisible particles. Therefore, DY events tend to have little missing transverse momentum, which is exploited by a lower limit on $p_{\mathrm{T}}^{\mathrm{miss}}$ to reduce the DY contribution. The Drell-Yan process contributes most for an invariant mass of the leptons which is close to zero ($< 20\,\mathrm{GeV}$) or around the Z boson mass of 91 GeV. This phase space is therefore excluded.

5.3 Systematic Uncertainties

The systematic uncertainties considered in this thesis can be divided into experimental uncertainties and uncertainties originating from the modeling of signal or background events. In both cases, the uncertainties on event distributions are derived by shifting each individual uncertainty to then determine the deviation from the nominal value in each bin of the considered distribution. The systematic uncertainties are explained in more detail in [58].

5.3.1 Experimental Uncertainties

b Tagging

The b tagging uncertainty is derived by shifting the b tagging scale factors (see Sec. 4.2) by a recommended value. To account for correlations between the periods of data taking, the uncertainty is broken down into correlated and uncorrelated components.

Background Estimation

The background processes are modeled using MC simulations. The normalization of the processes is shifted by its estimated uncertainties and the effect on the final distributions is determined.

Jet Energy Scale

The uncertainty originating from applying the recommended jet energy scale corrections (see Sec. 4.2) is provided by the JetMET group [59] as a function of p_{T} and η and split into 7 different sources. For each of the individual sources, the jet energy scale is varied up and down by the corresponding uncertainty. The variation in the jet momenta is propagated on the missing transverse momentum, hence, the jet energy scale uncertainty is a dominant uncertainty in the $\mathrm{t\bar{t}}$ measurement.

Jet Energy Resolution

The jet energy resolution in simulation has to be adapted to match the recorded data. The uncertainty resulting from this process is determined by shifting the jet energy resolution within its uncertainty in two different η regions, as recommended by the JetMET group [60].

Lepton Selection

The lepton scaling factors employed in this thesis are also shifted within their uncertainties. Since the scaling factors are derived based on Drell-Yan events, but are applied predominantly to top events, an additional uncertainty of 1% (0.5%) for electrons (muons) is included to account for the different topologies.
Furthermore, lepton scale and resolution corrections result in uncertainties, which are derived by shifting the corrections by their uncertainty. The resulting change in the lepton momenta is propagated to the missing transverse momentum as well.

Luminosity

The correlated and uncorrelated uncertainties on the luminosity for all periods of data taking are implemented according to the recommendations (see [61]). The luminosity uncertainty ($\sim 1-2\,\%$) is a dominant uncertainty in the differential $\mathrm{t\bar{t}}$ measurement.

L1 Prefiring

The weights compensating for the prefiring of L1 triggers (see Sec. 4.2) are shifted within their uncertainties to obtain the resulting uncertainty.

HEM15/16 Issue

Since two HCAL modules were not available for a significant part of the 2018 data taking period, an additional systematic uncertainty is included. Here, the energy of jets in the affected phase space is rescaled according to recommendations [62].

Pileup

The pileup weights (see Sec. 4.2) rescale the pileup distribution in simulation to match the recorded data. The rescaling is based on the total inelastic proton-proton cross section of 69.2 mb [63]. To measure the connected uncertainty, this cross section is varied by ±4.6%.

Trigger Efficiency

The trigger efficiency in data and MC is measured and scale factors are calculated to compensate for differences (see [58]). The associated uncertainty is derived by shifting the scale factors within their uncertainties, which are in turn derived by taking into account run and phase space dependencies, trigger correlations and statistics.

Unclustered Energy

The uncertainty connected to the unclustered energy in the events is derived by shifting the deposited energy of photons, charged and neutral hadrons according to their resolution. The results are propagated to $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and the effect of the varied $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ on the event distributions is determined.

5.3.2 Modeling Uncertainties

b Semileptonic Branching Fraction

The b jet energy response is significantly influence by the choice of the semileptonic branching fraction. Based on the semileptonic branching fraction and its uncertainty given by the PDG [64], this effect is accounted for using weights on generator level and the corresponding uncertainties are propagated to the event distributions.

Color reconnection

The color reconnection is modeled in the nominal $\mathrm{t\bar{t}}$ sample based on multi parton interactions [65]. The connected uncertainty is estimated by running the analysis with three different $\mathrm{t\bar{t}}$ simulations and using the envelope of the results.

Initial and Final State Radiation

The initial and final state radiation (ISR and FSR) in QCD events is dependent on the strong coupling constant α_S. Its uncertainty is propagated by introducing per-event weights to shift the scale of ISR and FSR individually.

ME-PS Matching

The matrix element (ME) and parton shower (PS) are matched in POWHEG with a damping function, dependent on the constant h_{damp}. The associated uncertainty is propagated to the event distributions by running the analysis for two different $\mathrm{t\bar{t}}$ samples with different h_{damp} settings.

Matrix Element Scale

The uncertainty of the renormalization scale μ_R and factorization scale μ_F in the matrix element generators is derived by varying μ_R and μ_F by a factor of two up and down, both individually and simultaneously. The envelope of the variations is taken as the final uncertainty.

Parton Distribution Function

The uncertainty originating from the modeling of the parton distribution function (PDF) is estimated by applying per-event weights in the dileptonic $\mathrm{t\bar{t}}$ sample, rescaling the event distributions while preserving the normalization. The α_S value used in the PDF is also shifted within its uncertainty.

Top Mass

The top mass of $m_{\mathrm{t}} = 172.5\,\mathrm{GeV}$ used in the nominal $\mathrm{t\bar{t}}$ sample is shifted up and down by 3 GeV in alternative $\mathrm{t\bar{t}}$ simulations The connected uncertainty is propagated on the results by scaling the shifts down linearly, assuming a top mass uncertainty of 1 GeV.

Top p_{T} Reweighting

The uncertainty related to the top p_{T} weights (see Sec. 4.2) are estimated by repeating the analysis without the p_{T} reweighting and determining the effect on the event distributions.

Underlying Event

The settings of the underlying event tune are shifted in two alternative $\mathrm{t\bar{t}}$ samples. These are used to determine the connected uncertainty of the underlying event tune on the results.

6 Deep Neural Networks

Machine learning techniques have quickly risen in importance in the last decade. They are present in many consumer products, but have also found use in fundamental research, for instance in particle physics. This large progress in machine learning can, amongst others, be attributed to the development of deep neural networks (DNNs) [66–68], a multivariate learning technique. Conceptually, a machine learning algorithm is an algorithm that improves its performance with experience, i.e. data. This is fundamentally different to conventional algorithms, which are developed manually. The goal of a machine learning algorithm is generally to extract features from raw data. This is done by mapping an input vector of fixed size (e.g. an image) to an output vector of fixed size (for example probabilities of each category one tries to classify the images by). In case of a feed-forward neural network, as used in this thesis, information is propagated only in one direction through this map. To extract complex features from the input, the mapping from input to output is built in layers, with each layer extracting increasingly complex features. A neural network with more than one such layer is called a deep neural network.
There are mainly two types of tasks that are performed using DNNs, namely classifications and regressions. In a classification task, the input is related with a certain probability to each of several discrete classification categories (for example image recognition). In a regression, the output is given by one or multiple, indiscreet numbers, which could for example be the age of the person in an input picture. In this thesis, a regression is performed, where several kinematic variables of a proton-proton collision event are used as inputs to determine a correction on $p_{\mathrm{T}}^{\text{miss, Puppi}}$ to estimate the amount of $p_{\mathrm{T}}^{\text{miss, gen}}$ in that event more accurately.
In the following, the mathematical foundations of the DNN structure and training process are outlined. Subsequently, the network settings, described by the so-called hyperparameters, are detailed. Lastly, a Bayesian search algorithm is outlined, which is used to optimize the hyperparameters of the DNN.

6.1 Mathematical Foundation

Given a data set, where each of N input vectors $x_i \in \mathbb{R}^n$ corresponds to one of N output vectors $y_i \in \mathbb{R}^m$, a neural network aims to approximate the mapping function $f(x) = y$ with the model $f^*(x, \theta) = y^*$ by adjusting the free parameters θ. To adjust θ systematically, an objective function (loss) $J(\theta)$ is introduced. The loss function is a quantitative measure for the agreement between model and data. It is minimal for θ, which describes the mapping function best. The loss function can for instance be given by the mean squared error (MSE):

$$J_{MSE}(\theta) = \frac{1}{s_b} \sum_{i=1}^{s_b} \left(f^*(x_i, \theta) - y_i \right)^2 , \tag{6.1}$$

where the parameter s_b (batch size) denotes the number of input vectors, the loss is evaluated over in each iteration. For the parameter adjustment, a method called stochastic gradient descent can be employed. This method minimizes $J(\theta)$ by iteratively updating θ in the opposite direction of the loss gradient $\nabla_\theta J$:

$$\theta \to \theta - \alpha_s \cdot \nabla_\theta J , \tag{6.2}$$

with the step size α_s. Since the parameter space is generally very large, the calculation of the loss gradient with a deterministic algorithm is too slow to be practical. Therefore, the nondeterministic algorithm stochastic gradient descent [69] is employed, which uses elements of randomness in its decisions to speed up the optimization process. To remove bias, the parameters θ are also initialized randomly, which synergizes well with the stochastic search algorithm [70].

Once the model evaluation and parameter adaption has been performed for all batches in the data set, one so-called epoch is complete.
For the gradient descent method to work with reasonable computational resources and invested time, the model f^* is best chosen in such a way, that it can be evaluated quickly and efficiently. Such a model could be a linear model $y^* = Wx+b$, where the weight matrix $W \in \mathbb{R}^{m\times n}$ and the offset $b \in \mathbb{R}^m$ are free parameters. Most functions cannot be well described with such a simple linear model, hence, in a DNN, multiple linear operations are chained together:

$$g^{(1)} = W^{(1)}x + b^{(1)} \tag{6.3}$$
$$y^* = W^{(2)}g^{(1)} + b^{(2)} \tag{6.4}$$
$$\Rightarrow \quad y^* = W^{(2)}\left(W^{(1)}x + b^{(1)}\right) + b^{(2)} \tag{6.5}$$
$$= \underbrace{W^{(2)}W^{(1)}}_{W}x + \underbrace{W^{(2)}b^{(1)} + b^{(2)}}_{b} \tag{6.6}$$

However, as shown above, such a chain results in a model, which is still linear. Therefore, after each linear operation a non-linear activation function σ (see Sec. 6.2) is applied to each element:

$$h^{(1)} = \sigma(g^{(1)}) = \sigma\left(W^{(1)}x + b^{(1)}\right) \tag{6.7}$$
$$\Rightarrow \quad y^* = W^{(2)}h^{(1)} + b^{(2)} \tag{6.8}$$
$$= W^{(2)}\sigma(W^{(1)}x + b^{(1)}) + b^{(2)} \tag{6.9}$$
$$\Rightarrow \quad y^* \neq Wx + b \tag{6.10}$$

An example of such a model is illustrated as a network in Fig. 6.1. Each vector $h^{(j)}$, obtained by application of the linear model and the non-linear activation, defines a hidden layer. The number of layers of a network thus corresponds to the number of times, linear model and activation function are applied.

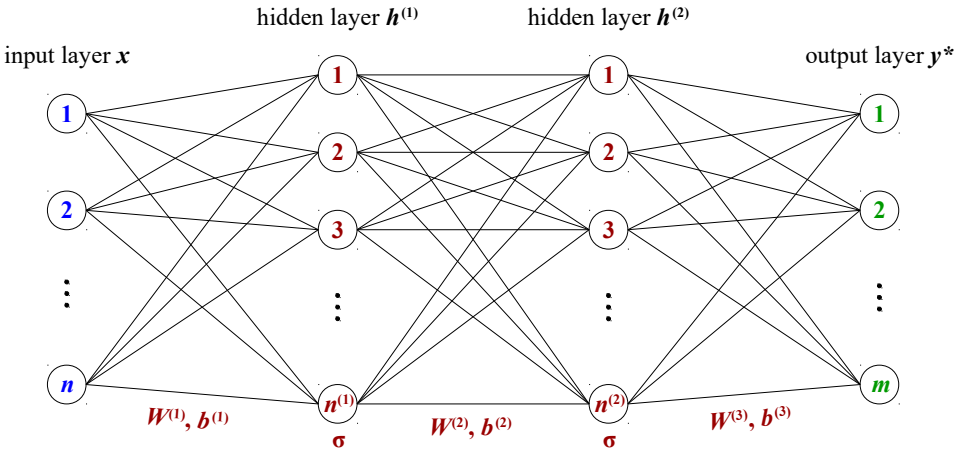


Figure 6.1: Illustration of a deep neural network with two hidden layers. Each layer is made up by nodes, inputs are shown in blue, parts of the calculation in red and outputs in green.

The components of the vectors $h^{(j)}$ are called nodes. The number of nodes per layer is given by the dimensions of the weight matrices $W^{(j)}$. These are fixed during training, but can be

otherwise chosen freely. Only the first (last) weight matrix is partly restricted in size by the dimensionality of the inputs x (outputs y^*).
Parameters such as the number of nodes and layers, which are fixed during the training process, but can be adapted to improve the network performance, are named *hyperparameters*. The hyperparameters of a DNN are discussed in detail in Sec. 6.2.

Since the number of nodes and layers and thus the number of parameters θ can be chosen arbitrarily large, the network can describe the data set used for training to arbitrary precision. Therefore, if the model is trained for too many epochs, it memorizes the training sample instead of generalizing its features. This phenomenon is called overtraining and is one of the big challenges when training DNNs. To monitor overtraining, the given data is split into a training sample, a validation sample, and a test sample. The network is trained based on the training sample, but in each epoch the loss of the validation sample is evaluated as well. If the loss is decreasing in the training sample, but stagnating or increasing in the validation sample, the network is overtraining and the training process should be stopped. This so-called early stopping is one method to reduce overtraining, others include *dropout* and *regularization*, which are explained in the following section. The test set is not used at all while training to preserve an unbiased data sample to evaluate the network performance after training.
In this thesis, DNNs are trained and their performance is evaluated using the KERAS interface of the TensorFlow package [71] in python. The division of the data sample into training, validation and test sample is done in fractions of 60%, 20% and 20%, respectively.

6.2 Hyperparameters

All parameters of a DNN that are fixed during training, but can be adapted outside the training process to improve the network performance, are called hyperparameters. Since it is a priori unclear, which hyperparameters suit the problem at hand, the optimization of hyperparameters is an important step to achieve good results. Since the size of the hyperparameter space, that is searched for the ideal parameter combination, increases exponentially with the number of hyperparameters, some hyperparameters are excluded from the optimization process. In the following, all hyperparameters included in the optimization are explained, followed by the hyperparameters of the loss function and early stopping algorithm, which were not adjusted in the scope of the optimization.

6.2.1 Learning Rate

The learning rate of a network is given by the step size α_s of the gradient descent method (see Eq. 6.2). If chosen too large, the loss function can oscillate around the minimum or not converge at all. If chosen too small, the process takes excessively many epochs and the probability increases for the loss to converge in a local minimum.
To improve performance and efficiency of a DNN, the learning rate can also be chosen to be variable during training. The step size is then decreased over the course of training to enhance the convergence of the algorithm. In this thesis, the Adam optimizer [72] is used, which adapts the learning rate automatically. As an input, the optimizer requires a starting learning rate r_l, which is the hyperparameter used in the optimization constrained to the range

$$r_l \in [0.00001, 0.05] \tag{6.11}$$

To account for the large range of this parameter, its logarithm $r_l' = \ln r_l$ is used in the optimization procedure.

6.2.2 Network Structure

The size of a neural network is defined by the number of layers (depth) and number of nodes (width). Usually, larger networks are better suited to model complex features, while a simpler feature extraction is achieved more efficiently by small networks. The number of hidden layers in this thesis is restricted to:

$$n_{\text{layer}} \in \{1, 2, 3, 4, 5, 6, 7, 8\} . \tag{6.12}$$

To choose a suited number of nodes per layer, the position of each layer has to be taken into consideration. Layers closer to the output layer should have less nodes, to improve the mapping to the output vector, which usually consists of few components. In this thesis, the standard values for the number of nodes per layer in the case of eight layers are given by

$$n_{\text{node}}^{\text{norm}} = (1024, 1024, 1024, 1024, 1024, 512, 256, 128) , \tag{6.13}$$

where the first value corresponds to the first hidden layer. If fewer layers are chosen, the first elements of this vector are discarded, e.g. resulting in $n_{\text{node}}^{\text{norm}} = (512, 256, 128)$ if there are three layers. The hyperparameter associated with the number of nodes per layer, that is optimized in this thesis is the ratio

$$r_{\text{node}} = \frac{n_{\text{node}}}{n_{\text{node}}^{\text{norm}}} , \qquad r_{\text{node}} \in \{1/8, 1/4, 1/2, 1, 2, 4\} . \tag{6.14}$$

For a fixed number of layers a value of $r_{\text{node}} = 2$ would therefore double the number of nodes in each layer.

6.2.3 Activation Function

Activation functions σ introduce non-linearity to the network, as explained in Sec. 6.1. Since the concept of DNNs is to perform numerous iterative calculations very rapidly, an important feature of activation functions is a short evaluation time. A popular choice, which is motivated by biological considerations in neuroscience [73], is the Rectified Linear Unit (ReLU):

$$\sigma_{\text{ReLU}}(x) = \max(0, x) . \tag{6.15}$$

In this thesis, the closely related function leaky ReLU is used, which is defined as

$$\sigma_{\text{leaky ReLU}}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha x & \text{if } x < 0 \end{cases} , \tag{6.16}$$

with the leakage α, which is a hyperparameter in this optimization and restricted to the interval

$$\alpha \in [0, 0.4] . \tag{6.17}$$

6.2.4 Dropout

Dropout aims to reduce overtraining by randomly turning off a fraction of the nodes in each step during the training process. The added noise to the process enforces training of redundant representations. This helps to decrease the reliance of the DNN on single nodes and thus improves

generalization and reduces overtraining.
The dropout fraction is a hyperparameter in this optimization and ranges within

$$p_{\text{drop}} \in [0.1, 0.5]\,. \tag{6.18}$$

6.2.5 Regularization

Similar to dropout, regularization is a method to reduce the reliance of the network on single nodes. The concept of regularization is to apply penalties on large weights in the weight matrices during training. One defines the regularized loss function $J^{\text{reg}}(\theta)$ as

$$J^{\text{reg}}(\theta) = J(\theta) + \lambda \cdot \Omega(\theta)\,, \tag{6.19}$$

where λ is the regularization weight and $\Omega(\theta)$ a measure for the magnitude of the weights. The hyperparameter λ is included in the hyperparameter optimization due to its large impact on the training process. If chosen too large, only Ω is minimized during training and no reasonable model is trained. If chosen too small, no regularization occurs. The regularization weight is restricted to the range

$$\lambda \in [0.0001, 0.2]\,. \tag{6.20}$$

A typical choice for the regularization function Ω, that is also used in this thesis, is the $L2$ norm, defined as

$$\Omega(\theta) = ||\theta||_2 = \theta_1^2 + \theta_2^2 + \dots\,. \tag{6.21}$$

Similar to the learning rate, the logarithm $\lambda' = \ln \lambda$ is used in the hyperparameter optimization.

6.2.6 Batch Size

For the gradient descent method (see Eq. 6.2), the loss function has to be evaluated and its gradient has to be determined. Calculating the loss over the entire training data set $\{x_i, y_i\}$ would provide the most accurate parameter update, but is computationally too expensive. Therefore, the training sample is split into random subsets, called batches and the loss gradient is determined successively in each separate batch. If the training sample does not divide evenly by the batch size, the last batch has a different size. The batch size is a hyperparameter in this thesis and restricted to

$$s_{\text{b}} \in [250, 10000]\,. \tag{6.22}$$

The upper limit on the batch size is required due to computation time and memory restrictions of the machine used for training. The logarithm $s_b' = \ln s_b$ is used in the hyperparameter optimization due to its large range.

6.2.7 Loss Function

The mean squared error introduced in Eq. 6.1 is not the only function suitable as loss function. The loss function is not a parameter of the hyperparameter optimization, but the most common loss functions were tried out exemplary, showing that the so-called logcosh loss is best suited for the problem at hand. The logcosh loss is defined as

$$J_{\text{logcosh}}(\theta) = \frac{1}{s_b} \sum_{i=1}^{s_b} \log\left(\cosh\left(D_i(\theta)\right)\right) \qquad \text{with } D_i(\theta) = f^*(x_i, \theta) - y_i\,, \tag{6.23}$$

where the number of summands is given by the batch size s_b. Since the function $\log(\cosh(D))$ is well approximated by D^2 for small D and by $|D|$ for large D, the logcosh can be seen as a compromise between the mean squared error (Eq. 6.1, D^2) and the mean error ($|D|$).

6.2.8 Early Stopping

One of the most effective techniques to reduce overtraining is to stop the training process once the validation loss stagnates compared to the training loss. To account for loss fluctuations, the loss is monitored for more than one epoch and the process is only stopped, once a clear trend is visible. The algorithm used in this thesis checks the following three requirements after each epoch:

1. The validation loss J^{vali} has not decreased within the last 10 epochs.
2. The relative generalization error $(J^{\text{vali}} - J^{\text{train}})/J^{\text{train}}$ exceeds 2% for ten epochs in a row
3. The absolute generalization error $J^{\text{vali}} - J^{\text{train}}$ increases for ten epochs in a row

If one of the requirements is met, the training is stopped and the weights of the epoch with the smallest validation loss are restored.

6.3 Bayesian Hyperparameter Optimization

The hyperparameters of a neural network can have significant impact on the performance of the network. However, suitable hyperparameters are a priori unclear and difficult to extrapolate from the network objective. Correlations between the hyperparameters further complicate finding a suitable set. Therefore, a consistent and well-structured approach to optimize the hyperparameters is needed.
There are various methods that can be employed to optimize the hyperparameters of a DNN. The most straight-forward of those is a grid search, where one hyperparameter is adapted while fixing all others. While this method is simple and robust, the number of performance evaluations needed to cover a significant part of the hyperparameter space increases exponentially with the number of hyperparameters. Since the performance evaluation of a set of hyperparameters requires the network to be trained with these settings, the evaluation is the most computationally expensive step in a hyperparameter optimization. A full grid search with the 7 hyperparameters optimized in this thesis is therefore not feasible.
A more efficient searching method would be to include the results of previous hyperparameter evaluations in the determination of the next set of hyperparameters. In such an approach, the regions of the hyperparameter space most likely to yield good results can be focused on in the search. This is done by the Bayesian optimization algorithm explained in the following.

6.3.1 Bayesian Search Algorithm

Bayesian search [74–76] is a method which aims to find the maximum point h_{opt} of a function $f(h)$, which is expensive to evaluate. As the name suggest, Bayes' theorem is central to this method. For a model M of f and evaluated data $\mathcal{D}$, Bayes' theorem relates the posterior probability $P(M|\mathcal{D})$ with the probability $P(\mathcal{D}|M)$ of observing $\mathcal{D}$ given M and the prior probability $P(M)$:

$$P(M|\mathcal{D}) \propto P(\mathcal{D}|M) \cdot P(M)\,, \tag{6.24}$$

In each optimization step, the posterior probability is calculated and used to update the model M. Based on the model, the so-called utility function is derived. The maximum of the utility

function represents the sample point, which would be most beneficial to be evaluated next.
In the scope of the Bayesian hyperparameter optimization, $\mathcal{D}$ corresponds to network evaluations with hyperparameters h. The true distribution $f(h)$ of the network performance as a function of the hyperparameters is modeled using a Gaussian process, outlined in the following paragraph. Since a Gaussian process not only yields a function, but a probability distribution of possible functions, the function with the highest probability can be used as M and its prior probability $P(M)$ is already described by the Gaussian process. The utility function can be defined based on the Gaussian process as well.

A Gaussian process yields the probability distribution of a collection of random variables (such as $\mathcal{D}$) as a joint distribution of all those variables in a continuous domain (such as h). In each point h in the domain, the probability distribution is described by a normal distribution with given mean and covariance. Hence, for the whole domain, a function of mean values $m(h)$ can be defined, which is continuous due to the properties of the Gaussian process. Similarly, the function of standard deviations $\sigma(h)$ can be defined.
An example of a Gaussian process using a one dimensional h is shown in Fig. 6.2. The true distribution $f(h)$ is chosen arbitrarily. It can be modeled by the mean function $m(h)$, which is shown with the corresponding standard deviation.
The utility function determines the next point to be evaluated. In this case, the utility function is given by the so-called upper confidence boundary (UCB), defined as

$$\mathrm{UCB}(h) = m(h) + \kappa \cdot \sigma(h)\,, \tag{6.25}$$

with a constant parameter κ. For the decision, which sample point to evaluate next, both exploration (sampling in areas with high uncertainty) and exploitation (sampling in areas with large expected $f(h)$) should be taken into account. A large value of κ (~ 10) favors exploration, a small value ($\kappa \sim 0.1$) results in more exploitation.

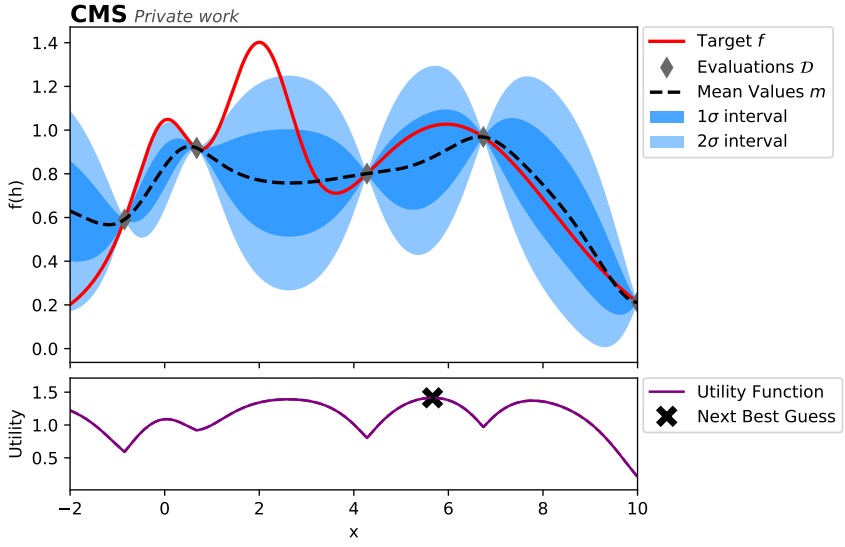


Figure 6.2: Illustration of a Gaussian process with 2 initial data points after 2 optimization steps and resulting sample point evaluations (top) and the corresponding utility function given by UCB with $\kappa = 2.5786$ (bottom). The true distribution $f(h)$ (red, solid line) is modeled by the mean function m (black, dashed line). The uncertainty of the model is illustrated by the standard deviation intervals $m(h) \pm \sigma(h)$ (dark blue) and $m(h) \pm 2\sigma(h)$ (light blue).

6.3.2 Implementation

In this thesis, the Bayesian hyperparameter optimization is performed using the BayesOpt library [77]. As utility function in the optimization, the UCB is used with $\kappa = 2.5786$, which is the default value in BayesOpt. The target function f to be maximized is defined as the negative validation loss $-J^{\text{val}}_{\text{logcosh}}$. The search space h is given by the hyperparameter space and is thus multidimensional. While the Gaussian process is suitable to describe such a domain, the search space is increased significantly in comparison to a one dimensional domain. Therefore many optimization steps and thus many network evaluations are needed. This process cannot be parallelized, because in each optimization step, all previous DNN evaluations are taken into account. The optimization steps therefore have to be performed iteratively. As starting points of the Gaussian process (intial set of data $\mathcal{D}$), the validation losses of networks trained with randomly guessed hyperparameters are used. The Bayesian optimization process and the set of hyperparameters resulting from the optimization are shown in Sec. 8 together with the performance of the corresponding DNN.

7 DNN-based $p_{\mathrm{T}}^{\mathrm{miss}}$ Correction

The reconstructed quantities $p_{\mathrm{T}}^{\mathrm{miss}}$ and $|\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}, \text{nearest } \ell)|$ (see Sec. 5.1.5) serve as proxies for the observables $p_{\mathrm{T}}^{\nu\nu}$ and $|\Delta\phi(p_{\mathrm{T}}^{\nu\nu}, \text{nearest } \ell)|$ in the differential $\mathrm{t\bar{t}}$ measurement. An unfolding procedure is applied on the reconstructed quantities, to account for detector and acceptance effects, and to reduce bin-to-bin migrations. For this procedure to work reliably, good reconstruction resolution is important, reducing the event migrations that need to be corrected.
The precision of the measurement therefore depends on the bias and resolution of the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ reconstruction. Since the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ resolution, for the reconstruction with PUPPI and PF algorithm, is insufficient in parts of the phase space (see Sec. 8.4), it needs to be corrected. This correction is derived using a DNN, which is optimized in a Bayesian hyperparameter optimization as explained in Sec. 6.3. In the following, the structure and input features of the DNN, which is optimized in this thesis, are detailed.

Firstly, a baseline DNN is presented, which is used for general performance studies and as a starting point for the hyperparameter optimization. Secondly, an event reweighting process is explained, which aims to mitigate network bias resulting from an unbalanced training sample. Subsequently, Shapley values of the DNN are shown, which sort the input features of a DNN according to their impact on the training process. Some of the input features might be inaccurately described in the MC simulations, which are used to train the DNN. To ensure that such inaccuracies do not bias the network, the input features must be validated, a process summarized in Sec. 7.3.

7.1 Baseline DNN

The purpose of the DNN is to correct $p_{\mathrm{T}}^{\mathrm{miss}}$ and $|\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}, \text{nearest } \ell)|$, which are used as proxies for the observables of the differential $\mathrm{t\bar{t}}$ measurement, for detector effects. Thus, the network aims to approximate the amount of generated missing transverse momentum in each event, which can be different to $p_{\mathrm{T}}^{\nu\nu}$ in $\mathrm{t\bar{t}}$ events, as b jets can include additional neutrinos. As targets of the DNN, the x and y component of the generated missing transverse momentum ($p_{x/y}^{\mathrm{miss,\,gen}}$) could be used, from which a correction on both $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ and $|\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss,\,gen}})|$ can be calculated directly. However, the differences to the corresponding quantities reconstructed by PUPPI ($p_{x/y}^{\mathrm{miss,\,Puppi}} - p_{x/y}^{\mathrm{miss,\,gen}}$) are used as targets instead. This improves the training process, since these differences are distributed around zero and the relative deviation between different events is thus much larger than with the absolute momenta, facilitating the distinction between events with under- and overvalued reconstructed $p_{\mathrm{T}}^{\mathrm{miss}}$.
The baseline DNN was developed by optimizing its hyperparameters in a preliminary grid search, where only one variable is adjusted and all others are fixed to an arbitrarily chosen standard value. This method is imprecise, since it does not account for correlations between hyperparameters and its results only serve as a starting point for the more coherent Bayesian search explained in Sec. 6.3. The hyperparameters of the baseline DNN, derived by the grid search, are summarized in Table 7.1. The optimization ranges of the hyperparameters for the Bayesian search were also determined with the grid search.
As inputs of the baseline DNN, 54 variables (see Appx. 5) are used. This number is reduced significantly in the input feature validation explained in Sec. 7.3.

Table 7.1: Summary of the hyperparameters of the baseline DNN with the corresponding optimization ranges used in the bayesian search. The hyperparameters are explained in detail in Sec. 6.2.

Hyperparameter	Baseline Value	Optimization Range
Learning Rate r_l	0.0001	$[0.00001, 0.05]$
Number of Layers n_{layer}	6	$\{1, 2, 3, 4, 5, 6, 7, 8\}$
Node Factor r_{node}	1	$\{1/8, 1/4, 1/2, 1, 2, 4\}$
ReLU Leakage α	0.2	$[0, 0.4]$
Dropout p_{drop}	0.35	$[0.1, 0.5]$
Regularization Weight λ	0.05	$[0.0001, 0.2]$
Batch Size s_b	5000	$[250, 10000]$

7.2 Event Reweighting

A challenge of the DNN-based $p_{\mathrm{T}}^{\mathrm{miss}}$ correction is the imbalanced distribution of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ in the training sample. Events with $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} < 200\,\mathrm{GeV}$ make up more than 99% of the data set (see Fig. 7.1), so that the events with high $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ have a small impact on the training process and are thus described inaccurately by the network (see Fig. 7.2a). To solve this problem, the $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ distribution is reweighted bin-wise, such that the resulting distribution is approximately flat in the relevant range from 0 GeV to 400 GeV. The weights in each bin i are determined as

$$w_i^p = \frac{n_{\mathrm{tot}}^{\mathrm{events}}}{n_{\mathrm{bins}} \cdot n_{\mathrm{i}}^{\mathrm{events}}}\,, \tag{7.1}$$

where $n_{\mathrm{tot}}^{\mathrm{events}}$ and $n_{\mathrm{i}}^{\mathrm{events}}$ denote the number of events in total and in bin i, respectively, and the number of bins is given by n_{bins}. For large n_{bins}, the number of events in bins of high $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ becomes too small, resulting in too large weights for a stable training process. For a small number of bins, the distribution is not flat, creating bias and reducing the effectiveness of the reweighting. A well suited binning for this $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ range was found to be 50 equidistant bins with a width of 8 GeV, in all periods of data taking. The original and the reweighted $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ distribution are shown exemplary in Fig. 7.1 for the 2018 training sample.

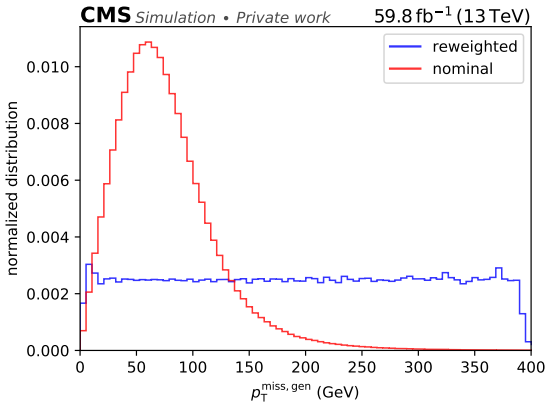
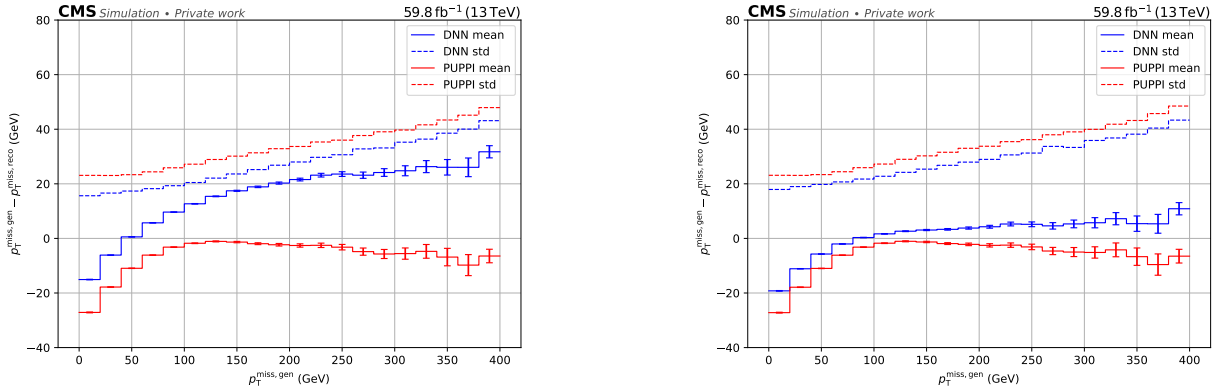


Figure 7.1: Distribution of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ in the 2018 training sample in red and in blue the reweighted distribution, both shown in 200 equidistant bins. The weights were derived using 50 equidistant bins from 0 GeV to 400 GeV, hence, the reweighted distribution is not perfectly flat.

The reweighting is working as intended, resulting in an approximately flat distribution. The weights w_i are implemented in the training process as weighting factors in the loss function:

$$J_{\text{logcosh}}(\theta) = \frac{1}{s_b} \sum_{i=1}^{s_b} w_i \cdot \log\left(\cosh\left(D_i(\theta)\right)\right), \tag{7.2}$$

with s_b, $D_i(\theta)$ defined as in Sec. 6.2. Each weight is computed as the product of the $p_{\text{T}}^{\text{miss, gen}}$ reweighting factor w_i^p and the corresponding SF weight (see Sec. 4.2). The latter is required to compensate for differing efficiencies between data and MC. The MC weights cannot be included here, since these can be negative, which defeats the purpose of the loss function. The influence of MC, p_{T} and pileup weights on the training process was studied and showed no significant differences, so that these weights are disregarded in the loss calculation.
The effect of the $p_{\text{T}}^{\text{miss, gen}}$ reweighting on the network performance is illustrated in Fig. 7.2, where resolution and bias of the baseline DNN is shown for the networks trained with (Fig. 7.2b) and without (Fig. 7.2a) the reweighting in comparison to the reconstruction (reco) with PUPPI. This is illustrated for the training sample, since it includes more events, but the result for the test sample shows the same characteristics (see Appx. 4).



(a) Without $p_{\text{T}}^{\text{miss, gen}}$ reweighting

(b) With $p_{\text{T}}^{\text{miss, gen}}$ reweighting

Figure 7.2: Mean value (solid) and standard deviation (dashed) of the $p_{\text{T}}^{\text{miss, gen}} - p_{\text{T}}^{\text{miss, reco}}$ distribution as a function of $p_{\text{T}}^{\text{miss, gen}}$, shown for reconstruction with PUPPI (red) and baseline DNN (blue) in the 2018 training sample. The last bin includes events with $p_{\text{T}}^{\text{miss, gen}} > 400\,\text{GeV}$. The red curves on the left and on the right are identical.

The mean values derived without reweighting show a significant bias towards positive values for large $p_{\text{T}}^{\text{miss, gen}}$ due to the under representation of such events in the training sample. In this phase space, the positive effect of the $p_{\text{T}}^{\text{miss, gen}}$ reweighting is clearly visible, as the mean values of the DNN calculation using the reweighting is much closer to the ideal value of zero. The resolution of the DNN, illustrated by the standard deviation, degrades slightly for bins of small $p_{\text{T}}^{\text{miss, gen}}$ through use of the reweighting, while it is mostly unchanged for large $p_{\text{T}}^{\text{miss, gen}}$. However, the bias reduction caused by the reweighting is clearly the dominant effect. Furthermore, the resolution of $p_{\text{T}}^{\text{miss, DNN}}$ derived using the $p_{\text{T}}^{\text{miss, gen}}$ reweighting, is still significantly higher than the resolution of $p_{\text{T}}^{\text{miss, Puppi}}$ in the whole phase space. For small values of $p_{\text{T}}^{\text{miss, gen}}$, a bias of the mean values is visible for both PUPPI and DNN reconstruction. This bias is caused by the correlation of the quantities displayed on the x and y axes. For small $p_{\text{T}}^{\text{miss, gen}}$, the difference $p_{\text{T}}^{\text{miss, gen}} - p_{\text{T}}^{\text{miss, reco}}$ is directly biased towards negative values. For both PUPPI and DNN reconstruction a step is visible in the last bin of the mean values, since it includes all events with $p_{\text{T}}^{\text{miss, gen}} > 400\,\text{GeV}$.

7.2.1 Performance Fluctuations

Training a neural network is a nondeterministic process (see Sec. 6.1). Therefore, two networks, that are independently trained with exactly the same data and hyperparameters will yield different results. These performance fluctuations of DNNs are significantly increased by large sample weights as for example introduced by the $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ reweighting. By repeating the training process of a DNN multiple times, one can derive the standard deviation of the resulting validation loss to obtain a quantitative measure for the performance fluctuations. Comparing this measure for networks trained with and without $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ weights yields relative standard deviations of the validation loss around 0.33% and 0.022%, respectively, showing an increase by a factor of ~ 10 caused by the $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ reweighting. This corresponds to a fluctuation of the mean difference $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} - p_{\mathrm{T}}^{\mathrm{miss,\,DNN}}$ of around 0.23 GeV. Therefore, multiple methods other than $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ reweighting were tested to balance the data sample, such as oversampling and undersampling, but $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ reweighting still achieved overall the best performance and is therefore used throughout this thesis.

7.3 Input Features

To remove bias from the DNN due to differences between the MC simulation used during training and recorded data, all input features are validated. This validation process is done by performing a goodness-of-fit (GOF) test of all input feature distributions in simulation to the corresponding input feature distributions in recorded data. Since many input features are correlated, the GOF tests are additionally conducted in 2D distributions of all possible combinations of two input features. Therefore, the number of goodness-of-fit tests that have to be performed is proportional to the number of input features in quadrature ($n_{\mathrm{GOF}} \propto n_{\mathrm{feat}}^2$). As it is not feasible to perform $\sim 54^2$ GOF tests, the number of input features has to be restricted. In order to maintain the network performance, only less impactful features to the training process are discarded, which is determined by so-called Shapley values, explained in the following section. Subsequently, the goodness-of-fit determination is outlined and the finalized set of input features is presented.

7.3.1 Shapley Values

Shapley values [78, 79] can be used as a game theoretic approach to quantify the influence of single input features to the network output. The Shapley value S of input feature f_0 is derived as a mean value of Shapley instances S_i:

$$S(f_0) = \frac{1}{n_S} \sum_{i}^{n_S} S_i(f_0)\,. \tag{7.3}$$

The Shapley instances S_i are calculated for each event i in a subset of $n_S = 10000$ events from the training sample as

$$S_i(f_0) = \frac{1}{n_{\mathrm{feat}}!} \sum_{\mathcal{F} \subseteq I \setminus f_0} \left[y_i^*(\mathcal{F} \cup f_0) - y_i^*(\mathcal{F}) \right]\,, \tag{7.4}$$

where $\mathcal{F}$ are all possible subsets of input features I without including f_0 and y_i^* is an estimation of the baseline DNN output trained with a subset of the input features. The difference $[y_i^*(\mathcal{F} \cup f_0) - y_i^*(\mathcal{F})]$ is thus an approximation of the difference in the baseline DNN output trained with a subset of input features between including and excluding f_0.
The determination of the shapley values, including the approximation of the baseline DNN

output for a subset of input features, is done in this thesis with Python's SHAP library [80]. The Shapley values of all initial 54 input features are computed and used to sort the input features by relevance. To evaluate how many features can be discarded, the performance of the baseline DNN is studied for different numbers of input features, shown in Fig. 7.3.

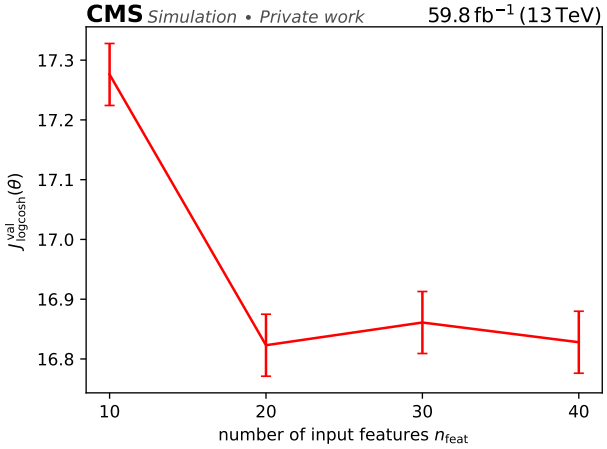


Figure 7.3: Validation loss of the baseline DNN in the 2018 simulation trained with different numbers of input features. Uncertainties on the loss values cover the performance fluctuations and are derived as explained in in Sec. 7.2.1.

Here, the validation loss $J^{\text{val}}_{\text{logcosh}}$ is determined for different numbers of input features (n_{feat}), always using the features with the highest Shapley values. A significant decrease in the network performance is visible when reducing n_{feat} from 20 to 10. Therefore, the first 20 input features by Shapley value, as shown in Fig. 7.4, are selected and further tested within the goodness-of-fit determination shown in Sec. 7.3.2.

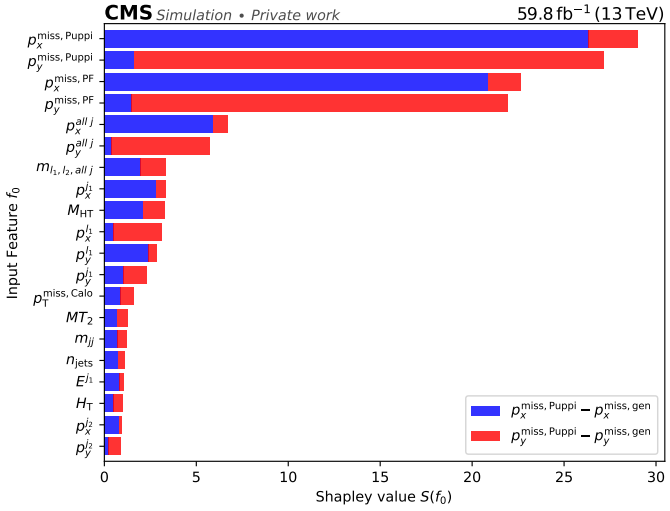


Figure 7.4: Shapley values of the first 20 input features, shown for both DNN outputs, $p_x^{\text{miss, Puppi}} - p_x^{\text{miss, gen}}$ (blue) and $p_y^{\text{miss, Puppi}} - p_y^{\text{miss, gen}}$ (red) in 2018 simulation.

The distribution of the Shapley values confirms that the x/y component of the $p_{\text{T}}^{\text{miss}}$ reconstructions with PUPPI and PF algorithms are the most important inputs for the correction of $p_{x/y}^{\text{miss, Puppi}}$. In contrast, $p_{\text{T}}^{\text{miss, Calo}}$ is much less impactful to the training process, as it cannot

be split into x and y component, due to its insufficient angular resolution (see Sec. 5.1.5). Most other input features describe energy or momentum of final state objects, increasing in relevance to the DNN training with larger energy contribution. The x and y component of the combined momentum of all jets therefore corresponds to the largest Shapley value, apart from the $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstructions. Since the Shapley values are only approximations of the input feature impact and are subject to statistical fluctuations, the small differences between the total Shapley values of for example $p_x^{j_1}$ and $p_y^{j_1}$ can be disregarded.

7.3.2 Goodness-of-fit determination

The goodness-of-fit tests are conducted using the combine tool [81] for 1D distributions of all 20 input features and for 2D distributions of all possible feature combinations to account for correlations between features. In each test, a fit of the simulated distribution to the data distribution is performed, by adjusting the signal strength μ and shifting the simulated distribution within its systematic uncertainties, described by the nuisance parameters ξ. The fit maximizes the likelihood

$$\mathcal{L}(\,\mathrm{data}\,|\,\mu,\xi) = \prod_i \left[\mathrm{Poisson}(\,N_i\,|\,\mu\cdot S_i(\xi) + B_i(\xi)\,)\right]\cdot p(\tilde{\xi}|\xi)\,, \tag{7.5}$$

where N_i, S_i and B_i denote data, signal and background yields in each bin i, respectively. The distribution of nuisance parameters ξ given the default values $\tilde{\xi}$, is accounted for by multiplying the underlying probability density function $p(\tilde{\xi}|\xi)$, based on a log-normal distribution [82]. The impact of the nuisance parameters is explained in more detail and shown exemplary for one of the 2D distributions in Appx 6.
As a goodness-of-fit measure, the *saturated model* method is used [83], which is similar to a χ^2 value, but can be calculated for arbitrary binned data. The test statistic of this model is defined as a likelihood ratio

$$\lambda_{\mathrm{sat}} = -2\ln\left(\frac{\max\mathcal{L}(\,\mathrm{data}\,|\,\mu,\xi)}{\max\mathcal{L}(\,\mathrm{data}\,|\,N_{\mathrm{sat}})}\right), \tag{7.6}$$

with the distribution of the saturated model N_{sat}. The saturated model is defined as a model, which describes the data perfectly in each bin $N_i = N_{\mathrm{sat},i}$, and therefore generally needs as many parameters as there are data points. The likelihood of the saturated model serves as a normalization to solve shortcomings of an absolute likelihood as a test statistic [83].
To exclude features, that are poorly described by simulation, a p-value [84] is used. This quantity compares the test statistic of the data ($\lambda_{\mathrm{sat}}^{\mathrm{data}}$) to the test statistic of toy distributions ($\lambda_{\mathrm{sat}}^{\mathrm{toy}}$), which are generated with the *frequentist* toy generation method [82] of the combine tool based on the MC simulation. The p-value in this thesis is defined as

$$p = \frac{\text{number of toys with } \lambda_{\mathrm{sat}}^{\mathrm{toy}} > \lambda_{\mathrm{sat}}^{\mathrm{data}}}{\text{total number of toys}}\,. \tag{7.7}$$

A large p-value implies a low λ_{sat} of the data compared to the toys and thus good agreement between data and simulation. With increasing number of toys, the distribution $\lambda_{\mathrm{sat}}^{\mathrm{toy}}$ approaches a χ^2 distribution with number of degrees of freedom n_{dof} equal to the number of bins in the feature distribution. The λ_{sat} distribution is shown exemplary for the input feature combination $p_x^{\mathrm{miss,\,Puppi}}$ vs $p_y^{\ell_1}$ in Fig. 7.5 together with a χ^2 fit. The number of freedoms resulting from the fit $n_{\mathrm{dof}} = 89.2 \pm 0.6$ is in good agreement with the expectation of $3\cdot 30 = 90$ for three channels (ee, eμ, $\mu\mu$) and 6×5 bins in the 2D distribution of $p_x^{\mathrm{miss,\,Puppi}}$ vs $p_y^{\ell_1}$. The 1D and 2D binnings of all input features are summarized in Appx. 7.

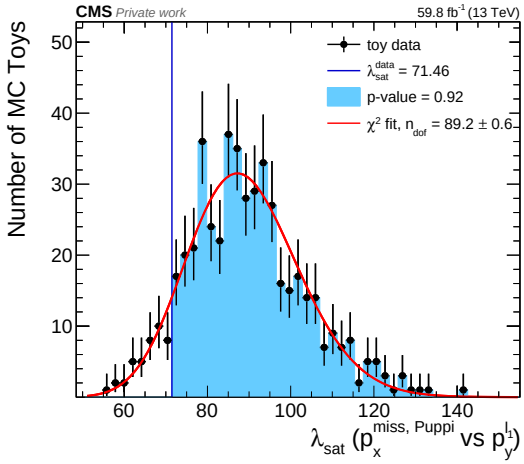


Figure 7.5: Distribution of λ_{sat} of 500 toys generated based on simulation for input feature combination $p_x^{\text{miss, Puppi}}$ vs $p_y^{\ell_1}$ with corresponding χ^2 fit in red using data from 2018. Poisson uncertainties of the toy distribution are displayed as black errorbars. The value of $\lambda_{\text{sat}}^{\text{data}}$ is illustrated as a blue line, all bins with higher λ_{sat} are shaded in light blue, illustrating the p-value.

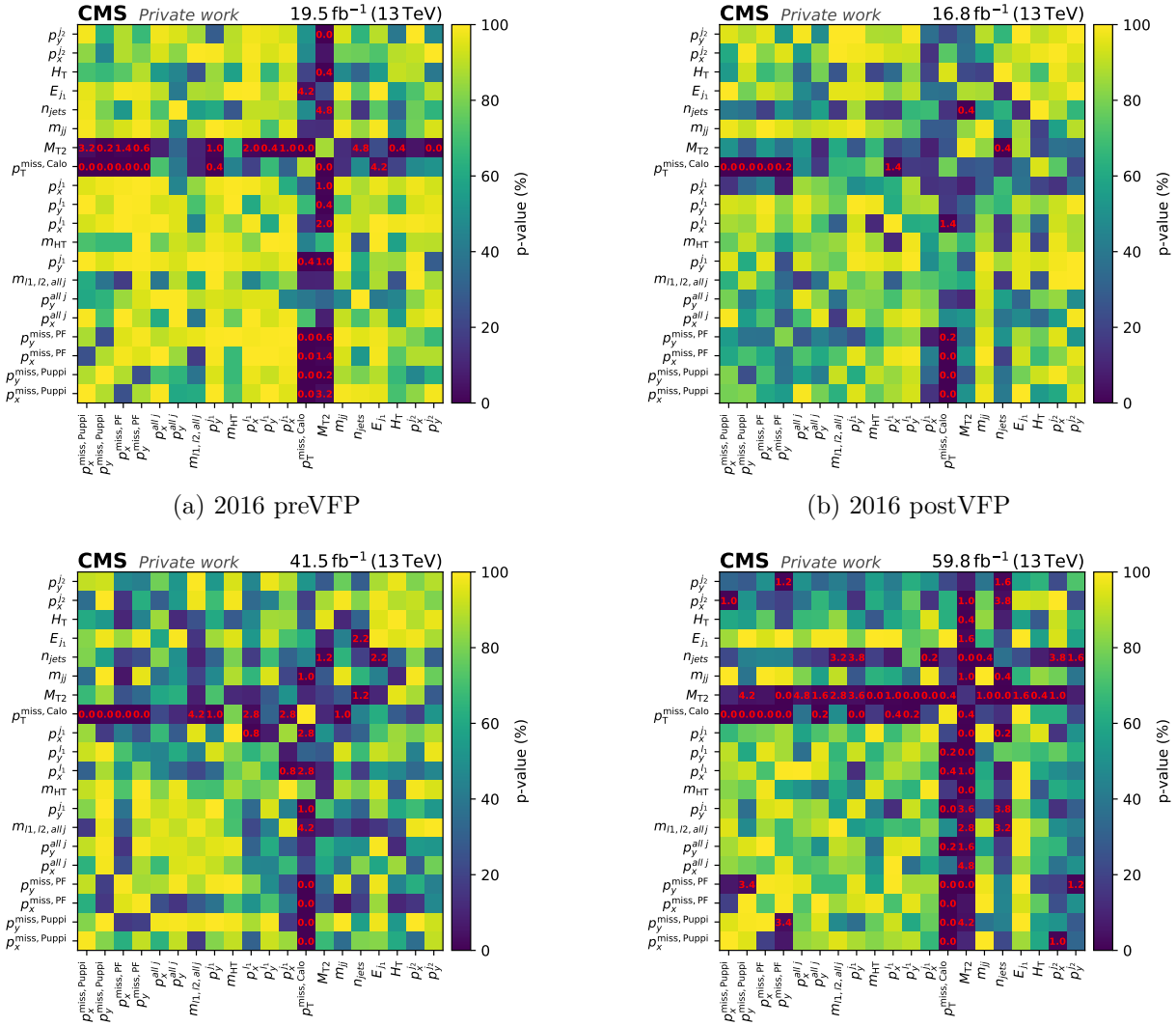
The test statistic λ_{sat} uses the signal- and background predictions, which maximize the likelihood shown in Eq. 7.5. The p-values are thus based on these so-called postfit distributions and evaluate if the MC prediction matches the recorded data distributions, taking into account all systematic uncertainties. If the data and MC distributions are not in good agreement, the fit fails resulting in a low p-value.

The feature exclusion is done based on the p-values of the 1D and 2D distributions of all input features, considering all periods of data taking. These p-values are illustrated in Fig. 7.6.
All off-diagonal elements in the two-dimensional p-value distributions are mirrored, as no distinction is made between e.g. the distributions $p_x^{\text{miss, Puppi}}$ vs $p_y^{\ell_1}$ and $p_y^{\ell_1}$ vs $p_x^{\text{miss, Puppi}}$.
Overall, the p-values show good agreement between data and simulation of most inputs features, with all 1D distributions and most 2D distributions yielding p-values above 5%. There are clear correlations between the periods of data taking. For all periods, most 2D distributions of $p_{\text{T}}^{\text{miss, Calo}}$ are described inaccurately by the MC simulation. The other features, which need to be excluded are M_{T2} and n_{jets}, as most clearly visible by the p-values for 2018 data. When excluding these three features, only a few p-values remain below 5% (see App. 8):

- $p = 3.4\%$ for $p_y^{\text{miss, Puppi}}$ vs $p_y^{\text{miss, PF}}$ in 2018
- $p = 1.0\%$ for $p_x^{\text{miss, Puppi}}$ vs $p_x^{j_2}$ in 2018
- $p = 1.2\%$ for $p_y^{\text{miss, PF}}$ vs $p_y^{j_2}$ in 2018
- $p = 0.8\%$ for $p_x^{j_1}$ vs $p_x^{\ell_1}$ in 2017

These feature combinations were studied in more detail. The features in each feature pair are strongly correlated. For instance $p_x^{j_2}$ influences $p_x^{\text{miss, Puppi}}$ directly and due to the kinematics of the $\text{t}\bar{\text{t}}$ decay, events with large missing transverse momentum almost always include large jet momenta as well. This leads to bins of large $p_x^{\text{miss, Puppi}}$ and low $p_x^{j_2}$ being almost empty, which can cause problems in the fitting procedure. When fitting the 2D distributions for all decay channels separately, the p value is computed to be above 5% for each pair. Additionally, since numerous goodness-of-fit tests are performed (210 in each period), it is to be expected, that

some combinations are excluded due to fluctuations. For these reasons, none of the features in these pairs were excluded, resulting in the final set of input parameters shown in Table 7.2.



(a) 2016 preVFP

(b) 2016 postVFP

(c) 2017

(d) 2018

Figure 7.6: Illustration of p-values for all periods and all input features. Elements on the diagonal are the results of the 1D distributions, all off diagonal elements show the results for the 2D distributions. Each bin with p-value below 5% is marked by the corresponding p-value in red font.

Table 7.2: Final set of 17 input features (see Sec 5.1.6), sorted by Shapley values.

$p_x^{\text{miss, Puppi}}$,	$p_y^{\text{miss, Puppi}}$,	$p_x^{\text{miss, PF}}$,	$p_y^{\text{miss, PF}}$,	$p_x^{\text{all } j}$,	$p_y^{\text{all } j}$,
$m_{\ell_1,\ell_2,\text{all } j}$,	$p_y^{j_1}$,	m_{HT},	$p_x^{\ell_1}$,	$p_y^{\ell_1}$,	$p_x^{j_1}$,
m_{jj},	E_{j_1},	H_{T},	$p_x^{j_2}$,	$p_y^{j_2}$	

8 Results

In this section, the results of the Bayesian hyperparameter optimization are presented. The performance of the obtained DNN is analyzed and compared between the periods of data taking. The resolution and bias of $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction using the DNN is studied in comparison to reconstructions with PUPPI and PF algorithm. Subsequently, the impact of the DNN on the event migration is shown as well as the performance of the DNN for background and BSM events.

8.1 Bayesian Optimization

The Bayesian hyperparameter optimization (see Sec. 6.3) is performed using the 2018 data sample for 200 optimization steps with 9 initial random guesses to start off the Gaussian process. The target of the maximization is the negative validation loss $-J^{\mathrm{val}}_{\mathrm{logcosh}}$. For better visibility, it is treated in the following as a minimization with $+J^{\mathrm{val}}_{\mathrm{logcosh}}$ as the target. The validation loss as a function of the optimization step is shown in Fig. 8.1. The DNN evaluations cluster mainly in two groups, one in the lower region $16.5 < J^{\mathrm{val}}_{\mathrm{logcosh}} < 17.5$, one in the upper region $20.5 < J^{\mathrm{val}}_{\mathrm{logcosh}} < 21.5$. An explanation for this is that the networks in the upper region are prone to overtraining and the early stopping thus triggers already at the start of the training process, leading to larger validation losses. Similarly, networks with many layers and nodes per layer, but small learning rates tend to yield large validation losses. In contrast, the training of the DNNs in the lower group seems to have worked well, yielding significantly smaller validation losses.

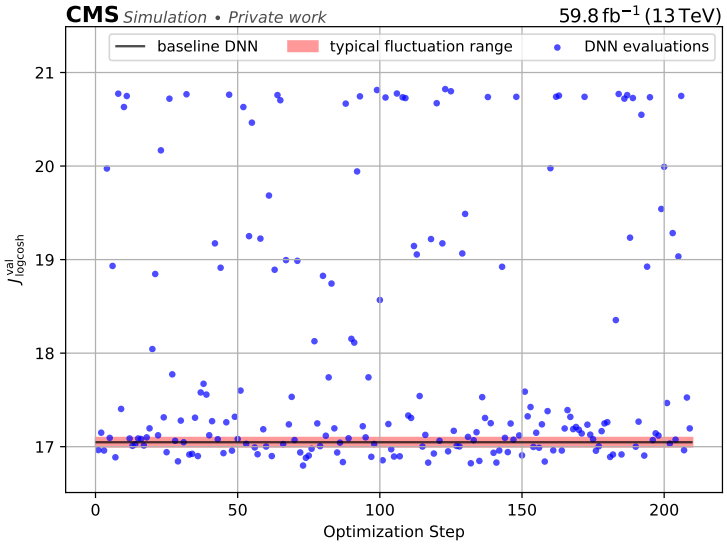


Figure 8.1: Visualization of the Bayesian optimization process for 2018. Each point represents a trained DNN with different hyperparameters. The optimization target, namely the validation loss, is shown on the y axis as a function of the optimization step. The performance of the baseline DNN is indicated by a black line, a typical performance fluctuation range is visualized by a red shaded rectangle.

Overall, no clear trend is visible as a function of the optimization step. One reason for this are the performance fluctuations of the DNNs (see Sec. 7.2.1), which can result in significantly different validation losses for very similar hyperparameters. This implies that the loss as a function of the hyperparameters is not a continuous function, which hinders the optimization process.

Another reason for this wide validation loss distribution is the large scale of the searched hyperparameter space. Since a large search space needs to be covered, many evaluations must serve the purpose of exploration instead of exploitation.
The set of hyperparameters, which yields the smallest validation loss in the course of this optimization (denoted as DNN 1) is shown in Table 8.1. Notably, DNN 1 only has 2 layers, a compact network structure thus suffices for the correction of the $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction.

Table 8.1: Hyperparameters (see Sec. 6.2) of DNN 1-4, DNN 1 denoting the best performing DNN from the Bayesian optimization, DNN 2 denoting the baseline DNN. DNN 3-4 are networks with significantly differing performances, used for the comparison between the periods (Fig. 8.2). $\langle J^{\mathrm{val}}_{\mathrm{logcosh}} \rangle$ is determined as the mean value over 5 network evaluations.

	r_{learn}	n_{layers}	r_{node}	α	p_{drop}	λ	s_{b}	$\langle J^{\mathrm{val}}_{\mathrm{logcosh}} \rangle$ 2018
DNN 1	$4.4 \cdot 10^{-04}$	2	2	0.12	0.35	0.0026	666	16.91
DNN 2	$1.0 \cdot 10^{-04}$	6	1	0.2	0.35	0.05	5000	17.05
DNN 3	$8.7 \cdot 10^{-05}$	4	0.125	0.066	0.47	0.00087	5194	17.82
DNN 4	$1.3 \cdot 10^{-05}$	3	1	0.34	0.25	0.039	965	18.09

To evaluate, how the performances of the DNNs differ between the periods of data taking, networks with significantly differing performances in 2018 are also trained using simulations of other periods. To mitigate the influence of performance fluctuations, the networks are trained 5 times and the mean values of the validation loss are compared. The results are shown in Fig. 8.2.

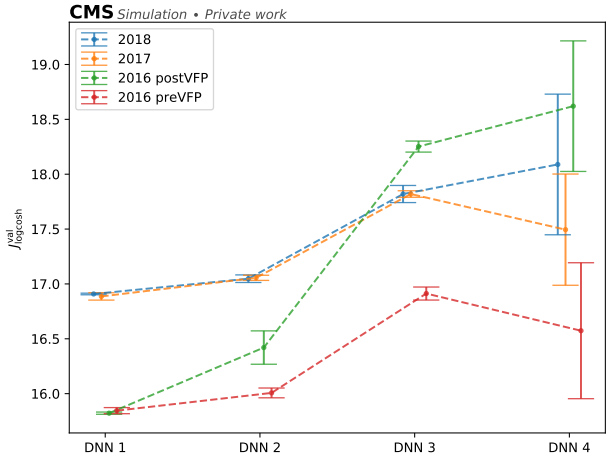


Figure 8.2: Mean value of five evaluations of the validation loss for four different networks and all periods of data taking. The error bars show the corresponding uncertainties on the mean values. The hyperparameters of DNN 1-4 are summarized in Table 8.1.

For all periods of data taking, DNN 1 performs best, showing the smallest validation loss with the smallest performance fluctuations. The performance of the baseline DNN (DNN 2) is slightly worse for all periods. The validation losses of DNN 3 and 4 are significantly larger in all periods. DNN 4 also shows considerably larger performance fluctuations. The training of DNN 4 either worked well, resulting in a validation loss of $J^{\mathrm{val}}_{\mathrm{logcosh}} \sim 16\,(17)$ for 2016 (2017 and 2018) or was

stopped early at losses around 19-19.5. This could be due to the very low learning rate of DNN 4, which can cause the early stopping to trigger in the beginning of the training process if no significant improvement is visible in the first 10 epochs.
During the Runs of the LHC in periods 2016 preVFP and 2016 postVFP, fewer simultaneous bunch collisions were induced, resulting in less pileup [34]. Therefore, the initial resolution of $p_{\mathrm{T}}^{\mathrm{miss,\,Puppi}}$ and $p_{\mathrm{T}}^{\mathrm{miss,\,PF}}$ are better in these periods. This has a direct influence on the DNN performance, especially notable for DNN 1 and DNN 2, where the network performs significantly better in those periods compared to 2017 and 2018.

Overall, the DNN performance is strongly correlated between the periods of data taking. The hyperparameter optimization is therefore not repeated for other periods, and the hyperparameters of DNN 1 from the 2018 optimization are used for 2016 preVFP, 2016 postVFP and 2017 as well.

8.2 Network Resolution

The goal of the DNN-based $p_{\mathrm{T}}^{\mathrm{miss}}$ correction is to achieve a better resolution of the top measurement observables than with PUPPI and PF reconstruction. The resolution can be evaluated by comparing the different $p_{\mathrm{T}}^{\mathrm{miss}}$ and $\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ reconstructions with the corresponding generator-level quantities. This is illustrated in Fig. 8.3, where the distributions of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} - p_{\mathrm{T}}^{\mathrm{miss,\,reco}}$ and $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,gen}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,reco}})$ are shown for reconstruction with DNN 1, PUPPI and PF in the 2018 test sample. In the $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} - p_{\mathrm{T}}^{\mathrm{miss,\,reco}}$ distribution, the DNN correction yields the best performance, achieving a mean value closest to zero and thus the least bias, as well as the smallest standard deviation and thus the highest resolution. The mean value of the $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,gen}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,reco}})$ distribution is slightly shifted towards negative values for the DNN reconstruction compared to PUPPI and PF, which are centered around zero. However, the width of the DNN distribution is significantly smaller, improving the resolution by about 13%.
The resolution in the other data taking periods shows the same effects (see Appx. 9). Overall, the best resolution can be achieved for both observables and all periods by employing the DNN-based $p_{\mathrm{T}}^{\mathrm{miss}}$ correction.

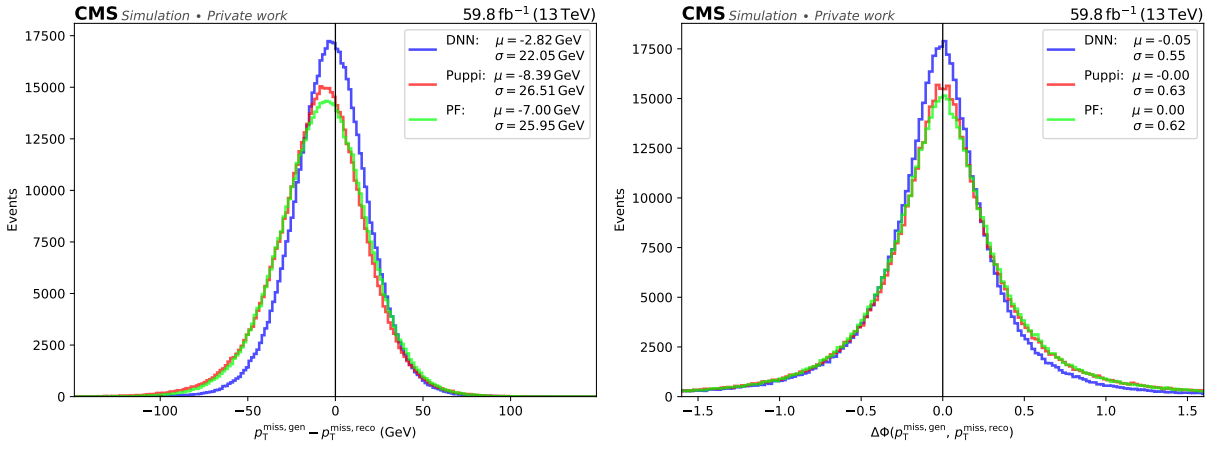


Figure 8.3: Distribution of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} - p_{\mathrm{T}}^{\mathrm{miss,\,reco}}$ (left) and $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,gen}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,reco}})$ for DNN 1 (blue), PUPPI (red) and PF (green) reconstruction in the 2018 test sample. Mean value μ and standard deviation σ of each distribution are displayed in the legend.

Since the DNN resolution is of particular interest in the region of high $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ for the top measurement, the mean value and standard deviation of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} - p_{\mathrm{T}}^{\mathrm{miss,\,reco}}$ are also studied as a function of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$. This is shown in Fig. 8.4 for DNN 1, PUPPI and PF reconstruction in the 2018 test sample.

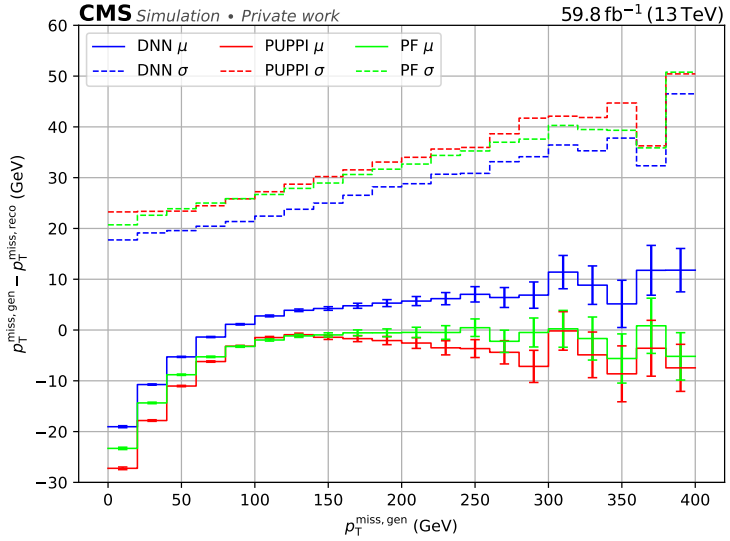


Figure 8.4: Mean value (solid) and standard deviation (dashed) of the $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} - p_{\mathrm{T}}^{\mathrm{miss,\,reco}}$ distribution as a function of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$, shown for reconstruction with PUPPI (red), PF (green) and DNN 1 (blue) in the 2018 test sample. The last bin includes overflow events (all events with $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} > 400\,\mathrm{GeV}$).

The mean distribution of DNN 1 is similar to the distribution of the baseline DNN (see Fig. 7.2b). As before, a bias is visible for small $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ values due to the correlation of the chosen variables on the x and y axes. For large $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$, the mean difference for both DNN 1 and PUPPI reconstructions deviates from zero by the same order of magnitude of $5 - 10\,\mathrm{GeV}$. Here, the mean values of the DNN distribution are slightly above zero, underestimating the amount of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$, while the amount of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ is overestimated by the PUPPI reconstruction. The reconstruction with PF shows the least bias in this $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ range.
The DNN reconstruction achieves the highest resolution, shown by the smallest standard deviation of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} - p_{\mathrm{T}}^{\mathrm{miss,\,reco}}$ in the whole phase space. Between PF and PUPPI reconstruction, the resolution is at similar level.

8.3 Performance for Background and BSM Processes

The performance of the DNN is also studied for the most dominant background processes and for an exemplary BSM sample, to ensure that the network does not memorize the kinematics of the $\mathrm{t\bar{t}}$ process, but generalizes the reconstruction to all processes in the considered phase space. This is illustrated in Fig. 8.5, where, analogously to Fig. 8.3 the distributions of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} - p_{\mathrm{T}}^{\mathrm{miss,\,reco}}$ and $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,gen}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,reco}})$ are shown for reconstruction with DNN 1, PUPPI and PF in a BSM sample and a sample of the single top quark production (single t). The considered BSM sample models a supersymmetric stop production, where the stops decay to top quarks and neutralinos, as shown in Fig. 2.2. A stop mass of 525 GeV and neutralino mass of 350 GeV are presumed in this sample. The resolution for other backgrounds, namely Drell-Yan process, semileptonic $\mathrm{t\bar{t}}$ decay and dileptonic $\mathrm{t\bar{t}}$ decay into taus, is shown in Appx. 10.

For both single t and stop pair production, a positive effect of the DNN-based $p_{\mathrm{T}}^{\mathrm{miss}}$ correction is visible. In $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} - p_{\mathrm{T}}^{\mathrm{miss,\,reco}}$ and $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,gen}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,reco}})$ distributions, the DNN achieves the lowest standard deviation. In case of the $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} - p_{\mathrm{T}}^{\mathrm{miss,\,reco}}$ distribution, the mean value can even be shifted closer to zero. This confirms that the DNN has improved the $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction in the considered phase space, without receiving significant bias through training only with a sample of the dileptonic $\mathrm{t\bar{t}}$ decay.

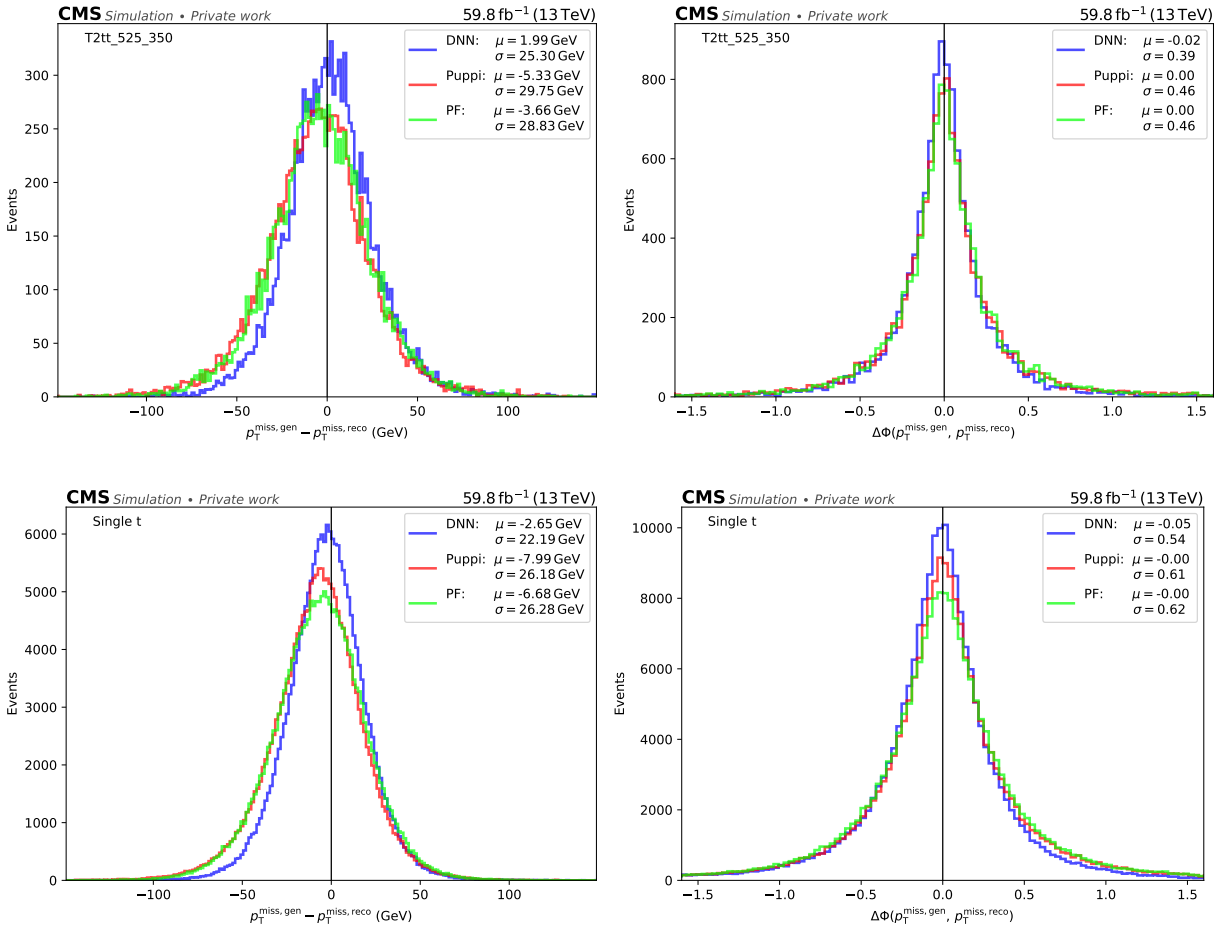


Figure 8.5: Distribution of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} - p_{\mathrm{T}}^{\mathrm{miss,\,reco}}$ and $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,gen}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,reco}})$ for DNN 1, PUPPI and PF reconstruction for a single t sample (top) and a BSM sample (bottom). The BSM process is a stop pair production (see Fig. 2.2) with stop mass of 525 GeV and neutralino mass of 350 GeV. Mean μ and standard deviation σ of the distributions are displayed in the legend. Both the single t and the BSM sample are generated for the 2018 data taking period.

8.4 Impact on Event Migrations

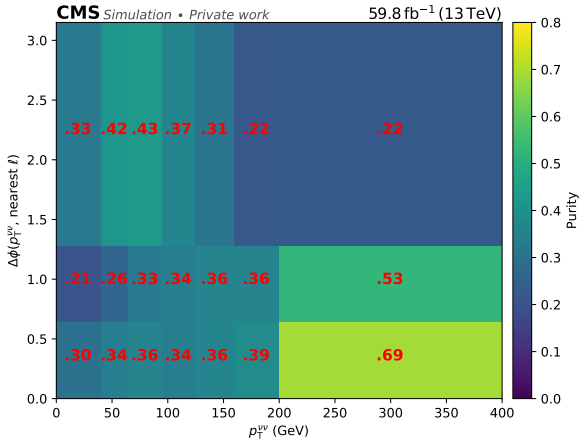
One central motivation for the development of the DNN are the event migrations in the binning of the differential $\mathrm{t\bar{t}}$ measurement (see Sec. 2.4). The event migrations can be quantified by the following three quantities:

The **purity** of bin i is defined as the number of events generated and correctly reconstructed in i, divided by the number of events reconstructed in i.

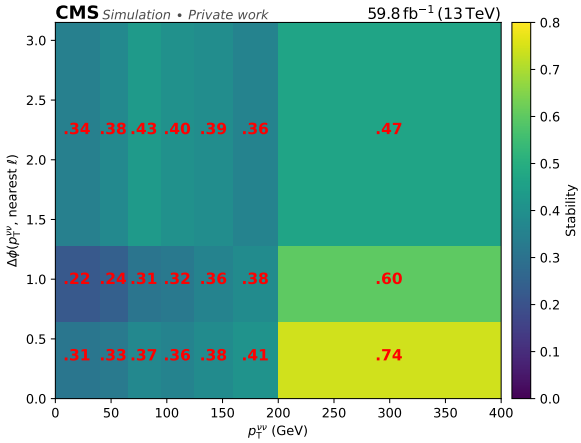
The **stability** of bin i is defined as the number of events generated and correctly reconstructed in i, divided by the number of events generated in i.

The **response matrix** element ij is defined as the number of events reconstructed in bin i and generated in bin j. In this thesis, the elements are normalized with the number of events reconstructed in bin i, for illustrative purposes. The diagonal elements of the response matrix are thus identical with the purity.

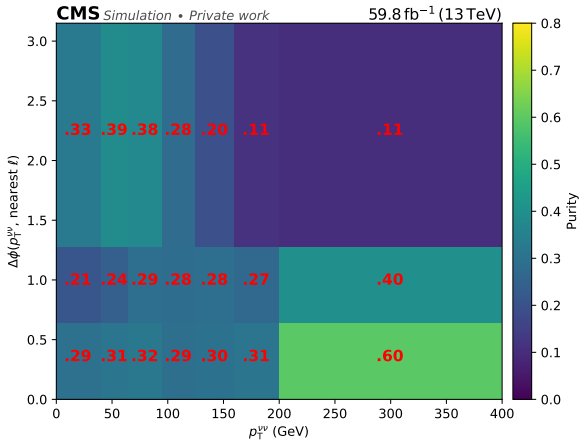
For all these definitions, the binning of the generated events is defined by the truth-level quantities $p_{\mathrm{T}}^{\nu\nu}$ and $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\nu\nu}$, nearest ℓ) in the optimized binning of the differential $\mathrm{t\bar{t}}$ measurement (see Table 2.1).
Purity and stability are displayed in Fig. 8.6 for reconstruction with PUPPI, PF and DNN 1. The purity can be increased significantly by employing the DNN, especially for large $p_{\mathrm{T}}^{\nu\nu}$, where the purity now exceeds 20% in all bins. The stability on the other hand mostly improves for small $p_{\mathrm{T}}^{\nu\nu}$, surpassing 20% in all bins through the DNN reconstruction as well. The stability slightly decreases for the largest $p_{\mathrm{T}}^{\nu\nu}$ bins, but the relative decrease is much less significant than the relative increases in purity and stability.
The event migrations are overall reduced, as also reflected in the response matrix, which tends towards a more diagonal matrix for the DNN compared to PUPPI and PF reconstructions, as shown in Appx. 11. In summary, the DNN-based $p_{\mathrm{T}}^{\mathrm{miss}}$ correction successfully reduces event migrations and is therefore a valuable addition in the differential $\mathrm{t\bar{t}}$ measurement, improving the robustness of the unfolding procedure.



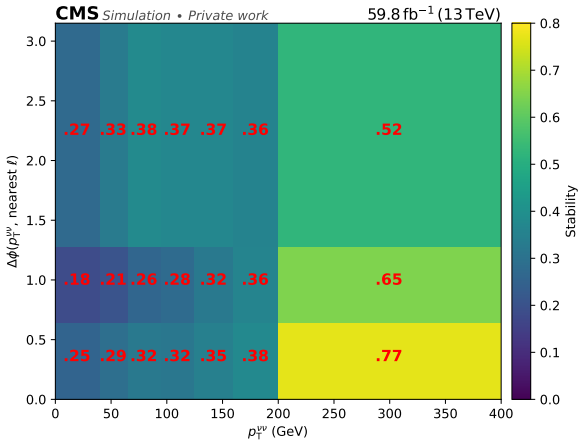
(a) Purity DNN



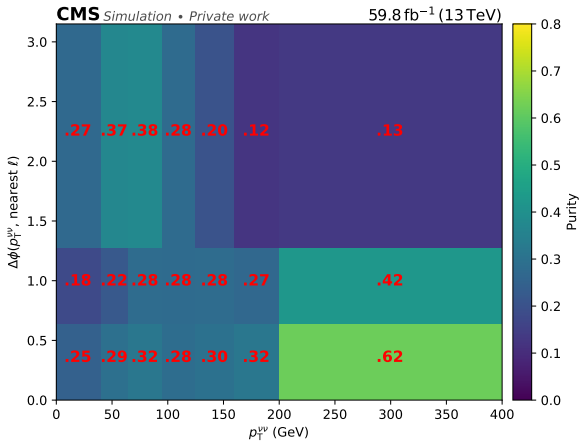
(b) Stability DNN



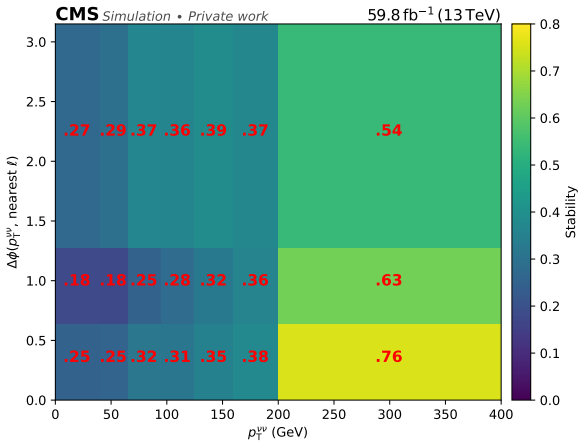
(c) Purity PUPPI



(d) Stability PUPPI



(e) Purity PF



(f) Stability PF

Figure 8.6: Purity and stability for $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction with DNN 1 (top), PUPPI (middle) and PF (bottom), shown for the full nominal $\mathrm{t\bar{t}}$ sample in period 2018.

9 Summary and Outlook

In this thesis, a deep neural network was developed, aiming to improve the resolution and reduce the bias of the $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction in order to mitigate event migrations in the $p_{\mathrm{T}}^{\mathrm{miss}}$-dependent binning of a differential cross section measurement of the dileptonic $\mathrm{t\bar{t}}$ production.
The analysis used data recorded at the CMS experiment at a center-of-mass energy of 13 TeV during Run 2 of the LHC, as well as corresponding MC simulations of the $\mathrm{t\bar{t}}$ process and its backgrounds, split into four periods of data taking.
The DNN uses kinematics of an event as input features, including $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstructions by the PUPPI and PF algorithms, to correct the x and y component of $p_{\mathrm{T}}^{\mathrm{miss,\,Puppi}}$ towards the amount of $p_{\mathrm{T}}^{\mathrm{miss}}$ generated in the event. Using these corrections, the variables $p_{\mathrm{T}}^{\mathrm{miss,\,DNN}}$ and $|\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss,\,DNN}}, \text{nearest } \ell)|$ were derived, which can serve as proxies for the observables of the $\mathrm{t\bar{t}}$ measurement $p_{\mathrm{T}}^{\nu\nu}$and $|\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\nu\nu}, \text{nearest } \ell)|$ in the unfolding procedure.
To develop a suitable DNN, a baseline network was trained for general performance studies and as a starting point of the network optimization process. To correct for the bias of the network for events with large $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$, caused by an under representation of this phase space in the training sample, a $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$-dependent reweighting procedure was introduced. The weights aim to balance the $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ distribution in the training sample, such that events from an under represented phase space have a larger impact on the training process. The network bias was compared between training with and without the $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ weights, showing a significant bias reduction caused by the reweighting procedure.
In the next step, the input features of the network were sorted by their impact on the training process according to their Shapley values. The compactness of the DNN was increased by reducing the initial 54 input features to include only the first 20 inputs by Shapley values, while the DNN performance showed not significant difference due to this input feature reduction. The 20 input features were then validated by performing goodness-of-fit tests between data and simulation of all 1D feature distributions and all 2D distributions of feature combinations to account for correlations. Overall, the tests yielded good agreement between data and simulation. Only three features needed to be excluded due to insufficient agreement in a few 2D distributions.
Based on the finalized set of input features, a Bayesian search algorithm was employed to find a optimized set of hyperparameters of the network. The algorithm showed no clear convergence, which might be caused by the large hyperparameters space, which needed to be covered, or due to performance fluctuations of the network, that were amplified by the $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ reweighting process. The network with the lowest loss in the hyperparameter optimization, could however still improve the performance of the baseline DNN significantly, for all periods of data taking.
Comparing the resolution and bias of the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction using the DNN-based correction to reconstructions using the PUPPI or PF algorithms, showed a considerably improved resolution for both the absolute value and ϕ angle of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$. The bias of the reconstruction of the absolute value could be reduced while maintaining bias in the same order of magnitude for the angle reconstruction of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$. The same was true for the DNN performance in single top production and BSM events, indicating that the network indeed corrects for detector effects in the $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction instead of memorizing kinematics of the $\mathrm{t\bar{t}}$ process.
Lastly, the event migrations in the binning of the differential $\mathrm{t\bar{t}}$ measurement were compared for the reconstruction with PUPPI, PF and the DNN. Purity and stability could be improved, exceeding 20% in all bins through application of the DNN. Therefore, less event migrations need to be corrected in the unfolding procedure, increasing its robustness. This directly affects the binning optimization, which is done in the differential $\mathrm{t\bar{t}}$ measurement [58], yielding a 2D binning of 21 bins when using DNN-based corrction, instead of 16 bins without the correction. These additional five measuring points in the differential cross section measurement can thus be

attributed to an improved $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction in a $\mathrm{t\bar{t}}$-like phase space, through application of the developed deep neural network.

The neural network could be improved by acquiring more MC statistics in the phase space of large $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ to render the $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ reweighting unnecessary and thus to stabilize the training process and reduce performance fluctuations. This could also improve the success of the Bayesian hyperparameter optimization. Additionally, more statistics could reduce the network bias in this phase space. Furthermore, the performance of different multivariate methods could be evaluated, such as recurrent and convolutionary neural networks, which are for example used for b-tagging with the DeepJet algorithm [52].
Conceptually, a multivariate method to improve the $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction in a certain phase space has been shown to be a suitable addition in a $p_{\mathrm{T}}^{\mathrm{miss}}$-dependent cross section measurement and should be considered in similar analyses as well. Especially with the perspective of future runs, e.g. in the planned high-luminosity LHC [85], the pileup will increase significantly, which likely leads to a degraded $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution. This would motivate a DNN-based $p_{\mathrm{T}}^{\mathrm{miss}}$ correction, customized to the phase space considered in an analysis.

Appendix 1: CMS Data Sets

The CMS data sets were recorded in the periods 2016 preVFP (era B-F), 2016 postVFP (era F-H), 2017 and 2018. Complementary to dilepton PDs, single lepton PDs are used. To avoid overlap, events from single lepton PDs are only accepted, if they are not included in the dilepton PDs. The paths of all used data samples are summarized in Tables A1.1 - A1.4.

Table A1.1: 2016 preVFP data samples used for the analysis.

PD	Path
/DoubleMuon	/Run2016B-ver2_HIPM_UL2016_MiniAODv2-v1/MINIAOD /Run2016{C,D,E,F}-HIPM_UL2016_MiniAODv2-v1/MINIAOD
/DoubleEG	/Run2016B-ver2_HIPM_UL2016_MiniAODv2-v3/MINIAOD /Run2016{C,D,E,F}-HIPM_UL2016_MiniAODv2-v1/MINIAOD
/MuonEG	/Run2016B-ver2_HIPM_UL2016_MiniAODv2-v2/MINIAOD /Run2016{C,D,E,F}-HIPM_UL2016_MiniAODv2-v2/MINIAOD
/SingleMuon	/Run2016B-ver2_HIPM_UL2016_MiniAODv2-v2/MINIAOD /Run2016{C,D,E,F}-HIPM_UL2016_MiniAODv2-v2/MINIAOD
/SingleElectron	/Run2016B-ver2_HIPM_UL2016_MiniAODv2-v2/MINIAOD /Run2016{C,D,E,F}-HIPM_UL2016_MiniAODv2-v2/MINIAOD
/MET	/Run2016B-21Feb2020_ver2_UL2016_HIPM-v1/MINIAOD /Run2016{C,D,E,F}-21Feb2020_UL2016_HIPM-v1/MINIAOD

Table A1.2: 2016 postVFP data samples used for the analysis.

PD	Path
/DoubleMuon	/Run2016{F,G}-UL2016_MiniAODv2-v1/MINIAOD /Run2016H-UL2016_MiniAODv2-v2/MINIAOD
/DoubleEG	/Run2016{F,G,H}-UL2016_MiniAODv2-v1/MINIAOD
/MuonEG	/Run2016{F,G,H}-UL2016_MiniAODv2-v2/MINIAOD
/SingleMuon	/Run2016{F,G,H}-UL2016_MiniAODv2-v2/MINIAOD
/SingleElectron	/Run2016{F,G,H}-UL2016_MiniAODv2-v2/MINIAOD
/MET	/Run2016{F,G,H}-UL2016_MiniAODv2-v2/MINIAOD

Table A1.3: 2017 data samples used for the analysis.

PD	Path
/DoubleMuon	/Run2017E-UL2017_MiniAODv2-v2/MINIAOD /Run2017{B,C,D,F}-UL2017_MiniAODv2-v1/MINIAOD
/DoubleEG	/Run2017{B,D,E}-UL2017_MiniAODv2-v1/MINIAOD /Run2017{C,F}-UL2017_MiniAODv2-v2/MINIAOD
/MuonEG	/Run2017{B,C,D,E,F}-UL2017_MiniAODv2-v1/MINIAOD
/SingleMuon	/Run2017{B,C,D,E,F}-UL2017_MiniAODv2-v1/MINIAOD
/SingleElectron	/Run2017{B,C,D,E,F}-UL2017_MiniAODv2-v1/MINIAOD
/MET	/Run2017{B,C,D,E,F}-UL2017_MiniAODv2-v1/MINIAOD

Table A1.4: 2018 data samples used for the analysis.

PD	Path
/DoubleMuon	/Run2018{A,B,C,D}-UL2018_MiniAODv2-v1/MINIAOD
/EGamma	/Run2018{A,B,C,D}-UL2018_MiniAODv2-v1/MINIAOD
/MuonEG	/Run2018{A,B,C,D}-UL2018_MiniAODv2-v1/MINIAOD
/SingleMuon	/Run2018{A,B,C,D}-UL2018_MiniAODv2-v1/MINIAOD
/MET	/Run2018A-UL2018_MiniAODv2-v2/MINIAOD /Run2018{B,C,D}-UL2018_MiniAODv2-v1/MINIAOD

Appendix 2: High-level Triggers

All triggers that are employed in this thesis are listed in Tables A2.1 - A2.3 for all data taking periods.

Table A2.1: Triggers used to collect 2016 data.

Channel	Eras	Path
e^+e^-	B-H	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
	B-H	HLT_Ele27_WPTight_Gsf_v
$\mu^+\mu^-$	B-H	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v
	B-H	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v
	B-G	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_v
	B-G	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_v
	B-H	HLT_IsoMu24_v
	B-H	HLT_IsoTkMu24_v
$e^\pm\mu^\mp$	B-H	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
	B-H	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v
	B-G	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v
	B-G	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v
	B-H	HLT_IsoMu24_v
	B-H	HLT_IsoTkMu24_v
	B-H	HLT_Ele27_WPTight_Gsf_v

Table A2.2: Triggers used to collect 2017 data.

Channel	Eras	Path
e^+e^-	B-F	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
	B-F	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v
	B-F	HLT_Ele32_WPTight_Gsf_L1DoubleEG_v
$\mu^+\mu^-$	B	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v
	C-F	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8_v
	B-F	HLT_IsoMu27_v
$e^\pm\mu^\mp$	B-F	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v
	B-F	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
	B-F	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v
	B-F	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v
	B-F	HLT_IsoMu27_v
	B-F	HLT_Ele32_WPTight_Gsf_L1DoubleEG_v

Table A2.3: Triggers used to collect 2018 data.

Channel	Eras	Path
e^+e^-	A-D	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
	A-D	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v
	A-D	HLT_Ele32_WPTight_Gsf_v
$\mu^+\mu^-$	A-D	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8_v
	A-D	HLT_IsoMu24_v
$e^\pm\mu^\mp$	A-D	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v
	A-D	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
	A-D	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v
	A-D	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v
	A-D	HLT_IsoMu24_v
	A-D	HLT_Ele32_WPTight_Gsf_v

Appendix 3: Simulated Samples

The MC samples used for modeling of the SM and BSM processes are summarized in the following. The full sample path is constructed as shown in Tab. A3.1, where for *{process}* and *{version}*, the appropriate process path and corresponding version are inserted, respectively, which are listed in Tab. A3.2.

Table A3.1: Data paths of simulated samples. For *{process}* and *{version}*, information from Tab. A3.2 has to substituted. For the complete data path, `/MINIAODSIM` has to be added after *{version}* for each period.

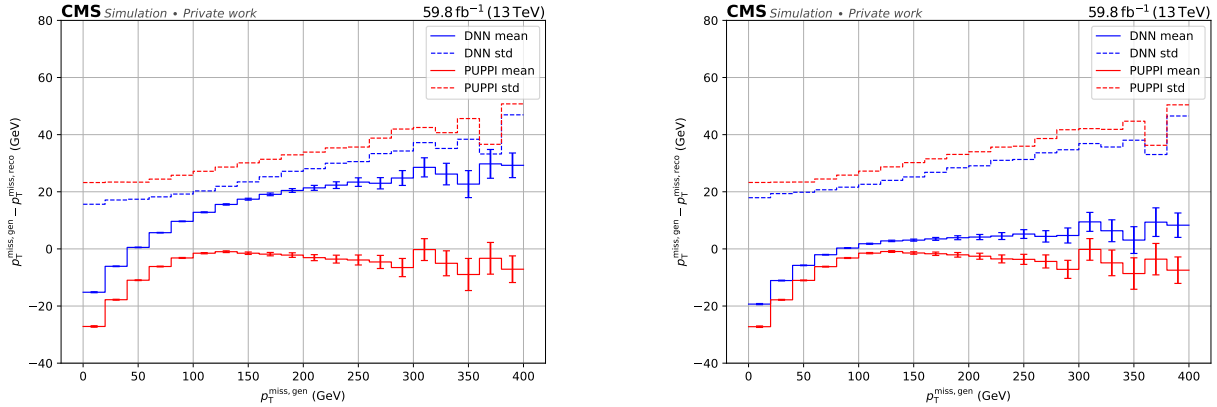
Sample Path	Period
/*{process}*/RunIISummer20UL16MiniAODAPVv2-106X_mcRun2_asymptotic_preVFP_v11-*{version}*	2016 preVFP
/*{process}*/RunIISummer20UL16MiniAODv2-106X_mcRun2_asymptotic_v17-*{version}*	2016 postVFP
/*{process}*/RunIISummer20UL17MiniAODv2-106X_mc2017_realistic_v9-*{version}*	2017
/*{process}*/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-*{version}*	2018

Table A3.2: Simulated samples used in the analysis for all data taking periods.

Process Path	Version in period				σ [pb]
	2016 preVFP	2016 postVFP	2017	2018	
/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/	v1	v1	v1	v1	89.05
/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/	v1	v1	v2	v1	377.96
/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/	v1	v1	v1	v2	366.91
/TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8/	v1	v1	v2	v2	831.76
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/	v1	v1	v2	v2	6077.22
/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/	v1	v1	v1	v1	18610
/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/	v2	v1	v1	v1	61526.7
/ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/	v1	v1	v1	v1	19.56
/ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/	v1	v1	v1	v1	19.56
/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/	v1	v1	v1	v1	10.32
/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/	v3	v3	v1	v1	136.02
/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/	v3	v3	v1	v1	80.95
/WW_TuneCP5_13TeV-pythia8/	v1	v1	v1	v1	118.7
/WZ_TuneCP5_13TeV-pythia8/	v1	v1	v1	v1	47.13
/ZZ_TuneCP5_13TeV-pythia8/	v1	v1	v1	v2	16.523
/TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8/	v1	v1	v1	v1	0.5297
/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8/	v1	v1	v1	v2	0.2529
/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/	v2	v1	v1	v1	0.2043
/TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/	v2	v1	v1	v1	0.4062

Appendix 4: Influence of Event Reweighting in Test Sample

The influence of the $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ reweighting is clearly visible in the test sample, as shown in Fig. A4.1. There are no systematic differences visible to the results in the training sample shown in Sec. 7.2, the graphs can thus be interpreted analogously.



(a) Without $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ reweighting

(b) With $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$ reweighting

Figure A4.1: Mean value (solid) and standard deviation (dashed) of the $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} - p_{\mathrm{T}}^{\mathrm{miss,\,reco}}$ distribution as a function of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}}$, shown for reconstruction with PUPPI (red) and the baseline DNN (blue) in the 2018 test sample. The last bin includes overflow (all events with $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} > 400\,\mathrm{GeV}$).

Appendix 5: Input Features

All 54 initial input features are listed in Table A5.1 with their definition.

Table A5.1: Input features of the baseline DNN with brief definitions.

Variable	Description
$p_x^{\text{miss, Puppi}}$	See Sec. 5.1.6
$p_y^{\text{miss, Puppi}}$	See Sec. 5.1.6
$p_x^{\text{miss, PF}}$	See Sec. 5.1.6
$p_y^{\text{miss, PF}}$	See Sec. 5.1.6
$p_x^{\text{all } j}$	See Sec. 5.1.6
$p_y^{\text{all } j}$	See Sec. 5.1.6
$m_{\ell_1,\ell_2,\text{all } j}$	See Sec. 5.1.6
$p_x^{j_1}$	See Sec. 5.1.6
M_{HT}	See Sec. 5.1.6
$p_x^{l_1}$	See Sec. 5.1.6
$p_y^{l_1}$	See Sec. 5.1.6
$p_y^{j_1}$	See Sec. 5.1.6
$p_T^{\text{miss, Calo}}$	See Sec. 5.1.6
MT_2	See Sec. 5.1.6
m_{jj}	See Sec. 5.1.6
n_{jets}	See Sec. 5.1.6
E_{j_1}	See Sec. 5.1.6
H_{T}	See Sec. 5.1.6
$p_x^{j_2}$	See Sec. 5.1.6
$p_y^{j_2}$	See Sec. 5.1.6
$n_{\text{Interactions}}$	Number of reconstructed vertices
$p_{\text{T}}^{\text{all } j}$	Vectorial p_{T} sum over all jets
$\sigma(p_{\text{T}}^{\text{miss, Puppi}})$	Uncertainty on $p_{\text{T}}^{\text{miss, PF}}$ as given by the PF algorithm
$p_y^{l_2}$	See Sec. 5.1.6
$\Delta\phi(p_{\text{T}}^{\text{miss, Puppi}}, \text{nearest } j)$	Distance in ϕ between $p_{\text{T}}^{\text{miss, Puppi}}$ and closest jet
$m_{\ell\ell}$	Invariant mass of both leptons
$p_{\text{T}}^{\ell_1,\ell_2,\text{all } j}$	Vectorial p_{T} sum over all jets and leptons
$\Delta\phi(p_{\text{T}}^{\text{miss, Puppi}}, j_1)$	Distance in ϕ between $p_{\text{T}}^{\text{miss, Puppi}}$ and leading jet
E_{j_2}	Energy of the subleading jet
$\Delta\phi(p_{\text{T}}^{\text{miss, PF}}, j_1)$	Distance in ϕ between $p_{\text{T}}^{\text{miss, PF}}$ and leading jet
$\Delta\phi(p_{\text{T}}^{\text{miss, Puppi}}, \text{leading b jet})$	Distance in ϕ between $p_{\text{T}}^{\text{miss, Puppi}}$ and leading b-tagged jet
$\Delta\phi(p_{\text{T}}^{\text{miss, PF}}, \text{leading b jet})$	Distance in ϕ between $p_{\text{T}}^{\text{miss, PF}}$ and leading b-tagged jet
$p_x^{l_2}$	See Sec. 5.1.6
M_T	Transverse mass of leading lepton and $p_{\text{T}}^{\text{miss}}$ (see Sec. 5.1.6)
E_{ℓ_1}	Energy of leading lepton
$\Delta\phi(\ell_1, j_1)$	Distance in ϕ between leading jet and lepton
E_{ℓ_2}	Energy of subleading lepton
$\Delta\phi(p_{\text{T}}^{\text{miss, PF}}, j_2)$	Distance in ϕ between $p_{\text{T}}^{\text{miss, PF}}$ and subleading jet
$\Delta\phi(\ell_1, \ell_2)$	Distance in ϕ between leading and subleading lepton
η_{j_1}	Pseudorapidity η of leading jet
$\Delta\phi(p_{\text{T}}^{\text{miss, PF}}, \text{furthest } j)$	Distance in ϕ between $p_{\text{T}}^{\text{miss, PF}}$ and jet furthest from it
η_{ℓ_2}	Pseudorapidity η of subleading lepton
$\Delta\phi(p_{\text{T}}^{\text{miss, Puppi}}, \text{furthest } j)$	Distance in ϕ between $p_{\text{T}}^{\text{miss, Puppi}}$ and jet furthest from it
$\Delta\phi(\ell_1, \text{leading b jet})$	Distance in ϕ between leading lepton and leading b jet
$\Delta\phi(p_{\text{T}}^{\text{miss, PF}}, \text{nearest } j)$	Distance in ϕ between $p_{\text{T}}^{\text{miss, PF}}$ and nearest jet
$\Delta\phi(j_1, j_2)$	Distance in ϕ between leading and subleading jet
$\text{flavor}(\ell_2)$	Flavor of subleading lepton, value 1 (2) for e (mu)
$\text{flavor}(\ell_1)$	Flavor of leading lepton, value 1 (2) for e (mu)
$\Delta\phi(p_{\text{T}}^{\text{miss, Puppi}}, j_2)$	Distance in ϕ between $p_{\text{T}}^{\text{miss, Puppi}}$ and subleading jet
η_{ℓ_1}	Pseudorapidity η of leading lepton
η_{j_2}	Pseudorapidity η of subleading jet
$\sigma(p_{\text{T}}^{\text{miss, PF}})$	Uncertainty of $p_{\text{T}}^{\text{miss, PF}}$ as given by PF algorithm
loose ℓ veto	Value of 0 or 1 if the lepton veto would trigger for loose lepton ID
$p_{\text{T}}^{\ell_1,\ell_2}/p_{\text{T}}^{\text{all } j}$	Ratio between combined transverse momenta of leptons and jets

Appendix 6: Nuisance Parameters

The adaption of the nuisance parameters in the maximum likelihood fit using COMBINE are shown exemplary for one variable combination in Fig. A6.1. The so-called asimov data-set is used, which describes the data as the sum of the simulated signal and background processes, so no recorded data is used in this fit, and a signal strength of $r = 1$ is expected. Each nuisance parameter θ (denoted as ξ in Sec. 7.3.2) corresponds to one distribution of systematic uncertainties (see Sec. 5.3). During the fitting procedure of combine, the MC distribution is shifted by a fraction of this distribution. For instance, the nuisance parameter associated to the top mass (MTOP) is pulled by a fraction of its shape, the impact of this nuisance pull on the resulting signal strength is illustrated on the right in Fig. A6.1. This is performed with all nuisance parameters, only the first 30 parameters, sorted by their impact, are shown exemplary.
The results seem reasonable, as expected the signal strength is exactly 1 through use of the asimov data set. The nuisance parameters with the largest impacts are similar to the expectations. The signal strength can be constrained considerably by the luminosity (`lumi`). Nuisances associated to momentum and energy of the top quark are impactful aswell, similar to nuisances regarding the object identification (`MUON_ISO`, `ELECTRON_ID`, `BTAGBC`) and regarding the cross sections of the dominant background processes (`DY_xsec`, `ST_xsec`, `TTother_xsec`).

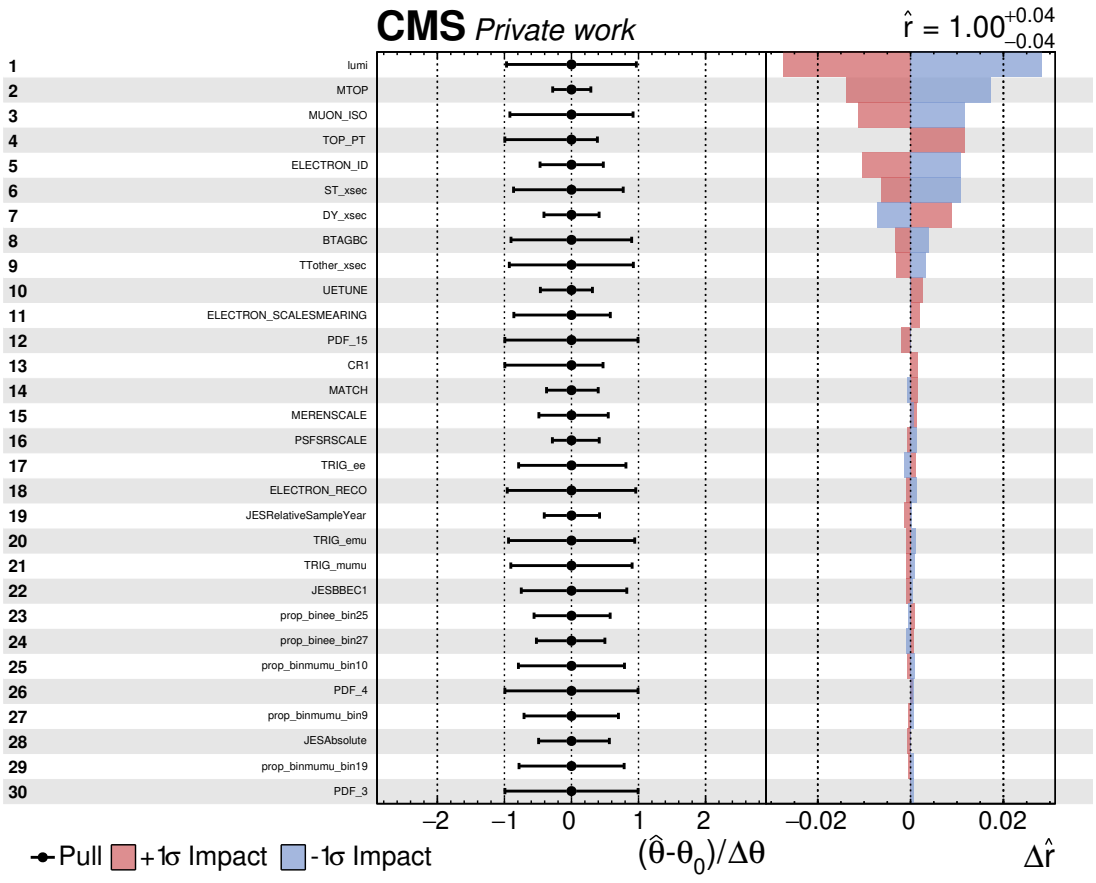


Figure A6.1: Nuisance pulls in maximum likelihood fit of the asimov sample for input feature combination $p_x^{\text{miss, Puppi}}$ vs $p_x^{\ell_1}$ in period 2018. Nuisance parameters are denoted as θ, the signal strength is denoted as r.

Appendix 7: Binnings of Final Input Features

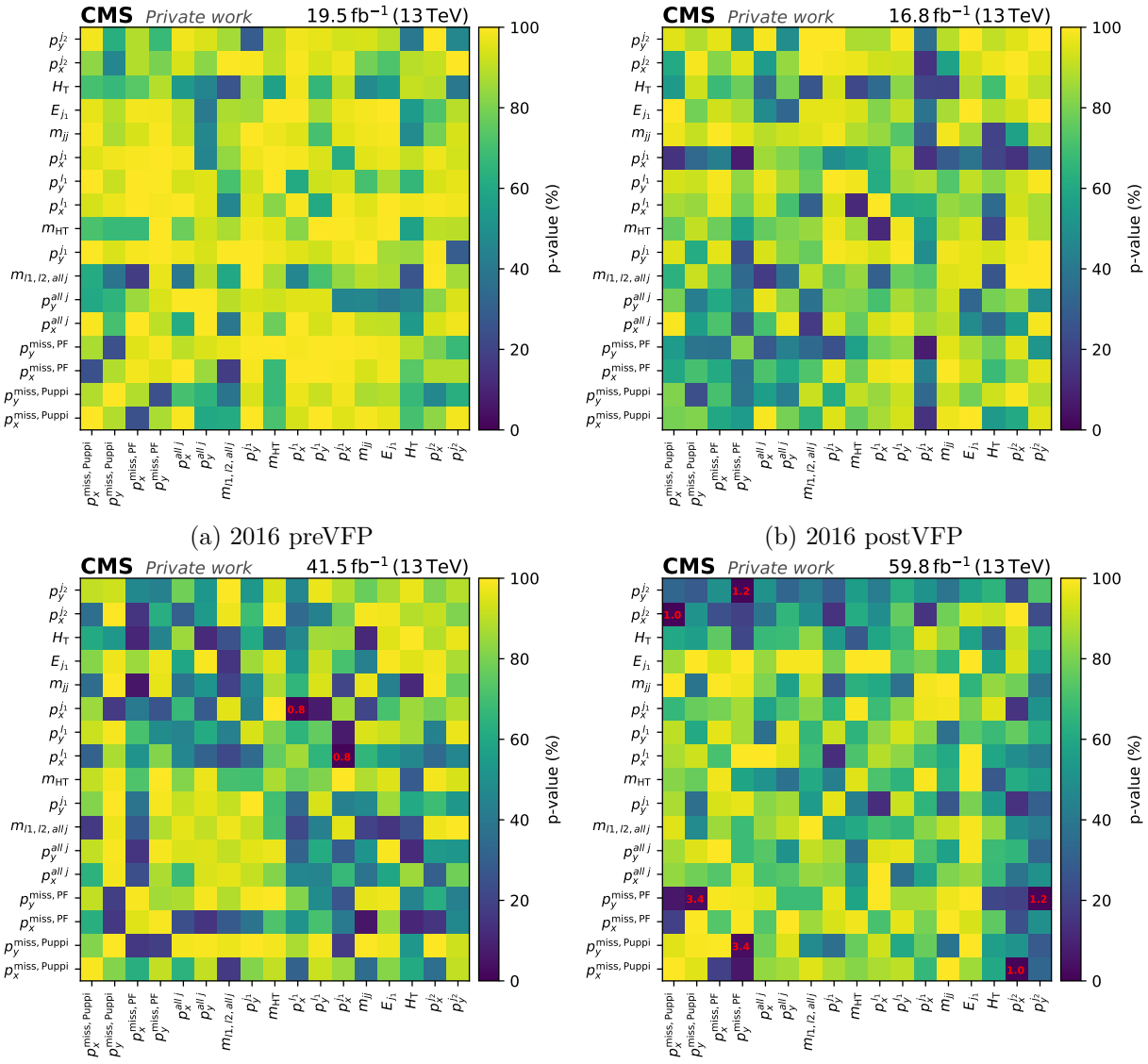
The 1D and 2D binnings of all 20 input features tested within the GOF determination (see Sec. 7.3) are summarized in A7.1.

Variable	1D binning (GeV)	2D Binning (GeV)
$p_x^{\text{miss, Puppi}}$	$(-150, 150, 30)$	$(-150, 150, 6)$
$p_y^{\text{miss, Puppi}}$	$(-150, 150, 30)$	$(-150, 150, 6)$
$p_x^{\text{miss, PF}}$	$(-150, 150, 30)$	$(-150, 150, 6)$
$p_y^{\text{miss, PF}}$	$(-150, 150, 30)$	$(-150, 150, 6)$
$p_x^{\text{all}\,j}$	$(-150, 150, 30)$	$(-150, 150, 6)$
$p_y^{\text{all}\,j}$	$(-150, 150, 30)$	$(-150, 150, 6)$
$m_{\ell_1,\ell_2,\text{all}\,j}$	$(120, 1800, 28)$	$(0, 1800, 6)$
$p_y^{j_1}$	$(-400, 400, 40)$	$(-400, 400, 5)$
m_{HT}	$(60, 1800, 29)$	$(0, 1800, 6)$
$p_x^{\ell_1}$	$(-250, 250, 25)$	$(-250, 250, 5)$
$p_y^{\ell_1}$	$(-250, 250, 25)$	$(-250, 250, 5)$
$p_x^{j_1}$	$(-400, 400, 40)$	$(-400, 400, 5)$
m_{jj}	$(60, 1800, 29)$	$(0, 1800, 6)$
E_{j_1}	$(40, 500, 23)$	$(0, 500, 5)$
H_{T}	$(50, 1200, 23)$	$(0, 1200, 6)$
$p_x^{j_2}$	$(-250, 250, 25)$	$(-250, 250, 5)$
$p_y^{j_2}$	$(-250, 250, 25)$	$(-250, 250, 5)$
$p_{\text{T}}^{\text{miss, Calo}}$	$(0, 200, 40)$	$(0, 200, 5)$
M_{T2}	$(0, 160, 40)$	$(0, 160, 5)$
$n_{\text{jets}} \cdot \text{GeV}$	$(2, 8, 7)$	$(2, 8, 7)$

Table A7.1: Binnings of the DNN input features used in the GOF determination (see Sec. 7.3). The binnings are denoted in the form $(\textit{start}, \textit{stop}, \textit{number of bins})$, where the first and last bin include overflow.

Appendix 8: Final Set of Input Features

The p-values for the final set of input features and their pairwise combinations are shown in Fig. A8.1 for all periods of data taking. As described in Sec. 7.3, only few outliers have p-values below 5%.



(a) 2016 preVFP

(b) 2016 postVFP

(c) 2017

(d) 2018

Figure A8.1: Illustration of p-values of the final set of input features in all periods of data taking. Elements on the diagonal are the results of the 1D distributions. All off diagonal elements show the results for the 2D distributions. Each bin with p-value below 5% is marked by the corresponding p-value in red font.

Appendix 9: Performance in 2016 and 2017

The resolution of DNN 1 is shown in Fig. A9.1 for periods 2016 preVFP, 2016 postVFP and 2017. The distribution are overall very similar to the 2018 results presented in Sec 8.2, and can be interpreted equivalently. Overall, the DNN improves the $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction for all periods of data taking.

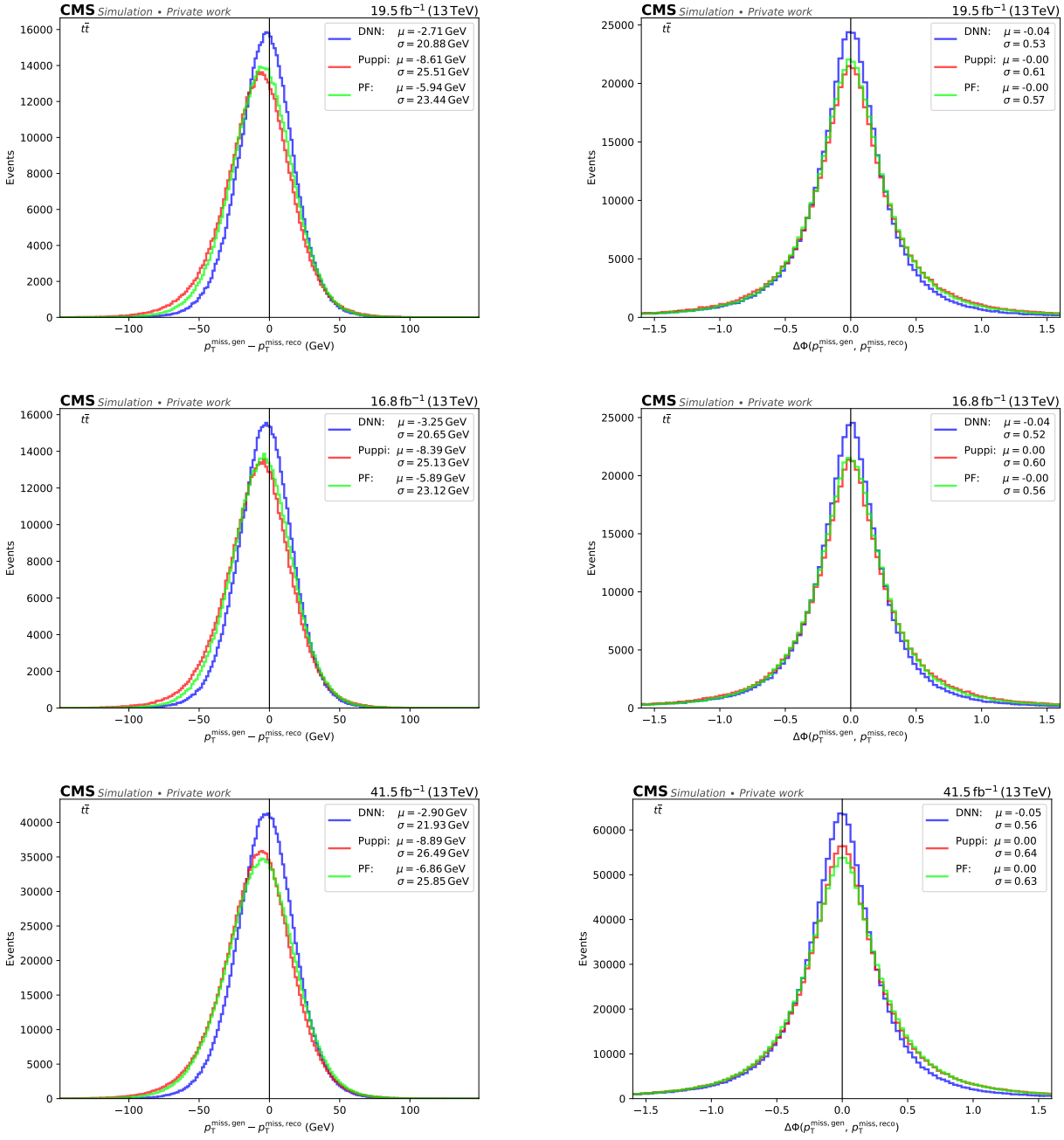


Figure A9.1: Distribution of $p_{\mathrm{T}}^{\mathrm{miss,\,gen}} - p_{\mathrm{T}}^{\mathrm{miss,\,reco}}$ and $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss,\,gen}}, \vec{p}_{\mathrm{T}}^{\mathrm{miss,\,reco}})$ for DNN 1, PUPPI and PF reconstruction in periods 2016 postVFP (top), 2016 postVFP (middle) and 2017 (bottom). Mean μ and standard deviation σ of the distributions are displayed in the legend.

Appendix 10: Resolution for Background Events

The resolution of DNN 1 is compared to the resolution of PUPPI and PF reconstruction in Fig. A10.1 for background processes Drell-Yan and $t\bar{t}$ decay with tauonic and semileptonic final states, similarly to the illustrations in Sec. 8.3.

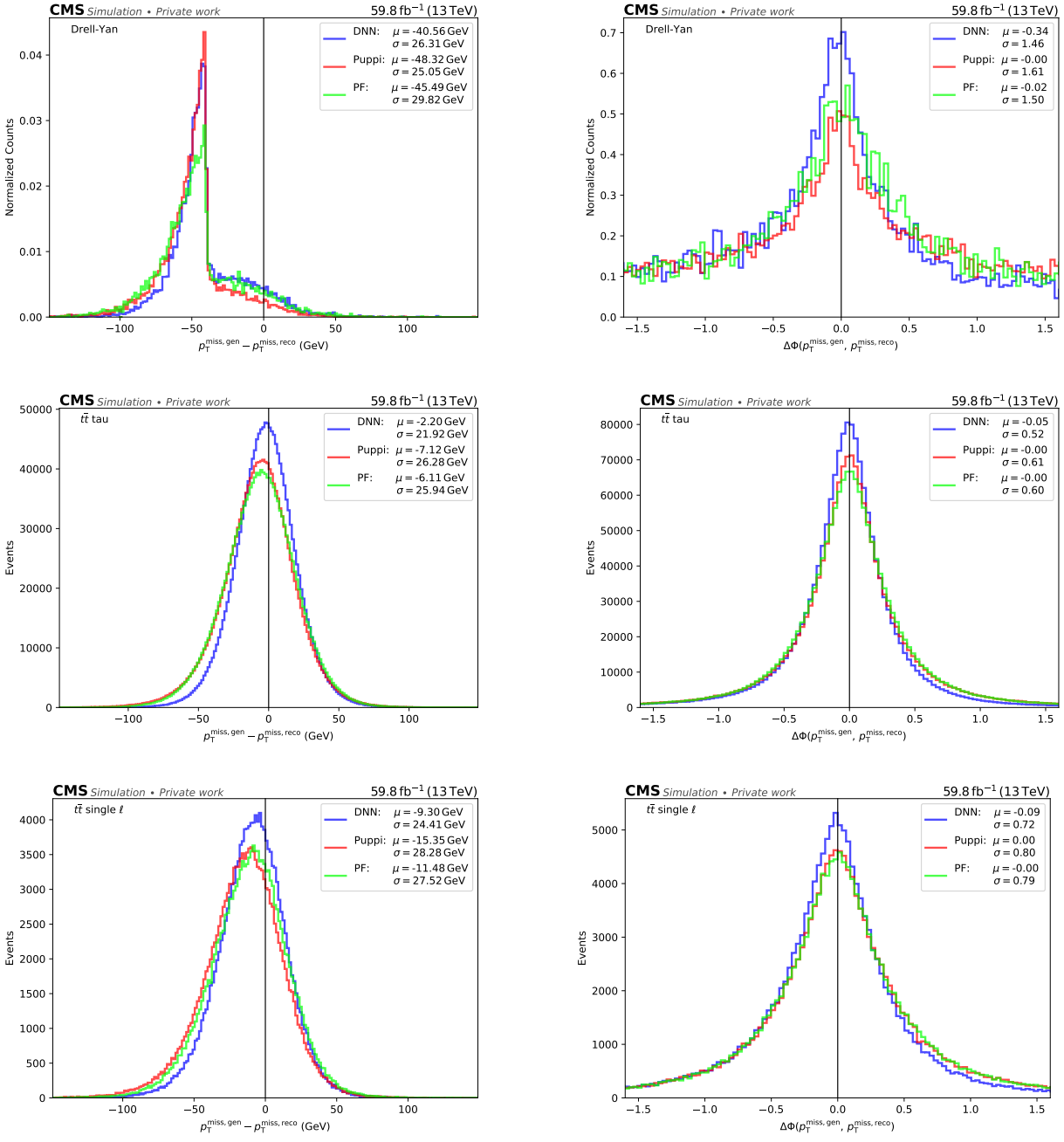


Figure A10.1: Distribution of $p_T^{\text{miss, gen}} - p_T^{\text{miss, reco}}$ and $\Delta\phi(\vec{p}_T^{\text{miss, gen}}, \vec{p}_T^{\text{miss, reco}})$ for DNN 1, PUPPI and PF reconstruction for a sample of the Drell-Yan process (top), dileptonic $t\bar{t}$ decay into at least one tau (middle) and semileptonic $t\bar{t}$ decay (bottom). Mean μ and standard deviation σ of the distributions are displayed in the legend.

Since the contribution of eμ events is very small in the Drell-Yan sample, the $p_{\mathrm{T}}^{\mathrm{miss,\ Puppi}}$ requirement of 40 GeV for ee and $\mu\mu$ events in the event selection (see Sec. 5.2) has a much more direct influence, leading to an edge in the distribution at around -40 GeV. This effect is amplified by the fact that due to its kinematics, the Drell-Yan process generally includes small amounts of $p_{\mathrm{T}}^{\mathrm{miss,\ gen}}$. For comparability between the different reconstructions, only for the performance illustration of the Drell-Yan sample, additionally to the $p_{\mathrm{T}}^{\mathrm{miss,\ Puppi}}$ cut, the same requirement of 40 GeV is applied on $p_{\mathrm{T}}^{\mathrm{miss,\ DNN}}$ and $p_{\mathrm{T}}^{\mathrm{miss,\ PF}}$.The DNN shows smaller standard deviations and, in case of the $p_{\mathrm{T}}^{\mathrm{miss,\ gen}} - p_{\mathrm{T}}^{\mathrm{miss,\ reco}}$ distribution, less bias. In the $\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss,\ gen}}, p_{\mathrm{T}}^{\mathrm{miss,\ reco}})$ distribution, the bias of the DNN distribution is increased.
The resolution in the $\mathrm{t\bar{t}}$ samples with semileptonic and tauonic final states is similar to the performance of the signal process and single t process. The resolution is overall increased significantly, the bias is reduced for $p_{\mathrm{T}}^{\mathrm{miss,\ gen}} - p_{\mathrm{T}}^{\mathrm{miss,\ reco}}$, while the bias in the $\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss,\ gen}}, p_{\mathrm{T}}^{\mathrm{miss,\ reco}})$ distribution increases slightly.
Overall, the performance for the DNN is still similar or better than for PUPPI and PF reconstruction, and receives no significant bias through training only in a dileptonic $\mathrm{t\bar{t}}$ sample.

Appendix 11: Response matrices

The response matrices (see 8.4 for definition) are shown for reconstruction with PUPPI, PF and DNN 1 in Fig A11.1. The response of DNN 1 is closer to the ideal case of a diagonal matrix, the event migration is thus overall reduced compared to reconstruction with PUPPI and PF.

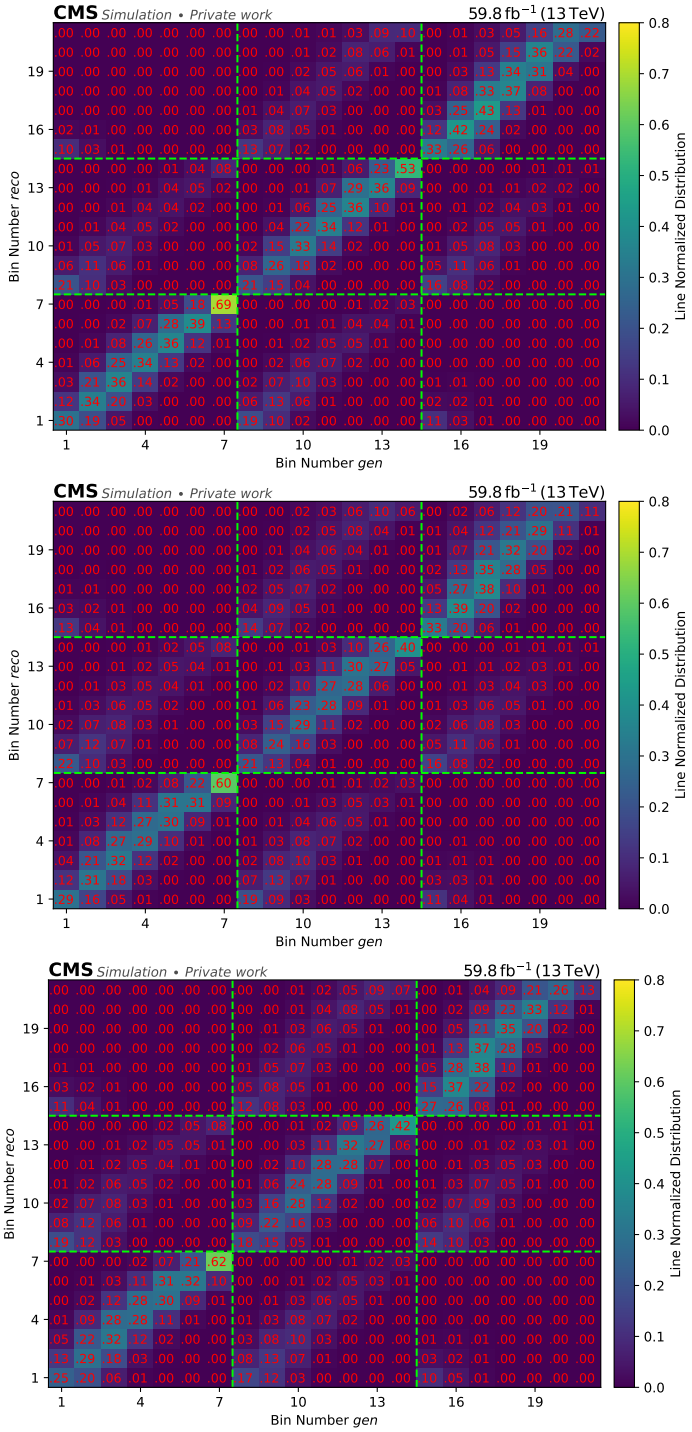


Figure A11.1: Line-normalized response matrices for DNN (top), PUPPI (middle) and PF (bottom) reconstructions in 2018 simulation. The binning is shown in Tab. 2.1, the dashed green lines illustrate bins of same $|\Delta\phi(p_{\mathrm{T}}^{\nu\nu},\ \text{nearest}\ \ell)|$.

References

[1] W. Hollik, "Quantum field theory and the Standard Model", in *High-energy physics. Proceedings, 17th European School, ESHEP 2009, Bautzen, Germany*. 2010. arXiv:1012.3883.

[2] S. Novaes, "Standard model: An Introduction", in *10th Jorge Andre Swieca Summer School: Particle and Fields*. 1, 1999. arXiv:hep-ph/0001283.

[3] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008), doi:10.1088/1748-0221/3/08/S08001.

[4] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012), doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[5] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012), doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[6] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008), doi:10.1088/1748-0221/3/08/S08003.

[7] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008), doi:10.1088/1748-0221/3/08/S08004.

[8] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS", *Eur. Phys. J. C* **81** (2021), doi:10.1140/epjc/s10052-021-09538-2, arXiv:2104.01927.

[9] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV", , CERN, Geneva (2018).

[10] CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV", , CERN, Geneva (2019).

[11] Wikipedia contributors, "Standard Model — Wikipedia, The Free Encyclopedia". https://en.wikipedia.org/wiki/Standard_Model. Accessed 07.09.2022.

[12] J. Greensite, "The Confinement problem in lattice gauge theory", *Prog. Part. Nucl. Phys.* **51** (2003), doi:10.1016/S0146-6410(03)90012-3, arXiv:hep-lat/0301023.

[13] M. Persic, P. Salucci, and F. Stel, "The universal rotation curve of spiral galaxies I. The dark matter connection", *Monthly Notices of the Royal Astronomical Society* **281** (1996), doi:10.1093/mnras/278.1.27, arXiv:https://arxiv.org/abs/astro-ph/9506004.

[14] D. Clowe et al., "A Direct Empirical Proof of the Existence of Dark Matter", *The Astrophysical Journal* **648** (2006), doi:10.1086/508162.

[15] G. Bertone, D. Hooper, and J. Silk, "Particle dark matter: Evidence, candidates and constraints", *Phys. Rept.* **405** (2005), doi:10.1016/j.physrep.2004.08.031, arXiv:hep-ph/0404175.

[16] Planck Collaboration, "Planck 2015 results - I. Overview of products and scientific results", *A&A* **594** (2016), doi:10.1051/0004-6361/201527101.

[17] ATLAS Collaboration and CMS Collaboration, "Combined Measurement of the Higgs Boson Mass in *pp* Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments", *Phys. Rev. Lett.* **114** (May, 2015), doi:10.1103/PhysRevLett.114.191803.

[18] M. Gonzalez-Garcia and Y. Nir, "Neutrino Masses and Mixing: Evidence and Implications", *Rev. Mod. Phys.* **75** (2003), doi:10.1103/RevModPhys.75.345, arXiv:hep-ph/0202058.

[19] The Super-Kamiokande Collaboration, "Evidence for an Oscillatory Signature in Atmospheric Neutrino Oscillations", *Phys. Rev. Lett.* **93** (Sep, 2004), doi:10.1103/PhysRevLett.93.101801.

[20] Super-Kamiokande Collaboration, "Evidence for the Appearance of Atmospheric Tau Neutrinos in Super-Kamiokande", *Phys. Rev. Lett.* **110** (May, 2013), doi:10.1103/PhysRevLett.110.181802.

[21] S. P. Martin, "A Supersymmetry primer", *Adv. Ser. Direct. High Energy Phys. 21* (2010), doi:10.1142/9789814307505_0001, arXiv:hep-ph/9709356.

[22] R. Barbieri and G. F. Giudice, "Upper Bounds on Supersymmetric Particle Masses", *Nucl. Phys. B* **306** (1988), doi:10.1016/0550-3213(88)90171-X.

[23] R. Lysák, "Charge Asymmetry in Top Quark Pair Production", *Symmetry* **12, 1278** (August, 2020).

[24] V. Khachatryan et al., "Search for direct pair production of supersymmetric top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 8$ TeV", *The European Physical Journal C* **76** (08, 2016), doi:10.1140/epjc/s10052-016-4292-5.

[25] W. Venus, "A LEP summary", *PoS* **HEP2001** (2001), doi:10.22323/1.007.0284.

[26] Wikipedia contributors, "Large Hadron Collider — Wikipedia, The Free Encyclopedia". https://en.wikipedia.org/w/index.php?title=Large_Hadron_Collider. Accessed 24.09.2022.

[27] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008), doi:10.1088/1748-0221/3/08/S08002.

[28] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008), doi:10.1088/1748-0221/3/08/S08005.

[29] CMS Collaboration, "People statistics". https://cms.cern/collaboration/people-statistics. Accessed 20.09.2022.

[30] CMS Tracker Group of the CMS Collaboration, "The CMS Phase-1 Pixel Detector Upgrade", *JINST* **16** (Dec, 2020), doi:10.1088/1748-0221/16/02/P02027, arXiv:2012.14304.

[31] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017), doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[32] CMS Collaboration, "The CMS trigger in Run 2", *PoS* **EPS-HEP2017** (2017), doi:10.22323/1.314.0523.

[33] M. Jansova, "Search for the supersymmetric partner of the top quark and measurements of cluster properties in the silicon strip tracker of the CMS experiment at Run 2. ",. Presented 27 Sep 2018.

[34] CMS Collaboration, "Public CMS Luminosity Information - Multi-year plots". https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/228. Accessed 07.09.2022.

[35] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006), doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[36] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003), doi:10.1016/S0168-9002(03)01368-8.

[37] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007), doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[38] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010), doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[39] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014), doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[40] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $\mathcal{O}(\alpha_S^4)$", *Physical Review Letters* **110** (2013), doi:10.1103/physrevlett.110.252004.

[41] M. Czakon and A. Mitov, "NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels", *Journal of High Energy Physics* **2012** (dec, 2012), doi:10.1007/jhep12(2012)054.

[42] M. A. et al., "Top quark pair differential cross sections in the dilepton channel at 13 TeV", *CMS Analysis Note AN-2015/309 (unpublished)* (2016).

[43] CMS Collaboration, "Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV", *Physical Review D* **95** (2017), doi:10.1103/physrevd.95.092001.

[44] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017), doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[45] CMS Collaboration, "Baseline muon selections for Run-II". https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2. Accessed 14.07.2022.

[46] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC", *Journal of Physics G: Nuclear and Particle Physics* **31** (2005).

[47] CMS Collaboration, "Cut Based Electron ID for Run 2". https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2. Accessed 14.07.2022.

[48] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", *Phys. Rev. D* **48** (Oct, 1993), doi:10.1103/PhysRevD.48.3160.

[49] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008), doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[50] CMS Collaboration, "Recommended Jet Energy Corrections and Uncertainties For Data and MC". https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECDataMC. Accessed 14.07.2022.

[51] CMS Collaboration, "Jet Identification for the 13 TeV UL data". https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVUL. Accessed 14.07.2022.

[52] E. Bols et al., "Jet Flavour Classification Using DeepJet", *JINST* **15** (2020), doi:10.1088/1748-0221/15/12/P12012, arXiv:2008.10519.

[53] CMS collaboration, "Missing transverse energy performance of the CMS detector", *Journal of Instrumentation* **6** (sep, 2011), doi:10.1088/1748-0221/6/09/p09001.

[54] CMS collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8TeV$", *Journal of Instrumentation* **10** (feb, 2015), doi:10.1088/1748-0221/10/02/p02006.

[55] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup per particle identification", *Journal of High Energy Physics* **2014** (oct, 2014), doi:10.1007/jhep10(2014)059.

[56] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ using the CMS detector", *Journal of Instrumentation* **14** (jul, 2019), doi:10.1088/1748-0221/14/07/p07004.

[57] C. Lester and D. Summers, "Measuring masses of semi-invisibly decaying particle pairs produced at hadron colliders", *Physics Letters B* **463** (sep, 1999), doi:10.1016/s0370-2693(99)00945-4.

[58] V. Botta, L. Feld, D. Meuser, P. Nattland, M. Teroerde, "Measurement of the dileptonic ttbar differential cross section in a BSM phase space", *CMS Analysis note (unpublished)* **AN-22-038** (2022).

[59] CMS Collaboration, "Jet energy scale uncertainty sources". https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources. Accessed 08.09.2022.

[60] CMS Collaboration, "JER Scaling factors and Uncertainty for 13 TeV". https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution#JER_Scaling_factors_and_Uncertai. Accessed 08.09.2022.

[61] Luminosity Physics Object Group, "Year-to-year combinations and correlations". https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM#LumiComb. Accessed 08.09.2022.

[62] CMS Collaboration, "HEM15/16 issue uncertainties". https://hypernews.cern.ch/HyperNews/CMS/get/JetMET/2000.html. Accessed 08.09.2022.

[63] CMS Collaboration, "Utilities for Accessing Pileup Information for Data - Recommended cross section". https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData#Recommended_cross_section. Accessed 08.09.2022.

[64] Particle Data Group Collaboration, "Review of Particle Physics", *Prog. Theor. Exp. Phys.* **083C01** (2020), doi:10.1088/1674-1137/40/10/100001.

[65] CMS Collaboration, "CMS PYTHIA 8 colour reconnection tunes based on underlying-event data", 2022. doi:10.48550/ARXIV.2205.02905.

[66] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning". MIT Press, 2016. http://www.deeplearningbook.org.

[67] N. Ketkar, "Deep Learning with Python". Apress Berkeley, CA, 2017. doi:https://doi.org/10.1007/978-1-4842-2766-4.

[68] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature* **521** (May, 2015), doi:10.1038/nature14539.

[69] N. Ketkar, "Stochastic Gradient Descent", pp. 113–132. Apress, Berkeley, CA, 2017. doi:10.1007/978-1-4842-2766-4_8.

[70] J. Brownlee, "Machine Learning Mastery". Independently published, 2021.

[71] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems", 2015. Software available from tensorflow.org.

[72] S. Ruder, "An overview of gradient descent optimization algorithms", doi:10.48550/ARXIV.1609.04747.

[73] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks", *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)* (2011).

[74] B. Shahriari et al., "Taking the Human Out of the Loop: A Review of Bayesian Optimization", *Proceedings of the IEEE* **104** (2016), doi:10.1109/JPROC.2015.2494218.

[75] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian Optimization of Machine Learning Algorithms", 2012. doi:10.48550/ARXIV.1206.2944.

[76] E. Brochu, V. M. Cora, and N. de Freitas, "A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning", 2010. doi:10.48550/ARXIV.1012.2599.

[77] F. Nogueira, "Bayesian Optimization: Open source constrained global optimization tool for Python", 2014.

[78] E. Winter, "Chapter 53 The shapley value", in *Handbook of Game Theory with Economic Applications*, volume 3, pp. 2025–2054. Elsevier, 2002.

[79] A. Roth, "The Shapley Value: Essays in Honor of Lloyd S. Shapley". Cambridge University Press, 1988.

[80] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions", in *Advances in Neural Information Processing Systems 30*, I. Guyon et al., eds., pp. 4765–4774. Curran Associates, Inc., 2017.

[81] J. S. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra", 2011. doi:10.48550/ARXIV.1103.0354.

[82] ATLAS Collaboration, CMS Collaboration, LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011", , CERN, Geneva (2011).

[83] R. D. Cousins, "Generalization of Chisquare Goodness-ofFit Test for Binned Data Using Saturated Models , with Application to Histograms". https://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf, 2013.

[84] L. Demortier, "P values and nuisance parameters", in *Statistical issues for LHC physics. Proceedings, Workshop, PHYSTAT-LHC, Geneva, Switzerland, June 27-29, 2007*, p. 23. 2008. doi:10.5170/CERN-2008-001.

[85] I. Zurbano Fernandez et al., "High-Luminosity Large Hadron Collider (HL-LHC): Technical design report", doi:10.23731/CYRM-2020-0010.