ETP-KA/2017-25

Karlsruher Institut für Technologie

Multivariate Analysemethoden zur Vorhersage der fehlenden transversalen Energie am CMS Experiment

Zur Erlangung des akademischen Grades eines

MASTER OF SCIENCE

bei der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)
eingereichte

MASTERARBEIT

von

B.Sc. Nicola Zäh

Tag der Abgabe: 1. September 2017

Referent: Prof. Dr. Günter Quast
Korreferent: Priv.-Doz. Dr. Roger Wolf

Erklärung der selbständigen Anfertigung der Masterarbeit

Hiermit erkläre ich, dass ich die vorliegende Arbeit mit dem Titel

»*Multivariate Analysemethoden zur Vorhersage der fehlenden transversalen Energie am CMS Experiment*«

selbständig und unter ausschließlicher Verwendung der angegebenen Hilfsmittel angefertigt habe.

Nicola Zäh
Karlsruhe, den 1. September 2017

ETP-KA/2017-25

Karlsruhe Institute of Technology

Multivariate analysing techniques for the prediction of the missing transverse energy at the CMS experiment

For obtaining the academic degree

MASTER OF SCIENCE

at the Department of Physics
of the Karlsruhe Institute of Technology (KIT)

MASTER'S THESIS

of

B.Sc. Nicola Zäh

Handed in: 1st September 2017

Reviewer: Prof. Dr. Günter Quast
Second Reviewer: Priv.-Doz. Dr. Roger Wolf

Contents

CHAPTER 1

Introduction

Initiated by Alan Turing's formulation of the "Turing test" in 1950, scientists were determined to develop an artificial intelligence being able to autonomously learn from information or data by utilizing computing resources. With the invention of the world wide web in the 1980s at the European Organization for Nuclear Research (CERN) in Geneva, data was commonly shared and increasingly gathered to be harnessed by data scientist around the world. Today's computing resources have reached a level at which storing and processing of data is feasible, resulting in many applications of so-called multivariate analysis (MVA) techniques in every part of human life.
Also at CERN, the starting point of this worldwide spread of information, MVA methods are increasingly used for physics analysis. With its Large Hadron Collider (LHC), over 100 Petabytes of data have been recorded over the last 20 years, portraying the ideal basis for the application of an MVA algorithm. The LHC is a ring collider for proton-proton collisions, currently operated at a center of mass energy of 13 TeV. The Compact Muon Solenoid (CMS) represents one of four detectors at which these collisions are measured in order to gain deeper knowledge about the fundamental principles of particle physics and the search for yet unknown particles.
The CMS detector can however not detect every kind of particles. Particles such as neutrinos do not interact with any detector component and are therefore not directly measurable, though they are of importance for many physics analyses. An important quantity, to gain information about these non-measurable particles, is the missing transverse energy ($\not{E}_T$). It relies on the basic physical principle of momentum conservation. Beam particles show negligible initial momentum in the plane perpendicular to the beam axis. Hence, the transverse momenta of the particles produced by a collision have to be cancelled out. A non-vanishing sum of the measured transverse momenta can therefore be used to gain information about undetected particles, which are missing in this momentum addition.
The accuracy of missing transverse energy ($\not{E}_T$) is determined by detector resolution limits on reconstructing individual particles. By using an MVA technique, utilizing all information available about the structure and characteristics of an event, it is possible to improve the resolution of $\not{E}_T$ in order to get more precise information about non-measurable particles.

Boosted Decision Trees (BDTs) are commonly used for classification in the field of particle physics. A similar technique, gradient boosted regression trees (GBRTs), are used in the context of this thesis for the $\not\!\!E_T$ calculation.
An overview about the experimental setup of CMS and its particle reconstruction is given in chapter 2. The MVA techniques are described in chapter 3, followed by an introduction to $\not\!\!E_T$ in chapter 4. The MVA approach applied to $\not\!\!E_T$ is introduced in chapter 5 with a validation on events in the absence and presence of non-measurable particles. Its impact on a typical analysis as well as the performance on real data is also shown.

CHAPTER 2

Particle Physics at the Compact Muon Solenoid

2.1. The Standard Model of Particle Physics

Three of the four fundamental forces are described in the theory of the Standard Model of Particle Physics. An elaboration about the standard model can be found in [1–3]. This theory represents a re-normalisable quantum field in which excitations are interpreted as particles. These particles can be distinguished in bosons and fermions. Bosons as mediators of the fundamental forces have integer spin whereas fermions are characterized with half-integer spin and represent massive particles.
The Lagrange formalism and Lagrangian density $\mathcal{L}_{\mathrm{SM}}$ are used to describe field dynamics of this theory and thus describe the dynamics of all particles in the Standard Model. Similar to classical mechanics, problems can be solved by using generalized coordinates, simplifying problems by using intrinsic symmetries.
Objects in this formalism are described by Lorentz vectors. These four-vectors are invariant under relativistic transformation (Lorentz-transformation). This is important as particle physics at high energies implies relativistic effects.
Even though the Standard Model does not incorporate e.g. the full theory of gravitation or neutrino oscillations, it has proven successful at predicting the outcome of modern particle physics experiments.

2.2. Higgs Boson - Production and Decay

The Higgs mechanism was formulated independently in 1964 by P. Higgs [4] and F. Englert and R. Brout [5]. It explains the non-vanishing masses of the gauge vector bosons $W^{\pm}$ and Z by electroweak symmetry breaking. The mechanism also introduces an additional boson, called the Higgs boson. This boson was discovered at the LHC in 2012 [6, 7]. An elaboration about the Higgs mechanism can be found in [8, 9].
The Higgs boson has four main production modes: gluon-gluon-fusion ($gg \rightarrow H$), vector boson fusion ($qq \rightarrow H$), the associated production with a Z or $W^{\pm}$ boson and the associated top-pair production ($t\bar{t}H$) mode. The gluon-gluon-fusion is the dominant production mode for the Higgs boson with a cross section of 50 pb at 13 TeV at the LHC.

The Higgs boson is produced by the interaction of two gluons via a ferminoic loop. The Higgs boson has many decay channels, one of these being the leptonic $H \to \tau\tau$ decay, studied in this thesis. With a branching ratio of 22.5 %, the $H \to \tau\tau \to \mu\tau_h$ final state plays an important role for validating the Higgs mass, as the muon is well measured and enhances the Higgs signal sensitivity.

2.3. The Compact Muon Solenoid Experiment

The CMS experiment is part of the LHC at CERN in Geneva, Switzerland. The CMS experiment is one of four detectors located around the LHC. With its pre-accelerators, the LHC is currently the largest synchrotron with two proton beams colliding at a center of mass energy of 13 TeV and a $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ instantaneous luminosity.

2.3.1. Coordinate system

The CMS coordinate system is defined as a right-handed system centered at the nominal interaction point. The x-axis points to the centre of the LHC ring. The y-axis points upwards, perpendicular to the ring collider plane. Hence, the transverse plane is described by the x-y plane of the CMS coordinate system. In spherical coordinates, the azimuthal angle ϕ is measured with respect to the x-axis in the x-y or transverse plane. The polar angle θ, defined with respect to the z-axis, is often replaced by the pseudo-rapidity

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{2.1}$$

The metric

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}, \tag{2.2}$$

is used to describe angular distances between two objects in the detector.

2.3.2. Detector elements

Figure 2.1 and 2.2 show a transverse and longitudinal view of the CMS detector system, respectively. The individual components from the inside to the outside are:

- **Tracker system**: This innermost detector part consists of silicon semi-conductor pixel detectors, followed by strip detectors. This set-up allows the reconstruction of tracks with up to 10 μm accuracy [10]. The transverse momentum of charged particles can also be calculated from the curvature of the associated tracks in the 4 T magnetic field with an accuracy of a few percent [11, 12].

- **Electromagnetic calorimeter**: This part of the detector is made of lead tungstate crystals, serving as both absorber and scintillator. Electrons and photons deposit most of their energy in this part of the detector.

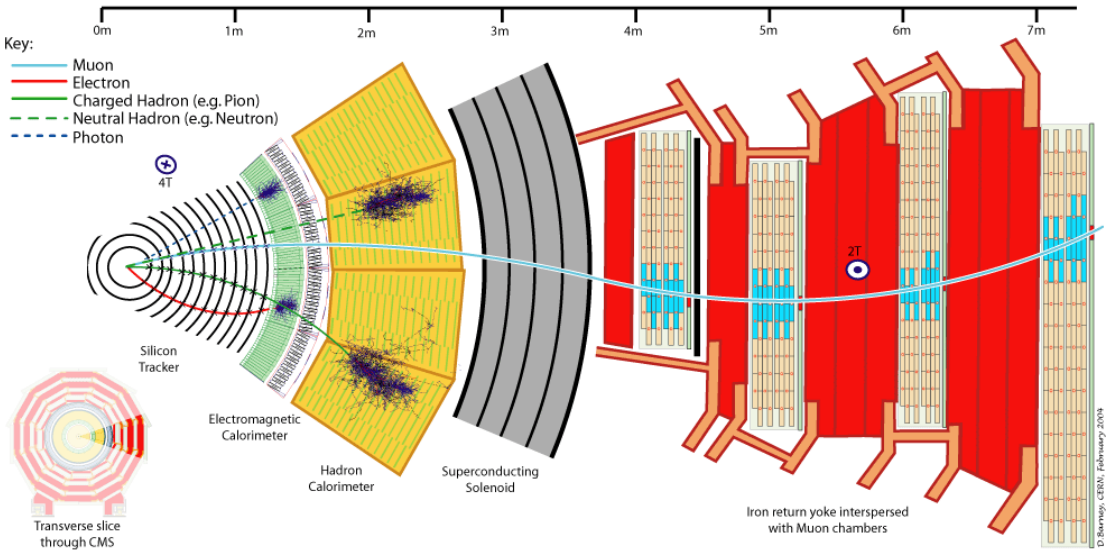


Figure 2.1.: Transverse view of the CMS detector. The technical components and typical particle tracks are shown. Figure is taken from [13].

- **Hadronic calorimeter**: Hadrons passing the electromagnetic calorimeter are measured in this detector sub-system. It consists of brass absorbers and plastic scintillators [10].
- **Superconducting solenoid**: The 4 T magnetic field inside the detector is created by this solenoid. It is surrounded by a steel return-yoke, returning the magnetic flux outside the magnet.
- **Muon chambers**: This outermost part of the detector system is built for the measurement of muons, the only measurable particles that escape the detector.

2.3.3. Particle reconstruction

Particles at CMS are reconstructed using the particle-flow (PF) approach [15]. PF candidates are reconstructed by combining signals from different sub-detector systems. Tracks are reconstructed in an iterative process, depending on the quality of each reconstructed PF candidate. Starting from tracks fulfilling tight requirements, the respective signals are removed for the following iteration with loosened requirements. This process ensures a high reconstruction efficiency in particle reconstruction. Stable particles deposit their energy in several calorimeter cells which are summarized as clusters. These Particles follow different reconstruction criteria:

- Muon reconstruction: Muons are the best-measured particles at CMS, as they interact with all detector sub-systems. Reconstruction algorithms match tracks in

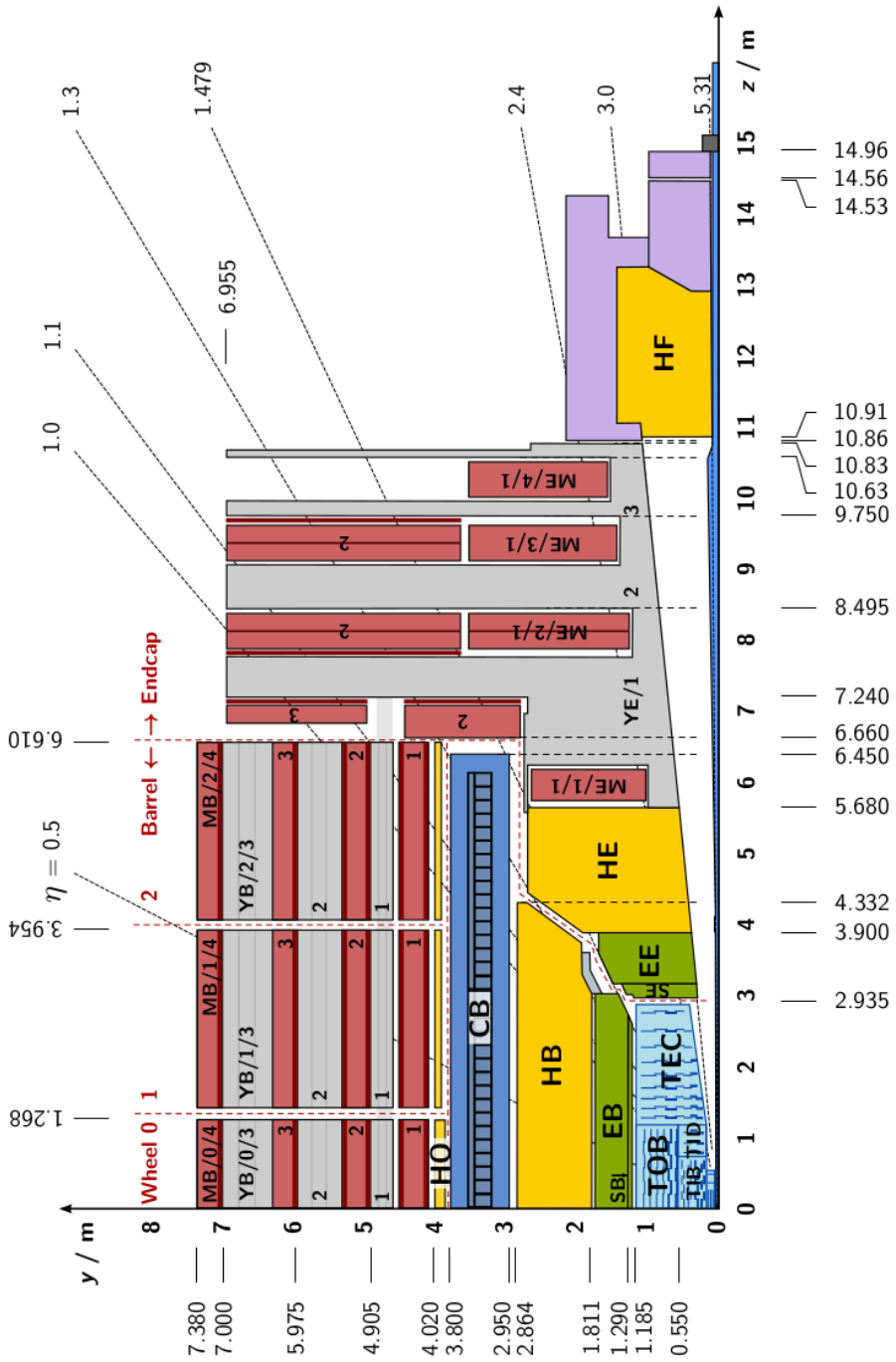


Figure 2.2.: Longitudinal view of the CMS detector. Spatial dimensions of the technical components as a function of y, z and η: inner tracking system (light blue), electromagnetic calorimeter (green), hadronic calorimeter (yellow), superconducting solenoid (blue), muon chambers (red), iron return yoke (gray). Figure is taken from [14].

the muon chambers and in the tracker to reach a muon reconstruction efficiency rate of 99,5 % with a resolution around a few percent [12].

- Electron and Photon reconstruction: Both of the particles deposit most of their energy in the electromagnetic calorimeter. The clustered calorimeter deposits (and the reconstructed electron track) are combined to reconstruct an electron or photon object.
- Jet reconstruction: Jets from quarks and gluons show a broad energy resolution distribution and are reconstructed using the anti-k_T algorithm [16]. This algorithm uses a metric to measure the distance between adjacent candidates to decide if both candidates originate from the same particle and will consequently be clustered together as part of the same jet.
- Tau reconstruction: Due to their high mass of 1.78 GeV, Tau leptons decay into lighter particles and are therefore only measurable through their decay products, clustered as jets.

CHAPTER 3

Machine Learning

3.1. Introduction

Machine learning was described already in 1959 by Arthur Samuel as giving "computers the ability to learn without being explicitly programmed" [17]. Machine learning can thus be understood as applied statistics with an emphasis on using computers to estimate complex functions and relationships by reference to the given information or data.
A machine learning algorithm needs a task or target to be able to learn from data. An algorithm has learned a task if its performance at the task has improved with the given information. Information or data is usually represented as a collection of features, measured from an event or object. A task describes the manner in which one of these collections is processed by the algorithm. This input- or feature-vector $\vec{x}$ is defined as $\vec{x} \in \mathbb{R}^n$ with x_i representing the i-th element of this vector. Machine learning is typically used for different types of tasks:

- **Classification**: A classification task represents a function $f : \mathbb{R}^n \to \{1, .., k\}$ to be learned, assigning the input to one of k categories. The output $\hat{y} = f(\vec{x})$ can represent a probability distribution for each category. This type of task is widely used in object recognition. The pixels of an image depict the feature vector, whereas the output assigns different probabilities to different object classes. This technique is also used in particle physics, using hits in the tracker system in the transverse plane, to create a two-dimensional image of the detector. This input is used to train a machine learning algorithm to determine the probability of each track in this image to belong to a specific particle. Another popular use of classification is for distinguishing between signal and background processes. For instance in the Higgs analysis, an input vector, featuring important quantities of an event, is fed into an algorithm, whose target it is to predict if this particular event contains a Higgs boson or not [18].

- **Regression**: In contrast to classification, the trained function of this approach has to map its input on a continuous output $f : \mathbb{R}^n \to \mathbb{R}$. This approach is used for predictions on e.g. algorithmic trading or insurance claims. Despite its occasional use in particle physics, it will be applied in the context of this thesis.

- **Transcription**: Transcription transforms differently structured data into textual form. Speech recognition transforms audio data into text, whereas Google Street View transforms photographs of house numbers into written numbers [19]. A related task would be machine translation, for which the input already consists of textual form.

- **Sampling**: One can use machine learning algorithms to synthesize new data according to the structure of existing data. The task of the algorithm is to exploit the structure and form of its input data in order to sample new data with comparable properties. It is also possible to apply this in particle physics when it comes to generating additional simulated event data based on obtained data from the experiment or simulation.

- **Denoising**: In this type of task, the machine learning algorithm is given a corrupted data sample, aiming to obtain a clean input sample as an output. One example of this is "unfolding", in which one tries to reverse the effect of combined uncertainty sources being propagated to the measured corrupted sample.

All of these methods need a specific performance measurement in order to be able to evaluate and compare different algorithms. Classification or transcription tasks can be assigned an accuracy rate, measuring the fraction of how many input samples were classified or transcribed correctly. Regression tasks require a different performance metric, since its output is continuous. This so-called loss function will be introduced in the following section 3.2.
In order to guarantee an unbiased estimate for performance, it is necessary to evaluate the algorithm on a statistically independent test sample. This sample cannot be used for learning and is only used as an independent measurement of generalization capacity of the algorithm. Generalization is the ability of a trained algorithm to perform equally well on previously unknown data. This is directly associated with overfitting, which represents a scenario in which the algorithm has memorized not only the structure of a dataset but also its single events, loosing the generalization ability on unknown data.
This behaviour can be mitigated by regularization techniques, aiming to avoid the algorithm overfitting its training data. Some regularization techniques are introduced in the following section 3.3.1. For very high statistics, the overfitting effect is mitigated by the fact that the algorithm will not be able to learn dataset specific features, as these features will be statistically cancelled out.
An important distinction has to be made between supervised and unsupervised learning.

Unsupervised learning represents a technique in which the algorithm has to identify useful features from a dataset. For instance, a clustering algorithm aims to divide data in different clusters in order to learn about useful information categories in the data.

Supervised learning involves data samples which are associated with a target. The algorithm has to learn structure or features from the dataset in order to make predictions about the target value of each sample.

Very often, both of these methods are combined. A possible scenario would be an unsupervised algorithm, transforming the given input into a set of abstract variables, which are then used to make prediction on a target of a supervised learning algorithm.
In this thesis, a supervised regression machine learning algorithm is used, introduced in the following chapter 3.2.

3.2. Regression

Regression analysis is the mapping of a multivariate input $\vec{x} \in \mathbb{R}^n$ on a continuous output $y = g(\vec{x})$. A machine learning algorithm aims to find a function $f(\vec{x})$, approximating the function values y by its prediction $\hat{y} = f(\vec{x})$. The function $f(\vec{x})$ is optimized by minimizing a loss function $L(\hat{y}, y)$. A typical loss function is the squared-loss function

$$L(\hat{y}, y) = \sum_{i=1}^{N} (\hat{y} - y)^2, \tag{3.1}$$

summing over the squared difference of all N input samples. Figure 3.1 shows two possible regression approximations of a given dataset.
Another loss function which is more robust in terms of outlying target values, is the Huber loss function [20]

$$L(\hat{y}, y) = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{for } (y - \hat{y}) \leq \delta \\ \delta(|y - \hat{y}| - \delta/2) & \text{for } (y - \hat{y}) > \delta \end{cases}, \tag{3.2}$$

with δ representing a threshold beyond which outlying events only have linear effect on the loss function.

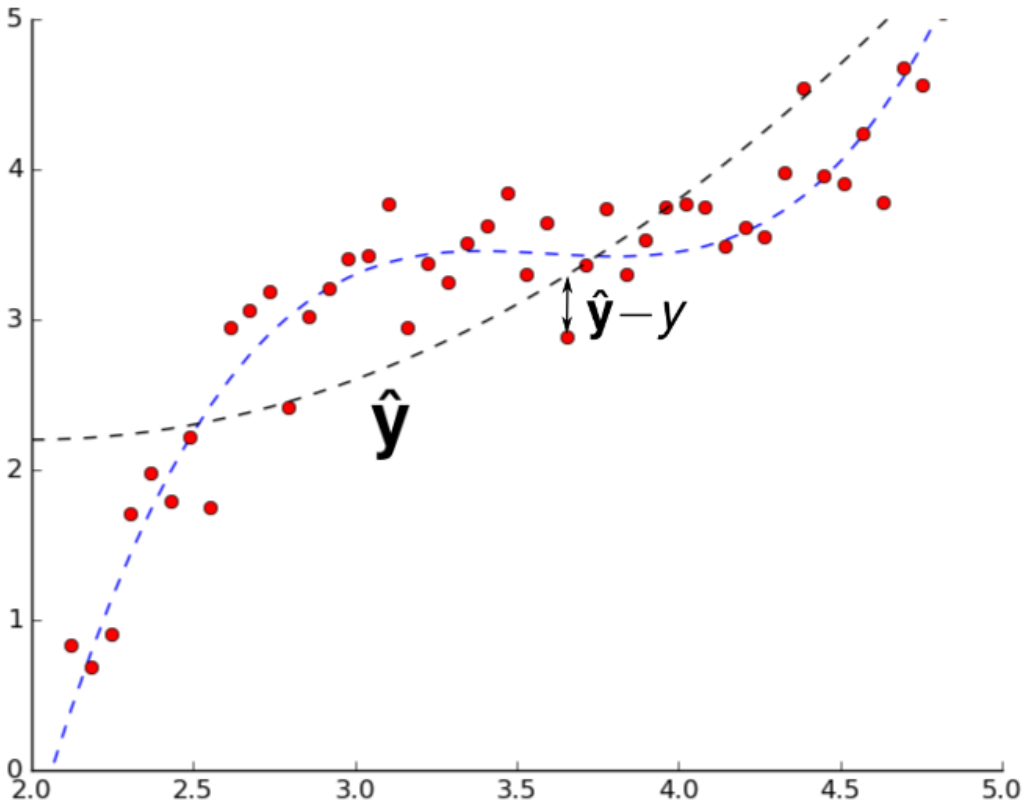


Figure 3.1.: Illustration of two different regression approximations of a given dataset. The loss is calculated as the sum of squared difference between target and prediction. The loss for the black regression approach is higher than for the blue one, as the data points y in red show a larger distance to the approximated function $\hat{y} = f(x)$ in black than to the blue approximation.

3.3. Method

Finding and optimizing a suitable function $f(\vec{x})$ for approximating the original distribution $y = g(\vec{x})$ plays an important role in machine learning. In the following, two approaches will be introduced and discussed shortly.

3.3.1. Gradient Boosted Regression Trees

Whereas boosted decision trees are used for classification in particle physics, a GBRT algorithm follows a similar approach for regression.
Decision trees classify data into different categories by applying a cut on a certain input variable. This input variable is chosen to be the variable on which a cut yields the best possible separation power regarding the target variable. The cut value is determined accordingly. However, a single cut is often not enough to sufficiently categorize data. Additional cuts on other variables are therefore applied, aiming to further discriminate on the target variable. This number of succeeding cuts is defined as the **depth** of a decision tree.
With a typical depth of three, a decision tree can only separate the data in eight different segments and therefore lacks separation power for complex datasets. This is handled by a so-called **boosting** technique. By training an additional tree on top of the predictions of the first tree, one aims to "boost" the overall performance of the first classifier. This procedure is repeated until a sufficient separation power is achieved. These additional trees are weighted with smaller relative weights, as additional trees aim to further increase the separation power of the preceding trees.
As boosted decision trees separate data into two (or more) target classes, a GBRT predicts a function $h(\vec{x})$ by dividing the target values y into different regions, based on a gradient-boosting technique. A first tree is trained, approximating the data with

$$h_1(\vec{x}) = \sum_{s=1}^{S} \mu_s I(\vec{x} \in A_s), \tag{3.3}$$

with S standing for the number of possible regions A_s, and μ_s representing the average value of all target values in that region. $I(\vec{x} \in A_s)$ is one, if $\vec{x}$ is contained in A_s and zero otherwise. See figure 3.2a for an exemplary application of two cuts on a target variable x, resulting in three different regions.
Boosting is now applied by calculating so-called pseudo-residuals

$$r_{i,s} = -\left(\frac{\partial L(\hat{y}_{i,u-1}, y_i)}{\partial \hat{y}_i}\right), \tag{3.4}$$

for each event i, based on the predictions of the previous trees, $u-1$. These pseudo-residuals are used as regression targets for the next tree h_u, aiming to improve the previous one and so forth (fig 3.2b). Starting with $f_1(\vec{x}) = h_1(\vec{x})$, trees are finally combined with

$$f_u(\vec{x}) = f_{u-1}(\vec{x}) + \nu h_u(\vec{x}), \tag{3.5}$$

with ν representing the previously mentioned weight parameter or **learning rate** of an additional tree. A high learning rate accelerates the training process, leading however to possible fluctuations if the chosen step size is too large to converge towards the global minimum. A low learning rate counters this effect at the cost of slow convergence of the algorithm.
Different criteria can be defined to stop the training of additional trees. In order to be insensitive to outliers, a minimum number of events per regions is defined, preventing overfitting. In addition, each cut has to significantly improve the effect on the target value. If either of these two criteria cannot be fulfilled, no further trees will be trained. Another stopping criterion is the maximum number of trees, in case a training did not converge within a reasonable computing time. This is, however, suboptimal as training will not yet have converged or is fluctuating around possible minima.
Decision trees are commonly used due to their robustness and the transparency of the training process. It is possible to evaluate the chosen variables and their cut values in order to obtain deeper knowledge about underlying structures in studied datasets. The actual GBRT algorithm used in this thesis was developed by Joshua Bendavid (California Institute of Technology) as part of the software framework of CMS.

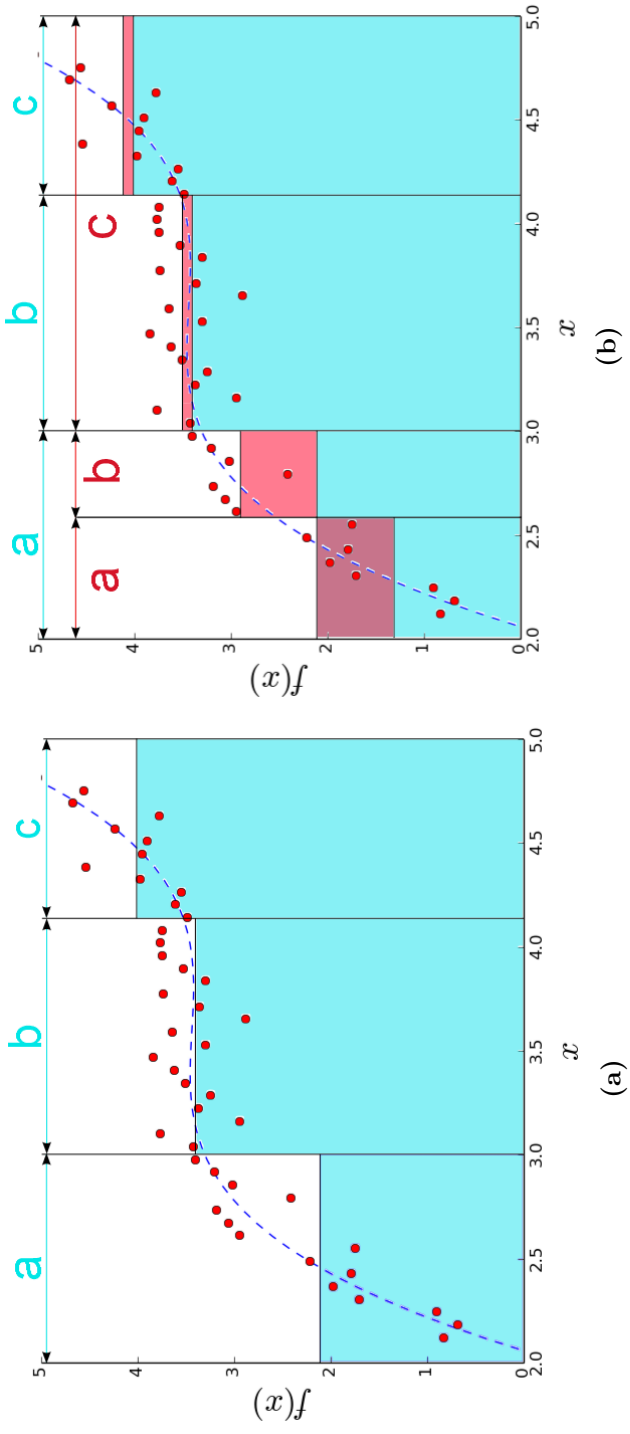


Figure 3.2.: Illustration of the first (left) and second (right) training iteration of a gradient boosted regression tree. First, the data points are divided in three different regions aiming to minimize the variance of data points within the respective area. It then predicts a constant value, based on the average function value. A second step aims to minimize the calculated pseudo-residuals, based on the difference between the first prediction and the actual function value. The data is again divided in three different regions, aiming to minimize these pseudo-residuals. Further training iterations aim to minimize the prediction of all previous trees. The choice of dividing the data into three regions is model dependent and was chosen here for illustrating purposes.

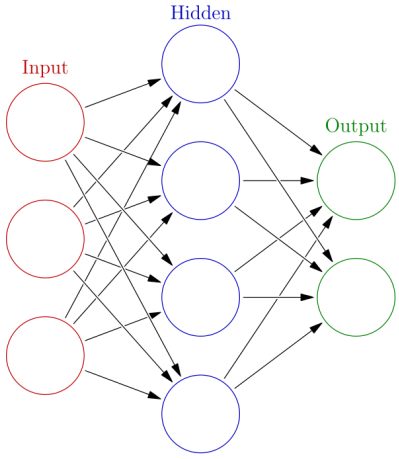


Figure 3.3.: Illustration of a typical neural networks structure with three input, four hidden and two output nodes. Figure is taken from [23].

3.3.2. Neural Networks

Neural networks are inspired by the biological functionality of the human brain. Analogous to axons in the brain, these networks are made of artificial neurons, connected like synapses. Neurons can therefore communicate with each other and transmit and process information. They take continuous states between zero and one, depending on their input and propagate their states to following neurons. Each neuron is assigned an individual weight to prioritize important information aspects. These neurons are typically organized in layers, a first layer processing the input vector $\vec{x}$, several hidden layer to generate the function $f(\vec{x})$ and a last output layer finally providing the prediction $\hat{y}$ (figure 3.3). All neurons are connected linearly, while the activation, calculating the state of each neuron, can be non-linear. These non-linear activation functions enable the modelling of any complex function, expandable by adding additional hidden layers or varying their size. Overtraining can be prevented by randomly excluding nodes in the training, similar to the brain "forgetting" some information [21]. More information about neural networks can be found in [22].
Regression is realized with several hidden layers containing non-linear activation functions, providing an estimate for the linear output neuron, which predicts $\hat{y} = f(\vec{x})$. This prediction is optimized by an appropriate loss function $L(\hat{y}, y)$. Appendix B.1 introduces a custom loss function dealing with periodic target values. Appendix B.2 evaluates the performance of a neural network compared to a GBRT, regarding the scalar correction of the MVA$\not{E}_T$, introduced in chapter 5.

CHAPTER 4

Missing Transverse Energy

4.1. Introduction

The quantity of missing transverse energy ($\not\!\!E_T$) is based on the physical principle of momentum conservation. It is defined in the transverse plane, the plane perpendicular to the beam axis. Colliding particles have no initial momentum in this transverse plan. Therefore, each transverse directed momentum of a particle originating from a collision has to be compensated by other particles from the same collision. Thus $\not\!\!E_T$ is defined as the negative sum of the transverse momenta of all reconstructed particles. This excludes particles which can not be measured by the detector e.g. neutrinos. Consequently the particles' sum of transverse momenta defines the genuine $\not\!\!E_T$. However, a non-zero $\not\!\!E_T$ does not necessarily comply with undetected particles. Detector misreconstruction effects play an important role and can originate from:

- Particle misidentification: Tracking algorithms reconstruct tracks from detector signals and assign a particle to each track. Misidentified particles can lead to wrong calculations of $\not\!\!E_T$.
- Detector resolution: The tracking algorithms are limited by the resolution of the detector. E.g. detector signals cannot always be doubtlessly assigned to a certain track or can be distorted by detector malfunctions or misalignments. This results in a reconstruction uncertainty for each particle, being propagated into an uncertainty on $\not\!\!E_T$.
- Beam misalignment: The center of the detector system is not necessarily in perfect agreement with the beam interaction point. This leads to a non-uniform distribution of the angular distribution of $\not\!\!E_T$.
- Detector coverage: A small region ($|\eta| > 5$) in the area around the beam axis is not covered by any part of the detector. Particles in this region are therefore not detectable and may imbalance the $\not\!\!E_T$ calculation.

Every reconstructed particle at CMS is denoted as a particle-flow (PF) candidate. The PF algorithm uses all information from the CMS (sub-) detectors to create a collection

of particle-flow candidates, including muons, electrons, photons, neutral and charged hadrons, pooled together as jets. The quantity $\not{E}_T$ is then defined as

$$\vec{\not{E}}_T = -\sum_{\text{PF cand.}} \vec{p}_T, \tag{4.1}$$

with $\vec{p}_T$ representing the transverse momentum of the respective PF candidate.
The $\vec{\not{E}}_T$-vector is given in polar coordinates with the magnitude $\not{E}_T$ and the respective angle in the transverse plane $\phi_{\not{E}_T}$. Another important parameter is the scalar sum of all transverse momenta, $\sum E_T$, providing an estimator of how much transverse energy is involved in the current evaluation of $\not{E}_T$.
The $\not{E}_T$ variable is calculated from all PF candidates. Wrong descriptions of these candidates lead to an incorrect description of the $\not{E}_T$. Input particles can therefore be scaled to compensate for detector resolution effects. These scale corrections are called jet energy corrections. More information regarding jet energy corrections can be found in [24].
This $\not{E}_T$, obtained by the PF candidates and corrected by the jet energy corrections, is denoted as PF$\not{E}_T$.

4.2. Additional Missing Transverse Energy definitions

Only charged particles can be directly measured by the detector via electromagnetic interactions (for further information see [1]). This allows the reconstruction algorithms to assign a track to these charged particles and thus reconstruct their origin. From this, interaction points can be reconstructed, called primary vertices. One distinguishes between the main primary vertex and pile-up vertices. The main primary vertex is defined as the vertex to which the quadratic sum of the transverse momenta of all particle tracks associated with it, is the largest. All other vertices are considered as pile-up, not being of primary interest for this event.
Neutral particles, clustered as jets, can be assigned to either the primary vertex or pile-up vertices. Neutral particles which cannot be clustered to jets are considered a third category.
Regarding charge and interaction vertex particles can be allocated to five categories: Charged particles from primary vertex (Charged PV) or pile-up (Charged PU) and neutral particles, clustered as jets from primary vertex (Neutral PV) and pile-up (Neutral PU). The neutral unclustered particles (Neutral Unclustered) form an additional category.
Based on these five categories, additional $\not{E}_T$ definitions can be formed:

- PF$\not{E}_T$: This definition takes all particles into account
- Track $\not{E}_T$: Only charged particles from the primary vertex are used for this definition.
- No PU $\not{E}_T$: This definition only accounts for particles from the primary vertex.

Table 4.1.: Overview of different types of $\not{E}_T$ definitions and the respective types of included particles.

$\not{E}_T$ **definition**	**Charged PV**	**Charged PU**	**Neutral PV**	**Neutral PU**	**Neutral Unclustered**
PF$\not{E}_T$	✓	✓	✓	✓	✓
Track $\not{E}_T$	✓				
No PU $\not{E}_T$	✓		✓		
PU Corrected $\not{E}_T$	✓		✓		✓
PU $\not{E}_T$		✓		✓	
PUPPI $\not{E}_T$	(✓)		(✓)		

- PU Corrected $\not{E}_T$: This $\not{E}_T$ definition adds unclustered neutral particles to the "No PU $\not{E}_T$".
- PU $\not{E}_T$: This definition consists of all charged and neutral clustered particles from pile-up.

An unequivocal assignment to either primary vertex or pile-up is not always possible for neutral particles. A new method was therefore proposed in 2014, called the pile-up-per-particle-identification (PUPPI) $\not{E}_T$ [25]. This method assigns a probabilistic value to each particle, giving the probability of this particle to come from pile-up. This probability is derived by using information about pile-up density of the whole event and the area around the particle and vertex information retrieved from the tracks of charged particles. Table 4.1 gives an overview over all $\not{E}_T$ definitions. All these $\not{E}_T$ definitions are to be evaluated regarding their performance on calculating the $\not{E}_T$ in the following chapter.

4.3. Performance

4.3.1. Response and Resolution

In order to be able to compare these different $\not{E}_T$ definitions, the quantities response and resolution can be defined. The $\not{E}_T$ response is defined as

$$\not{E}_T \,\text{Response} = \left\langle \frac{\text{Prediction}}{\text{Target}} \right\rangle = \left\langle \frac{\not{E}_T}{\not{E}_T{}^{True}} \right\rangle, \tag{4.2}$$

and $\not{E}_T$ resolution as

$$\not{E}_T \,\text{Resolution} = \sigma\left(\text{Prediction} - \text{Target}\right) = \sigma\left(\not{E}_T - \not{E}_T{}^{True}\right). \tag{4.3}$$

Prediction and Target are terms from machine learning and refer to the predicted or calculated value of the algorithm and the nominal or true value of the quantity in question. In this case, prediction refers to the calculated $\not{E}_T$ value, according to the specific $\not{E}_T$ definition and target is the true value of the $\not{E}_T$. Based on this equation, the question

arises what the true $\not{E}_T$ value is. To answer this, one has to distinguish between simulated events, based on theory assumptions, and events, measured by the detector. For simulated events, the true values of all quantities are known and therefore easily accessible. Data events are limited to measurable quantities of which the true value is not known.
Nevertheless it is possible to filter for well understood events in which the true value of a quantity can be estimated from theoretical models. As it comes to $\not{E}_T$, which is a sum of all reconstructed transverse momenta, one can inspect events containing only reconstructable particles, resulting in a nominal ${\not{E}_T}^{True}$ value of 0. The $\not{E}_T$ value is then an indicator for the detector performance. Well measurable and understood is the Z boson decay into two muons. Muons are the most efficiently identifiable and measurable particles at CMS[11, 12]. See figure 4.1 for schematic Z event topologies with and without real $\not{E}_T$.

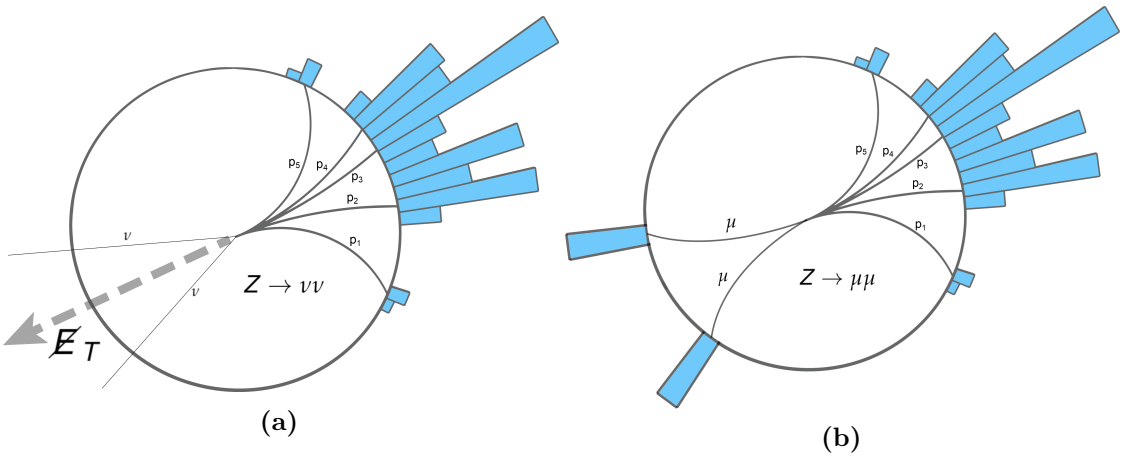


Figure 4.1.: Two schematic event topologies of a Z decay to neutrinos (${E_T}^{True} > 0$) (left) and to muons (${E_T}^{True} = 0$) (right). Note that the signal bars indicate the deposit in the hadronic calorimeters and are therefore not proportional to the transverse energy as muons escape the detector.

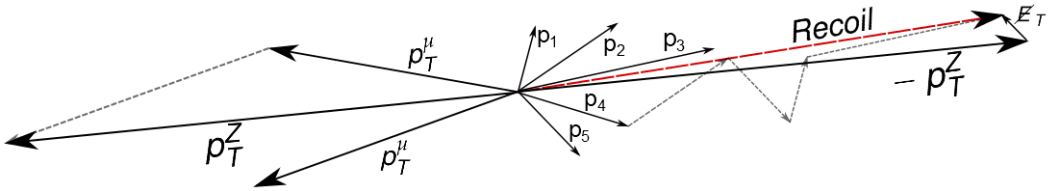


Figure 4.2.: Transverse momentum addition for the $Z \to \mu\mu$ decay illustrated in figure 4.1b. The discrepancy between calculated recoil and (negative) boson momentum lead to a non-zero $\not{E}_T$.

The main part of the incorrect $\not{E}_T$ in the $Z \to \mu\mu$ decay comes from particles carrying the recoil of the Z (p_1 to p_5 in figure 4.1). The transverse momentum of the two muons

is precisely measured and is therefore a good foundation for comparison of the $\not{E}_T$ algorithms. The recoil is defined as the sum of the transverse momenta of all reconstructed particles except for the muons. The transverse momentum of the Z boson is reconstructed as the sum of the four-momenta of the two muons.
The negative recoil is now compared to the reconstructed boson momentum. Depending on the $\not{E}_T$ definition, all pile-up particles are included in the recoil. See figure 4.2 for a schematic recoil calculation.

In order to compare recoil and the negative transverse momentum of the Z boson p_T^Z, the recoil is split into components parallel and perpendicular to p_T^Z, denoted as $U_{||}$ and $U_\perp$ (fig. 4.3). The performance of the $\not{E}_T$ definitions can now be analysed separately in both components on a given validation sample with

$$\begin{aligned} \text{Response}_{||} &= \left\langle \frac{U_{||}}{-p_T^Z} \right\rangle, \\ \text{Resolution}_{||} &= \sigma(U_{||} + p_T^Z), \\ \text{Resolution}_\perp &= \sigma(U_\perp). \end{aligned} \tag{4.4}$$

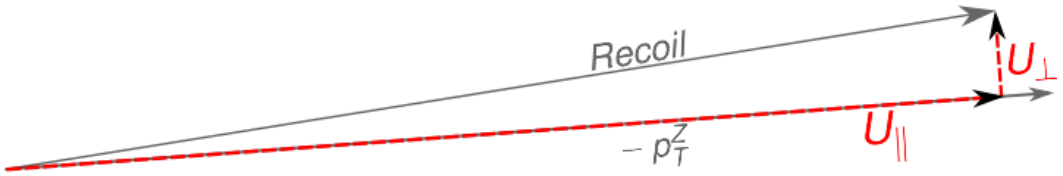


Figure 4.3.: Construction of $U_{||}$ and $U_\perp$ based on the projection of the recoil on the transverse momentum of the Z boson p_t^Z.

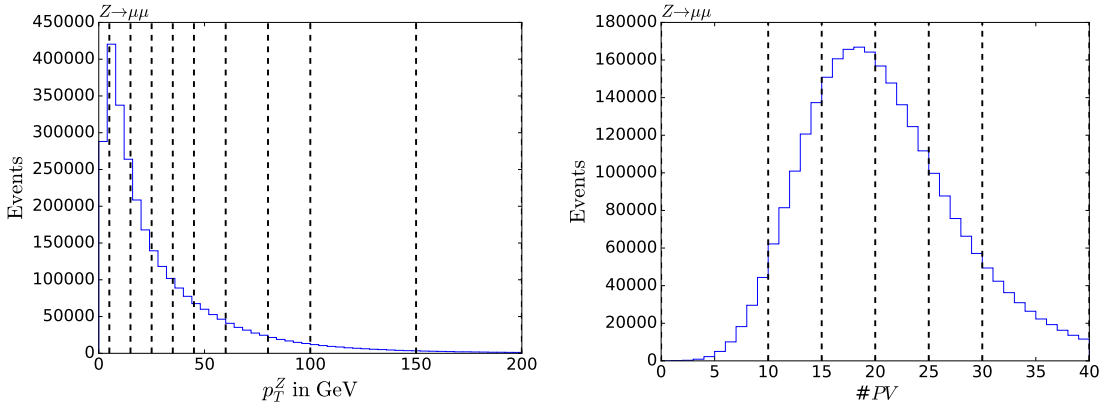
Response is a measure of how correct the $\not{E}_T$ definition is. If the expected value $\left\langle U_{||}/(-p_T^Z) \right\rangle$ is around 1, the $\not{E}_T$ algorithm is estimating the magnitude of p_T^Z correctly, averaged over the data sample. A response below 1 indicates an under-estimation of the recoil, meaning that e.g. some particles are missing to reconstruct the complete recoil of the Z boson.
Resolution is a measurement on how accurate the $\not{E}_T$ definition is. The difference between the calculated and the nominal value gives an estimate for the variance of the method.

$$\text{Response}_{\text{Corrected}} = \frac{\text{Resolution}_{||}}{\text{Response}_{||}}, \tag{4.5}$$

provides a combined measurement, accounting for both close-to-unity response and low resolution.

Both, response and resolution, behave differently for specific collision characteristics. The transverse Z boson momentum, p_T^Z, is a measure for the structure of an event. High

p_T^Z events are defined by a high transverse momentum of the boson and as a consequence high recoil values. Tracks of reconstructed particles are often better described and the response reaches one for good $\not{E}_T$ definitions.
Another quantity is the number of pile-up vertices. This quantity gives an estimate on how densely the particles are distributed. A good $\not{E}_T$ algorithm has to still perform well under high pile-up scenarios. The distributions of p_T^Z and #PV are shown in figure 4.4, motivating the regions of interest for validation techniques introduced in the following chapter. Simulated data was generated in 2016 and is based on the instantaneous luminosities at the LHCin 2016.



(a) **(b)**

Figure 4.4.: Distribution of the transverse momentum of the Z boson p_T^Z (left) and the number of pile-up vertices #PV (right). p_T^Z decreases exponentially, while the number of pile-up vertices has its maximal value around 18 vertices. The dashed black lines indicate the event partitioning in the following plots, adjusted to the event distribution of the considered quantity. Events within a partition were averaged for validation.

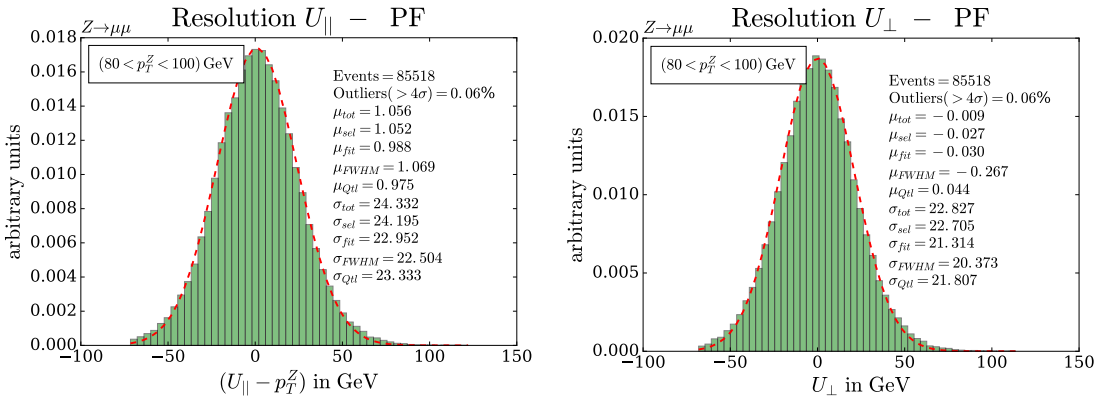
4.3.2. Validation techniques

As stated in equation 4.4, response and resolution are obtained by determining the mean and standard deviation respectively of all events assigned to a defined region as shown in figure 4.4. Under the assumption of the particles being normal distributed, one can fit a gaussian distribution function and obtain μ_{fit} and σ_{fit} (figure 4.5). As can be seen in figure 4.6a, a gaussian fit is not always the correct choice and would especially overestimate the resolution for non-gaussian distribution shapes. Another approach, tackling the standard deviation over-estimation consistently, is the full-width-half-maximum method. The relationship to the standard deviation of a gaussian distribution is given by $\sigma_{FWHM} = \frac{\text{FWHM}}{2.355}$. Nevertheless, one can see in figure 4.6b that this method struggles with sparsely populated event regions, as it relies on reasonable step sizes to determine the maximum and the interpolation function.
A universal approach is to evaluate the empiric mean μ_{tot} and standard deviation σ_{tot} of

the whole event sample. This has however drawbacks as outliers introduce biases towards shifted means and larger variances.
A possible solution is to exclude these outliers. This can be done by excluding events exceeding a certain distance to the mean, μ_{tot}, determined over the whole sample. This distance was chosen to be four standard deviations σ_{tot}, representing an interval including 99,99% of all data points for normal distributed data. The μ_{sel} and σ_{sel} of this truncated method have proven to be robust and reasonable for most scenarios. However, the truncated mean, even with excluding outlying events, is still sensitive to highly asymmetric distributions. An estimator, mostly ignoring shape distributions, is the 50 % quantile, $P(50)$, for the mean and

$$\sigma_{Qtl} = \frac{P(84,13) - P(15,87)}{2}. \tag{4.6}$$

As symmetries and shape distributions are important for the understanding of section 5.5, all plots in this section, concerning response and resolution determination, were done with the truncated method. All other plots, concerning performance comparison of $\not\!\!E_T$ definitions, were done using the quantile method.

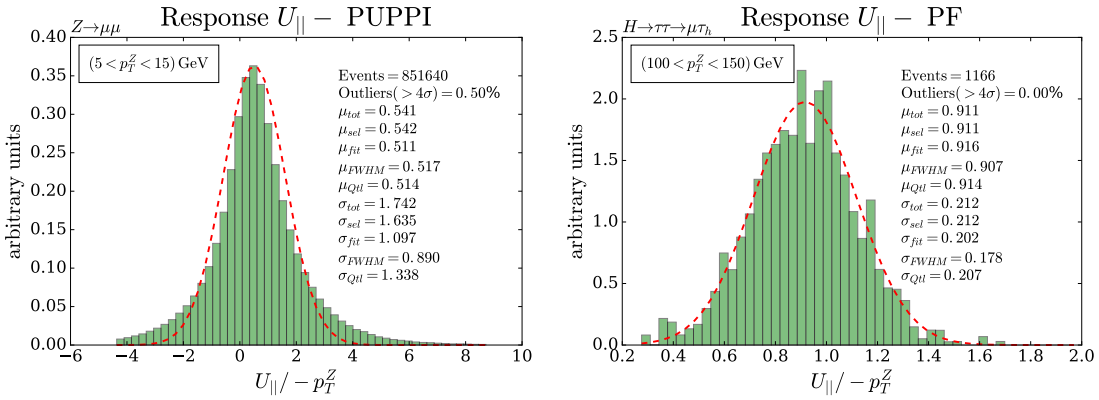


(a) **(b)**

Figure 4.5.: Distribution shape for the resolution determination for both parallel (left) and perpendicular (right) components of the recoil in the bin ($80\,\text{GeV} < p_T^Z < 100\,\text{GeV}$). Both shapes are normally distributed in this case.

4.3.3. Comparison

Figures 4.7, 4.8, 4.9 and 4.10 show the performance of the $\not\!\!E_T$ definitions defined in equation 4.4.
All left-sided plots are given as a function of the transverse momentum of the Z boson, p_T^Z, the right-sided plots as a function of the number of reconstructed primary vertices, #PV. Note that only events with a $p_T^Z > 60\,\text{GeV}$ are considered for plots with respect to #PV,



(a) **(b)**

Figure 4.6.: Distribution shape for the response determination. On the left a distribution based on very few events. This especially happens for small data samples and sparsely populated event partition. On the right a non-gaussian-like shape.

as otherwise the plots would be dominated by low-p_T^Z events. These low-p_T^Z events are poorly described, having an statistically broad influence on the response and resolution parameters, resulting in worse validity of the validation parameters. For instance, figure 4.11 shows negative response values for low p_T^Z, which means that the recoil vector points in the other direction than the boson momentum vector.
All $\not\!\!E_T$ definitions have an increasing response over p_T^Z, reaching a constant level of response for events with $p_T^Z > 60\,\text{GeV}$. Figure 4.12 shows the distribution of the fraction of the sum of transverse momenta of each $\not\!\!E_T$ definition over the sum of all transverse momenta.
The $\not\!\!E_T$ characteristics can be summarized as

- PF $\not\!\!E_T$: Together with pile-up-per-particle-identification (PUPPI)$\not\!\!E_T$, this definition has the best-to-unity response of all $\not\!\!E_T$ definitions. It has a fairly low increase in resolution with p_T^Z ($< 5\,\text{GeV}$ in the considered range of $(5 < p_T^Z < 200)\,\text{GeV}$) but a quite great dependence on #PV in resolution (up to 13 GeV) compared to Track $\not\!\!E_T$(up to 3 GeV) or Puppi $\not\!\!E_T$ (up to 6 GeV).

- Puppi $\not\!\!E_T$: With its response close to one, this $\not\!\!E_T$ definition has one of the best resolution performances. It shows low dependence on pile-up (up to 6 GeV) and is only inferior to Track $\not\!\!E_T$ in resolution as a function of p_T^Z. It also has the smallest, on average, fraction of $\sum E_T$, indicating that 90 % of the transverse momenta of an event are represented by particles coming from pile-up.

- Track $\not\!\!E_T$: Even though this $\not\!\!E_T$ definition has a quite good resolution performance for low-p_T^Z regimes and little dependence on pile-up, it shows only 50% of the expected response. This means that half of all particles representing the recoil are charged.

- PU $\not{E}_T$: As it is only considering particles from pile-up, the recoil is consistently excluded, resulting in a response of zero. It has comparably good performance on the perpendicular resolution component ($< 20\,\text{GeV}$), depending only on pile-up effects. The dependency on pile-up is consequently the greatest (up to 25 GeV) and worst for high pile-up regimes.
- No PU $\not{E}_T$: Trying to exclude particles coming from pile-up, it shows inferior performance on the perpendicular resolution part compared to other $\not{E}_T$ definitions. The differentiation between pile-up and main primary vertex is thus not as effective as for PUPPI$\not{E}_T$.
- PU Corrected $\not{E}_T$: This $\not{E}_T$ definition adds the neutral unclustered particles to the "No PU $\not{E}_T$" definitions. As the differences to the "No PU $\not{E}_T$" in resolution and response are not significant, one can conclude that unclustered neutral particles are not having a great influence on $\not{E}_T$. Nevertheless, the unclustered particles carry 50% of the $\sum E_T$, following figure 4.12. Their small effect on the $\not{E}_T$ is due to their randomly directed small transverse momenta statistically evening out for the high amount of neutral unclustered particles.

Taking all these effects into account, considering advantages and disadvantages of all $\not{E}_T$ definitions, one can conclude that most $\not{E}_T$ definitions have advantages in a specific category but suffer from performance deficiencies in others.
This is the reason why the idea arose to use a multivariate analysing method to profit from the specific benefits of each $\not{E}_T$ definition in order to create a multivariate $\not{E}_T$ definition with close to unity response and superior resolution abilities.

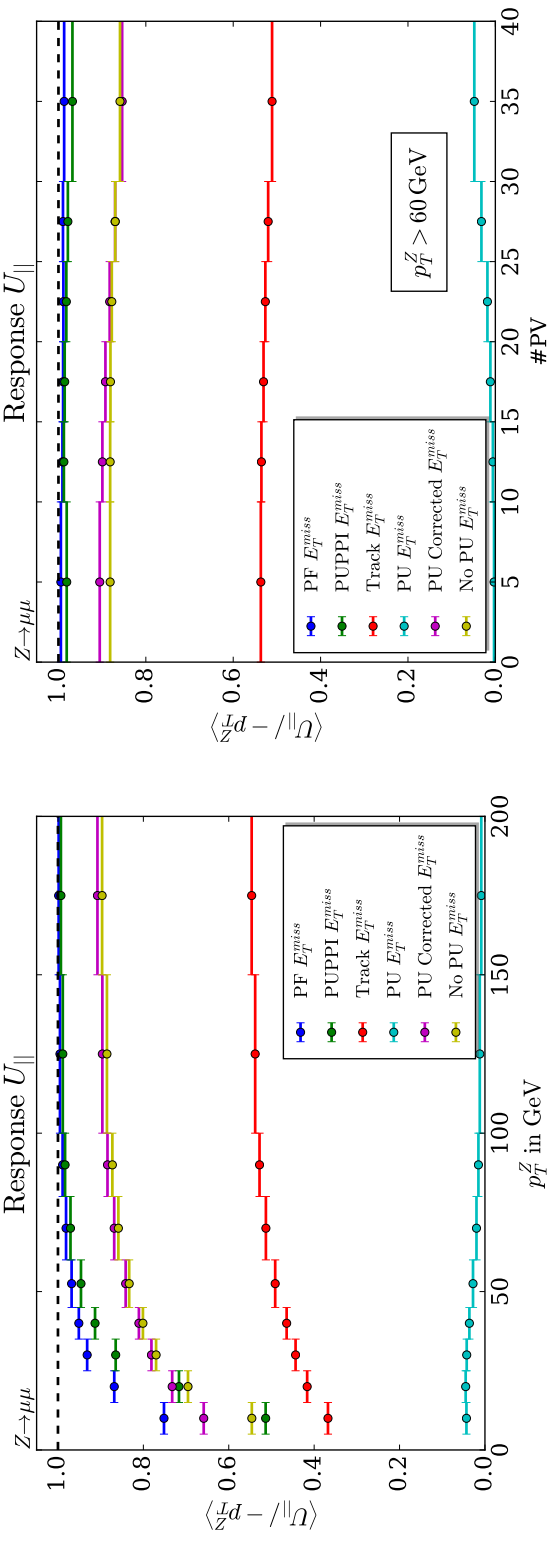


Figure 4.7.: Response plot for a simulated $Z \to \mu\mu$ data sample with roughly 6 M events. On the left side in dependence of the transverse momentum of the Z boson p_T^Z and on the right side as a function of number of pile-up vertices #PV. The y-range of all response and resolution plots in this thesis is adjusted to the region of interest, which excludes a starting point of zero for most of them.

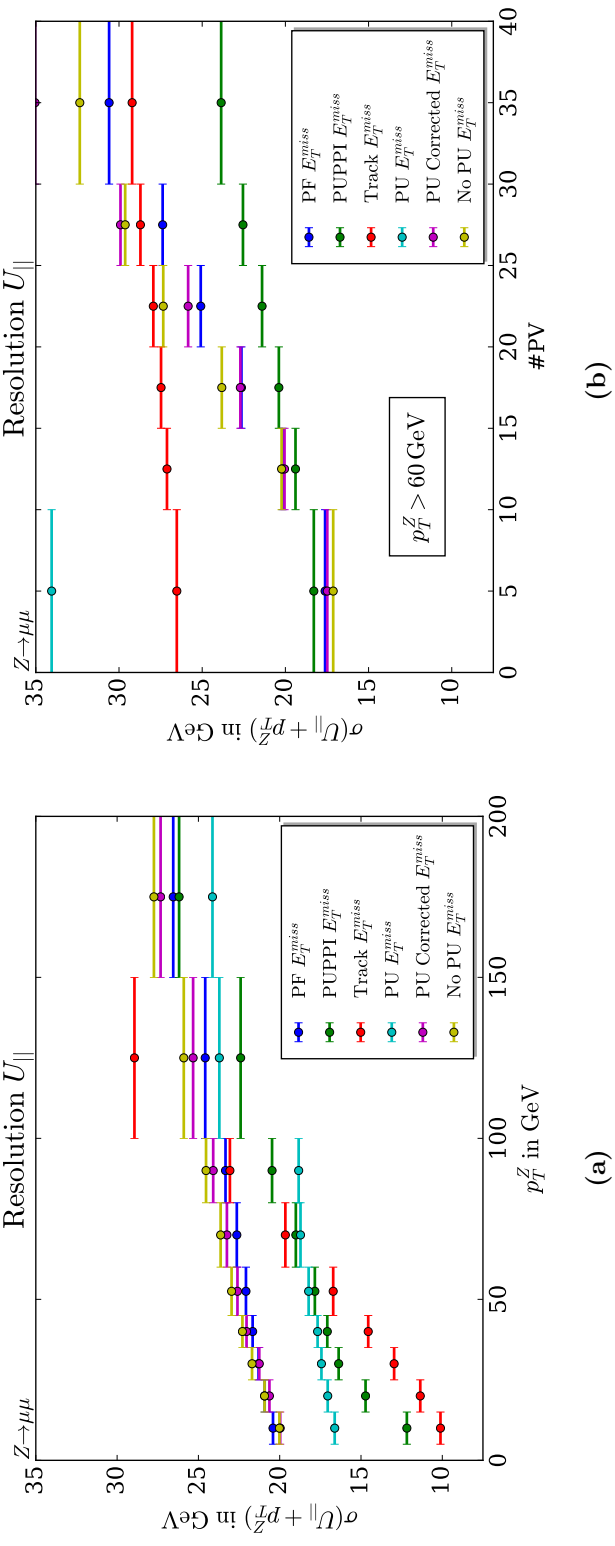


Figure 4.8.: Longitudinal resolution plot for a simulated $Z \to \mu\mu$ data sample with roughly 6 M events. On the left side in dependence of the transverse momentum of the Z boson, p_T^Z, and on the right side as a function of number of pile-up vertices #PV.

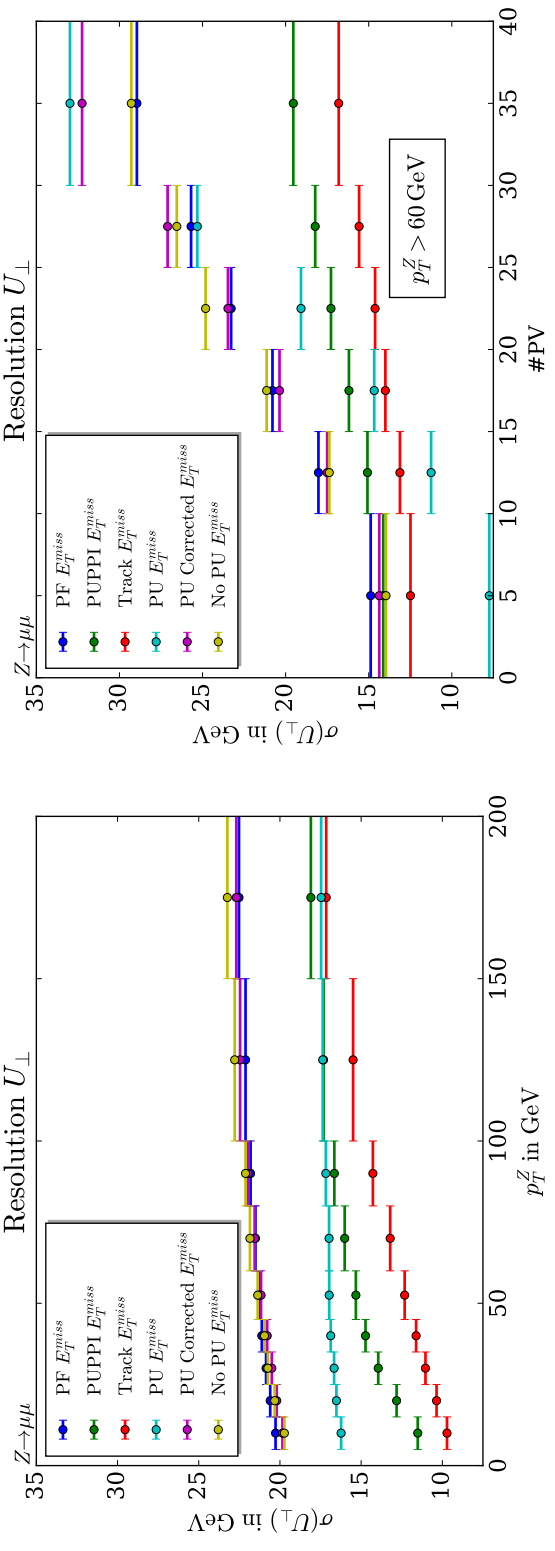


Figure 4.9.: Perpendicular resolution plot for a simulated $Z \to \mu\mu$ data sample with roughly 6 M events. On the left side in dependence of the transverse momentum of the Z boson, p_T^Z, and on the right side as a function of number of pile-up vertices #PV.

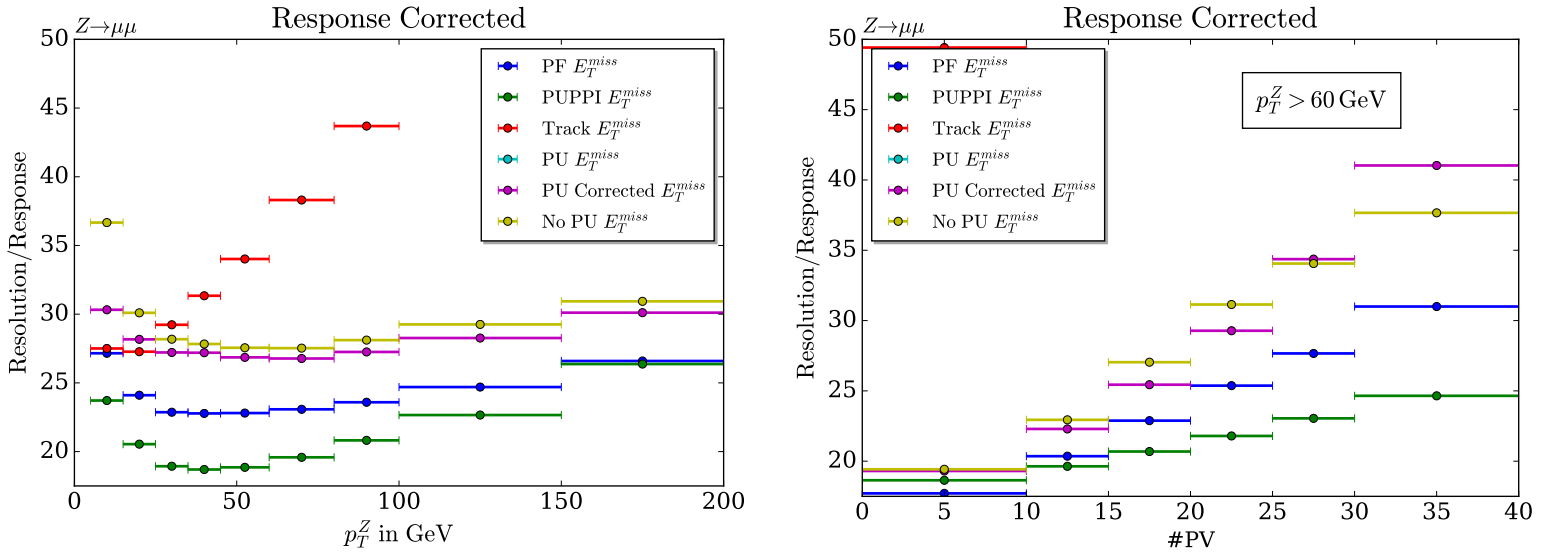
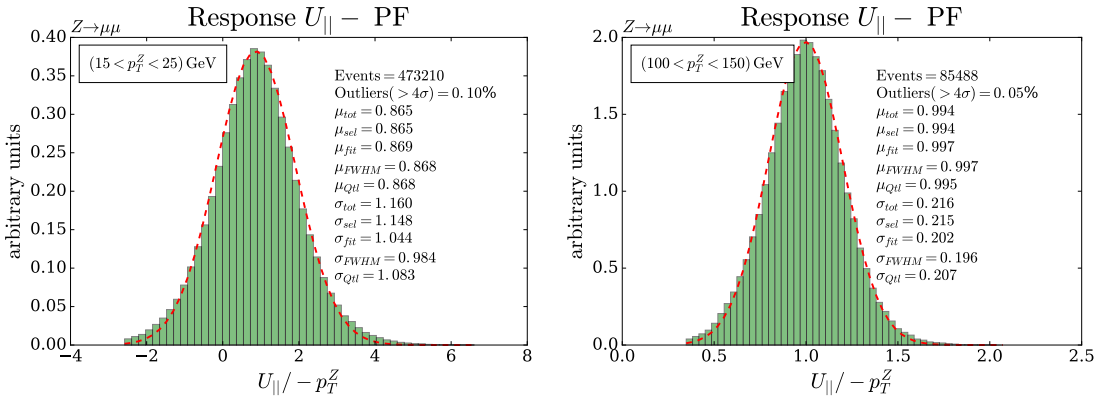


Figure 4.10.: Response corrected plot for a simulated $Z \to \mu\mu$ data sample with roughly 6 M events. On the left side in dependence of the transverse momentum of the Z boson, p_T^Z, and on the right side as a function of number of pile-up vertices #PV. One can see that only PUPPIand PF $\not{E}_T$ yield a low corrected response, profiting from both close-to-unity response and low resolution.



(a) **(b)**

Figure 4.11.: Examples of response distributions for different p_T^Z regions. In the low-p_T^Z regime (left), the standard deviation σ_{sel} is higher than μ_{sel}. On the right, the same plot for high-p_T^Z regimes, with a relatively small standard deviation of the determined response value.

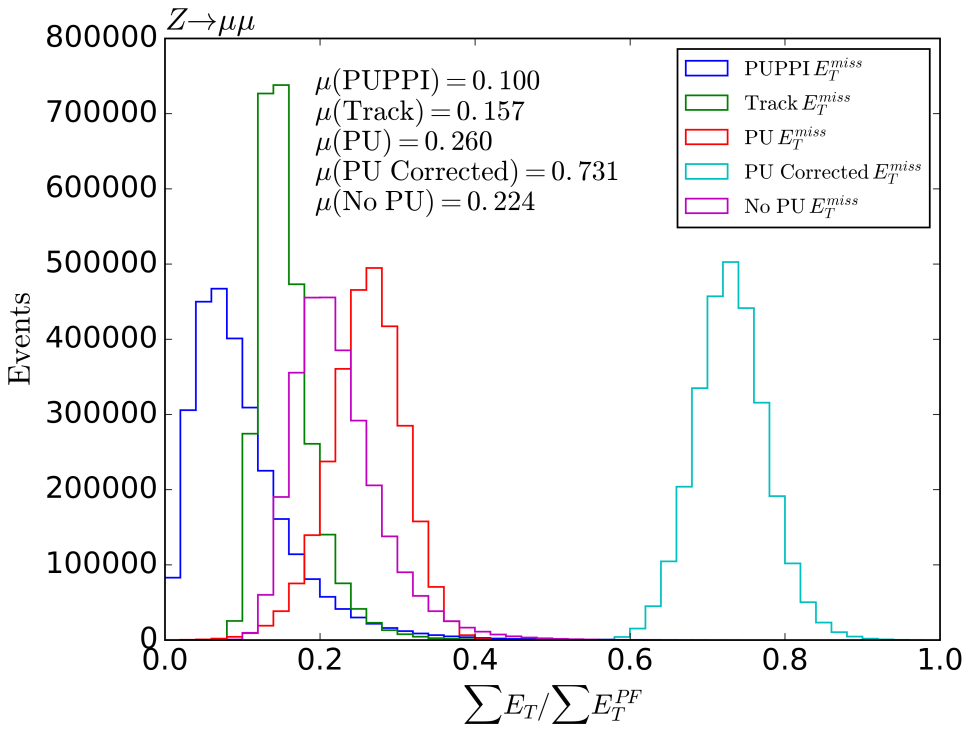


Figure 4.12.: Distribution of $\sum E_T / \sum E_T^{PF}$ for all $\not{E}_T$ definitions.

CHAPTER 5

Multivariate Missing Transverse Energy determination

In chapter 4, various $\not\!\!E_T$ definitions were introduced, each performing superior in different performance aspects but suffering in others. To harness the advantages of all $\not\!\!E_T$ definitions, a multivariate analysis (MVA) approach has been developed at CMS and will be explained in this chapter.
The ideal $\not\!\!E_T$ vector only consists of the transverse momenta of all non-measurable decay products from the main primary vertex. In section 4.3, the recoil definition was introduced, distinguishing between decay products of a particle and particles carrying the recoil of this particle. Figure 4.1b and 4.2 show this definition for the $Z \to \mu\mu$ decay. The decay of a particle in e.g. muons or neutrinos is purely probabilistic and does not influence the particles carrying the recoil. Excluding all decay particles from an event, one can therefore not differentiate between events from e.g. a $Z \to \mu\mu$ or $Z \to \nu\nu$ decay, as the recoil structure stays the same.
The $Z \to \mu\mu\,(ee)$ decay has two outstanding characteristics. First, all particles from this decay are measurable and the true $\not\!\!E_T$ of this event is therefore zero. Second, muons and electrons are the best measured particles at the CMS detector [11, 12]. Any $\not\!\!E_T$ unequal to zero originates therefore from reconstruction inefficiencies of particles in the recoil. In order to mitigate resolution effects on the $\not\!\!E_T$, this recoil can be corrected. In the $Z \to \mu\mu\,(ee)$ decay, this is done by comparing it to the reconstructed transverse Z boson momentum p_T^Z, obtained by adding up the four-momenta of the muons or electrons respectively (compare figure 4.2).
The p_T^Z variable represents the ideal environment for the application of a machine learning algorithm by providing a target to which the recoil can be compared to. Based on training information gained from p_T^Z, the MVA$\not\!\!E_T$ algorithm corrects the recoil on an event-by-event basis.

5.1. Input

The MVA$\not\!\!E_T$ algorithm is based on a GBRT method, as introduced in chapter 3.3.1. It is provided all $\not\!\!E_T$ definitions introduced in chapter 4.2 as input. As explained before, this

method relies on the fact that the decay process is irrelevant for the recoil. All bosonic decay products are subtracted from these definitions, thus omitting all information about the boson. This results in a recoil definition, U, for each algorithm, defined with

- ϕ_U: The angular direction of the recoil in the CMS coordination system.
- p_T^U: The magnitude of the recoil in direction of ϕ_U.
- $\sum E_T / \sum_{PF} E_T$: The fraction of the sum of all transverse momenta of the corresponding $\not{E}_T$ definition over the sum of the transverse momenta of all PF candidates. This provides an estimator for the MVA algorithm on how significant the impact of the corresponding $\not{E}_T$ definition is for each event.

Additionally the following variables are passed to the GBRT:

- $p_T^{Jet}, \phi_{Jet}, \eta_{Jet}$: Transverse direction ϕ_{Jet}, momentum p_T^{Jet} and direction in the longitudinal beam direction plane η_{Jet} of the five leading jets with $p_T^{Jet} > 15\,\text{GeV}$.
- N_{Jets}: The number of jets with a threshold of $p_T^{Jet} > 15\,\text{GeV}$.
- #PV: The number of reconstructed primary vertices as introduced in chapter 4.2.

5.2. Method

The MVA$\not{E}_T$ algorithm improves the $\not{E}_T$ resolution of an event by correcting its recoil. This correction is based on the PF$\not{E}_T$ definition and performed in two steps.

Angular correction

First, an angular correction rotates the PF recoil vector, $\vec{U}^{\text{PF}}$, by an angle ϕ, based on the nominal angular difference

$$T_1 = \phi_Z - \phi_U - \pi. \tag{5.1}$$

The constant factor π is subtracted as $\vec{p}_T^{\,Z}$ shows in the opposite direction of $\vec{U}^{\text{PF}}$. This allows a symmetric output of the machine learning algorithm around zero. See figure 5.1 for a schematic sketch of the angular correction. This $\phi-$corrected recoil vector U_ϕ^{MVA} is used for the second correction.

Scale correction

In a second step, this rotated recoil is multiplied with a scalar factor in order to stretch or bulge its parallel component, $U_{||}^{\text{MVA}}$, to match p_T^Z as illustrated in figure 4.3. For the machine learning algorithm, this nominal stretching factor is defined as

$$T_2 = \frac{-p_T^Z}{U_{||,\phi}^{\text{MVA}}}. \tag{5.2}$$

Figure 5.2 shows a possible scale correction.

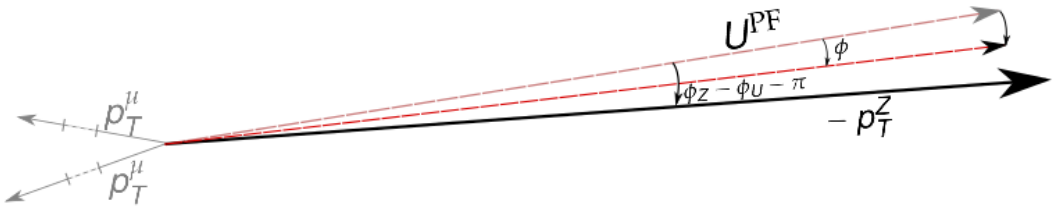


Figure 5.1.: Schematic angular correction of U^{PF}. The angle $\phi_Z - \phi_U - \pi$ represents the angular difference between the two vectors and the target for the multivariate regression algorithm. ϕ embodies a possible prediction of the algorithm.

Covariance estimation

A last training is performed to provide an event-by-event error estimation of the MVA$\not\!\!E_T$ calculation. This is done by predicting the covariance matrix

$$V(\vec{\not\!\!E}_T) = E\left(\left(\vec{\not\!\!E}_T - \vec{\not\!\!E}_T^{True}\right)\left(\vec{\not\!\!E}_T - \vec{\not\!\!E}_T^{True}\right)^T\right), \tag{5.3}$$

with the genuine missing transverse energy $\vec{\not\!\!E}_T^{True}$. As mentioned before, the genuine $\not\!\!E_T$ for $Z \to \mu\mu(ee)$ decays is estimated as zero, simplifying this equation. The diagonal elements of this matrix refer to the error estimation of $U_{||}$ and $U_\perp$. As both recoil components are assumed to be uncorrelated, off-diagonal elements are zero, resulting in two additional regression targets

$$T_3 = \sqrt{\frac{\pi}{2}}\frac{\left|p_{T,||}^Z - U^{\mathrm{MVA}}\right|}{U^{\mathrm{MVA}}}, \tag{5.4}$$

and

$$T_4 = \sqrt{\frac{\pi}{2}}\frac{\left|p_{T,\perp}^Z\right|}{U^{\mathrm{MVA}}}. \tag{5.5}$$

Both targets, T_3 and T_4, refer to an error estimation of the parallel and perpendicular component by projecting the boson momentum p_T^Z on the MVA recoil U^{MVA}. The target values are normalized by U^{MVA} and the constant factor $\sqrt{\frac{\pi}{2}}$, later to be applied with

$$V(\vec{\not\!\!E}_T) = \begin{pmatrix} (T_3)^2 \cdot U & 0 \\ 0 & (T_4)^2 \cdot U \end{pmatrix}. \tag{5.6}$$

This covariance is calculated in the coordinate system of the recoil and accounts for the error estimation of the MVA$\not\!\!E_T$ recoil. The covariance of the decay products is added to this estimation to ensure a correct $\not\!\!E_T$ covariance estimation.

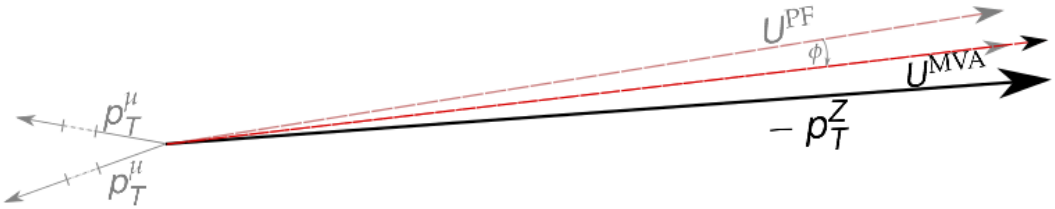
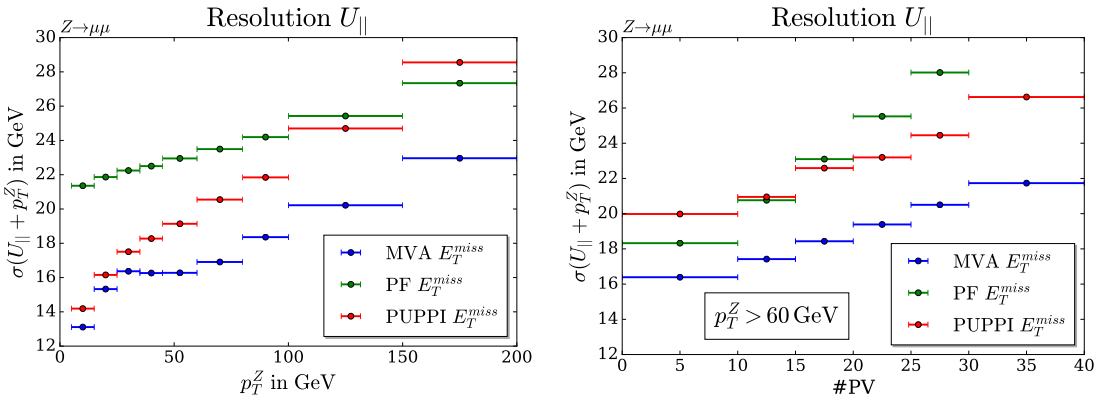


Figure 5.2.: Schematic scalar correction of the rotated U^{PF}. U^{MVA} represents the final angular and scalar corrected recoil vector.

5.3. Calculation

This multivariate method is based on reconstructed quantities. This implies that also the target value, p_T^Z, is a measured quantity. To ensure an unbiased training and to exclude reconstruction inefficiencies, the training events have to pass certain filters. A first criterion requires the reconstructed mass of the di-lepton system to be between 80 and 100 GeV. This assures that the muons or electrons are decay products of a Z boson. In about 10 % of all events, more than one combination of lepton-pairs fulfil this criterion. This means that different combinations of di-lepton pairs can be found. These events are skipped to avoid possibly wrong training targets. The samples used for training of the MVA method are simulated events, therefore leptons, wrongly reconstructed as other particles, are also discarded.
As shown in figure 4.4a, the bosonic transverse momentum has a falling spectrum. By simply using this p_T^Z-distribution for training, the low p_T^Z-area is relatively denser populated and therefore an MVA algorithm would tend to favour this region, introducing an emphasis towards low p_T^Z regimes. This focus can be avoided by applying weights to each event such that high p_T^Z events are weighted with a relatively higher weight than low p_T^Z events. This event reweighting was applied in a way that the resulting p_T^Z spectrum was uniformly distributed.
Nevertheless, very few events exist for the $p_T^Z > 150$ GeV region. In order to avoid too high weights for events in this sparsely populated region, special simulated high p_T^Z event samples exist. These samples exist up to transverse momenta of 2500 GeV and were added to the training data. In total, 8.427.529 events were used to train the GBRT algorithm. An important point for validation of the MVA method is a statistically independent test sample for performance measurement and avoidance of over-training. This sample is supplied by another simulation campaign and contains 4.066.203 events, forming a statistically sound sample for a measurement of performance.
The training was performed with a maximum number of 500 trees, a shrinkage rate of 0.05 and 250 minimum events per node (see chapter 3.3.1 for further explanation). This ensured a stable and converging training process of the GBRT.

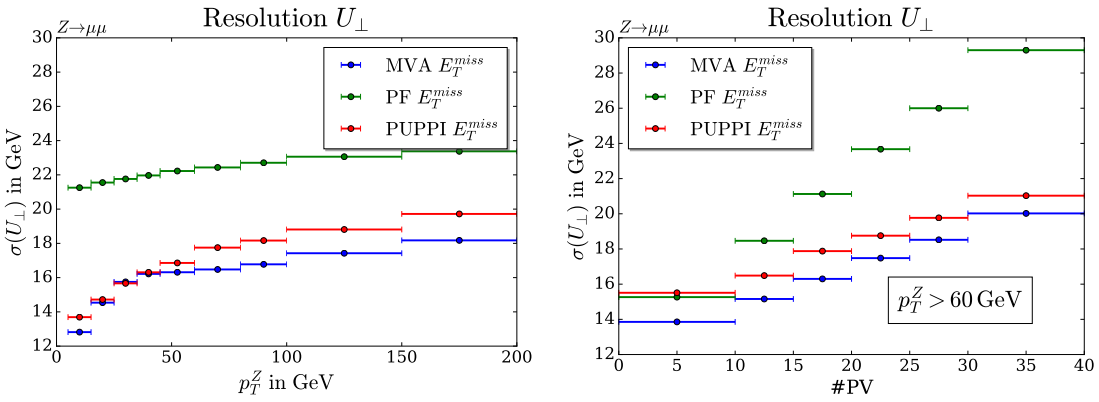


(a) **(b)**

Figure 5.3.: Resolution of MVA, PF and PUPPI $\not\!\!E_T$ as a function of p_T^Z (left) and #PV (right). MVA$\not\!\!E_T$ shows a 5 GeV resolution gain versus PF$\not\!\!E_T$ as a function of p_T^Z and increasing resolution gain of up to 9 GeV versus #PV. MVA$\not\!\!E_T$ shows a similar 6-8 GeV resolution gain with PUPPI$\not\!\!E_T$ versus PF$\not\!\!E_T$ in low p_T^T-regions, however profiting from better-defined event topologies for higher p_T^Z, resulting in a performance gain of up to 6 GeV versus both $\not\!\!E_T$ definitions. Both, PUPPI and MVA $\not\!\!E_T$, have a similar dependence on pile-up, with MVA$\not\!\!E_T$ showing a constant 4 GeV resolution difference to PUPPI$\not\!\!E_T$.

5.4. Performance on events with no genuine $\not\!\!E_T$

Figure 5.3, 5.4 and 5.5 show the resolution and response of MVA$\not\!\!E_T$ compared to PF and PUPPI $\not\!\!E_T$ for the $Z \to \mu\mu$ decay. Other $\not\!\!E_T$ definitions from section 4.2 were omitted as they either lack close-to-unity response or show inferior resolution capabilities. Figure 5.6 shows the response corrected plot.
All plots were done with the above mentioned validation set and passed the filters introduced in the previous section.
MVA$\not\!\!E_T$ features a gain in resolution versus PF and PUPPI $\not\!\!E_T$. It shows a constant 5 GeV gain versus PF$\not\!\!E_T$ as a function of p_T^Z and up to 9 GeV (31 %) resolution gain with increasing pile-up.
As for PUPPI$\not\!\!E_T$, the resolution margin is bigger for the parallel than for the perpendicular component. The reason for that lies within the importance of the angular and scalar correction of the recoil vector for each component. As will be shown in section 5.5, the perpendicular component is dominated by the correctness of the angular direction of the recoil vector. Small deviations from the boson's momentum direction, result in high resolution losses especially in high p_T^Z-regions. PUPPI$\not\!\!E_T$ shows a good performance on estimating the angular orientation by showing a 3-7 GeV resolution gain versus PF$\not\!\!E_T$ but lacks capability of estimating the scalar component of the recoil. This behaviour can be seen in figure 5.3a, with PUPPI$\not\!\!E_T$ being highly dependent on p_T^Z, showing worse resolution than PF$\not\!\!E_T$ for high p_T^Z and MVA$\not\!\!E_T$ outperforming both $\not\!\!E_T$ definitions by 2-5 GeV.

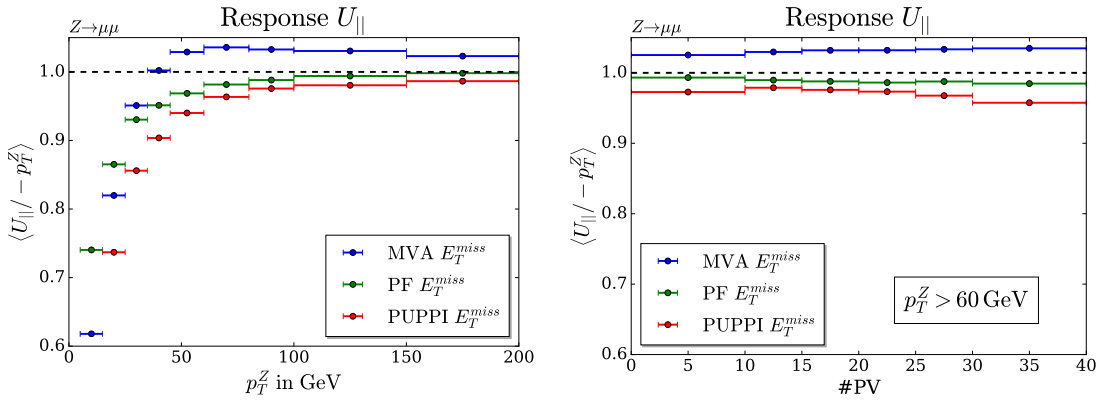


(a) **(b)**

Figure 5.4.: Resolution of MVA, PF and PUPPI $\not{E}_T$ as a function of p_T^Z (left) and #PV. MVA$\not{E}_T$ shows a 5 GeV resolution gain versus PF$\not{E}_T$ as a function of p_T^Z and increasing gain of up to 9 GeV versus #PV. MVA$\not{E}_T$ shows comparable angular correction performance to PUPPI$\not{E}_T$, resulting in a similar perpendicular resolution performance. For high p_T^Z, MVA$\not{E}_T$ shows a 2 GeV resolution gain, also observable in the right plot as a function of #PV.

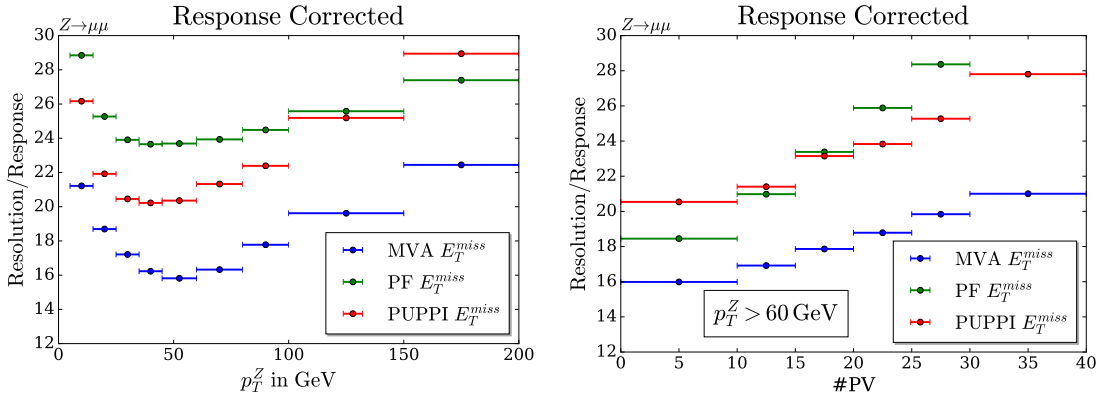
In terms of response, MVA$\not{E}_T$ reaches unity response quicker than PF and PUPPI $\not{E}_T$. However the response is about 2-3 % above unity for events with $p_T^Z > 60$ GeV. This behaviour is unexpected considering the fact that an MVA method was used to correct the PF recoil in order to scale it to the true p_T^Z-vector. This effect was therefore studied and will be discussed in the following chapter.
Figure 5.6 shows the resolution corrected by the response and illustrates the overall resolution gain of MVA$\not{E}_T$ versus PF and PUPPI $\not{E}_T$ with 2.5 GeV difference for low p_T^Z regions and up to 9 GeV for high p_T^Z events.

Figure 5.7 shows Resolution$_{||}$ for $Z \to \mu\mu$ in comparison to $Z \to ee$. All $\not{E}_T$ definitions show a 1-2 GeV resolution gain in ee compared to $\mu\mu$. However, all $\not{E}_T$ definitions follow the same trend in both channels. This is expected as the recoil structure of both decay channels is not dependent on the decay products and the $\not{E}_T$ calculation should thus not be affected by this decay. The shift between both channels can be introduced by different reconstruction resolutions for electrons and muons, resulting in possible resolution differences for both particles. Additional plots for a comparison of $Z \to \mu\mu$ and $Z \to ee$ can be found in appendix A.1.
Modern machine learning algorithms such as neural networks were also tested in the context of this thesis. If adjusted correctly, these neural networks performed equally well in terms of resolution and response, however not being able to significantly surpass the GBRT in correcting the $\not{E}_T$ of an event. Appendix B introduces two possible machine learning scenarios, indicating possible pitfalls when dealing with (modern) machine learning algorithms.



(a) **(b)**

Figure 5.5.: Response of MVA, PF and PUPPI $\not{E}_T$ as a function of p_T^Z (left) and #PV (right). A response of 2-3 % above unity can be observed for MVA $\not{E}_T$.



(a) **(b)**

Figure 5.6.: Response corrected plot of MVA, PF and PUPPI $\not{E}_T$ as a function of p_T^Z (left) and #PV (right). MVA $\not{E}_T$ has the lowest response corrected behaviour, however profiting from the fact that it has a slightly above unity response, leading to a lower response corrected value.

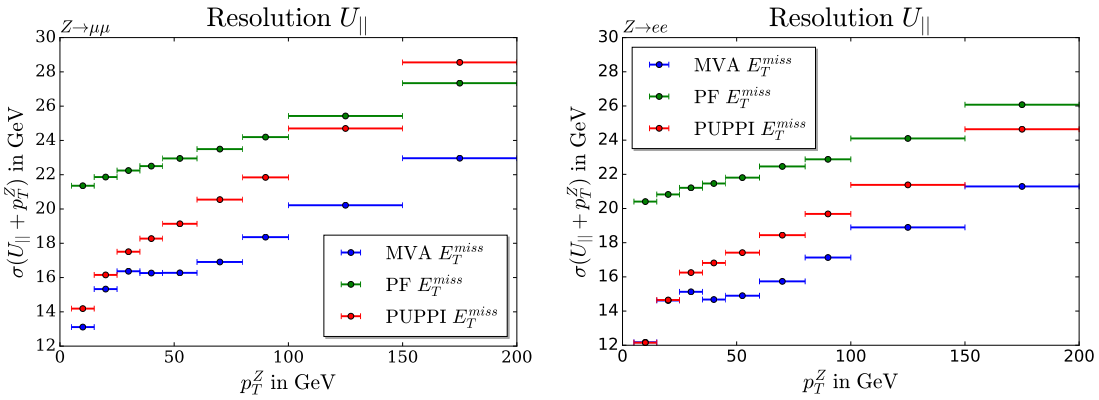


Figure 5.7.: Resolution$_{||}$ of MVA and PF $\not{E}_T$ for the $Z \to \mu\mu$ (left) and $Z \to ee$ decay (right). Whereas PF and MVA$\not{E}_T$ show the same resolution in both channels, PUPPI$\not{E}_T$ has a smaller dependence on pile-up for the $\mu\mu$-channel, resulting in an up to 4 GeV resolution difference for high p_T^Z.

5.5. Symmetries in response determination

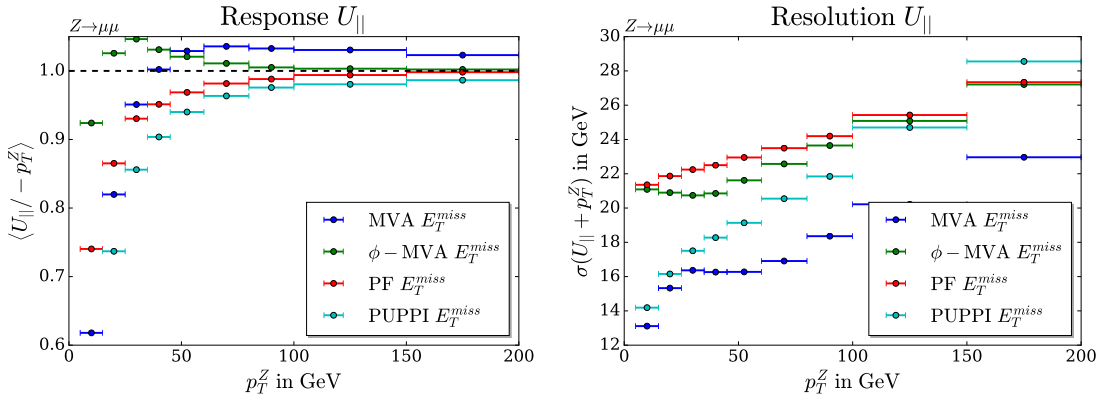
Figure 5.5 shows a 2-3 % response above unity for the MVA$\not{E}_T$. This effect was studied and found to be introduced by the scalar correction of the MVA approach. Compare figure 5.8 for performances of MVA$\not{E}_T$ and an only angular-corrected $\phi-$MVA$\not{E}_T$. The angular corrected $\not{E}_T$ shows even closer-to-unity response behaviour than PF$\not{E}_T$. Nevertheless, the scalar correction can not be omitted, as the parallel component shows no gain in resolution without the scalar correction (fig. 5.8b).

Training with excluded low p_T^Z events

A possible explanation for this may originate from the low-p_T^Z region in which events are averagely under-estimated (compare figure 5.5a) and need an up-scaling to reach unity-response. In order to possibly eliminate this bias, a training was performed excluding events with $p_T^Z < 30$ GeV. Figure 5.9 shows response and resolution of this training. Beside the fact that low p_T^Z-events were handled incorrectly, with a response of up to 60 % above unity, it was not possible to avoid above-unity response in any region.

Training with limited scale corrections

Assuming this bias not being introduced by the MVA method, a possible cause of trouble could be its target, T_2. Figure 5.10 shows the target distribution for a high p_T^Z region. The highly asymmetric shape of the distribution introduces a bias towards, on average, scale values above unity. A possible approach to eliminate this bias is an additional filter



(a) **(b)**

Figure 5.8.: Response as a function of p_T^Z with the angular corrected MVA$\not\!\!E_T$ (left). Whereas MVA$\not\!\!E_T$ shows its 2-3 % above unity response, the only ϕ-corrected $\not\!\!E_T$ yields a close-to-unity response. Nevertheless, the resolution gain of 5 GeV for the parallel component is introduced by the scalar correction (right).

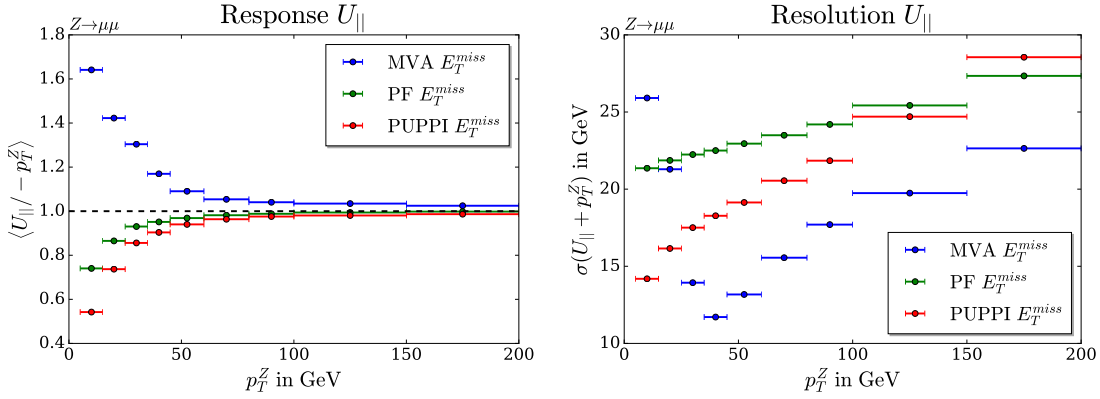


Figure 5.9.: Response as a function of p_T^Z of MVA$\not\!\!E_T$, trained on events with $p_T^Z > 30$ GeV. The effect of above-unity-response could not be eliminated by excluding these low-p_T^Z events from training. Consequently low p_T^Z were also not handled properly, resulting in worse resolution than PF$\not\!\!E_T$ for low p_T^Z.

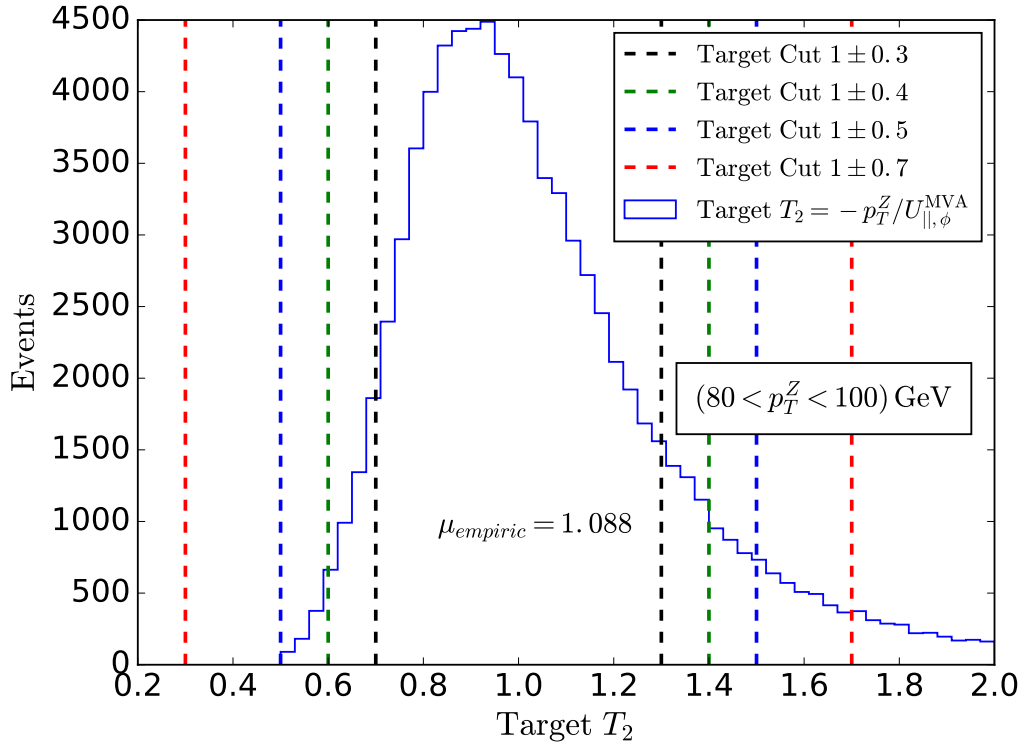


Figure 5.10.: Illustration of the target distribution, T_2, for the scale correction of the MVA method. The distribution shape is asymetric around 1, wherfore the empiric mean of the distribution, $\mu_{empiric}$, is above 1, introducing a bias towards overcorrection for events with $(80 < p_T^Z < 100)$ GeV. A similiar target distribution shape can be found for other p_T^Z-regimes.

to exclude events for training, whose target value exceeds too high scaling thresholds around one. This was done for different arbitrarily chosen thresholds as can be seen in figure 5.10. The results can be compared in figure 5.11. One can observe a mitigation of the above-unity behaviour for tighter cuts at the cost of gain in resolution. The observed 2-3 % response above unity could be gradually mitigated to about 1 % (0.5 %) for a target cut of $1 \pm 0.5(\pm 0.3)$. The decrease in resolution gain is visible for low-p_T^Z regions in which the initial gain of 8 GeV for the conventional MVA$\not{E}_T$ training was reduced to 4 and 2 GeV for the 1 ± 0.5 and ± 0.3 training respectively in the lowest p_T^Z region. The differences in resolution between different trainings become smaller for higher p_T^Z and are negligible for the highest region with 1 GeV difference, compared to the resolution gain versus PF$\not{E}_T$ of 6 GeV.

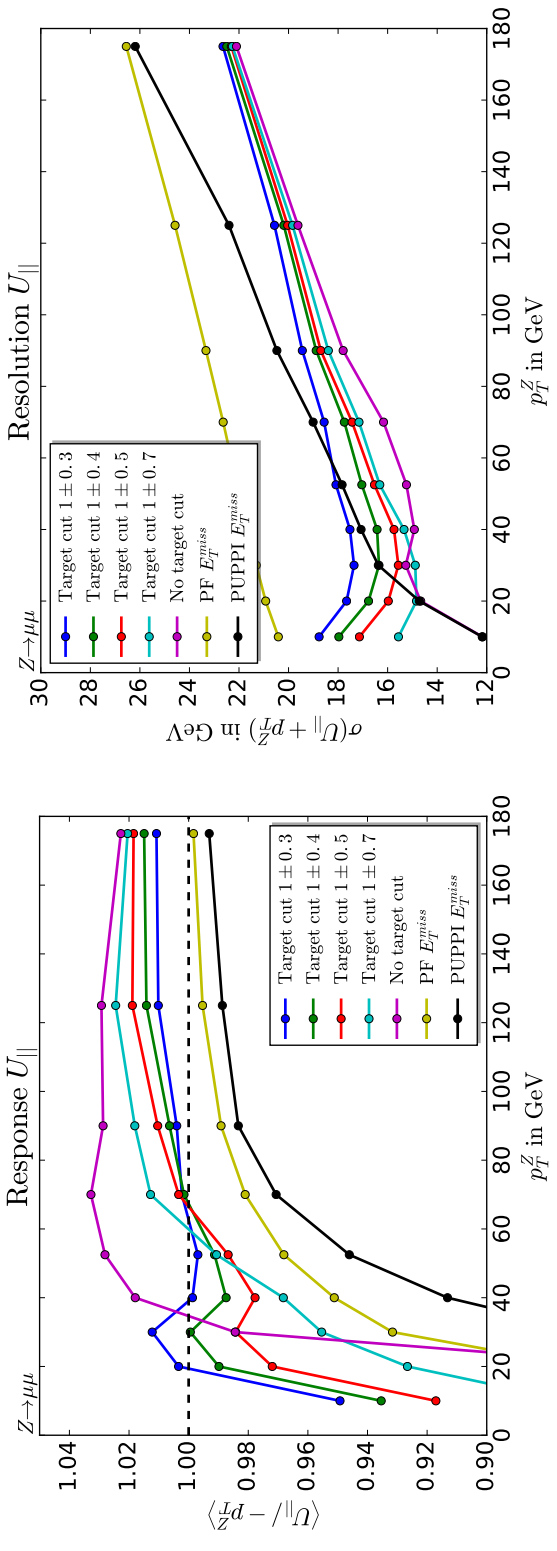


Figure 5.11.: Resolution as a function of p_T^Z of MVA$\not{E}_T$ trained on events with a maximal scale correction of $1 \pm (0.3, 0.4, 0.5, 0.7)$. Data points are connected for better visibilty in this plot. Resolution gain degrades with tightness of the cuts. This is mainly due to the fact that events, needing a high scale correction, are excluded in tighter target cuts. This effect is dominant in low-p_T^Z regions, as events with high p_T^Z need only small relative scale corrections. One can observe a closer-to-unity response with tighter target cuts, with a less than 1 % performance gap to unity-response for the training based on a target cut of 0.3 for events with $p_T^Z > 25\,\text{GeV}$.

Scalar target calculation

In order to fully understand the origin of this asymmetry in the target distribution T_2, it is helpful to examine the initial distributions the target is calculated from. The scalar target, T_2, was calculated as

$$T_2 = \frac{-p_T^Z}{U_{||,\phi}^{MVA}}, \tag{5.7}$$

with the distribution of the bosonic transverse momentum, p_T^Z, and $U_{||,\phi}^{\text{MVA}}$ representing the projection of the ϕ-corrected-vector on the p_T^Z-vector. The distribution of $U_{||,\phi}^{\text{MVA}}$ is illustrated in figure 5.12. While the p_T^Z distribution of the (weighted) events is uniform, the corresponding $U_{||,\phi}^{\text{MVA}}$ values are normal distributed around their respective p_T^Z. Constructing the target, T_2, by dividing p_T^Z by $U_{||,\phi}^{\text{MVA}}$, the left-sided tail of the normal distribution creates the asymmetric distribution, as small values in the denominator lead to high target values. This asymmetry introduces the bias of the MVA algorithm to predict on average too high scale factors and can not easily be removed.

Changing the target value to

$$T_2^{-1} = \frac{U_{||,\phi}^{\text{MVA}}}{-p_T^Z}, \tag{5.8}$$

would create a symmetric target function, since a normal divided by a uniform distribution results in a normally distributed shape. However this approach just postpones the introduction of the asymmetry, as the scaling would then be applied as a division rather than a multiplication

$$U^{\text{MVA}} = \frac{U_{\phi}^{\text{MVA}}}{T_2^{-1}}, \tag{5.9}$$

introducing the asymmetry by, again, dividing by a normal distribution. Further studies about shape alteration based on distribution divisions, can be found in the appendix A.3.

Inverse Response definition

The $\text{Response}_{||}$ was defined in equation 4.4 as

$$\text{Response}_{||} = \left\langle \frac{U_{||}}{-p_T^Z} \right\rangle. \tag{5.10}$$

The optimal response equals unity, a response of one. Intuitively one would expect the inverse of the response

$$\text{Response}_{||}^{-1} = \left\langle \frac{-p_T^Z}{U_{||}} \right\rangle, \tag{5.11}$$

also to be one for an optimal unity response behaviour. However this $\text{Response}_{||}$ definition, as a division of normal by uniform distribution, has a symmetric shape (fig. 5.13a). The

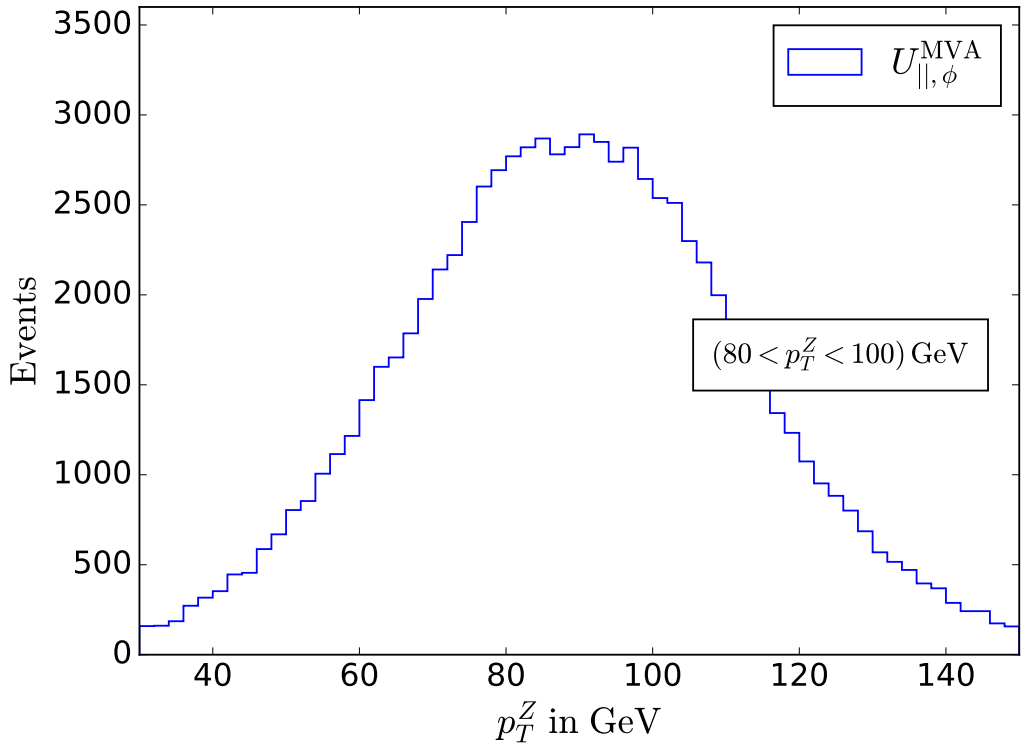
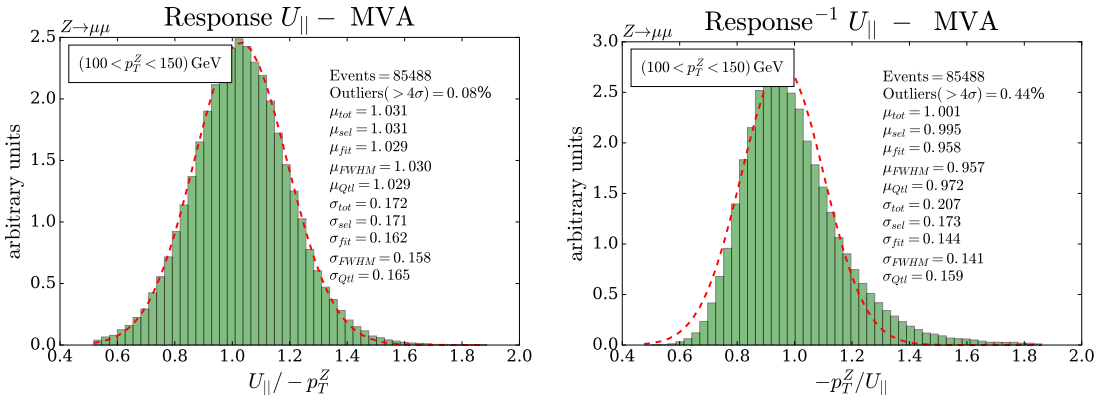


Figure 5.12.: Illustration of the distribution $U_{||,\phi}^{\mathrm{MVA}}$ for events with $(80 < p_T^Z < 100)\,\mathrm{GeV}$.



(a) **(b)**

Figure 5.13.: Distribution shapes for the Response (left) and Response^{-1} (right) determination of MVA$\not\!\!E_T$ for events with $(100 < p_T^Z < 150)$ GeV. Events for the response determination are normal distributed, resulting in negligible differences between all validation methods. Response^{-1} is asymmetrically distributed, yielding different response values for the validation methods. Whereas μ_{tot} and μ_{sel} emphasise outlying events, μ_{FWHM} and μ_{Qtl} focus on the centre of mass of the distribution.

Response$_{||}^{-1}$, as a division of uniform by normal distribution, has an asymmetric shape, similar to the target, T_2, of the MVA algorithm (fig. 5.13b). As can be seen in figure 5.14, MVA$\not\!\!E_T$ shows unity-response behaviour for the inverse response definition using the truncated method. However Figure 5.13b shows that other validation techniques such as the FWHM or quantile method calculate a below-unity response for this asymmetric distribution. This is due to the fact that few high response values are compensated by many low response values to achieve unity-response, resulting in a median value below one. To avoid asymmetric scaling effects, the median is therefore a better measurement to define the centre of mass of an asymmetric distribution. Figure 5.15 shows Response and Response^{-1} with this quantile mapping as introduced in chapter 4.3.2, showing that, in contrast to PF and PUPPI $\not\!\!E_T$, the MVA$\not\!\!E_T$ is scaling its recoil too high, due to its asymmetric training target. For further plots, the quantile method is used to evaluate the performance of the $\not\!\!E_T$ definitions.

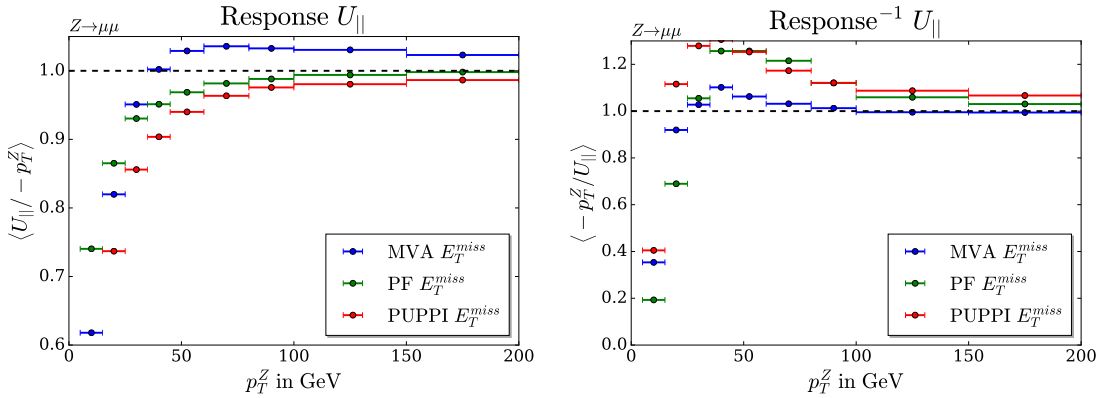


Figure 5.14.: Response (left) and Response^{-1}(right) of MVA, PF and PUPPI $\not{E}_T$ using the truncated method. MVA$\not{E}_T$ reaches unity response for the asymmetric inverse response definition.

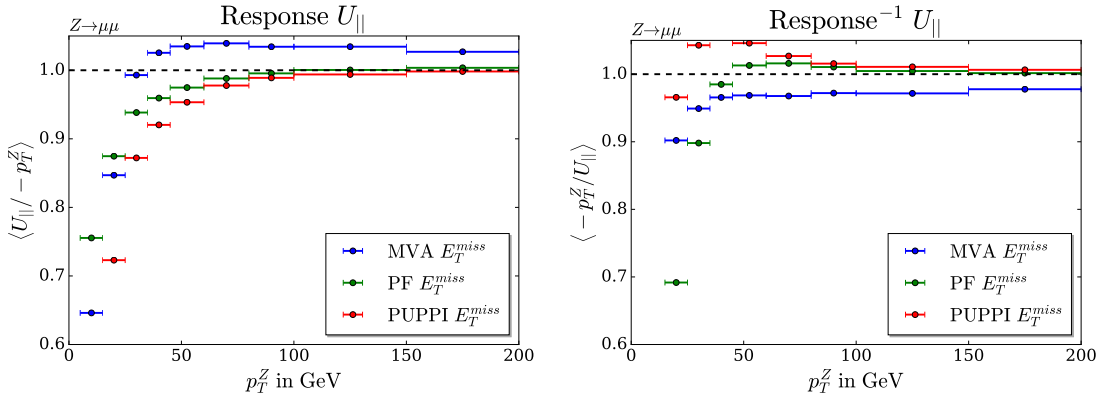


Figure 5.15.: Response (left) and Response^{-1}(right) of MVA, PF and PUPPI $\not{E}_T$ using the quantile method. Both, PF and PUPPI $\not{E}_T$, have close-to-unity response, whereas MVA$\not{E}_T$ shows above unity response for the response definition and vice versa for inverse response. This method indicates an on average too high scaling of the recoil for MVA$\not{E}_T$.

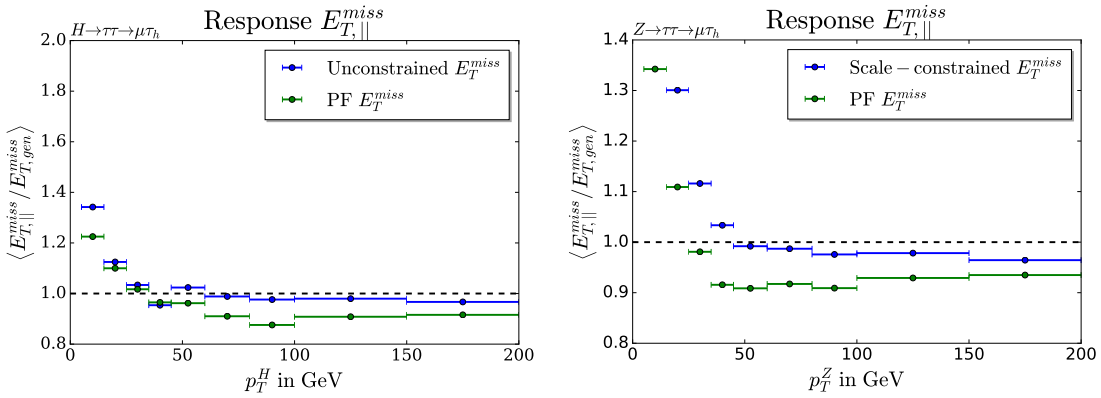
5.6. Performance on events with genuine $\not{E}_T$

Both $Z \to \mu\mu$ and $Z \to ee$ decays are special due to their visible decay products. $Z \to \tau\tau \to \mu\tau_h$ is a decay with genuine $\not{E}_T$, since the tau decay produces neutrinos. The transverse momentum of the boson, p_T^Z, reconstructed as the sum of the transverse momenta of its visible decay products, is in this case missing the momentum of the neutrinos. As for this decay the neutrinos are the only particles contributing to the $\not{E}_T$, p_T^Z can be approximated by adding the true $\not{E}_T$. This information is available as plots in this section were made with simulated events.

The response can now only be measured on the $\not{E}_T$, as the projection for splitting in parallel and perpendicular components was done on the reconstructed boson momentum. The $\not{E}_T$ response is defined, following its definition in equation 4.2, as

$$\not{E}_T\text{Response} = \left\langle \frac{\not{E}_T}{\not{E}_T^{True}} \right\rangle . \tag{5.12}$$

Plots in this section were done only with PF$\not{E}_T$, as PUPPI$\not{E}_T$ was not available for this data sample.
The above unity response, introduced by the asymmetric target distribution, T_2, could be mitigated in the previous chapter by applying a constraint to the scale-correction of the MVA$\not{E}_T$ training. In this $Z \to \tau\tau \to \mu\tau$ channel, the conventional training, denoted as unconstrained $\not{E}_T$ and a scale constrained training are validated. The training with a scale constraint of 30 % was chosen as it mitigated the above-unity response effect more consistently than less tight constraints.
Figure 5.16 shows the response for the unconstrained and constrained training set-ups with respect to p_T^Z. PF$\not{E}_T$ shows up to 35 % above-unity response for low p_T^Z, reaching a constant level of response for events with $p_T^Z > 35\,\text{GeV}$ around 8-10 % below unity. The unconstrained MVA$\not{E}_T$ training reaches unity response for the highest p_T^Z region with increasing above-unity response of up to 50 % for lower p_T^Z. The constrained set-up reaches close-to-unity response for events with $p_T^Z > 50\,\text{GeV}$. However, low p_T^Z events are described by an up to 70 % above-unity response.
Figure 5.17 shows the parallel resolution as a function of p_T^Z for an unconstrained MVA$\not{E}_T$. Compared to a constant 5 GeVresolution gain of MVA$\not{E}_T$ versus PF$\not{E}_T$ in the $\mu\mu$ channel, the resolution gain in $\mu\tau_h$ is maximal for events with $p_T^Z = 50\,\text{GeV}$ (6-7 GeV). Low p_T^Z events yield an only 2-3 GeV gain. This effect is possibly introduced by the jet clustering algorithm. The tau lepton momentum is reconstructed by clustering all particles within a certain cone. Low tau momenta can be distorted by other low p_T unclustered particles. If the clustering algorithm is assigning these particles to the hadronic tau jet, the tau momentum is reconstructed incorrectly and can thus distort the $\not{E}_T$ correction. This effect is negligible for higher p_T^Z as individual neutral unclustered particles carry small transverse momenta.
Figure 5.18 shows the perpendicular resolution as a function of p_T^Z. Both $\not{E}_T$ definitions show a comparable trend in both channels resulting in a constant 4-6 GeV resolution



(a) **(b)**

Figure 5.16.: $\not{E}_T$ Response of MVA and PF $\not{E}_T$ for the $Z \to \tau\tau \to \mu\tau$ decay for an unconstrained MVA$\not{E}_T$ training (left) and a scale-constrained training (right).

difference between MVA$\not{E}_T$ and PF$\not{E}_T$. Figure 5.19 shows the parallel resolution as a function of reconstructed primary vertices, #PV. In both channels, MVA$\not{E}_T$ yields a smaller dependence on pile-up resulting in an up to 8-9 GeV resolution gain with increasing pile-up. Both definitions show a shift of 2-3 GeV between $\mu\mu$ and $\mu\tau_h$. This effect can, similar to electrons, be introduced by different reconstruction efficiencies of the specific leptons.

Figure 5.20 and 5.21 show the parallel and perpendicular resolution of a scale-constrained MVA$\not{E}_T$ training. Compared to an unconstrained set-up, the scale-constrained $\not{E}_T$ shows negligible resolution differences for low p_T^Z. This can be explained by the fact that low p_T^Z events need high relative scale corrections to yield an improvement in $\not{E}_T$ resolution. High p_T^Z events need relative small corrections, resulting in the fact that this constrained $\not{E}_T$ shows a similar resolution trend to the unconstrained training for high p_T^Z.

The following chapter will evaluate the effects of different training set-ups on the reconstructed Higgs mass in the $H \to \tau\tau$ analysis.

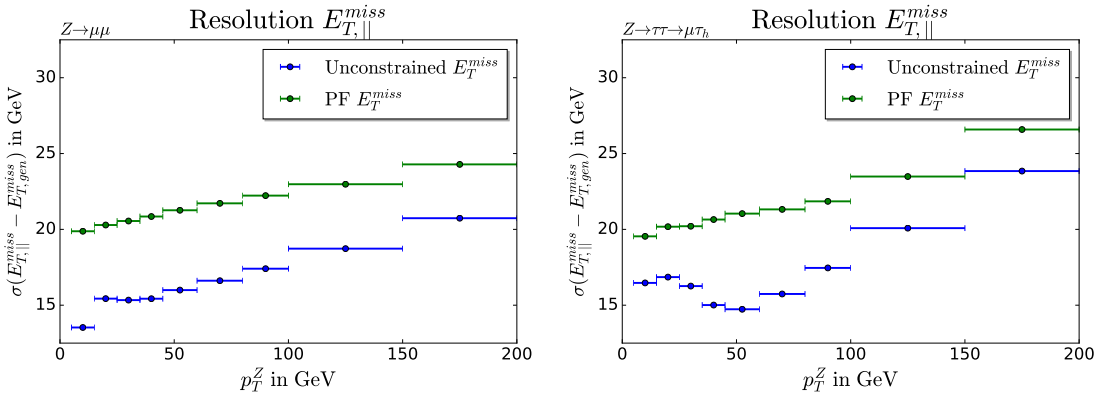


Figure 5.17.: $\not\!\!E_T$ Resolution$_{||}$ of MVA and PF $\not\!\!E_T$ for the $Z \to \mu\mu$ (left) and $Z \to \tau\tau \to \mu\tau$ decay (right). The MVA$\not\!\!E_T$ performance is based on a training without any restrictions.

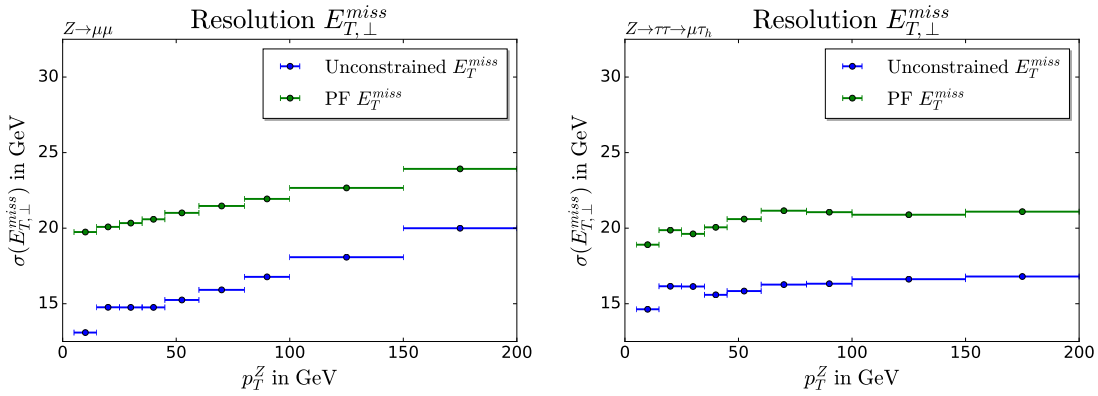


Figure 5.18.: $\not\!\!E_T$ Resolution$_{\perp}$ of MVA and PF $\not\!\!E_T$ for the $Z \to \mu\mu$ (left) and $Z \to \tau\tau \to \mu\tau$ decay (right). The MVA$\not\!\!E_T$ performance is based on a training without any restrictions.

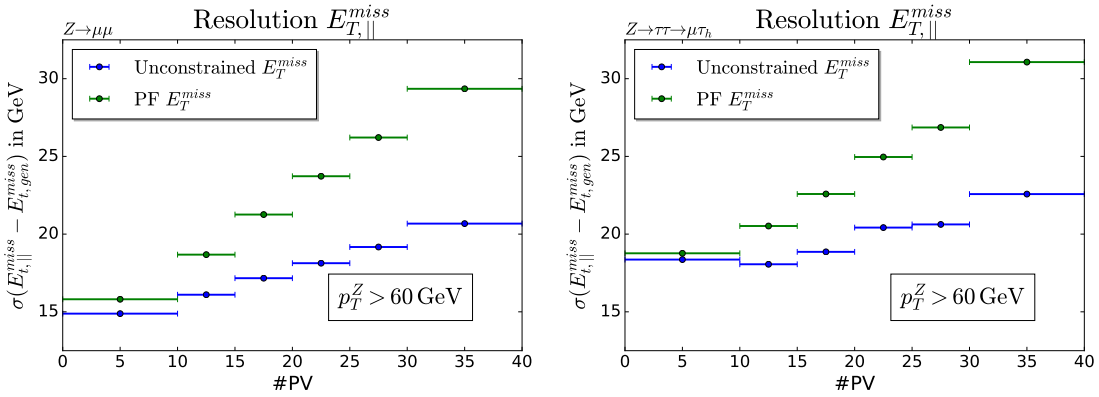


Figure 5.19.: $\not{E}_T$ Resolution$_{||}$ of MVA and PF $\not{E}_T$ for the $Z \to \mu\mu$ (left) and $Z \to \tau\tau \to \mu\tau$ decay (right) as a function of number of primary vertices, #PV.

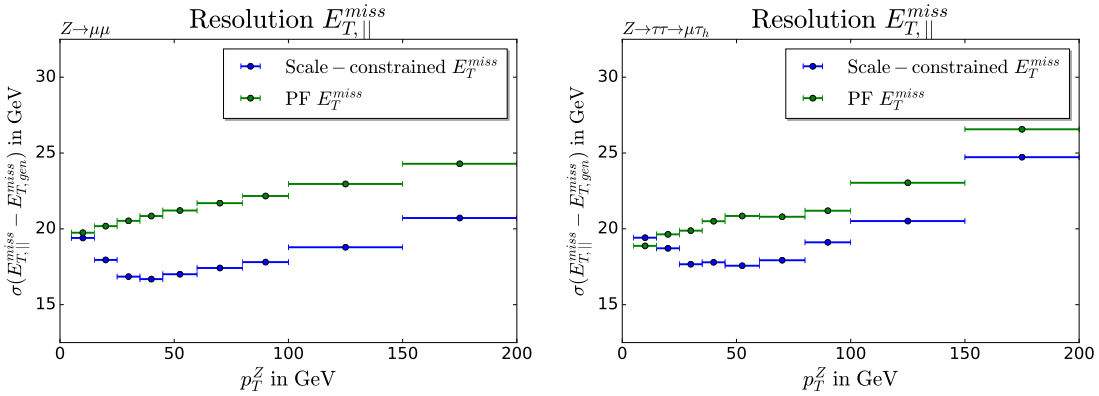


Figure 5.20.: $\not{E}_T$ Resolution$_{||}$ of scale-constrained $\not{E}_T$ and PF $\not{E}_T$ for the $Z \to \mu\mu$ (left) and $Z \to \tau\tau \to \mu\tau$ decay (right). The scale-constrained $\not{E}_T$ resolution performance is based on a training with constraint scalar corrections.

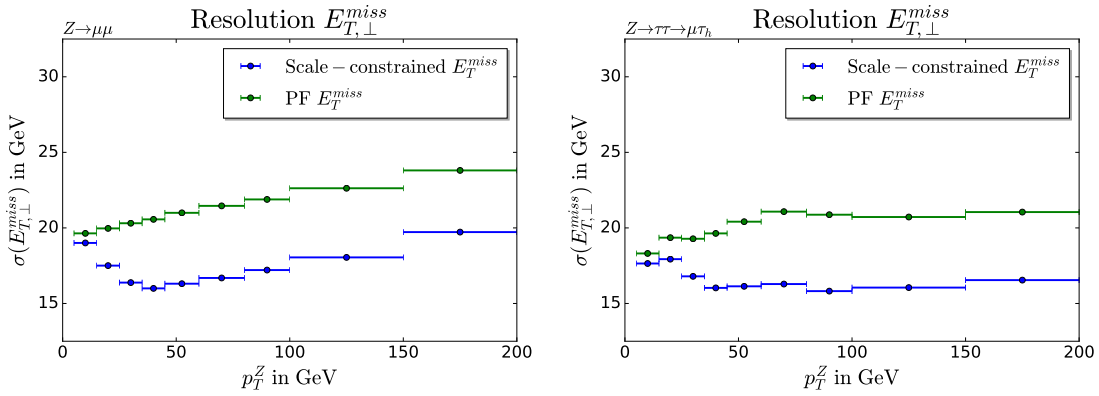


Figure 5.21.: $\not{E}_T$ Resolution$_\perp$ of scale-constrained and PF $\not{E}_T$ for the $Z \to \mu\mu$ (left) and $Z \to \tau\tau \to \mu\tau$ decay (right). Scale-constrained $\not{E}_T$ is based on a training with constraint scalar corrections.

5.7. Performance in the $H \to \tau\tau$ analysis

MVA$\not{E}_T$ was used in the $H \to \tau\tau$ analysis based on data taken in 2016 at CMS [26]. An algorithm reconstructs the Higgs mass based on many input variables of an event (for further information, see [27]). One of them is the respective $\not{E}_T$ of each event and its significance. This algorithm was run on a simulation sample with generated Higgs bosons from gluon-gluon fusion and a nominal mass of 125 GeV. In order to validate the impact of MVA$\not{E}_T$, this algorithm was first run with PF$\not{E}_T$ as $\not{E}_T$ definition and then substituted by MVA$\not{E}_T$. The reconstructed di-τ mass of the Higgs boson can be found in figure 5.22. To compare different $\not{E}_T$ performances, the empiric mean μ, median and standard deviation σ were calculated from these distributions and can be compared in table 5.1. Figure 5.23 and 5.24 show response and resolution of an unconstrained and scalar constrained MVA$\not{E}_T$ training. Perpendicular resolution plots are shown in the appendix A.2.

An overall shift to higher masses for all $\not{E}_T$ definitions is observable. The PF$\not{E}_T$ shift of 2.5 GeV could be mitigated to 1.1 GeV for the unconstrained MVA$\not{E}_T$ training. Also the resolution on the Higgs mass could be reduced by 0.6 GeV. The median shows a 1 GeV shift of PF$\not{E}_T$ which could be removed by MVA$\not{E}_T$. The median evaluates a lower value as the Higgs mass shows an asymmetric tail to higher mass values (figure 5.22). The scalar constraint training yields a lower mass than PF$\not{E}_T$, however not showing close-to-nominal mass prediction nor an improvement in resolution.

Analyses are based on so-called signal regions, excluding regions that could distort their result. Table 5.1 shows this region with $p_T^H > 50$ GeV, representing an area in which particles are well reconstructed. All $\not{E}_T$ definitions show a larger shift to higher masses than for the whole sample. Though both, PF and MVA$\not{E}_T$ show the same 1.5-3.5 GeV shift in the Higgs mass, MVA$\not{E}_T$ improves the resolution by 1.7 GeV (7.5%).

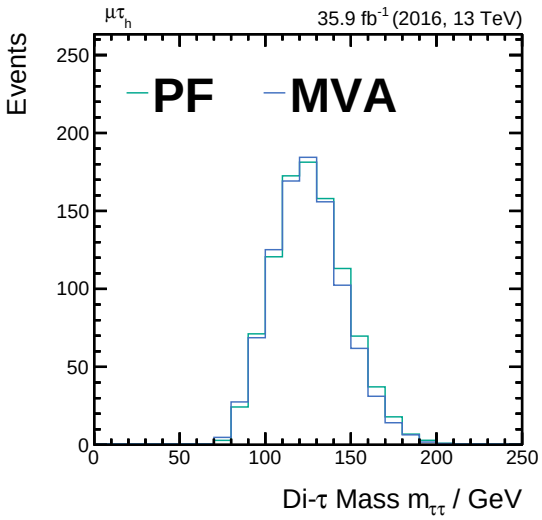


Figure 5.22.: Reconstructed mass of the Higgs boson for the $H \to \tau\tau \to \mu\tau$ decay produced with gluon-gluon fusion, using PF$\not{E}_T$ and MVA$\not{E}_T$. The boson was simulated with a nominal mass of 125 GeV. Median, empiric mean and standard deviation were calculated for different training setups and can be compared in table 5.1.

Possible explanations for this behaviour can be found in figure 5.23, showing the response and the parallel resolution of PF and MVA$\not{E}_T$. MVA$\not{E}_T$ reaches unity-response for $p_T^H > 50\,\text{GeV}$ with an up to 35 % above unity response for low p_T^H. PF$\not{E}_T$ yields only an up to 25 % above unity response for these low p_T^H region, however deviating up to 15 % below unity response for high p_T^H. MVA$\not{E}_T$ shows a 2 GeV parallel resolution gain for low p_T^H, increasing up to 4 GeV for higher p_T^H. This behaviour in resolution explains a better resolution gain of MVA$\not{E}_T$ over PF$\not{E}_T$ for $p_T^H > 50\,\text{GeV}$. Counter-intuitively to close-to-unity response in this signal region, both $\not{E}_T$ definitions show a larger shift in the Higgs mass than for the whole sample. Also, though $\not{E}_T$ in low p_T^H regions is on average overestimated, the algorithm calculates a lower Higgs mass. A possible explanation can be found in the significance of $\not{E}_T$, yielding an estimate for each event on how accurate the current $\not{E}_T$ calculation is. This information about the $\not{E}_T$ calculation is taken into account to determine to most likely Higgs mass. Therefore, above-unity response has not necessarily to result in an above-nominal predicted Higgs mass.
The scale-constrained training set-up performed worse in resolution than the PF$\not{E}_T$ and the conventional MVA$\not{E}_T$. Also, the conventional training predicted an on average closer-to-nominal Higgs mass. Future applications of MVA$\not{E}_T$ should therefore be based on the conventional MVA$\not{E}_T$ set-up.

Table 5.1.: Performance of different MVA$\not{E}_T$ trainings on the evaluation of the reconstructed nominal Higgs mass of 125 GeV in the $H \to \tau\tau$ analysis. The lower part of the table is calculated with $p_T^H > 50\,\text{GeV}$.

	Empiric Mean μ in GeV	**Median in GeV**	**Resolution σ in GeV**
PF$\not{E}_T$	127.5	126.0	22.04
MVA$\not{E}_T$ No constraint	126.1	124.9	21.44
MVA$\not{E}_T$ Scalar constraint	126.9	125.2	23.02
		$p_T^H > 50\,\text{GeV}$	
PF$\not{E}_T$	128.4	126.5	22.7
MVA$\not{E}_T$ No constraint	128.1	126.6	21.01
MVA$\not{E}_T$ Scalar constraint	129.3	127.1	23.54

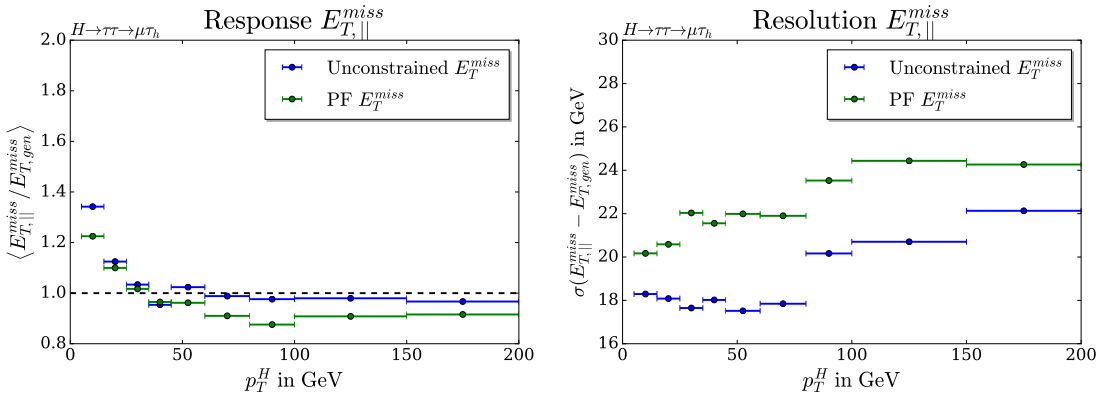


Figure 5.23.: Response and Resolution of an unconstrained MVA$\not{E}_T$ training in the $H \to \tau\tau \to \mu\tau$ decay produced with gluon-gluon fusion.

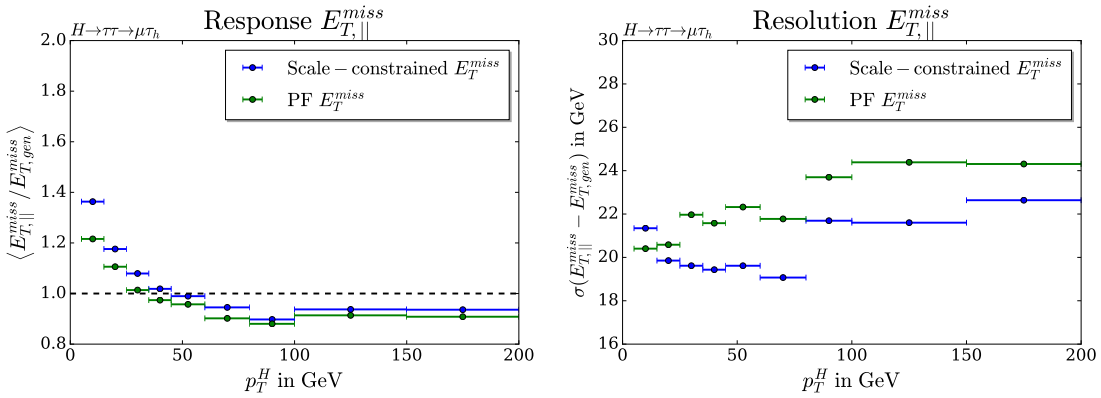


Figure 5.24.: Response and Resolution of a scalar constrained MVA$\not\!\!E_T$ training in the $H \to \tau\tau \to \mu\tau$ decay produced with gluon-gluon fusion.

5.8. Performance on data

As mentioned in section 5.3, the training and validation of MVA$\not\!\!E_T$ is based on reconstructed quantities. It is therefore possible to also validate its performance on real data in the $Z \to \mu\mu$ channel. This is best possible for the $\mu\mu$ channel as the genuine $\not\!\!E_T$ is zero and thus not needed to be added to the recoil.
However, due to mismodelling in the simulation, simulated and real events differ slightly as can be seen in figure 5.25. To mitigate these differences, so-called recoil corrections are applied, scaling the recoil of the simulated sample to match its corresponding real data recoil.
Figure 5.25 shows the resolution as a function of p_T^Z for both components. By applying these corrections, that have been produced for this thesis, it was possible to scale the simulated recoil to real data, also confirming the performance gain of MVA$\not\!\!E_T$ versus PF$\not\!\!E_T$ in real data.
The resolution with respect to the number of reconstructed primary vertices, #PV, is given in figure 5.26 yielding an up to 7 GeV resolution gain of MVA$\not\!\!E_T$ over PF$\not\!\!E_T$ with increasing pile-up for both simulated and real data.

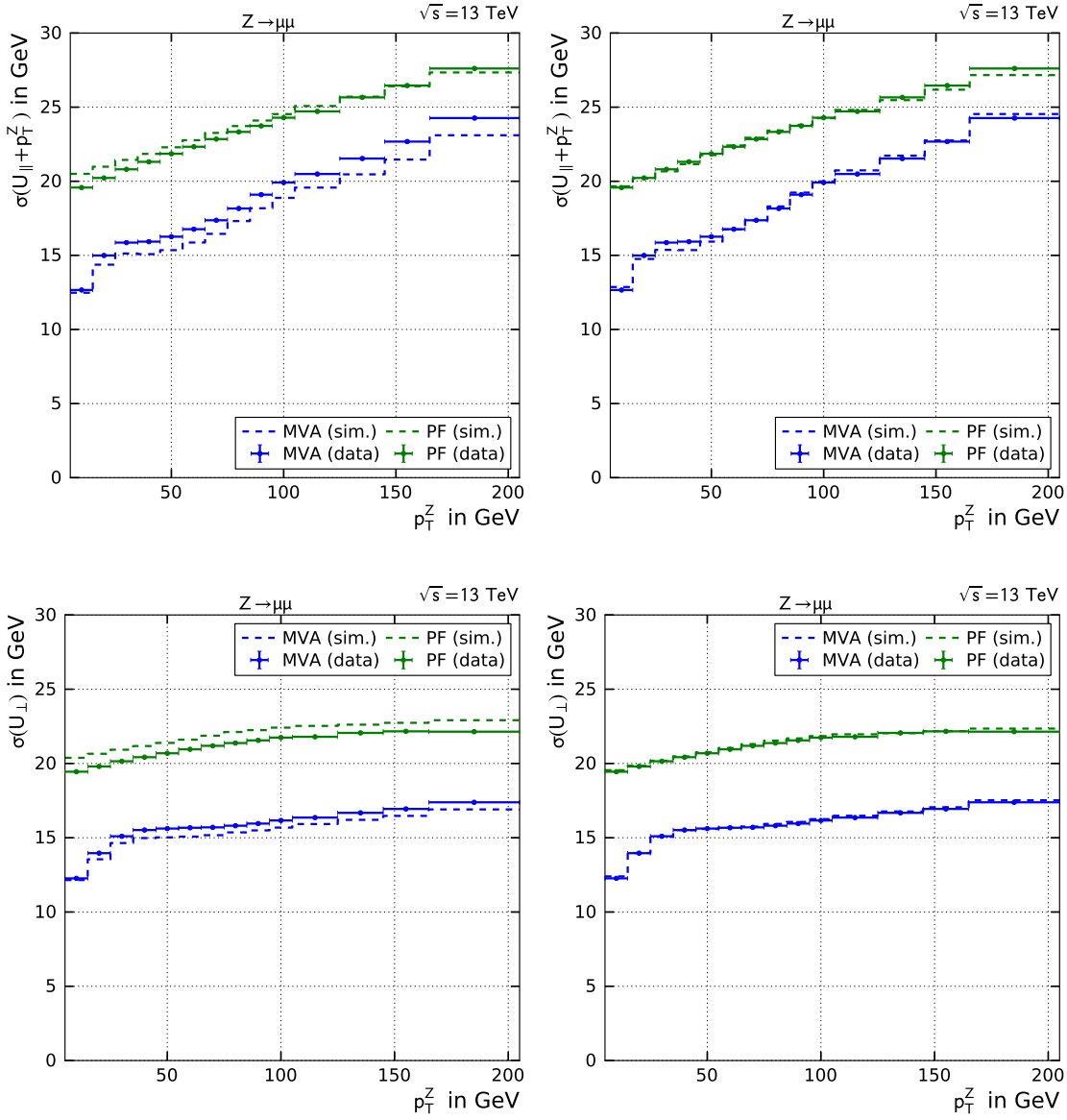


Figure 5.25.: Resolution plots of PF and MVA$\not\!\!E_T$ on simulated and real events for the $Z \to \mu\mu$ decay as a function of p_T^Z. Left sided plots are not recoil corrected, whereas right sided plots are recoil corrected. The 3-5 GeV gain of MVA$\not\!\!E_T$ over PF$\not\!\!E_T$ is also visible in real data. Differences between real data and simulated events by at most 1 GeV were consistently removed by applying recoil corrections.

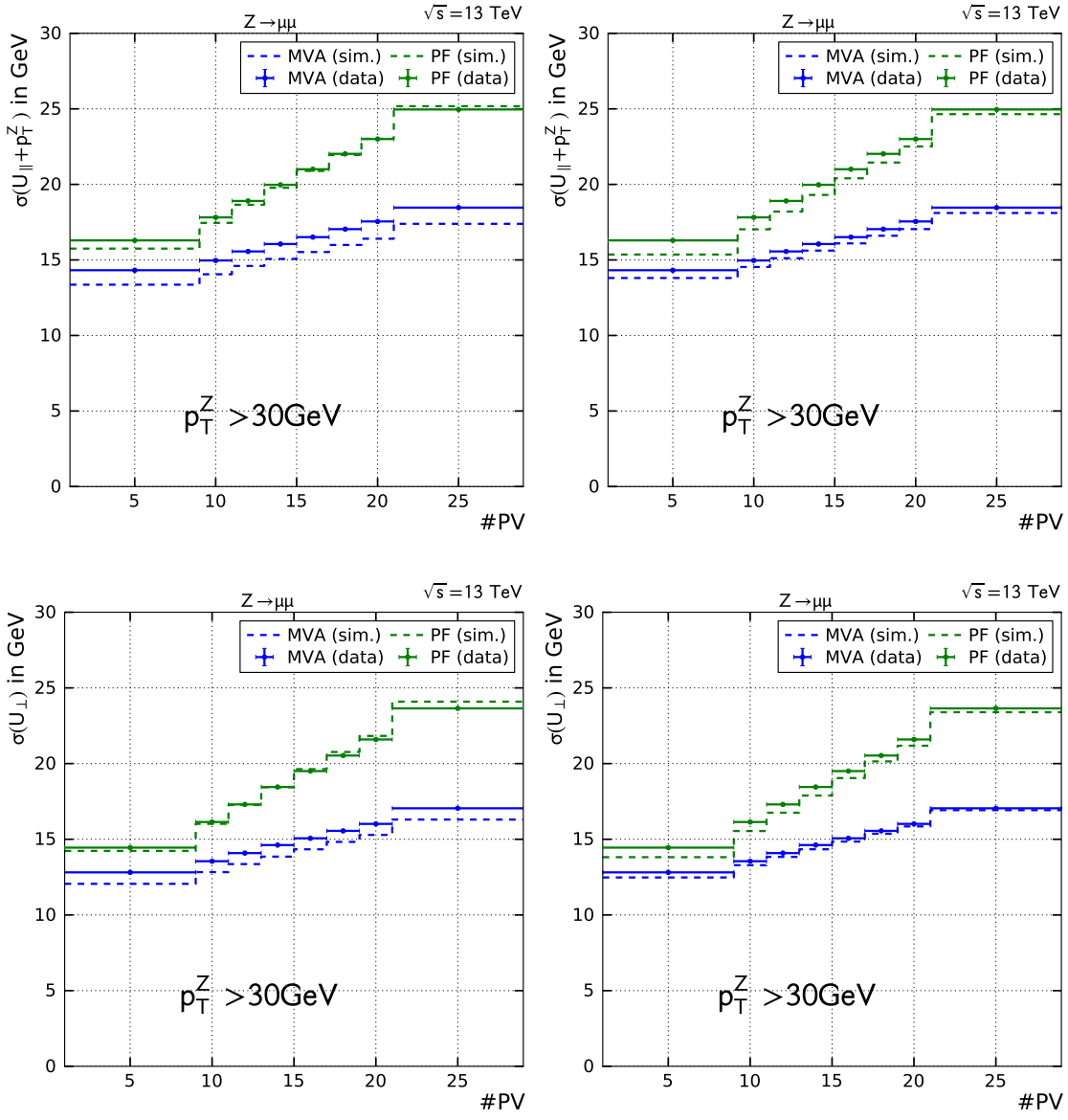


Figure 5.26.: Resolution plots of PF and MVA$\not\!\!E_T$ on simulated and real events for the $Z \to \mu\mu$ decay as a function of #PV. Left sided plots are not recoil corrected, whereas right sided plots are recoil corrected. MVA$\not\!\!E_T$ shows as well in data an up to 7 GeV performance gain versus PF$\not\!\!E_T$. Recoil corrections are modelling #PV dependence well for MVA$\not\!\!E_T$ but introduce a small discrepancy of up to 1 GeV between simulation and real data for PF$\not\!\!E_T$.

CHAPTER 6

Conclusion and Outlook

An MVA method was introduced to improve the $\not\!\!E_T$ resolution of an event at the CMS experiment. It is based on a GBRT algorithm and has been trained on $Z \to \mu\mu$ events, due to accurate muon reconstruction resolution and an expected $\not\!\!E_T$ of zero.
The validation on a statistically independent $Z \to \mu\mu$ sample has confirmed a resolution gain of 5-7 GeV (16-35 %) with respect to the conventional PF$\not\!\!E_T$ when resolved in p_T^Z. The pile-up dependence of MVA$\not\!\!E_T$ has proven to be significantly smaller than for PF$\not\!\!E_T$, resulting in a resolution gain of up to 9 GeV (31 %) for high pile-up regimes. Similar resolution performance was shown for $Z \to ee$, representing another decay channel with zero expected $\not\!\!E_T$.

This MVA approach is only dependent on the recoil structure of an event. Its performance on evaluating the $\not\!\!E_T$ of each event is thus independent of the decay channel. The $Z \to \tau\tau \to \mu\tau_h$ was therefore studied, representing a decay process with genuine expected $\not\!\!E_T$ carried by neutrinos. The superior resolution performance with respect to PF$\not\!\!E_T$ on estimating the $\not\!\!E_T$ of an event was also verified in this channel with small deviations between $\mu\mu$ and $\mu\tau_h$ for low p_T^Z. These differences can be explained with different reconstruction efficiencies of the decay products and are thus not caused by the MVA method.

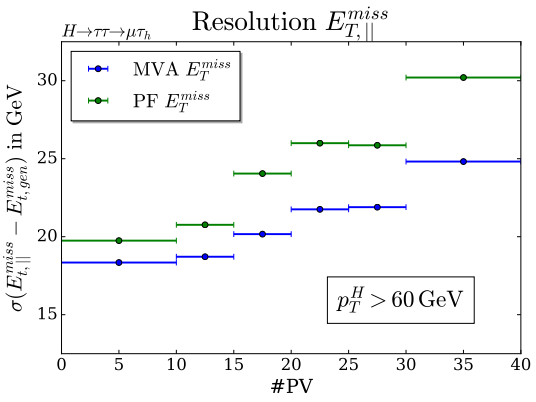


Figure 6.1.: $\not\!\!E_T$ Resolution as a function of reconstructed primary vertices, #PV, in the $H \to \tau\tau \to \mu\tau_h$ decay channel. MVA$\not\!\!E_T$ yields an up to 5 GeV resolution gain with respect to PF$\not\!\!E_T$ with increasing pile-up vertices.

The $H \to \tau\tau$ analysis plays an important role in studying the Higgs boson and its properties. The MVA$\not\!\!E_T$ was used in this analysis to determine the Higgs mass in the $H \to \tau\tau \to \mu\tau_h$ decay channel and compared to PF$\not\!\!E_T$. It was found to improve the resolution of the Higgs mass by 7.5 % compared to the conventional $\not\!\!E_T$ definition. Its resolution gain improves with increasing pile-up

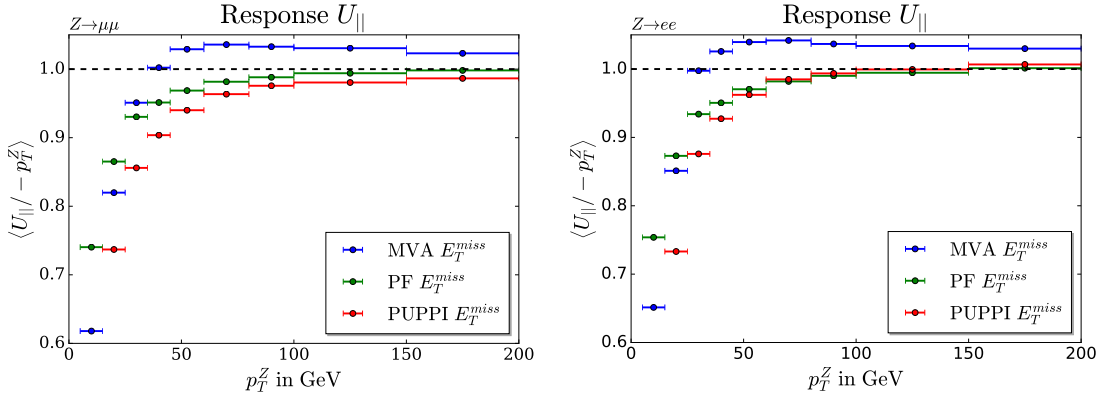
as can be seen in figure 6.1. This small pile-up dependence is important as future data taking periods at LHC with higher instantaneous luminosities will increase pile-up and portray an ideal environment for MVA$\not\!\!E_T$.
MVA$\not\!\!E_T$ was tested on real data in the $Z \to \mu\mu$ decay channel, confirming its constant resolution gain of 5 GeV as a function of p_T^Z and small pile-up dependence.
The studies of this thesis cleared the way for an appliance of MVA$\not\!\!E_T$ in on-going analyses. The current MVA$\not\!\!E_T$ is limited to di-lepton decays. An expansion to one or more than two decay products would create possible new applications in other analyses.

APPENDIX A

Additional plots for MVA$\not\!\!E_T$

A.1. Performance of MVA$\not\!\!E_T$ in the $Z \to ee$ channel

Figure A.1 and A.2 show the response and perpendicular resolution for the $Z \to \mu\mu$ and $Z \to ee$ channel as a function of p_T^Z. Figure A.3 and A.4 show the parallel and perpendicular resolution as a function of reconstructed primary vertices, #PV. All $\not\!\!E_T$ definitions show a 1-2 GeV down-shift between $\mu\mu$ and ee due to different reconstruction resolution effects for electrons and muons. This shift is larger for PUPPI$\not\!\!E_T$ (2 GeV) than for PF or MVA$\not\!\!E_T$ (1 GeV). Different handling of electrons and muons in the PUPPI algorithm can introduce these specific resolution effects.



(a) **(b)**

Figure A.1.: Response of MVA and PF $\not\!\!E_T$ for the $Z \to ee$ (left) and $Z \to \mu\mu$ decay (right). All three $\not\!\!E_T$ definitions show the same response behaviour in both decay channels.

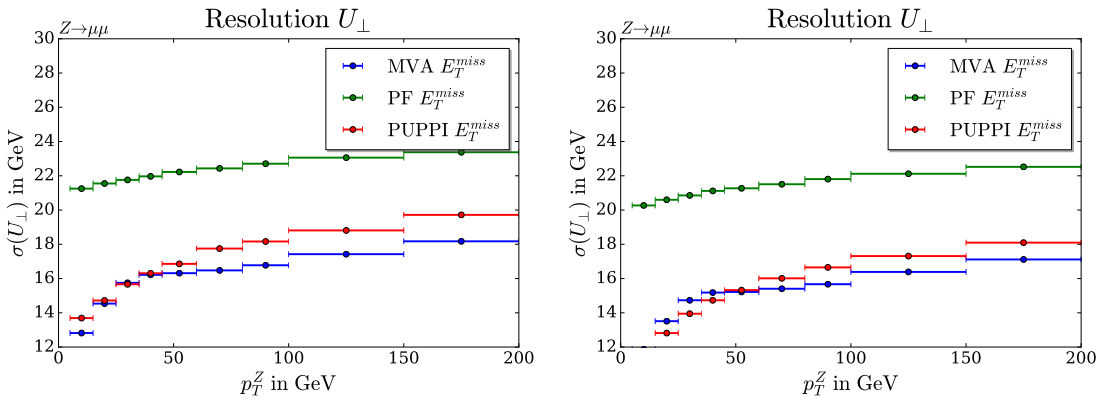


Figure A.2.: Resolution$_\perp$ of MVA and PF $\not\!\!E_T$ for the $Z \to \mu\mu$ (left) and $Z \to ee$ (right) decay. All $\not\!\!E_T$ definitions show a 1-2 resolution gain in the ee-channel compared to $\mu\mu$.

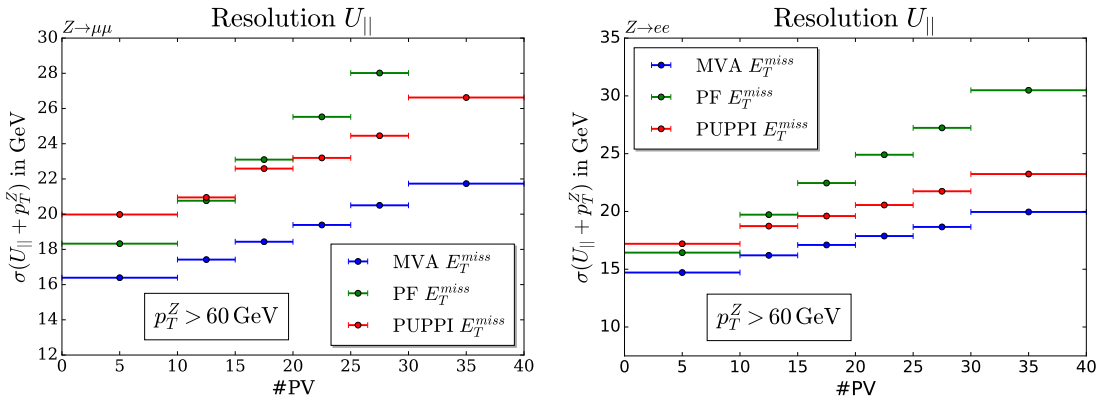


Figure A.3.: Resolution$_{||}$ of MVA and PF $\not\!\!E_T$ for the $Z \to \mu\mu$ (left) and $Z \to ee$ (right) decay as a function of #PV. Whereas PF and MVA$\not\!\!E_T$ yield a less than 1 GeV resolution difference between $\mu\mu$ and ee, PUPPI$\not\!\!E_T$ shows an up to 2.5 GeV resolution deficit in $\mu\mu$ compared to ee. The pile-up dependence is smaller for PUPPI and MVA$\not\!\!E_T$ resulting in a more than 9 GeV gain of MVA$\not\!\!E_T$ versus PF$\not\!\!E_T$ for high p_T^Z.

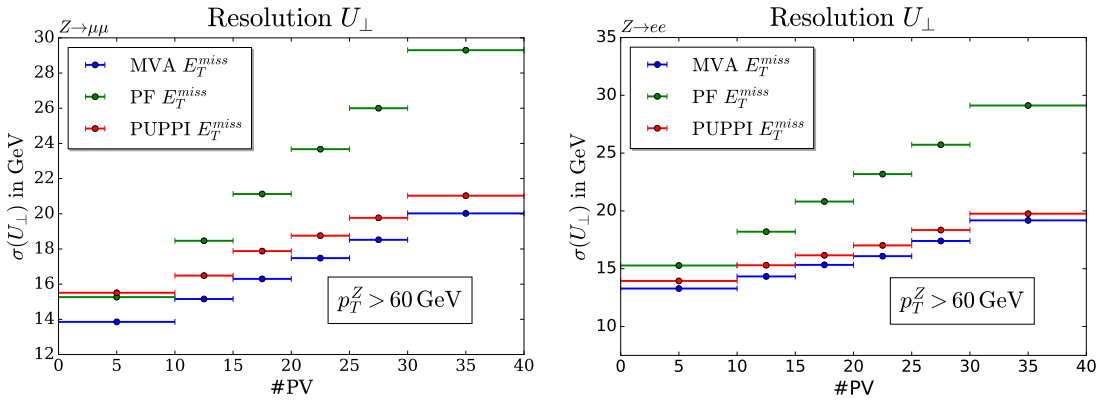
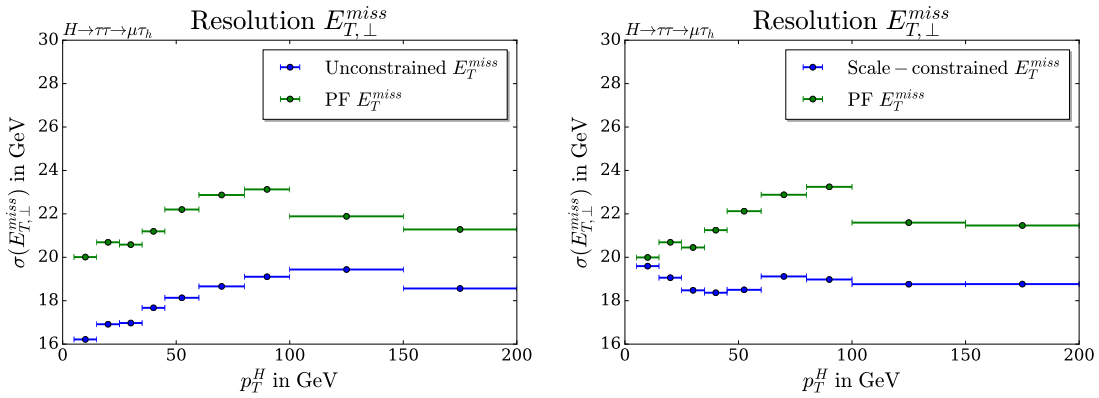


Figure A.4.: Resolution$_\perp$ of MVA and PF $\not{E}_T$ for the $Z \to \mu\mu$ (left) and $Z \to ee$ (right) decay as a function of #PV. This plot is similar to figure A.3, yielding a 1-2 GeV resolution gain in ee for all $\not{E}_T$ definitions. Pile-up dependence is again smaller for PUPPI and MVA$\not{E}_T$, confirming the 9-10 GeV (34 %) resolution gain of MVA$\not{E}_T$ over PF$\not{E}_T$.

A.2. Plots of the perpendicular resolution in the $H \to \tau\tau \to \mu\tau_h$ channel

Perpendicular resolution plots for the $H \to \tau\tau \to \mu\tau$ channel with simulated Higgs bosons from gluon-gluon fusion are shown in figure A.5. The unconstrained training set-up shows a constant 3-4 GeV resolution gain over PF$\not{E}_T$, whereas the scale-constrained training yields a negligible resolution difference for low p_T^Z. These events require relative high scale corrections larger than 30 %, resulting in a small gain for low p_T^Z as these events are insufficiently corrected.

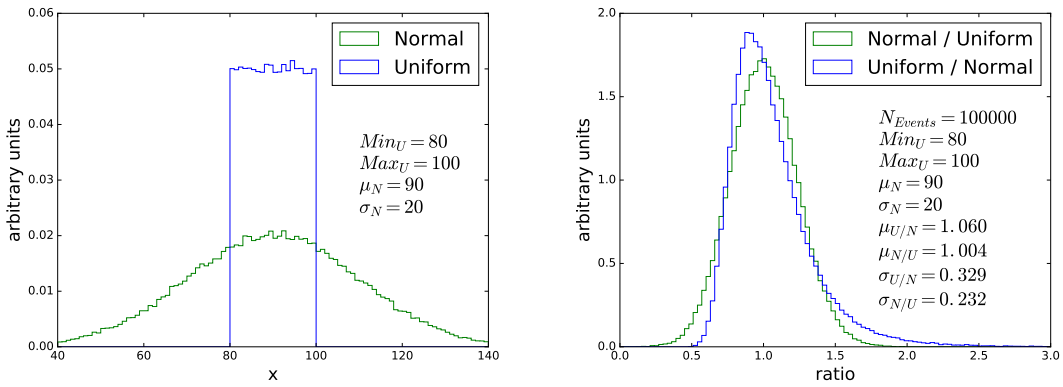


(a) **(b)**

Figure A.5.: Perpendicular resolution plots for the $H \to \tau\tau \to \mu\tau$ channel. The unconstrained MVA$\not\!\!E_T$ training (a) shows a constant 3-4 GeV resolution gain versus PF$\not\!\!E_T$ whereas the 30 % scale-constrained set-up (b) shows no resolution gain for low p_T^Z.

A.3. Toy study - Symmetries in distribution fractions

In this toy study, the ratio of two distributions is studied in dependence of the respective mean and standard deviation. One distribution is assumed to be uniform, the other one is normal distributed. Both distributions have the same mean value and the width of the uniform distribution is fixed to 20. In total, 100000 random data points were generated. Figure A.6a shows the input distributions. Figure A.6b illustrates the ratio of the respective distributions. The normal distribution divided by a uniform distribution is normal distributed around one. This is expected, as the normal distributed values, around 90 with a standard deviation of 20, are divided by a value in the range of 80 to 100, resulting in a statistically normal distributed shape. The other case, a uniform distribution divided by a normal distribution yields an asymmetric distribution with a mean value above one. This asymmetry is introduced by small values of the normal distribution. The uniformly distributed data points around 80 to 100 are divided by a normal distributed data points with a mean of 90 and standard deviation of 20. This, relative to the mean, high standard deviation leads to data points around zero, resulting in high values of the out-coming distribution shape.
Figure A.7 shows the effect of standard deviations. Data points are generated with the same uniform range of 80 to 100, whereas the standard deviation of the normal distribution is varied to 5 and 50. The small standard deviation yields no significant difference between both ratios. As the standard deviation of 5 compared to a mean of 90 is small, the generated values are drawn within a range comparable with the uniformly distributed data points, preventing an asymmetric distribution. A high standard deviation of e.g. 50 compared to a mean of 90, results in larger asymmetries, as normal distributed values close to zero introduce divergences in the resulting distribution.
The effect of varied mean values at a fixed standard deviation is shown in figure A.8. The same effect of asymmetry occurs for a mean value (30) in the range of the standard deviation (20). Changing the mean value to 290, reduces the effect of asymmetry.
This study shows the effects of asymmetries introduced by dividing by a normal distribution. The ratio of standard deviation over mean has to be large to mitigate these asymmetric effects. This factor is also dependent on the width of the uniform distributed data. Figure A.8 indicates a factor of 10 to be sufficient to reduce these effects for this set-up.



(a) **(b)**

Figure A.6.: Toy study for symmetries in distribution fractions. The first distribution is uniformly distributed between 80 and 100, whereas the second distribution is normal distributed around 90 with a standard deviation of 20.

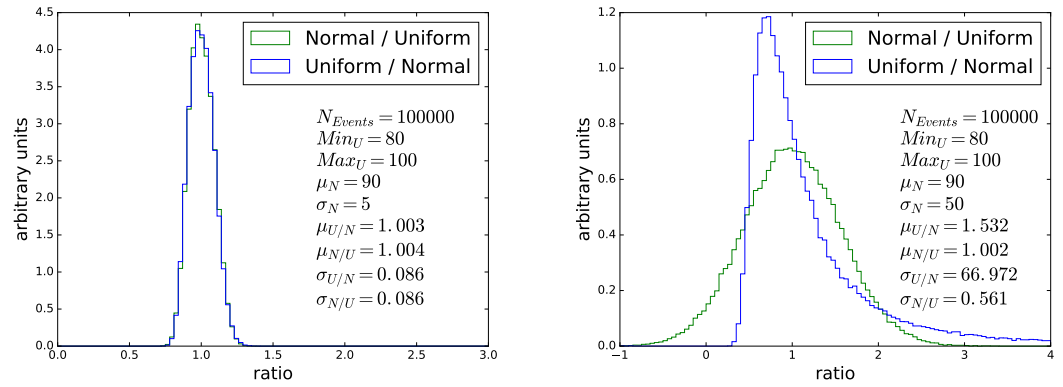


Figure A.7.: Toy study for symmetries in distribution fractions with large and small standard deviations. The first distribution is uniformly distributed between 80 and 100, whereas the second distribution is normal distributed around 90 with a standard deviation of 5 (left) and 50 (right).

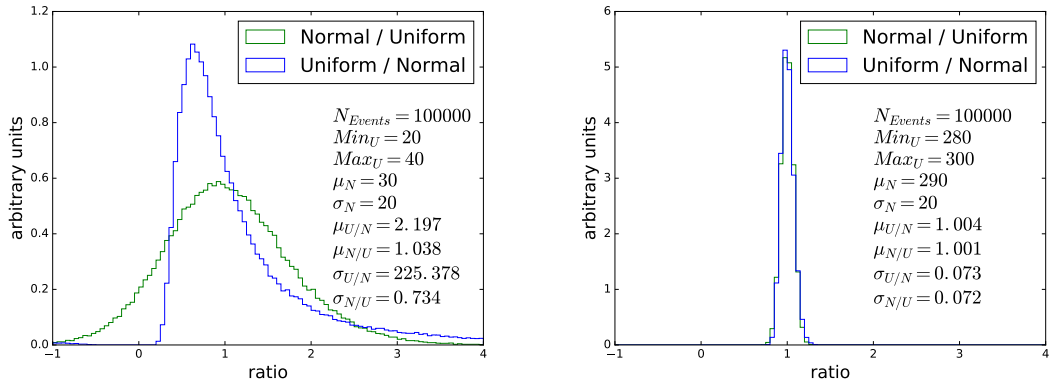


Figure A.8.: Toy study for symmetries in distribution fractions with large and small mean values. The first distribution is uniformly distributed between 20 and 40 (left) and 280 and 300 (right), whereas the second distribution is normal distributed around 30 (left) and 290 (right) with a standard deviation of 20.

APPENDIX B

Neural network performance on $\not\!\!E_T$ calculation

This chapter shows two problematic examples of appliances of a modern typical machine learning algorithm. The first section presents a problem that may occur when dealing with periodic variables, the second section shows the unusual performance of a trained neural network on the scalar correction of MVA$\not\!\!E_T$.

B.1. Custom loss functions for angular target values

As introduced in chapter 5.2, the recoil has to be corrected to fit the true boson momentum, p_T^Z. The first angular correction rotates the vector by an angle ϕ_{MVA} to approximate the nominal angular orientation of p_T^Z. Especially for high p_T^Z this required rotation is smaller than 30 ° but for low p_T^Z the required rotation can range from $-\pi$ to π as can be seen in figure B.1. From a physical point of view, a nominal rotation of $+\pi$ equals a rotation of $-\pi$ as the resulted rotated vector points in the same direction. A schematic sketch of this rotation can be found in figure B.2. However, for a machine learning algorithm, this nominal correction value or target is essential. It is optimized by minimizing the loss function introduced in chapter 3.2. For a nominal target value of $y = \pi$, the two physical correct rotations of $\hat{y} = \pm\pi$ result in the following squared loss $L(\hat{y}, y)$:

$$\begin{aligned} L(\pi, \pi) &= (\pi - \pi)^2 = 0, \\ L(-\pi, \pi) &= (-\pi - \pi)^2 = 4\pi^2 \gg 0. \end{aligned} \tag{B.1}$$

The trained neural network faces the largest possible loss for a rotation in the wrong direction. As events with a nominal $\pm\pi$ rotation are physically not distinguishable, the neural network learns in these cases to rather not perform any correction, resulting in an average loss of $L(0, \pi) = \pi^2$, compared to $(L(\pi, \pi) + L(-\pi - \pi))/2 = 2\pi^2$. This can be seen in figure B.3.
To avoid this behaviour, the target can be transformed to cartesian coordinates to avoid this periodic target value. However, for MVA$\not\!\!E_T$ this was not possible as the correction had to be applied in polar coordinates, to guarantee scalability for high p_T^Z events.
Another approach to avoid this behaviour is a custom loss function. This loss function can be designed in a way that differences between target and prediction of $\pm 2\pi$ result in a minimal loss as well and are thus not avoided by the machine learning algorithm. Figure

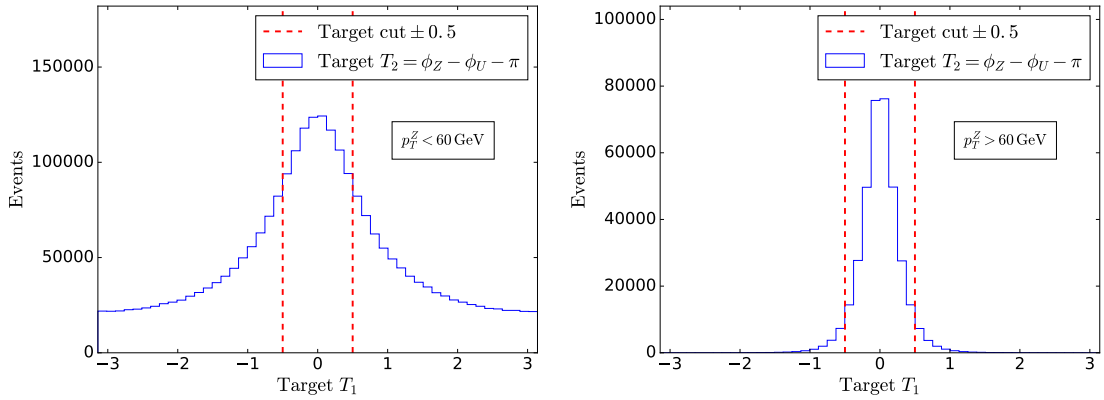


Figure B.1.: Illustration of the angular target distribution for events below $p_T^Z < 60\,\text{GeV}$ (left) and above. Whereas the nominal corrections for low p_T^Z events extend from $-\pi$ to π, corrections for high p_T^Z events are mostly in the range of -0.5 to 0.5. The dashed line of ±0.5 represents a maximal target rotation of ±28.6°.

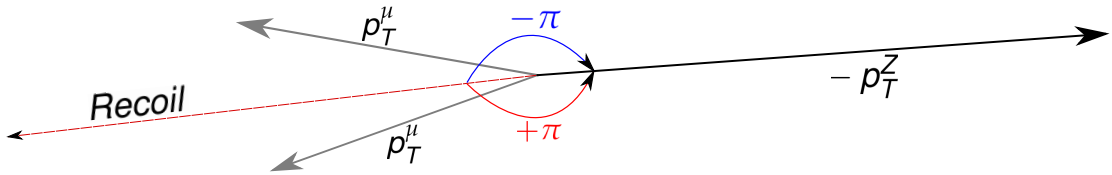
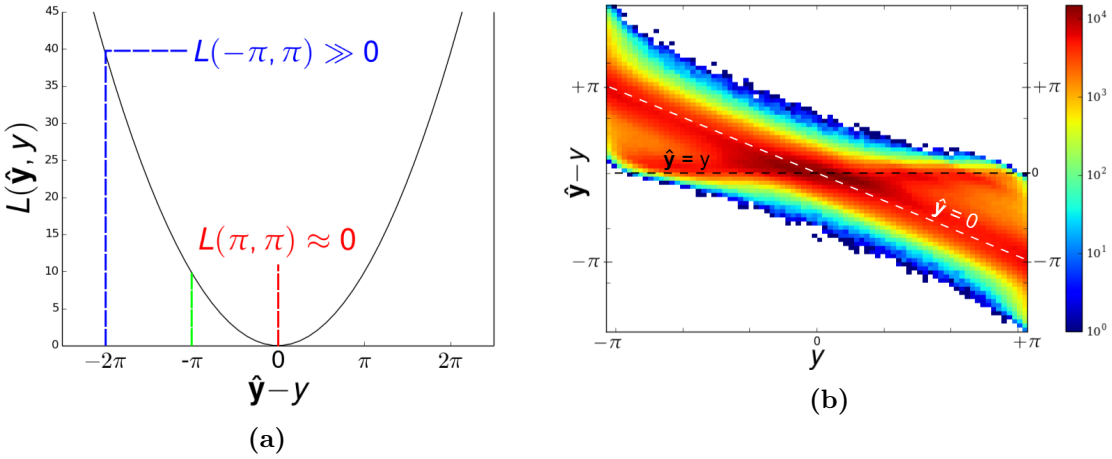


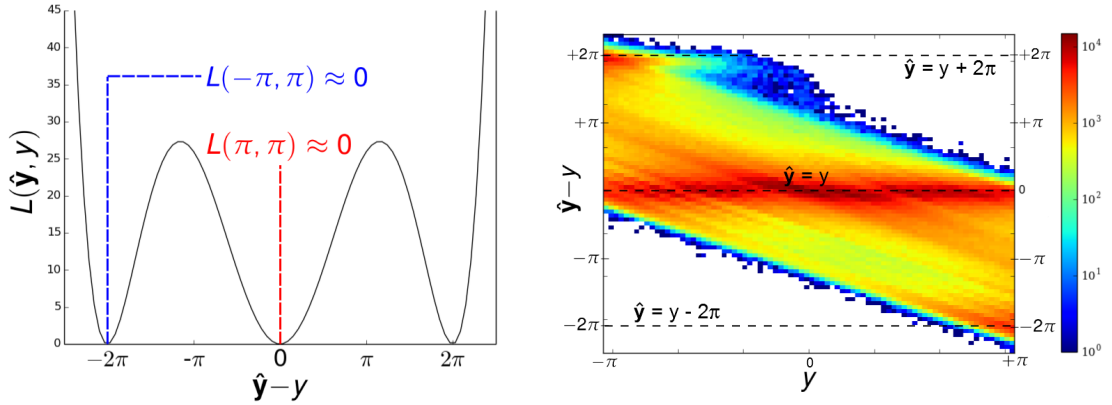
Figure B.2.: Possible angular correction scenario. The recoil vector has to be rotated by $\pm\pi$ to match the direction of p_T^Z.

B.4 shows this custom loss function as a multiplication of three parabolic functions with minima around zero and $\pm 2\pi$. It can be seen that the tendency of predicting zero for high nominal rotations could be repealed, resulting in an improvement of angular correction.



(a) (b)

Figure B.3.: Squared loss function with the two possible losses for a rotation of $\pm\pi$ on the left. The right plot shows a two-dimensional distribution of the difference between prediction and target as a function of the target value. The black dashed line indicates the preferred behaviour of the algorithm predicting the nominal target value in each event. The white dashed line indicates a zero prediction, meaning that the algorithm has predicted zero for events following this line.

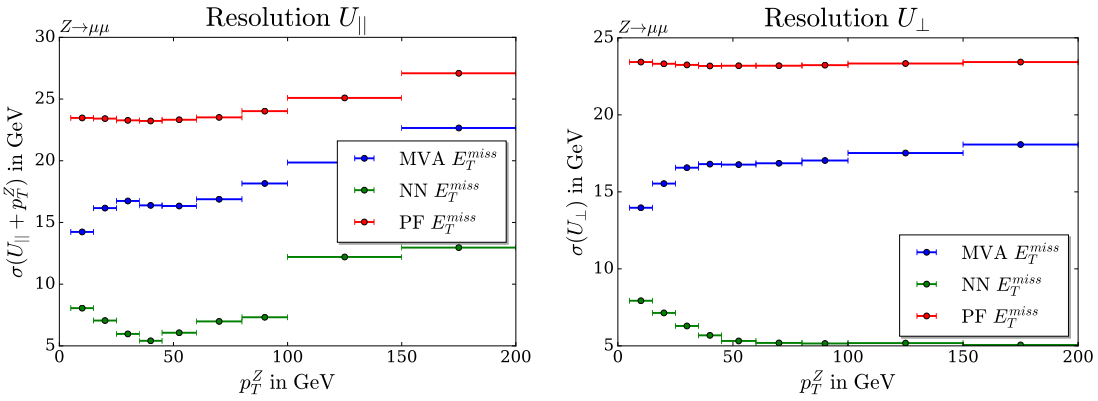


(a) (b)

Figure B.4.: Custom loss function with the two possible losses for a rotation of $\pm\pi$ on the left. Both scenarios result in a minimal loss for the algorithm. The right plot shows a two-dimensional distribution of the difference between prediction and target as a function of the target value. The black dashed line indicates the preferred behaviour of the algorithm predicting the nominal target value in each event or a difference of $\pm 2\pi$. Events are now mostly distributed around $\hat{y} = y$. Events with a high nominal rotation value are additionally distributed around $\hat{y} = y \pm 2\pi$.

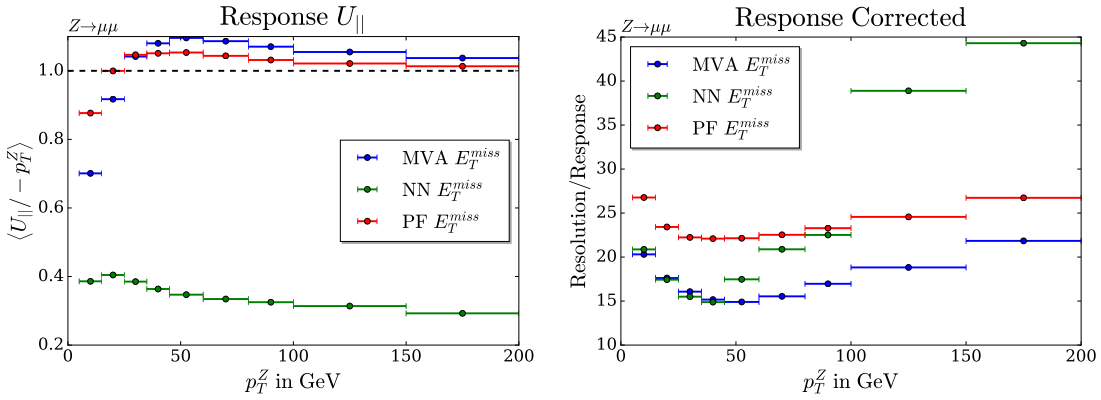
B.2. Evaluation on $Z \to \mu\mu$

This section shows the result of a neural network training on the scalar correction as introduced in section 5.2. The neural network was trained with 10 hidden layers with 500 neurons each. Figure B.5 shows the parallel and perpendicular resolution as a function of p_T^Z. NN$\not{E}_T$ yields an up to 80 % resolution gain compared to PF$\not{E}_T$, seeming to perform far superior to both PF$\not{E}_T$ and MVA$\not{E}_T$. However, figure B.6 shows response and response corrected for all three $\not{E}_T$ definitions. In contrast to MVA and PF$\not{E}_T$, NN$\not{E}_T$ shows an response around 30-40 % compared to unity-response. Combining both response and resolution, yields a comparable resolution for low p_T^Z with MVA$\not{E}_T$ but a worse response corrected value for high p_T^Z.
An explanation for this small resolution value and below-unity response can be found in figure B.7. Whereas the difference between nominal p_T^Z and parallel component of the recoil, $U_{||}$, is normal distributed around zero for PF$\not{E}_T$, the difference of both values is normal distributed around 46 GeV for NN$\not{E}_T$. Compared to PF$\not{E}_T$ the standard deviation of this distribution is small, leading to a small resolution value. However, this constant shift in resolution can also be observed in response, leading to a constant under-estimation of the recoil.
As mentioned before, the machine learning algorithm aims to minimize its loss function. If the loss is minimized by accurately predicting a value shifted from the target value more accurately and thus minimize its loss, this prediction is preferred over another approach in which it predicts less accurately the nominal value but at the cost of a larger variance and thus greater loss.
Both introduced scenarios reveal potential difficulties when dealing with modern machine learning algorithms. These algorithms are able to learn complex data structures and create predictions based on functions learned from its data. However, these functions may not lead to the intended behaviour of the trained algorithm, making it necessary to always precisely formulate the given task and determine a reasonable validation technique.



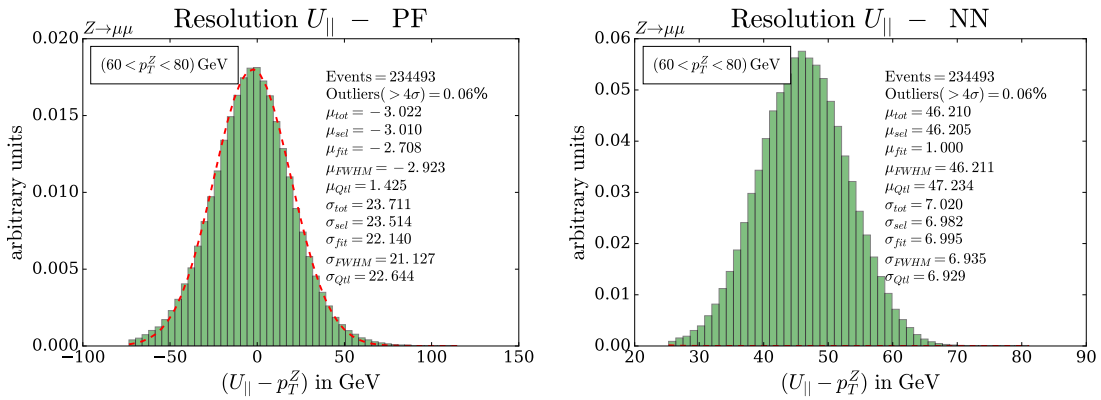
(a) **(b)**

Figure B.5.: Resolution$_{||}$ and Resolution$_\perp$ of MVA, NN and PF $\not{E}_T$ for the $Z \to \mu\mu$ decay.



(a) **(b)**

Figure B.6.: Response and Response corrected of MVA, NN and PF $\not\!\!E_T$ for the $Z \to \mu\mu$ decay.



(a) **(b)**

Figure B.7.: Event distribution for the resolution determination for PF (left) and NN $\not\!\!E_T$(right).

APPENDIX C

List of Figures

APPENDIX D

List of Tables

APPENDIX E

Bibliography

[1] Bogdan Povh et al. “Teilchen und Kerne: eine Einführung in die physikalischen Konzepte”. Springer-Verlag, 2013.

[2] David Griffiths. “Introduction to Elementary Particles”. Wiley-VCH GmbH & Co. KGaA, Weinheim, 2014. ISBN: 978-3527406012.

[3] Roger Wolf. “The Higgs Boson Discovery at the Large Hadron Collider”. Springer, 2015. Aufl. ISBN: 978-3-319-18512-5.

[4] Peter W. Higgs. “Broken symmetries, massless particles and gauge fields”. *eμ* Phys. Lett. 12.2 (1964), pp. 132–133.
DOI: 10.1016/0031-9163(64)91136-9.

[5] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. *eμ* Phys. Rev. Lett. 13 (9 1964), pp. 321–323.
DOI: 10.1103/PhysRevLett.13.321.

[6] Georges Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. *eμ* Physics Letters B 716.1 (2012), pp. 1–29.

[7] Serguei Chatrchyan et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. *eμ* Physics Letters B 716.1 (2012), pp. 30–61.

[8] Franz Mandl and Graham Shaw. “Quantum field theory”. 2. ed. Chichester: Wiley, 2010. ISBN: 978-0-471-49683-0; 978-0-471-49684-7.

[9] Ta-Pei Cheng and Ling-Fong Li. “Gauge theory of elementary particle physics”. Repr. Oxford science publications. Oxford [u.a.]: Clarendon Press, 2008. ISBN: 978-0-19-851961-4.

[10] CMS Collaboration. “The CMS Experiment at the CERN LHC”. *eμ* JINST 3, S08004 (Aug. 2008).
DOI: 10.1088/1748-0221/3/08/S08004.

[11] CMS collaboration et al. “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at sqrt (s)= 8 TeV”. *eμ* arXiv preprint arXiv:1502.02701 (2015).

[12] CMS collaboration et al. "Performance of CMS muon reconstruction in pp collision events at sqrt(s)= 7 TeV". *eμ* Journal of Instrumentation 7.10 (2012), P10002.

[13] "Interactive Slice of the CMS detector". URL: https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/RetrieveFile?docid=4172 (visited on 08/08/2017).

[14] Joram Berger. "Search for the Higgs Boson Produced via Vector-Boson Fusion in the Decay Channel H to Tau Tau". Karlsruhe, KIT, Diss., 2014. PhD thesis. 2014.

[15] CMS collaboration, CMS Collaboration, et al. *eμ* Particle-flow event reconstruction in CMS and performance for jets, taus and MET. 2009.

[16] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti k_T jet clustering algorithm". *eμ* Journal of High Energy Physics 2008.04 (2008), p. 063.

[17] Andres Munoz. "Machine Learning and Optimization". *eμ* URL: https://www.cims. nyu. edu/˜ munoz/files/ml_optimization. pdf [accessed 2016-03-02][WebCite Cache ID 6fiLfZvnG] (2014).

[18] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. "Searching for exotic particles in high-energy physics with deep learning". *eμ* arXiv preprint arXiv:1402.4735 (2014).

[19] Ian J Goodfellow et al. "Multi-digit number recognition from street view imagery using deep convolutional neural networks". *eμ* arXiv preprint arXiv:1312.6082 (2013).

[20] Peter J Huber et al. "Robust estimation of a location parameter". *eμ* The Annals of Mathematical Statistics 35.1 (1964), pp. 73–101.

[21] Nitish Srivastava et al. "Dropout: a simple way to prevent neural networks from overfitting." *eμ* Journal of machine learning research 15.1 (2014), pp. 1929–1958.

[22] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. "Deep Learning". http://www.deeplearningbook.org. MIT Press, 2016.

[23] Wikimedia Commons. *eμ* File:Colored neural network.svg — Wikimedia Commons, the free media repository. 2016.

[24] CMS collaboration et al. "Determination of jet energy calibration and transverse momentum resolution in CMS". *eμ* Journal of Instrumentation 6.11 (2011), P11002.

[25] Daniele Bertolini et al. "Pileup per particle identification". *eμ* Journal of High Energy Physics 2014.10 (2014), pp. 1–22.

[26] Friese Raphael and Günter Quast. "A new Multivariate Approach for the $H \to \tau\tau \to \mu\mu$ analysis". MA thesis. KIT, May 2013.

[27] Lorenzo Bianchini et al. "Reconstruction of the Higgs mass in H→ $\tau\tau$ events by dynamical likelihood techniques". *eμ* Journal of Physics: Conference Series. Vol. 513. 2. IOP Publishing. 2014, p. 022035.

Danksagung

Einen besonderen Dank möchte ich Herrn Prof. Dr. Günter Quast aussprechen für das Ermöglichen dieser Arbeit in seiner Arbeitsgruppe und die vielen konstruktive Rückmeldungen während den Gruppensitzungen. Herrn Dr. Roger Wolf möchte ich danken für die Übernahme des Korreferats und die vielen fachlichen Diskussionen, die zum Gelingen dieser Arbeit einen Großteil beigetragen haben.

Dr. Raphael Friese danke ich für die sehr kompetente und engagierte direkte Betreuung während des letzten Jahres. Ohne seine Hilfe wäre diese Arbeit nicht möglich gewesen.

Ebenfalls danken möchte ich meiner Arbeitsgruppe, insbesondere Stefan Wunsch und Sebastian Wozniewski, für die fachlichen Diskussionen, das offene und angenehme Arbeitsklima und die stete Hilfsbereitschaft hinsichtlich jeder Art von Problemen.

Dem Computing-Team und Administratoren des ETP danke ich für die Bereitstellung und Betreuung der internen und externen Computing-Infrastruktur.

Abschließend möchte ich die Rechenzentren BwForCluster NEMO in Freiburg und National Analysis Facility in Hamburg, auf denen ein Großteil der Berechnungen dieser Arbeit stattfanden, dankend nennen.