

Using symbolic regression to predict optimal observables for the measurement of the anomalous magnetic moment of the τ -lepton in PbPb collisions at the LHC

Bachelor Thesis

presented for the Degree of Bachelor of Science (B.Sc.) in Physics

Submitted by
Ruben Elias Bies

Supervisor:
Prof. Dr. Markus Schumacher

universität freiburg

University Freiburg
Faculty for Mathematics and Physics 3rd January 2024

Erklärung gem. der geltenden Prüfungsordnung:

Hiermit versichere ich, dass

1. ich die eingereichte Bachelorarbeit bzw. bei einer Gruppenarbeit meinen Anteil entsprechend gekennzeichneten Anteil der Arbeit selbständig verfasst habe,
2. ich keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommenen Inhalte als solche kenntlich gemacht habe und
3. die eingereichte Bachelorarbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens war oder ist.

Ort, Datum

Unterschrift

Abstract

This thesis investigates the determination of the anomalous magnetic moment of the τ -lepton in PbPb collisions at the LHC using Symbolic Regression. The anomalous magnetic moment a_ℓ for electrons and muons is among the most precisely measured physical quantities in the standard model (SM) of particle physics and could be sensitive to phenomena that are not explained by the SM. For the τ -lepton, its short lifetime prevents the measurement of the anomalous magnetic moment through the same methods as for the electron and muon, thus a different measurement technique is needed. The $\gamma\gamma \rightarrow \tau\tau$ process is observed in ultraperipheral PbPb collisions at the Large Hadron Collider (LHC), where the anomalous magnetic moment of the τ -lepton a_τ influences the production cross section as well as kinematic observables. One interesting class of kinematic quantities for this purpose is denoted as *optimal observables*. In this thesis, two versions of the optimal observables for a_τ are calculated. One depends on the helicities and parton level observables, it is called fully polarized optimal observable $OO_{full.pol.}$. The second averages over incoming and outgoing particle helicities and is called the helicity averaged optimal observable $OO_{hel.avg.}$. The simulation used in this thesis was generated to match the integrated luminosity of 2.0 nb^{-1} and centre of mass energy of 5.02 TeV from the data recorded on PbPb-collisions by the ATLAS experiment in 2015 and 2018. Events are selected in the signal region defined by one τ -lepton decaying leptonically (with a muon in the final state) and one decaying hadronically. A machine-learning method called Symbolic Regression is applied to predict the above-mentioned optimal observables. This method, carried out with the PySR tool, yields interpretable models in the form of analytic expressions of given input variables. A hyperparameter optimization was performed on models for $OO_{hel.avg.}$. The best constraint on a_τ yielding a confidence interval of $[-0.026, 0.012]$ at 95% confidence level was achieved with a model, trained on $OO_{full.pol.}$ and using final state kinematic observables as input. This result was extracted using the maximum likelihood method and outperforms the observable currently used in experiments, p_T^μ , by 4.6% for the simulated data.

Zusammenfassung

Diese Arbeit untersucht die Bestimmung des anomalen magnetischen Moments des τ -Leptons in PbPb-Kollisionen am LHC mit Hilfe von Symbolischer Regression. Das anomale magnetische Moment a_ℓ für Elektronen und Myonen gehört zu den am genauesten gemessenen physikalischen Größen im Standardmodell (SM) der Teilchenphysik und könnte empfindlich für Phänomene sein, die nicht durch das SM erklärt werden. Für das τ -Lepton ist es aufgrund seiner kurzen Lebensdauer nicht möglich, das anomale magnetische Moment mit denselben Methoden wie für das Elektron und das Myon zu messen, weshalb eine andere Messtechnik erforderlich ist. Der $\gamma\gamma \rightarrow \tau\tau$ -Prozess wird in ultraperipheren PbPb-Kollisionen am Large Hadron Collider (LHC) beobachtet, wo das anomale magnetische Moment des τ -Leptons a_τ sowohl den Produktionsquerschnitt als auch kinematische Messgrößen beeinflusst. Eine für diesen Zweck interessante Klasse von kinematischen Observablen wird als *optimale Observablen* bezeichnet. In dieser Arbeit werden zwei Versionen der optimalen Observablen für a_τ berechnet. Eine davon verwendet die Helizitäten und Partonlevel-Observablen, die als vollständig polarisierte (fully polarised) optimale Observable $OO_{full.pol.}$ bezeichnet wird. Die zweite Version mittelt über die Helizitäten der ein- und austretenden Teilchen und wird als helizitätsgemittelte (helicity averaged) optimale Observable $OO_{hel.avg.}$ bezeichnet. Die in dieser Arbeit verwendete Simulation erstellt, um der integrierten Luminosität von $2,0\text{nb}^{-1}$ und der Schwerpunktsenergie von $5,02\text{TeV}$ aus den vom ATLAS-Experiment in den Jahren 2015 und 2018 aufgezeichneten Daten zu entsprechen. Die Ereignisse werden in der Signalregion ausgewählt, die durch ein leptonisch zerfallendes τ -Lepton (mit einem Myon im Endzustand) und ein hadronisch zerfallendes definiert ist. Eine Methode des maschinellen Lernens namens Symbolische Regression wird angewandt, um die oben genannten optimalen Observablen vorherzusagen. Diese Methode, die mit dem PySR-Tool durchgeführt wird, liefert interpretierbare Modelle in Form von analytischen Ausdrücken für gegebene Eingangsvariablen. Eine Hyperparameter-Optimierung wurde für Modelle für $OO_{hel.avg.}$ durchgeführt. Die beste Einschränkung für a_τ mit einem Konfidenzintervall von $[-0,026, 0,012]$ bei einem Konfidenzniveau von 95%, wurde mit einem Modell erreicht, das auf $OO_{full.pol.}$ trainiert wurde und kinematische Endzustandsobservablen als Eingabe verwendet. Dieses Ergebnis wurde mit der Maximum-Likelihood-Methode extrahiert und führt zu einer Verbesserung gegenüber der derzeit in Experimenten verwendete Observable p_T^μ von 4,6% für die simulierten Daten.

CONTENTS

Introduction	7
1 Theoretical background	9
1.1 The Standard Model of Particle Physics	9
1.1.1 Fermions	10
1.1.2 Bosons	11
1.2 Anomalous magnetic dipole moments of leptons	11
1.3 The $\gamma\gamma \rightarrow \tau\tau$ process	12
1.3.1 Photon-Fermion coupling	13
1.3.2 Decay of the τ -lepton	13
2 The experimental setup	15
2.1 High energy physics experiments at the LHC	15
2.2 Coordinate systems and kinematics	15
2.3 Ultrapерipheral PbPb collisions	17
3 Optimal observables	18
3.1 Definition of optimal observables	18
3.2 Formalising optimal observables for a_τ	19
3.2.1 Helicity averaged optimal observable	19
3.2.2 Fully polarised optimal observable	19
4 Symbolic regression	21
4.1 Genetic programming for learning analytic expressions	21
4.2 Symbolic regression in python using PySR	22
4.2.1 The outer and inner loops of PySR	22
4.2.2 Distinguishing features compared to other SR applications	25
4.2.3 General overview of hyperparameters	27
5 The maximum likelihood method	29
5.1 Binned log-likelihood	29
5.2 Fit function and extracting confidence intervals	30
6 Event simulation chain	32
6.1 Tools used for Event Generation	32
6.2 Predicting kinematic distributions for specific a_τ values	33
6.3 Generated samples and their content	33
7 Analysis	35
7.1 Regression on the helicity averaged optimal observable	35

7.1.1	Hyperparameter search	36
7.1.2	Comparing symbolic models	37
7.1.2.1	Comparing by loss	37
7.1.2.2	Comparing by correlation	40
7.1.2.3	Comparing by confidence intervals	42
7.1.3	Interim conclusion	45
7.2	Regressing the fully polarised optimal observable	48
7.2.1	Special phase-space and helicity groups of $OO_{full.pol.}$	48
7.2.2	Selection of observables for regression	49
7.2.3	Choosing hyperparameters from earlier experience	50
7.2.4	Comparing symbolic models	51
7.2.4.1	Predicting $OO_{full.pol.}$ from the entire distribution	51
7.2.4.2	Predicting $OO_{full.pol.}$ from helicity groups 1 and 2	53
7.3	Final comparison of observables considering experimental viability	58
7.4	Recommendations for further studies of a_τ using symbolic regression	60
8	Conclusion	61
	References	63
	Acknowledgements	69
	Appendix	70

Introduction

Probing the microcosmos with modern technologies has been a quest in high-energy physics for decades. The standard model (SM) of particle physics [1, 2, 3] is currently the best theory describing experimental observations on a subatomic level. It describes the elementary particles that makeup matter and the four fundamental interactions. This theory was completed with the discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments at the Large Hadron Collider [4, 5]. There are however phenomena observed in experiments, that the standard model is not able to predict, such as dark energy [6] or the baryon asymmetry of the universe [7]. Therefore modern particle physics aims towards hints that indicate Beyond Standard Model physics.

The anomalous magnetic moment a_ℓ of charged leptons ℓ is a quantity with high sensitivity towards such effects [8]. For electrons and muons, it is under the most precisely known physical properties [9, 10] the latter one reporting a record tension of 4.2σ with its prediction [11]. This is a promising hint at new physics, that motivates further research into the topic as a potential source of new concepts for theories beyond the SM. In theory, the anomalous magnetic moment a_τ of the τ -lepton should yield the highest sensitivity for new physics [8]. However, the high mass and resulting short lifetime of the τ -lepton prevents its direct measurement by methods used for the electron or muon, to a precision that challenges its prediction of $a_\tau = 0.00117721(5)$ by the Standard Model [12]. The most precise current measurement of a_τ was conducted by the DELPHI collaboration [13] at the Large Electron Positron (LEP) collider to be $-0.052 < a_\tau < 0.013$ at 95% confidence level. This measurement cannot compete with the precision of the available prediction and therefore is not sufficient in testing the Standard Model.

In this thesis, a new method of extracting constraints on measurements of a_τ is investigated, using simulated data to match the measurements of the $\gamma\gamma \rightarrow \tau\tau$ process in ultraperipheral PbPb collisions by the ATLAS experiment at the Large Hadron Collider (LHC) [14]. The used data is simulated to match an integrated luminosity of 2.0 nb^{-1} and centre of mass energy of 5.02 TeV from the data taken at the ATLAS experiment in 2015 and 2018 [15]. This thesis uses the following decay mode combinations of the two τ -leptons: one decays leptonically with a muon and two neutrinos in the final state while the other decays into a charged pion, up to two neutral pions and one neutrino in the final state. In heavy ion collisions, a high photon flux is observed from the electromagnetic fields of the nuclei that pass each other with impact parameters larger than twice their radius. This photon flux results in light-by-light scattering that includes $\ell\ell$ production [16]. The cross-section of these processes scales with the fourth power of the proton number $Z = 82$ of the lead nuclei. In the τ -lepton production, the twice present vertex $\gamma\tau\tau$ is sensitive to the anomalous magnetic moment a_τ due to its influence on the photon-lepton coupling.

Current methods to extract constraints on a_τ use observables from final state particle kinematics such as p_T^μ of a detected final state muon from the τ -decay. Maximum likelihood fits are used on the observed distribution to estimate a_τ . The concept of *optimal observable* is a way of determining a parameter such as a_τ with maximised sensitivity [17, 18, 19]. They depend on matrix element calculations which in turn depend on parton-

level observables and helicities. Thus two optimal observables can be formalised: one that assumes helicities are known and is calculated for each event, it is called the fully polarised optimal observable $OO_{full.pol.}$, and one that averages the possible helicity combinations, called the helicity averaged optimal observable $OO_{hel.avg.}$, which is valid for all events equally [20]. The first is not usable in experiments, because helicities are difficult to reconstruct from final state kinematics.

The goal of this thesis is to use Machine Learning to learn optimal observables from particle kinematic observables. In contrast to many current trends to use non-interpretable neural networks, that can achieve great precision in modelling a problem from data, another method called Symbolic Regression is employed. This method learns models in the form of analytical expressions, by using a genetic algorithm that evolves populations of models over a number of generations via random mutations. The tool used for this study is PySR [21] by Miles Cranmer [22]. The observable $OO_{hel.avg.}$ will be learned from its parton observables as variables to conduct a small hyperparameter search over some options of PySR to gather an understanding of its behaviour. The results from this optimisation will be transferred to the regression of $OO_{full.pol.}$ from final state kinematic observables to obtain models applicable to experimental data. The sensitivity of different models will be compared by performing maximum-likelihood analysis to obtain confidence intervals on a_τ .

The structure of this Investigation is as follows: chapter 1 introduces the theoretical background that this study relies on, focusing on the calculation of the $\gamma\gamma \rightarrow \tau\tau$ process. The representative experimental setup, consisting of the Large Hadron Collider and the ATLAS experiment, for which data is simulated will be explained in chapter 2. Chapter 3 introduces the concept of optimal observables and how they are used for a_τ in particular. The Symbolic regression tool PySR is presented in chapter 4 and its algorithm to learn analytical expressions is explained in detail. Chapter 5 will describe the statistical analysis method of maximum likelihood that is used for extracting constraints on a_τ from observables. The data simulation process and tools involved are explained in Chapter 6. The results from learning optimal observables are discussed in chapter 7. The entire thesis will be summarised in chapter 8.

1 Theoretical background

This chapter will concisely introduce the basic physical principles necessary for this thesis. The first section 1.1 will give a brief overview of the general structure of the standard model of particle physics (SM). Section 1.2 describes the phenomenon of anomalous magnetic moments in leptons that is being investigated. the last section 1.3 will focus on the $\gamma\gamma \rightarrow \tau\tau$ process that this thesis studies, introducing the theory behind the photon-lepton coupling and defining the signal region of τ -lepton decay products that are relevant.

1.1 The Standard Model of Particle Physics

The standard model of particle physics [1, 2, 3] is to this day the best theory to describe the world at a subatomic level and since its inception, it was able to predict many discoveries like the W/Z bosons [23, 24] or the top quark [25]. Its core components can be seen in Figure 1.1, which are the fundamental particles we currently know to exist. It is the culmination of decades of experimental and theoretical research work, to find a universally applicable theory by describing particles as excitations of quantum fields [26] and their interaction via the three forces electromagnetic, strong and weak interaction, described by quantum-electrodynamics (QED), quantum-chromodynamics (QCD) and quantum flavourdynamics (QFD)[27].

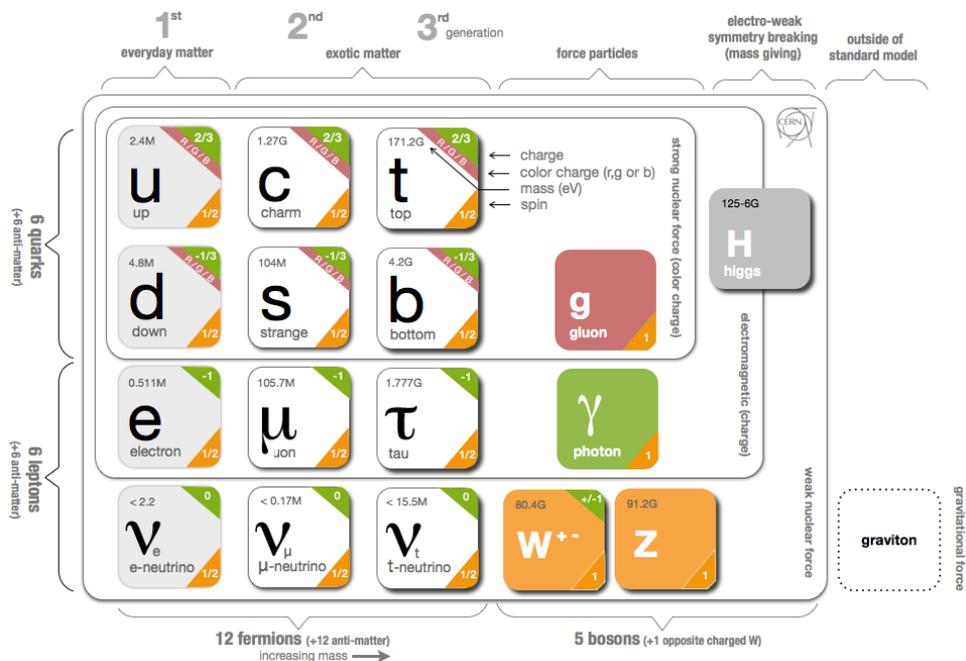


Figure 1.1: The standard model of particle physics with quarks, leptons and gauge-bosons and the Higgs field [28].

The particles we know today are separated into two groups. The fermions, which make up

the matter we encounter in our day-to-day life but also exotic matter that is only found in the early universe or inside of colliders. The second group are the bosons, three of which mediate the interaction between other particles and the fourth is the famous Higgs boson, needed to explain massive particles in the SM. These groups and their basic properties will be discussed in the following sections. Spin is an intrinsic property of elementary particles described as quantized angular momentum. It has a direction and an associated quantum number.

For this thesis the concept of helicity is important. Helicity is defined as the projection of the spin vector onto the vector of three-dimensional momentum [26].

$$h = \vec{p}\vec{s} \quad (1.1)$$

A particle with spin $\frac{1}{2}$ may have helicities of $h = +\frac{1}{2}$ or $h = -\frac{1}{2}$. It is a conserved property since momentum is also conserved. The helicities of incoming particles in any interaction will therefore influence the helicities of the outgoing particles. By continuation, helicities will influence the matrix elements and cross-sections of processes that are described by quantum field theories.

1.1.1 Fermions

Fermions are what make up matter in our universe as seen in figure 1.1. They have antiparticle counterparts with inverted charges. Fermions have a spin of $\frac{1}{2}$ and therefore obey Fermi-Dirac statistics, which means they behave according to the Pauli-principle and are indistinguishable [29]. They also have half-integer isospin I_W enabling them to interact via the weak force through the W^\pm bosons and weak charge Q_W for neutral current interactions of the weak force via the Z boson. There are three families of fermions, growing progressively heavier to the right of figure 1.1 [27]. Two of four species of fermions make up the quarks that have charges $-\frac{1}{3}$ or $\frac{2}{3}$. They are also the only fermions with a colour charge out of red, green or blue, enabling them to interact via the strong force, described by quantum chromodynamics (QCD). The other two species are called the leptons. There are the charged leptons, consisting of the electron, the muon and the τ -lepton. The electron has a mass of

$$0,51099895000(15) \text{ MeV}, \quad (1.2)$$

and is still stable. The muon has a mass of

$$m_\mu = 105,6583755(23) \text{ MeV} \quad (1.3)$$

resulting in a lifetime of

$$\tau_\mu = 2,1969811(22) \times 10^{-6} \text{ s} \quad (1.4)$$

and decays via the weak force. The τ -lepton has a mass and lifetime of

$$m_\tau = 1776,86(12) \text{ MeV} \quad (1.5)$$

$$\tau_\tau = 290,3(5) \times 10^{-15} \text{ s} \quad (1.6)$$

[30] and therefore decays very rapidly. Each of them has a corresponding neutrino, which make up the second lepton species. Neutrinos are not charged, almost massless, and interact only via the weak force [27]. Neutrinos are often produced in processes including other leptons of the same family, where they ensure the conservation of lepton flavour.

1.1.2 Bosons

There are 5 kinds of vector bosons that all have a spin of 1, they are called the gauge bosons and mediate one of three fundamental forces [27]. Gluons couple to the quarks via colour and anti-colour, acting as mediators for the strong force. They are electrically neutral and have no mass. Photons are bosons that couple to all charged particles and mediate for electromagnetic interaction via the electrical charge Q . They are electrically neutral themselves and do not have a colour charge or mass. The W^\pm/Z - bosons act as mediators for the weak force and couple to all fermions via weak charge Q_W and weak isospin I_W , they do not have colour charge. The W^\pm bosons have a charge of ± 1 , a weak isospin of $I_W = \pm 1$ and a mass of 80.377(12) GeV, while the Z boson is electrically neutral, has a weak isospin of $I_W = 0$ and a mass of 91.1876(21) GeV [30]. The only scalar boson with a spin of 0 is the Higgs particle that was discovered in 2012 [4, 5], it corresponds to the Higgs field, and through interaction with it, particles gain their masses. The Higgs bosons mass is 125,25(17) GeV, it is electrically neutral but interacts weakly.

Observations that are not described by the SM include dark energy [6], the neutrino masses [31] or quantum gravity [32]. We can test the ability of the SM to describe reality by measuring particle properties with precisions close to their predictions. The deviation of an experimental result to the underlying theory can therefore be quantified, which if significant, indicates new theories beyond the standard model (BSM).

1.2 Anomalous magnetic dipole moments of leptons

Every elementary particle that is electromagnetically charged and has a spin must have a magnetic moment as an intrinsic property. This influences the behaviour of particles inside electromagnetic fields as they will experience a force described by electrodynamics. The magnetic moment of leptons $\vec{\mu}$ is proportional to the spin and the gyromagnetic factor g and is defined as follows:

$$\vec{\mu} = g \frac{e}{2m} \vec{s} \quad (1.7)$$

where e is the elementary charge and m the mass of the lepton. Using the Dirac- and Pauli-equations together with the Gordon decomposition [29], the gyromagnetic factor is predicted to be $g = 2$ at Leading orders of the perturbative extension of QED. The anomalous magnetic moment of the lepton henceforth called a_ℓ , is defined as

$$a_\ell = \frac{g - 2}{2}. \quad (1.8)$$

Higher order corrections yield non-vanishing values of a_ℓ . The QED correction is $a_\ell = 2$, while other contributions like QCD have a smaller influence.

Influences from beyond the Standard Model (BSM), also manifest as a term $a_\ell(BSM)$ in a_ℓ . This term could couple (and therefore would be proportional) to the mass of a lepton

for supersymmetric BSM theories (SUSY) [8].

$$a_\ell(BSM) \propto \frac{m_\ell^2}{\Lambda^2} \quad (1.9)$$

where Λ denotes the energy scale at which the BSM effects contribute. Potential BSM physics from SUSY theories will be most notable for the τ -lepton. Because it is the heaviest lepton it has a high potential for sensitivity towards new physics.

Most recent experiments at the Fermilab National Accelerator Laboratory have shown a discrepancy between the measurements of a_μ for the muon [10] and the SM prediction of 4.2σ [11], which is by far the strongest hint for BSM contributions currently. But efforts to measure a_τ to a precision that challenges the SM prediction [12] of

$$a_\tau^{SM} = 0.00117721(5)$$

have so far been challenging due to the short lifetime of the τ -lepton. The best measurement stems from the Large Electron Positron Collider (LEP) by the DELPHI Collaboration in 2004[13]:

$$-0.052 < a_\tau < 0.013,$$

at 95% confidence level. The most recent measurements have been carried out by the ATLAS collaboration [33]. Results from the L3 and OPAL collaborations are also presented [34, 35]:

$$\begin{aligned} -0.057 < a_\tau^{ATLAS} < 0.024, \\ -0.068 < a_\tau^{OPAL} < 0.065, \\ -0.052 < a_\tau^{OPAL} < 0.058 \end{aligned}$$

These are still an order of magnitude away from the SM prediction.

1.3 The $\gamma\gamma \rightarrow \tau\tau$ process

In this thesis, the light scattering process in which two τ -leptons are produced will be investigated. This process occurs via electromagnetic interaction, where non vanishing values of a_τ will alter the cross-section and the kinematics of the τ -lepton in the final state. The contributing Feynman diagrams for the process are shown in figure 1.2. The $(\gamma\gamma\tau)$ vertex appearing in the $\gamma\gamma \rightarrow \tau\tau$ makes it sensitive to changes in a_τ .

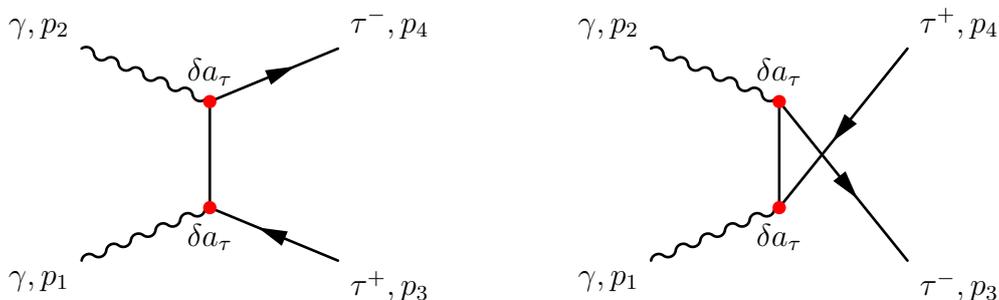


Figure 1.2: The Feynmann diagram for the $\gamma\gamma \rightarrow \tau\tau$ process at leading orgers of QED(Quelle).

1.3.1 Photon-Fermion coupling

The amplitude M for the $\gamma\gamma \rightarrow \tau\tau$ process can be calculated from the Feynman diagrams for t- and u-channel process depicted in figure 1.2. With photon momenta p_1, p_2 and τ^+ and τ^- momenta p_3 and p_4 respectively [36, 37]. The helicities $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ matching to the momenta as variables are also needed.

$$M = (-i)\epsilon_{1\mu}(\lambda_1)\epsilon_{2\nu}(\lambda_2)\bar{u}(p_3, \lambda_3) \left(i\Gamma^{(\gamma\tau\tau)\mu}(p_3, p_t) \frac{i(p_t + m_\tau)}{t - m_\tau^2 + i\epsilon} i\Gamma^{(\gamma\tau\tau)\nu}(p_t, -p_4) \right. \\ \left. + i\Gamma^{(\gamma\tau\tau)\nu}(p_3, p_u) \frac{i(p_u + m_\tau)}{u - m_\tau^2 + i\epsilon} i\Gamma^{(\gamma\tau\tau)\mu}(p_u, -p_4) \right) v(p_4, \lambda_4) \quad (1.10)$$

where the following relations are used: $p_t = p_2 - p_4 = p_3 - p_1$ and $p_u = p_1 - p_4 = p_3 - p_2$. There are the polarization vectors for the photons $\epsilon_1(\lambda_1), \epsilon_2(\lambda_2)$, the spinors of the anti- τ -lepton and the τ -lepton $\bar{u}(p_3, \lambda_3)$ and $v(p_4, \lambda_4)$ and the mandelstamm variables $u = p_u^2$ and $t = p_t^2$. The term of $\frac{i(p_t + m_\tau)}{t - m_\tau^2 + i\epsilon}$ is identified as the τ -lepton propagator. The vertex function $i\Gamma_\mu^{\gamma\tau\tau}(p', p)$ with p' being the incoming and p being the outgoing lepton momenta and $q = p' - p$ the momentum transfer, can be written as:

$$i\Gamma_\mu^{\gamma\tau\tau}(p', p) = -ie \left[\gamma_\mu F_1(q^2) + \frac{i}{2m_\tau} \sigma_{\mu\nu} q^\nu F_2(q^2) + \frac{1}{2m_\tau} \gamma^5 \sigma_{\mu\nu} q^\nu F_3(q^2) \right] \quad (1.11)$$

with the spin tensor $\sigma_{m\nu\nu} = \frac{i}{2}[\gamma_\mu, \gamma_\nu]$, where γ_μ, γ_ν refer to the Dirac-matrices. $F_1(q^2) = 1$ is the Dirac Factor. The Pauli form factor of the lepton $F_2(q^2)$, connects to the anomalous magnetic moment as $a_l = \frac{g-2}{2} = F_2(0)$. $F_3(q^2)$ denotes the electric dipole form factor that connects to the dipole moment d_ℓ via $F_3(q^2) = d_\ell \frac{2m_\tau}{e}$.

1.3.2 Decay of the τ -lepton

The average lifetime of τ -leptons is $\tau_\tau = 2.903 \times 10^{-13}$ s [30] resulting in a decay length $l = \beta\gamma c\tau_\tau = 87.03 \times 10^{-6}$ m [30], which is too short to reach current detectors in most experimental setups. τ -leptons are therefore identified only by their decay products. The τ -lepton decay occurs via the weak interaction and is categorised into two groups of decay modes. Either the τ -lepton decays into lighter leptons and two neutrinos, which is called leptonic decay or into hadrons - mainly pions or Kaons - and one neutrino, which is denoted as hadronic decay. The presence of neutrinos in the decay ensures the conservation of lepton number [27]. In table:1.1 the most common decay modes and their branching ratio for the τ -lepton are listed. Hadronic decays make up around 65% of the decays, whereas leptonic decays are only around 35%. Decays with one charged particle (hadronic or leptonic), among the final state particles, account for the majority of decays with 80%, also denoted as 1-prong decay, while decays with 3 charged particles amount to 20%, also denoted as 3-prong decay [30].

decay mode	branching ratio (%)
$\pi^- \pi^0 \nu_\tau$	25.494 ± 0.089
$\mu^- \bar{\nu}_\mu \nu_\tau$	17.394 ± 0.038
$e^- \bar{\nu}_e \nu_\tau$	17.817 ± 0.040
$\pi^- \nu_\tau$	10.817 ± 0.051
$\pi^- 2\pi^0 \nu_\tau$	9.260 ± 0.096
$2\pi^- \pi^+ \nu_\tau$	8.987 ± 0.051

Table 1.1: Most dominant τ decay modes and their branching ratios [30].

This thesis focuses on the decay modes $\mu^- \bar{\nu}_\mu \nu_\tau$ and $\pi^- n\pi^0 \nu_\tau$ and $n = 0, 1, 2$. We also allow for charge conjugated cases since there is one τ -lepton and one anti- τ -lepton in the $\gamma\gamma \rightarrow \tau\tau$ process. In total, the signal region used in this thesis is $\mu^\pm \pi^\pm n\pi^0$ also called "one muon and one track" (1M1T) in the final state particles that are observed. With this signal region a total branching ratio for $\mu^- \pi^+ n\pi^0$ and its charge conjugated case of

$$BR_{1M1T} = 2 \cdot BR_{\mu^-} (BR_{\pi^-} + BR_{\pi^- \pi^0} + BR_{\pi^- 2\pi^0}) = 0.1549 \quad (1.12)$$

is covered.

2 The experimental setup

This thesis uses simulated data to investigate the potential use of observables from machine learning. These simulations emulate events that might occur in studying the $\gamma\gamma \rightarrow \tau\tau$ process inside a detector at the European Organisation for Nuclear Research (CERN) [38]. The detector considered for measuring the collision products is the ATLAS detector [14], one of the four large detectors placed around the interaction points of the Large Hadron Collider (LHC) [39]. This chapter introduces concepts needed for the investigation of collider-simulated data. Section 2.1 presents the LHC and section 2.2 introduces coordinate systems of the ATLAS detector. Ultrapерipheral heavy ion collisions are discussed in more detail in section 2.3.

2.1 High energy physics experiments at the LHC

The LHC [39, 40] uses electric fields alternating at high frequencies to accelerate charged particles and magnetic fields to steer them along defined paths and create a beam consisting of separated bunches of particles. These bunches are initialised by a linear accelerator and then sent to several pre-accelerator rings before continuing to the largest one where they reach their ultimate collision velocity within less than 1% of the speed of light c . Superconducting magnets are used to guide the beam along the accelerator, for aligning the beam for collision, and for focusing and controlling beam position. They operate at temperatures near the absolute 0K and produce fields around 8T. Bunches are tightly packed with particles and are guided to collide at specific points around the accelerator where experiments are set up.

At the LHC there are four points where these beams are brought into collision, they mark the places where the Experiments at the LHC are located: ATLAS, CMS, LHC-b and ALICE [41, 42, 43, 44]. The LHC is currently running at centre-of-mass energies of $\sqrt{s} = 13.6$ TeV for pp collisions with a peak luminosity of 22×10^{-33} cms^{-1} [45]. For run 2 from 2015 to 2018. The pp collision data amounts to an integrated luminosity of 147.0fb^{-1} . Ultrapерipheral PbPb collisions are used to generate $\gamma\gamma$ events. PbPb data was collected at $\sqrt{s} = 5.0$ TeV and to an integrated Luminosity of 1.76nb^{-1} for run 2 [46]. The data collected during run 1 and run 2 in 2015 and 2018 for PbPb collisions at the LHC amounts to an integrated Luminosity of $\mathcal{L} = 2.0$ nb^{-1} [15], which is considered in this thesis.

2.2 Coordinate systems and kinematics

Detectors in high energy physics (HEP) are devices, capable of measuring almost all of the collision products and their decay products, with a solid angle of 4π . They reconstruct the paths particles take and their energies on small time scales.

The ATLAS detector at the LHC has a cylindrical shape, with a diameter of 25m, 44m in length and weighing almost 7000 tons [14]. It has an inner detector with a radius of 1.2 m inside a 2 T solenoid magnetic field, it is dedicated to capturing tracking information. The calorimeters that extend from around the inner detector to about 4 m, measure the energy deposited by electrons, photons and hadrons. The muon system is the outermost and largest part, immersed in an air-core toroidal magnetic field.

The Coordinate system of the ATLAS detector is a righthanded, cylindrical system. The x-axis is pointed towards the centre of the LHC. The direction of the beam in a counterclockwise motion defines the z-axis, with the azimuthal angle θ measured against this beam axis. The x-y or transverse plane defines the radial coordinate r and the polar angle ϕ is measured from the x-axis. In the following definitions for kinematics, natural units have been used with $\hbar = 1$ and $c = 1$ for the reduced plank constant \hbar and the speed of light c . The coordinate system is shown in figure 2.1.

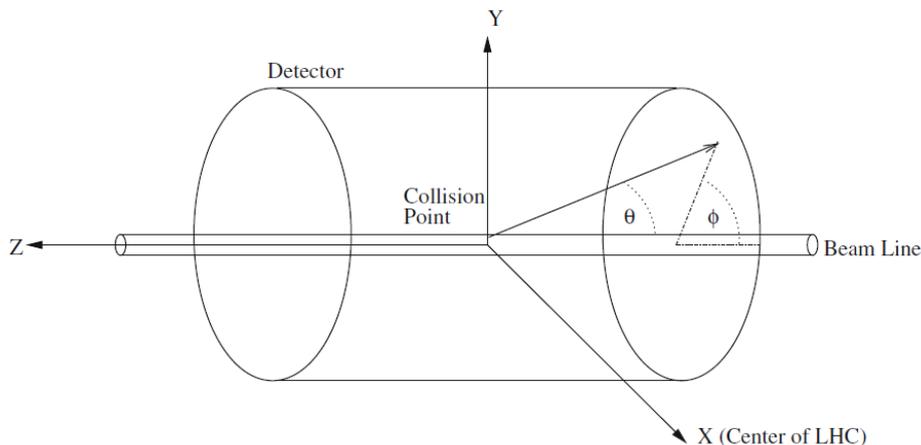


Figure 2.1: The coordinate system of the ATLAS detector [47], which will be used in this thesis.

In this coordinate system, the incoming photons in the rest frame of the $\gamma\gamma\tau\tau$ system will have the four-momenta

$$p_{\gamma_1}^{\vec{}} = (E, 0, 0, E), \quad (2.1)$$

$$p_{\gamma_2}^{\vec{}} = (E, 0, 0, -E), \quad (2.2)$$

with a center of mass energy of

$$\sqrt{s} = \sqrt{(p_{\gamma_1} + p_{\gamma_2})^2} = 2E. \quad (2.3)$$

For the outgoing τ -leptons, $\Phi_{\tau} = 0$ can be chosen because the system is symmetrical under rotation around the beam-axis. Because of energy conservation, the energy for each τ -lepton is known to be $E_{\tau} = E$. With the relation $E_{\tau} = \sqrt{m_{\tau}^2 + p_{\tau}^2}$, with the rest mass of the leptons m_{τ} and the 3-momentum p_{τ} , their four-momenta are calculated to be

$$p_{\tau_1}^{\vec{}} = (E_{\tau}, p_{\tau} \sin(\theta), 0, p_{\tau} \cos(\theta)), \quad (2.4)$$

$$p_{\tau_2}^{\vec{}} = (E_{\tau}, -p_{\tau} \sin(\theta), 0, -p_{\tau} \cos(\theta)), \quad (2.5)$$

where $\theta = \theta_{\tau_1} = \theta_{\tau_2}$ because of momentum conservation. We define the pseudorapidity η as

$$\eta = -\ln(\tan(\theta/2)), \quad (2.6)$$

The transverse energy E_T and the transverse momentum p_T for the τ -leptons can be defined as:

$$p_T = \sqrt{p_x^2 + p_z^2} = p_\tau \sin(\theta) \quad (2.7)$$

$$E_T = \sqrt{E_x^2 + E_y^2} = E_\tau \sin(\theta), \quad (2.8)$$

The cartesian momentum components using p_T, η and ϕ are given as:

$$p_x = p_T \cos(\phi), \quad (2.9)$$

$$p_y = p_T \sin(\phi), \quad (2.10)$$

$$p_z = p_T \sinh(\eta). \quad (2.11)$$

The centre-of-mass frame describes the collision process from a non-moving point of view. The detector will measure coordinates that are boosted along the beam axis via a Lorentz transformation since the particles move with relativistic velocities. Any variables in this thesis are regarded to be in the centre-of-mass frame.

2.3 Ultraperipheral PbPb collisions

Ultraperipheral collisions are characterized by impact parameters twice as large, as the radius of the colliding objects. At the LHC there are different kinds of particles used in these collisions, mostly ions like Gold, Lead or Xenon. PbPb collisions are used to generate high cross-sections of $\gamma\gamma$ interactions. The high impact parameters allow the electromagnetic fields of the colliding nuclei to interact with each other, which can be described as the emission and interaction of two quasi-real photons radiated of the original particles [16]. The advantage is that if the nuclei do not actually collide, there will be no additional background from QCD interactions. The EM fields of the nuclei will be stronger, depending on the proton number Z , therefore the cross-section of lead collisions will be larger compared to others, scaling with $\sigma \sim Z^4$. In the LHC, lead (Pb) is used due to its high proton number $Z = 84$ and therefore high cross-section for light scattering.

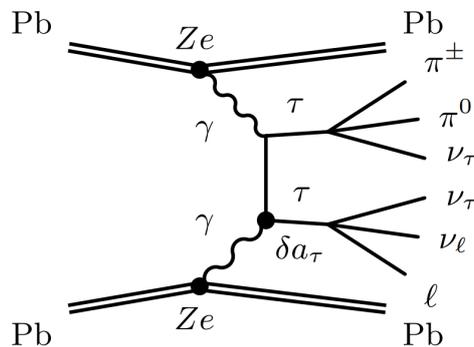


Figure 2.2: Feynman diagram of the production of $\tau\tau$ from ultraperipheral PbPb collisions [48].

Figure 2.2 shows the Feynman diagram of the production of $\tau\tau$ from ultraperipheral PbPb collisions, including one leptonic and one hadronic decay of the τ -leptons.

3 Optimal observables

Quantities such as a_τ are determined by analyzing the distributions of observables reconstructed from final state particle kinematics. Different observables will provide different sensitivity for a_τ and expected uncertainties Δa_τ . It has been proven, that there is a single best observable for a given process and a desired parameter called the optimal observable (OO). It has the smallest uncertainty per definition [17, 18, 19, 49]. Optimal observables have been used before by the ATLAS collaboration in a test of CP-invariance in vector-boson fusion production from the Higgs boson [50]. In this chapter, a general definition of OO is given in section 3.1. In section 3.2 the concept of OO is applied to determine a_τ from the $\gamma\gamma \rightarrow \tau\tau$ process.

3.1 Definition of optimal observables

The concept of optimal observables can be applied to processes, in case the squared matrix element can be written as a polynomial expansion in terms of the parameter of interest [17, 18]. Suppose the squared matrix element $|M|^2$ was in such a form for the parameter of interest p_i :

$$|M|^2 = |M_0|^2 + \sum_i p_i |M_i|^2. \quad (3.1)$$

The observable

$$OO_i = \frac{|M_i|^2}{|M_0|^2} \quad (3.2)$$

is the most efficient estimator for measuring p_i , denoted as optimal observable [19, 49]. For $\gamma\gamma \rightarrow \tau\tau$ production, the matrix element defined in section 1.3.1, can be written as series in $\lambda_i = a_\tau$:

$$M = M_0 + a_\tau M_1 + a_\tau^2 M_2 \quad (3.3)$$

resulting in a squared matrix element of

$$|M|^2 = |M_0|^2 + a_\tau \cdot 2\text{Re}\{M_0^* M_1\} + \dots, \quad (3.4)$$

where higher order terms have been neglected, assuming $a_\tau \ll 1$. The Optimal observable for the desired parameter a_τ is given by

$$OO = \frac{2\text{Re}\{M_0^* M_1\}}{|M_0|^2}. \quad (3.5)$$

3.2 Formalising optimal observables for a_τ

The evaluation of OO according to equation 3.5 requires that the matrix elements $|M_0|^2$ and $2Re\{M_0^*M_1\}$ can be reconstructed from the final state momenta. The matrix element in equation 1.10 depends on the momenta and helicities of the photons and τ -leptons. In experiments, helicities are difficult to obtain since an entire setup is needed to measure the spin of a particle [51]. In this section, the method of calculating the OO without knowing any helicities is presented first, then in case all the helicities are known.

3.2.1 Helicity averaged optimal observable

The $\gamma\gamma \rightarrow \tau\tau$ process has 16 possible helicity combinations of the four parton helicities $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ as seen for equation 1.10. However for the photons, helicities can not be reconstructed. For the τ -leptons, it is difficult to reconstruct helicities from final state particles. Hence the first method is to average all 16 possible matrix elements. This yields an optimal observable, which is called "helicity averaged optimal observable" or $OO_{hel.avg.}$. The helicities denoted as $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ for the four partons $\gamma_1, \gamma_2, \tau_1, \tau_2$ can have different possible values based on particle spin. $\lambda_{1,2}$ can have values of ± 1 for the photons and $\lambda_{3,4}$ are $\pm \frac{1}{2}$ based on the τ -lepton spin.

$$OO_{hel.avg.} = \frac{\sum_{\lambda_1, \lambda_2, \lambda_3, \lambda_4} 2Re\left\{\left(M_0^{\lambda_1, \lambda_2, \lambda_3, \lambda_4}\right)^* M_1^{\lambda_1, \lambda_2, \lambda_3, \lambda_4}\right\}}{\sum_{\lambda_1, \lambda_2, \lambda_3, \lambda_4} \left|M_0^{\lambda_1, \lambda_2, \lambda_3, \lambda_4}\right|^2} \quad (3.6)$$

The above expression will be the same for all events and only depend on parton kinematic observables. It has been recently calculated for the $\gamma\gamma \rightarrow \tau\tau$ process in [20], using the method described in section 1.3.1 for calculating the amplitudes.

$$OO_{hel.avg.} = \frac{4(\beta_\tau^2 \cos^2 \theta_\tau - 1)}{\beta_\tau^4(1 + \sin^4 \theta_\tau) - 2\beta_\tau^2 \sin^2 \theta_\tau - 1}, \quad (3.7)$$

where θ_τ is the azimuthal angle of the τ -lepton in the centre of mass rest frame, it will be the same for both τ -leptons due to momentum conservation. β is calculated via the lorentz-factor to be $\beta_\tau^2 = 1 - \frac{4m_\tau^2}{s}$ with the rest mass of the τ -leptons m_τ and the squared invariant mass s which is the mandelstam variable $s = (p_1 + p_2)^2 = (p_3 + p_4)^2$. With β being a value in the interval $[0, 1]$ and θ being restricted to $[0, \pi/2]$, one can calculate that equation 3.7 in total is restricted to the interval $[2, 4]$. This analytic expression will be used as a target for machine learning that is applied to the problem.

3.2.2 Fully polarised optimal observable

For simulated data, the particle helicities are known for each event. Therefore the exact Optimal observable can be calculated, which is different for each of the 16 helicity com-

binations in the $\gamma\gamma \rightarrow \tau\tau$ process. The resulting optimal observable is called the "fully polarized optimal observable" or $OO_{full.pol.}$.

$$OO_{full.pol.} = \frac{2Re \left\{ \left(M_0^{\lambda_1, \lambda_2, \lambda_3, \lambda_4} \right)^* M_1^{\lambda_1, \lambda_2, \lambda_3, \lambda_4} \right\}}{\left| M_0^{\lambda_1, \lambda_2, \lambda_3, \lambda_4} \right|^2} \quad (3.8)$$

There is no compact general analytic expression for this observable since the helicities influence the matrix elements from equation 1.10 directly. It can still be used for training in ML, to investigate its dependency on final state particle variables. It would be desirable if helicities influence of the final state in a way, that ML is able to learn the distribution of $OO_{full.pol.}$.

4 Symbolic regression

Symbolic regression (SR) describes a method of machine learning, which is able to learn analytic expressions via regression from (simulated) data [52]. This stands in contrast to other methods from ML such as neural networks (NN) where instead of searching for the best model, the goal is to optimize the parameters of a pre-defined model structure. However, deep learning in NNs has become the default numerical method in physics and many other fields. Their data-driven learning can achieve great precision in modelling specific problems as modern computing resources are enabling us to run these highly intensive algorithms on big amounts of data that are available nowadays [53]. The drawback is that NNs are not easily interpretable, as they can consist of millions of parameters like weights and biases, which makes traceability of predictions by NNs by humans very difficult. The latest developments in ML approach physics from the standpoint of algebraic formulae, which are more interpretable and theory-related [54]. SR has thus been explored as a way to re-discover known laws in physics from data [55, 56, 57, 58, 59], and is now establishing itself as a new and exciting tool for the physics community for obtaining approximations in form of analytical expressions for specific problems.

This chapter will give an introduction to SR in section 4.1 and then will explain the structure and algorithm of the tool called PySR, which is used in this thesis, in section 4.2.

4.1 Genetic programming for learning analytic expressions

The most common algorithm to achieve SR is via *genetic programming*, where the survival of the fittest models is simulated over several generations [52, 60]. Where fitness is some function that quantifies the prediction performance of a given model with respect to the targeted data. Genetic programming is an umbrella term for ML methods, that randomly change a set of models and evolve them towards a more effective version of themselves. When model fitness dictates the probability of reproduction, expressions with higher fitness pass their features on to the next generation. Over the course of many generations, a population of analytic expressions is going to be increasingly effective at solving the regression problem. The model fitness is typically measured by a loss function like in other ML applications and is re-evaluated for each model after each generation. A less complex solution is chosen at the cost of some fitness. This requires measuring and penalising higher complexity in the expressions.

4.2 Symbolic regression in python using PySR

PySR [61] is a SR tool for python [62], based on the Julia library SymbolicRegression [63]. It is open source, first published in 2020 and has already been successfully used in many different fields [64, 65, 66, 67, 56] including particle physics at the LHC [68, 69]. PySR is user-friendly and has many customisation options for a variety of problems. It is a genetic algorithm that evolves a set of expressions by randomly mutating them and applying a fitness function to measure performance compared to earlier versions of themselves.

In PySR and many other tools, expressions are internally represented as trees, with each operator being a node and variables or constants being so-called leaves. Nodes can have entire sub-trees, called branches attached to them whereas leaves signify the end of a branch. Expressions can be evaluated iteratively and it is easy to manipulate them by pruning or exchanging nodes and their branches. In figure 4.1 the expression tree for the analytical form of $OO_{hel.avg.}$ from equation 3.7 is shown.

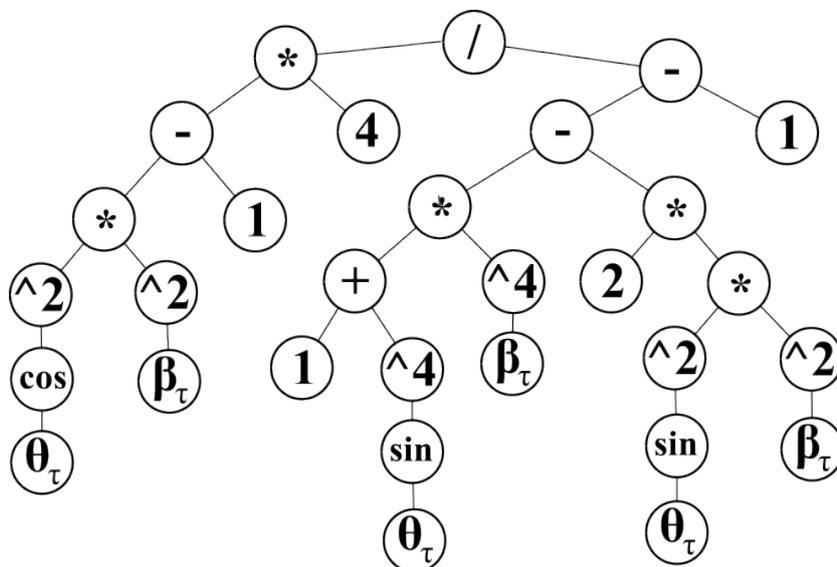


Figure 4.1: Example of an expression tree used by PySR for the analytic form of $OO_{hel.avg.}$.

This tree has a maximum depth of 8 and a complexity of 30. Depth is counted as the maximal number of nodes between the topmost node, called the root and leaf. Complexity is the total number of nodes. The default ways to determine depth and complexity is of assign all nodes the same value of 1 but this can change according to custom definitions. There are the binary operators $\{+,-,*,/\}$ as well as some unary operators like $\{x^2, \sin(x), \cos(x)\}$. The goal is to learn analytic expressions such as the equation 3.7 from data given the 2 variables β and θ of the τ -leptons.

4.2.1 The outer and inner loops of PySR

The structure of PySR is organised in inner and outer process loops [70]. The outer loop consists of generations that carry out the evolution process for each population thus leaving them in a fitter state than before. The inner loop consists of applying a number of

changes to the current population by selecting an individual for 'breeding' and replacing a member of the population with its new form.

At the start of the algorithm, PySR will initialise a number of populations, which are a set of individual models represented as expression trees. These individuals are initialised randomly but all at a set complexity of 3, which at the time this study was concluded was not tunable. The populations will evolve separately, going through several mutations. PySR will keep track of the fittest models at each possible complexity across populations by storing them inside a hall of fame (HOF) after each generation. After the HOF has been updated, PySR will perform migrations between populations, where each individual in each population has a certain probability to be exchanged with a random individual from another population or the HOF. This is done in order to diversify populations such that overfitting is avoided. The outer loop of PySR is represented in figure 4.2, where populations are represented as islands of expressions, evolving over the course of many generations.

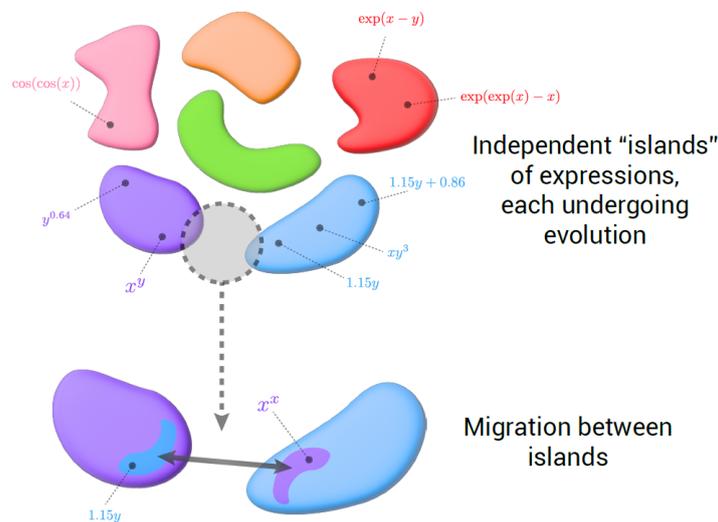


Figure 4.2: Outer loop of PySR: populations evolve separately from each other except when crossover events occur [70].

In the inner loop of PySR, displayed in figure 4.3, a given number of mutations will be carried out within each generation. For each so-called *cycle* of the inner loop, the population is randomly sampled to extract a subsample of the population and for each subsample, the individual for "breeding" is selected [70]. Selecting this individual from the subsample is done by a tournament function. The tournament will select the fittest expression with a certain probability, if not, the expression is removed from the subsample and this step is repeated. If only one individual is remaining, this will be selected. The type of "breeding" operation will be chosen randomly, corresponding to either mutation, crossover, simplification or optimisation of constants.

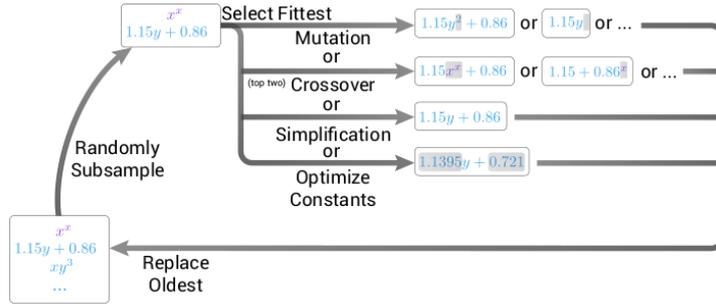
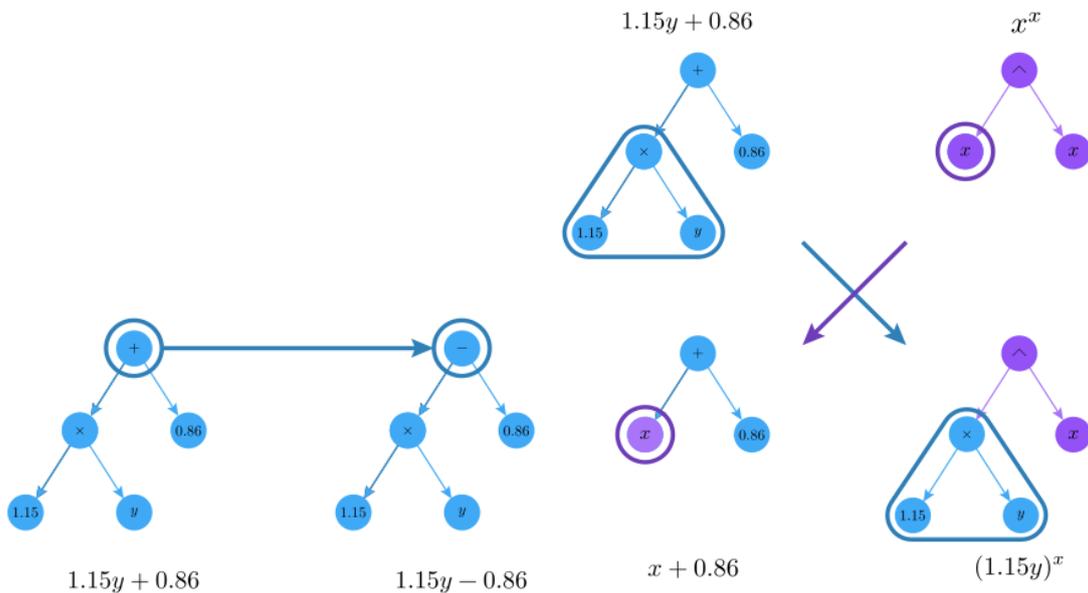


Figure 4.3: The inner loop of PySR: a population is separated into a set of samples, each undergoing the 'breeding' process via a random choice from mutation, crossover, simplification or optimisation [70].

Mutations are changes to an expression like adding, removing or altering nodes. An example of mutation is visualised in figure 4.4a. Crossover refers to randomly exchanging subtrees with another expression, selected by a second tournament on the same subsample (where crossover with itself is possible). If crossover is chosen, the two oldest members of the population get replaced by the two expressions. An example for crossover is displayed in figure 4.4b. Simplification will remove unnecessary complexity by applying a set of algebraic rules on the individual. For instance, the simplification operation might reduce the expression $x/x^2 + 1.3 + 0.7$ to $x^{-1} + 2$. The optimisation operation will run a classical optimisation algorithm on the constants of the selected expression for a few iterations. After this, the new expression will replace the oldest one from the population with a certain probability. How many cycles of this inner loop are carried out for each generation is a customisable option.



(a) Example of a mutation operation on an expression tree, changing a node [70]. (b) Example of a crossover operation on an expression tree, changing a branch [70].

Figure 4.4: Examples for the PySR operations mutation and crossover on expression trees.

By combining the inner and outer loops, PySR is a complex algorithm that tries to simulate a "survival of the fittest" among semi-independent populations of models. As

with evolution in nature, this process heavily relies on the randomness of mutation, initial state and many other factors. The same training will almost never yield the same results and can heavily fluctuate in its performance upon repetition. A number of parameters, steering the PySR configuration - such as mutation probabilities, etc.- also denoted as hyperparameters, can be set and be modified in addition .

4.2.2 Distinguishing features compared to other SR applications

There are several features of PySR, that distinguish it from other applications and enhance the flexibility of the algorithm. Since pySR is an open-source project, it is continuously developed further by the author and their contributors [21]. The ML community helps to test and further aid the team in optimizing the algorithm and its features. The most notable features are listed and shortly described here, roughly in order of importance to the functioning of PySR [70].

Punishing complexity: PySR has two ways of calculating the loss L of a given equation that defines fitness, where lower loss means better fitness. L will consist of the classical mean-squared error

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2, \quad (4.1)$$

where n is the number of events in the sample, y is the prediction and \hat{y} denotes the true values of the target. A custom loss function can be defined and used in stead of L_{MSE} . L has an additional term L_C that considers the complexity C of the given equation. L_C can be the product of C with a parsimony hyperparameter p , penalising higher complexities, resulting in

$$L = L_{MSE} + p \cdot C. \quad (4.2)$$

The other method is to penalise exponentially using the frequency f of complexity. The frequency measures, how often the complexity of the given equation is already present in the current population and will punish more common complexities. This gives the total loss as

$$L = L_{MSE} + \exp\{(f(C))\}. \quad (4.3)$$

Using frequency thus encourages searching a wide range of possibilities of the model space. A hyperparameter called adaptive parsimony scaling is used, to scale the exponential contribution of more complex models in addition to frequency.

Evolve-simplify-optimize: PySR performs simplification as well as optimisation regularly after all cycles of evolution from the inner loop have terminated. The reason to simplify at the end of the evolution process is that some forms of equations are only accessible via redundant steps. For example, $x \cdot x - x \cdot x$ would simplify to 0 but a mutation could change it to $x \cdot x - x \cdot y$, which is a new form entirely. Optimisation only happens after simplification, ensuring no optimisation on constants is performed that might be combined in simplification.

Simulated annealing: PySR can adjust the probability of accepting the new equation after the evolution via simulated annealing. Given an annealing temperature $T \in [0, 1]$, the probability for rejection is

$$p = \exp\left(\frac{L_n - L_o}{\alpha T}\right), \quad (4.4)$$

where L_n and L_o are the losses of the new and original equation respectively, such that all equations with higher loss than their original are rejected automatically with a probability > 1 . α is a hyperparameter that controls the scale of temperature, where $\alpha \rightarrow 0$ corresponds to rejecting all new equations with lower loss than their original. The annealing temperature T will change over the course of each generation according to

$$T = 1 - \frac{k}{n_c}, \quad (4.5)$$

where k is the current cycle of the inner loop and n_c denotes the total number of cycles the inner loop will carry out. This enables PySR to alternate between high temperatures at the start of each generation, which diversifies the population and lower temperatures which specialise the population at the end of each generation. Simulated annealing is not used by default.

Custom operators and constraints: A PySR model chooses from a pre-defined set of operators, that will be selected when creating or modifying equations. One can pass PySR custom operators as long as they are defined as either unary or binary functions of real numbers. One can also pass constraints on both custom and pre-set operators that control operator nesting. For instance, expressions like $\sin(\sin(\sin(x)))$ can be prevented by setting the nesting constraint of \sin and itself to 0. Other constraints can be selected to control the entire form of expressions like maximum depth and complexity.

Custom complexities and weights: In PySR almost all probabilities and many other hyperparameters can be customised. This includes that in the accounting for the complexity of an equation, the complexity of specific operators, or variables, or constants can be set to other values than the default of 1. Weights can be included in the training process to specify the importance of different events from the training data.

Model selection: PySR knows three different methods of selecting the final model from the HOF: 'accuracy', 'score' and 'best'. The method 'accuracy' will select the model with the lowest loss. The 'score' method will select the model with the highest score, which is defined as negated derivative of the log-lossfactoring in complexity. Smaller complexity is preferred over a slightly higher loss but a much smaller loss will justify selecting a higher complexity. 'best' will select the model with the highest score but only among expressions with a loss better than 1.5x the most accurate model.

Exporting models: Different representations of the final model can be extracted from PySR such as sympy, jax or pytorch. Extra mappings for custom operators and variable names need to be passed to PySR to access this feature. It is useful to visualise the different models and their analytic forms.

4.2.3 General overview of hyperparameters

The most important hyperparameters that are varied in this thesis for the results discussed in chapter 7 are listed in table 4.1, together with their default values. These default values are changed for particular studies, if specified explicitly.

name	default value	explanation
niterations	40	number of generations
ncyclesperiteration	550	number of mutations per generation
populations	15	number of parallel populations
population_size	33	number of expressions per population
model_selection	'best'	method to select final model
use_frequency	True	if True will use frequency

Table 4.1: Most important hyperparameters of PySR used in this thesis and their default values.

In table 4.2 the less important parameters for this thesis are listed together with their default values. Values that are modified from the default are listed in section 7.1 explicitly.

name	default value	explanation
maxsize	20	maximum nr. of nodes
maxdepth	none	maximum nr. of nodes from root to leave
adaptive_parsimony_scaling	20	punish higher complexity with frequency
weight_optimize	0	probability to optimize constants
parsimony	0.0032	factor to punish complexity
unary_operators	[]	if frequency is not used operators that take one input
binary_operators	['+', '-', '*', '/']	operators that take two inputs
annealing	False	if true, use annealing

Table 4.2: Hyperparameters, not used in the hyperparameter search, and their default.

A combination of these hyperparameters is used to define a PySR model [21] as follows:

```

1 from pysr import PySRRegressor
2
3 model = PySRRegressor(
4     populations = 15,
5     population_size = 33,
6     niterations = 40,
7     ncyclesperiteration = 550,
8     maxsize = 20,
9     maxdepth = None,
10    loss = 'L2DistLoss()',
11    weight_optimize = 0,
12    model_selection = 'best',

```

```

13     parsimony = 0.0032,
14     binary_operators = ['+', '-', '*', '/'],
15     unary_operators = [],
16     adaptive_parsimony_scaling = 20,
17     use_frequency = frequency,
18 )

```

The default base loss function is 'L2DistLoss()', which represents the mean squared error on the prediction.

The PySR model is trained calling the *fit* function as follows:

```

1 model.fit(X, y, variable_names=None)

```

It requires a list of events X of shape $(n_{samples}, n_{features})$, where $n_{samples}$ is the number of events in the training set and $n_{features}$ is the number of variables, given for the training process. The target Y of shape $(n_{samples})$ needs to be given as well [21]. This function will carry out the learning algorithm as described in section 4.2.1 and store the final model locally. Variable names need to be passed for visualisation if the equations need to be exported.

5 The maximum likelihood method

The method of maximum likelihood is used to find an estimation of a parameter of interest such as a_τ and its uncertainties from a limited data set [71, 68, 33]. Assuming a distribution $f(x; p)$ of a random variable x and parameter p that is unknown. If x has been measured repeatedly, resulting in a set $\{x_1, x_2, \dots, x_n\}$, the probability for the first measurement to be inside the interval $[x_1, x_1 + dx_1]$ can be defined as $f(x_1, p)dx_1$. Since the measurements of x are assumed to be independent, the probability that the values x_i are in $[x_i, x_i + dx_i]$ then given by

$$\prod_{i=1}^n f(x_i, p)dx_i. \quad (5.1)$$

Therefore a value p far away from the true value will result in a lower probability for the measured data, while values near the true value yield a high probability. The *Likelihood* function L is defined as

$$L(p) = \prod_{i=1}^n f(x_i; p). \quad (5.2)$$

By scanning the parameter space of p , the Likelihood function $L(p)$ can be evaluated and used to obtain an estimate of the true value \hat{p} by maximising L . Approximate confidence intervals at confidence levels 68%(95%), corresponding to $1\sigma(1.96\sigma)$ around \hat{p} , can be obtained from the values of p , where $\ln L(p)$ decreased by $0.5(1.92)$ with respect to the maximum [71].

The negative \ln of L is commonly used because it converts the products into sums and allows for minimising instead of maximising, which are both numerically easier to handle. Since \ln is a monotonously increasing function, the same value that maximise $L(p)$ will also minimise $-\ln L(p)$. This is called the negative-log-likelihood function.

$$-\ln L(p) = -\sum_{i=1}^n \ln(f(x_i, p)) \quad (5.3)$$

5.1 Binned log-likelihood

With large data sets, it is convenient to use histograms with N entries $\{n_1, \dots, n_N\}$ rather than computing the log-likelihood function for every value x_i . The expectation values $\{\mu_1, \dots, \mu_N\}$ in the histogram bins $\{1, \dots, N\}$ are defined as

$$\mu_i(p) = n_{tot} \int_{x_i^{min}}^{x_i^{max}} D(x; p)dx, \quad (5.4)$$

where x_i^{max} and x_i^{min} denote the upper and lower limits of bin i and $n_{tot} = \sum_{i=1}^N n_i$ is the total number of entries in the histogram. Assuming a Poisson distribution for each histogram bin

$$D(n_i; \mu_i(p)) = \frac{\mu_i^{n_i}(p)}{n_i!} \exp(-\mu_i(p)), \quad (5.5)$$

the extended negative-log-likelihood function is found to be

$$-\ln L(\mu(p); p) = \mu_{tot} - \sum_{i=1}^N n_i \ln(\mu_i(p)), \quad (5.6)$$

where μ_{tot} is the sum of all expected entries μ_i . The best estimate \hat{p} is defined by

$$-\ln L(\hat{p}) = -\ln L_{min} \quad (5.7)$$

The maximum of the negative-log-likelihood can be subtracted from its distribution, resulting in the difference in negative-log-likelihood ΔNLL which is used in this thesis. The best estimation and the confidence intervals become

$$\Delta NLL(\hat{p}) = 0 \quad (5.8a)$$

$$\Delta NLL(\hat{p} \pm 1\sigma) = 0.5 \quad (5.8b)$$

$$\Delta NLL(\hat{p} \pm 1.96\sigma) = 1.92 \quad (5.8c)$$

5.2 Fit function and extracting confidence intervals

After scanning the values of the parameter p , interpolation between the points of the ΔNLL curve is needed. In this thesis, the curve is fitted using a polynomial of degree 12. The fit uses the data analysis framework ROOT [72] and its built-in fitting method for graphs and histograms [73]. This fit algorithm applies the least-square (χ -square) method for fitting data without weights. Values for the best estimate of the parameter p can be extracted, by identifying the values defined by equations 5.8. The goodness of the polynomial fit is tested by determining χ^2/n_{df} where χ^2 is defined as

$$\chi^2 = \sum_i \frac{x_i - y_i}{\sigma_i} \quad (5.9)$$

where x_i denote the observed ΔNLL values and y_i the fit prediction, σ_i is the variance of the value x_i . Since no errors on the ΔNLL are extracted, the fitting tool [73] assumes them to be given by

$$\sigma_i = \frac{\sum_i (x_i - y_i)^2}{n_{df}} \quad (5.10)$$

The number of degrees of freedom n_{df} is given by $n_{df} = N - m$, where N denotes the number of points used in the fit and m , is the number of parameters in the fit function, in this case, they are $N = 50$ and $m = 12$. It is desired that $\frac{\chi^2}{n_{df}}$ be around 1, if it is much

larger, then the fit is not optimal. If it is much less than 1 then the fit is better than expected, given the estimated errors σ [71].

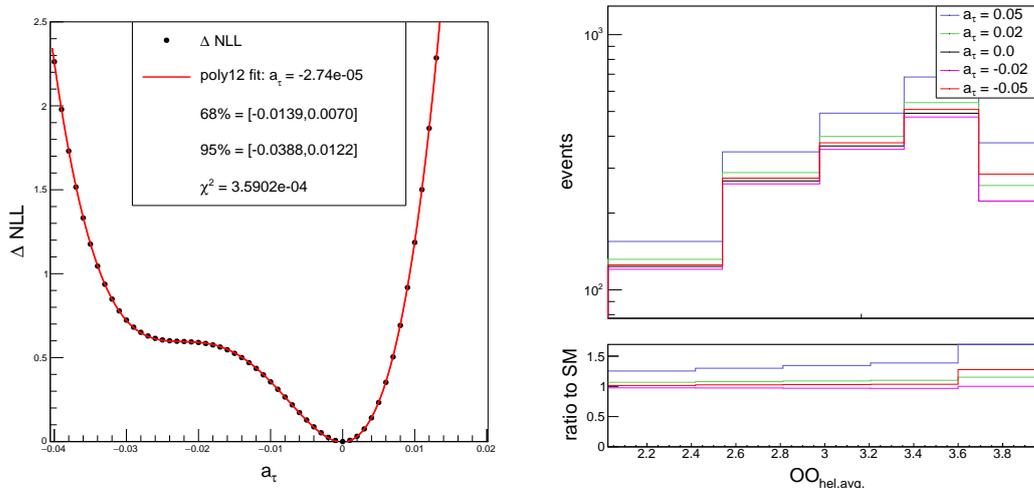


Figure 5.1: NLL fit example for the $OO_{hel.avg.}$ with a polynomial fit and an example for different histograms for calculating ΔNLL .

Figure 5.1 shows an example of the NLL method extracting a_τ from the helicity averaged optimal observable from equation 3.7 and its distribution. The relevant values of a_τ are in a range of $[-0.05, 0.05]$, where the step width was 0.001. The small χ^2 value indicates a better fit than expected. This would normally be a reason to verify that the errors σ_i are not being overestimated but for the purpose of this thesis, it is not a decisive factor. For the corresponding histograms, five values of a_τ are shown exemplarily. Similar figures for the particle level variable p_T^μ and the fully polarized optimal observable are given in appendix 8.1. The extracted confidence intervals and their length at 95% confidence level are:

$$-0.03878 < \hat{a}_\tau^{hel.avg.} < 0.01224, \quad range = 0.051019, \quad (5.11)$$

$$-0.00492 < \hat{a}_\tau^{full.pol.} < 0.00409, \quad range = 0.008987, \quad (5.12)$$

$$-0.02797 < \hat{a}_\tau^{p_T^\mu} < 0.01223, \quad range = 0.040196. \quad (5.13)$$

$$(5.14)$$

They are important comparison values for any learned observable. p_T^μ is currently the best experimental observable, since $OO_{full.pol.}$ requires knowing the helicities. They all agree with the SM prediction of $a_\tau = 0.0$.

The choice of histogram range and binning can influence the resulting confidence intervals. Empty bins in the predicted distribution will result in an error for the maximum likelihood fit when computing ΔNLL , since the logarithm of the expected bin contents is needed. For any maximum likelihood fits involving $OO_{hel.avg.}$ and its models, the variable range was constrained to $[2,4]$, covered by 5 equidistant bins. In the case of the $OO_{full.pol.}$ the variable range was chosen to only reach $OO_{full.pol.} = 500$ with logarithmic bin distribution in order to avoid empty bins. It was observed, that the bin limit of $OO_{full.pol.} = 1000$ can lead to a $\sim 10\%$ larger confidence interval. The distribution of p_T^μ is consistent enough, to set the bin maximum to the distribution maximum without generating empty bins while using a logarithmic bin distribution.

6 Event simulation chain

The basic method of Simulating random events is the Monte Carlo principle [74]. In physics, simulation is the common way of getting theoretical predictions for the experimental processes [68]. In this chapter, an overview of the simulation chain to generate the $\gamma \rightarrow \tau\tau$ events is given. The used tools are described in section 6.1 while the method of predicting kinematic distributions for specific a_τ values will be discussed in section 6.2.

6.1 Tools used for Event Generation

5 tools are involved in creating the sample of simulated events used in this thesis: gamma-UPC, Madgraph5, Pythia8, Photos and Rivet. Each of them has a different place and function in the simulation chain. Their function and usage will be explained briefly.

gamma-UPC is a library for simulating the emission of photons from ultraperipheral proton and nuclear collisions [75]. It calculates $\gamma\gamma$ processes using electric dipole or charge form factors. It can therefore simulate the PbPb collisions used in this thesis.

Madgraph5 is a framework for many kinds of SM and BSM phenomenons. It is used for the computations of cross-sections and the generation of scattering events at high momentum transfer. It can calculate leading order accuracy for user-defined Lagrangians and next to leading order in case of QCD corrections to SM processes [76]. Madgraph5 is used in this thesis for calculating the matrix elements and phase space of the $\gamma \rightarrow \tau\tau$ process using the four-momenta and helicities of the partons.

Pythia8 is a program for the generation of high-energy physics collision events. It can calculate a number of aspects, such as hard and soft interactions, parton distribution functions, initial- and final-state parton showers, multiparton interactions, fragmentation and decay [77]. For this thesis, it simulates the decay of the tau leptons.

Photos is a C++ algorithm for calculating bremsstrahlung in the decay of particles and resonances [78]. In this thesis, it was used to simulate the QED radiation of initial and final state particles.

Rivet is a system for validating Monte Carlo event generators. It includes a large set of experimental analyses for MC generator development, validation, and tuning [79]. Here it defines the fiducial selection of signal regions at particle level. For selecting the signal region, in addition to requiring one μ^\pm , one π^\pm and zero electrons, two conditions were set. The first is on the combined transverse momentum of muon and charged pion to be $p_T^{\mu^\pm+\pi^\pm} > 1$ GeV. The second condition is on the acoplanarity of muon and charged pion to be $A_\Phi^{\mu^\pm+\pi^\pm} > 0.1$. These are the same conditions used for the analysis from the ATLAS collaboration on a_τ [80].

6.2 Predicting kinematic distributions for specific a_τ values

For predicting the distribution of observables for different a_τ values needed for the ΔNLL fits, matrix element reweighting is used [81]. A weight for any event from the $\gamma\gamma \rightarrow \tau\tau$ process is given by

$$w = f_1(x_1)f_2(x_2)|M|^2\Omega_{PS} \quad (6.1)$$

where $f_i(x_i)$ are the parton distribution functions with the momentum fractions x_i given from the colliding photons to the τ -leptons. Ω_{PS} denotes the phase-space volume associated with the event. A weight relies on the squared matrix element $|M|^2$ that changes with the assumption of the value for a_τ . Therefore changing the weight is done by adding a ratio of matrix elements.

$$w_{new} = \frac{|M_{new}|^2}{|M_{old}|^2}w_{old}. \quad (6.2)$$

The matrix elements are calculated by Madgraph5 using the four-momenta and helicities of the partons as described in section 1.3.1. The values of a_τ for which weights will be calculated are in the interval of $[-0.1, 0.1]$. For values in $[-0.1, -0.05]$ and $[0.05, 0.1]$ a step width of 0.01 was chosen and for values in the interval $[-0.05, 0.05]$ the step width was set to 0.001 for better resolution from the likelihood method near the prediction of $a_\tau = 0.0$.

6.3 Generated samples and their content

Two samples have been generated for this thesis. They include particle four-momenta quantified by the columns p_x, p_y, p_z, e corresponding to the cartesian momentum and energy of each particle. For charged particles, the charge prefix is given by +1 or -1 to identify which particles came from the τ - and anti- τ -leptons. Since between 0 and 2 neutral pions are allowed in the signal region, the number of included π^0 n_{π^0} is given and their four-momentum in case they exist. Both samples also include the OOs that were introduced in section 3.2 for each event. Where $OO_{hel.avg.}$ will only depend on parton kinematics and $OO_{fullpol}$ will consider the helicity combinations for each event via the Matrix elements needed.

The SM sample has been generated assuming $a_\tau = 0.0$. Since this is a dedicated sample, each event has the same weight which is a constant. 45 million events were requested, out of which $\sim 115 \times 10^3$ were selected by Rivet to be in the 1 muon 1 track signal region. This sample will be used for training PySR. In all comparisons of a_τ estimates, one would expect $a_\tau \hat{=} 0.0$ to be the prediction.

The BSM sample has been generated for an $a_\tau = +0.06$. This sample has event weights for other a_τ values, which were obtained using the matrix element reweighting method described in section 6.2. 35 million events were requested, out of which $\sim 114 \times 10^3$ are in

the signal region. It contains the weights for all possible a_τ values from the reweighting method, and can therefore be used to acquire the corresponding observable distributions to calculate the ΔNLL curve.

Both samples are compared in figure 6.1 using the p_T^μ observable. For the BSM sample the weights for $a_\tau = 0.0$ were used for comparability. Both distributions are virtually the same except for statistical fluctuations at higher values of p_T^μ . This shows, that the BSM sample is fit for testing models trained on the SM sample.

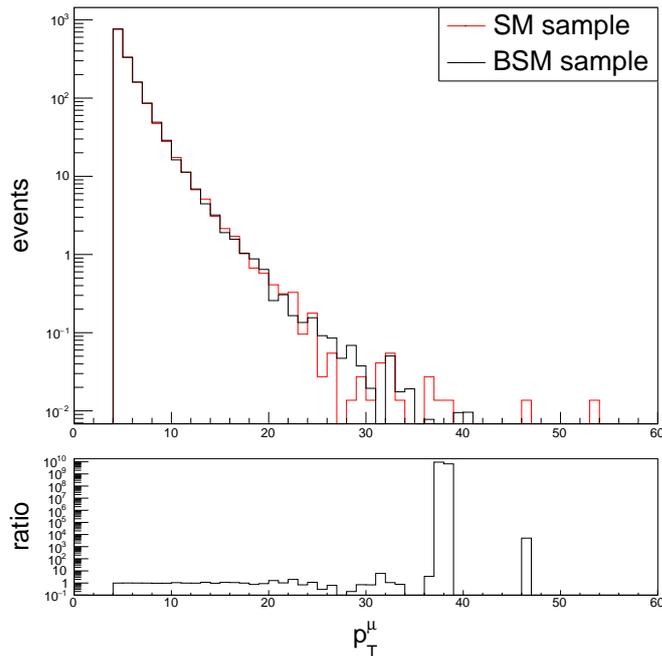


Figure 6.1: The kinematic particle property p_T^μ is shown for both the SM and the BSM sample by comparing with their ratio.

In this thesis, all distributions from events from either sample are normalized to the expected events for the integrated luminosity $\mathcal{L} = 2.0 \text{ nb}^{-1}$ from run 1 and run 2 data at the ATLAS experiment. Dividing the integrated luminosity by the number of generated events N_{events} yields the corresponding cross-section $\sigma(PbPb \rightarrow PbPb + \tau\tau)$ according to

$$\mathcal{L} = \frac{N_{events}}{\sigma(PbPb \rightarrow PbPb + \tau\tau)}. \quad (6.3)$$

The weight w is normalised by the calculated cross section to obtain the new weight corresponding to the expected events:

$$w_{norm} = w \frac{\mathcal{L}}{N_{events}}, \quad (6.4)$$

where w is the weight calculated by the matrix element for the event according to equation 1.10. N_{events} denotes the total amount of generated events. With this normalisation, it can be calculated that for $a_\tau = 0.0$, the SM and BSM samples a total of 1463 and 1472 events respectively are expected in the signal region.

7 Analysis

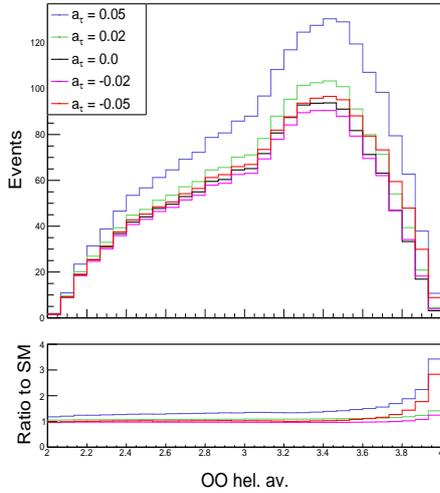
The PySR tool explained in chapter 4 can be used to learn observables from kinematic variables of the $\gamma\gamma \rightarrow \tau\tau$ process and its final state products. As targets for the learning process the optimal observables introduced in section 3.2 are used. ΔNLL fits will be performed on the model observables learned by symbolic regression for determining confidence intervals for a_τ as described in chapter 5. In section 7.1 the helicity averaged optimal observable $OO_{hel.avg.}$ is learned from parton level quantities. A group of hyperparameters of the PySR tool will be investigated for their influence on performance. The second section 7.2 will transfer the interim conclusion from the first section onto the learning of the fully polarised optimal observable $OO_{full.pol.}$ from final state particle quantities. In this thesis, the models are compared by their performance, first by ranking them by smallest mean loss. Then a comparison is drawn to the correlation between the model prediction and the true OO. Finally, the models are compared by sorting them by the smallest 95% confidence interval for a_τ obtained from the maximum likelihood fits. A final comparison between Observables learned via symbolic regression will be presented in 7.3.

7.1 Regression on the helicity averaged optimal observable

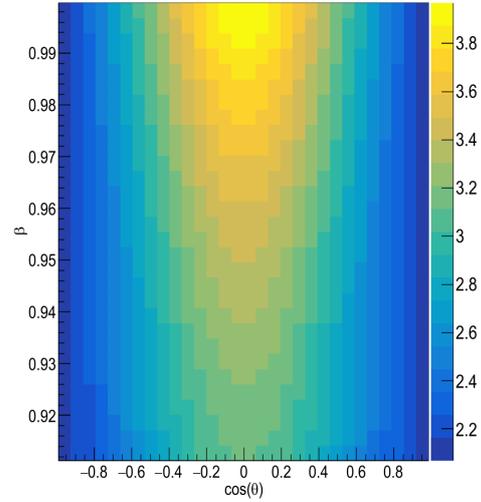
The helicity averaged optimal observable $OO_{hel.avg.}$ is a good target for ML because its analytic form is known. The symbolic models for the OO can be compared to its true analytic expression. From equation 3.7 one can see that $OO_{hel.avg.}$ depends on the two parton level observables $\beta = 1 - \frac{4m_\tau^2}{s}$ where s is the squared centre of mass energy and θ , defined as the angle between the four vector of the τ -leptons and the z -axis. The choice was made not to give PySR the unary operators \sin and \cos , but instead, give $\cos(\theta)$ as a model variable together with the default binary operators $['+', '-', '*', '/']$ and one unary operator $sqr(x) = x^2$. Therefore PySR might learn the trigonometric identity $\sin(\theta) = \sqrt{1 - \cos^2(\theta)}$ to model the dependency on $\sin(\theta)$. The only operator constraint set for PySR will be to not nest more than two squaring operators within themselves. The influence of using different fractions of the full SM sample on training will also be investigated. The corresponding hyperparameter will henceforth be called 'events_used'.

In figure 7.1a a distribution of $OO_{hel.avg.}$ is shown for different a_τ values. The sensitivity is represented as a ratio to the SM expectations for each a_τ value. It is highest near the upper limit of the distribution where not many events are present compared to the most populated region. A 2-dimensional representation of the optimal observable $OO_{hel.avg.}$ can be seen in figure 7.1b for both kinematic variables β and $\cos(\theta)$. The values for $OO_{hel.avg.}$ are highest for greater invariant di- τ masses near $\beta = 1$ and for values of $\cos(\theta)$ around 0. Combining this with the event distribution from figure 7.1a, one can gather that most events will lie in the middle and upper regions of figure 7.1b but not where the $OO_{hel.avg.}$

values are the highest. For smaller values of the observable, i.e. at higher values of $\cos(\theta)$, not many events are present.



(a) $OO_{hel.avg.}$ for different a_τ values with the ratio to SM expectations.



(b) Distribution of $OO_{hel.avg.}$ depending on β and $\cos(\theta)$.

7.1.1 Hyperparameter search

For the hyperparameter search performed to construct models for $OO_{hel.avg.}$, the relevant hyperparameters are defined in table 4.1. Their investigated values are presented in table 7.1. This combination results in a total of 486 different models. Their computing time varied between 5 to 30 minutes depending on the combination.

option name	searched values
events_used	[5000, 10000, 20000]
model_selection	['best', 'score', 'accuracy']
frequency_use	[True, False]
iterations	[10, 30, 50]
populations	[5, 10, 15]
population_size	[20, 30, 40]

Table 7.1: Parameters and their defined ranges for the hyperparameter search.

For the parameter 'events_used', three values were chosen that might induce a noticeable difference in performance. It was found no significant difference in performance between using 20000 and anywhere up to 100000 events. Therefore this upper value is chosen to avoid longer computation times. For 'model_selection' and 'frequency_use' all possible values were checked. For the rest, similar ranges were chosen around the default values to investigate their impact on the performance.

In table 7.2 the less important parameters that were changed from their default in table 4.2 are listed. These will not change for the remainder of this thesis. For 'maxsize'

and 'maxdepth', the values were chosen to be slightly higher than the expected values of size = 30 and depth = 8, of the true form of $OO_{hel.avg.}$ as in figure 4.1. This is to accommodate for exploration beyond the necessary complexity which helps with getting closer to the final form [21]. It was chosen, to use constant optimisation at the mutation stage, which manifests in the parameter 'weight_optimize' being non-zero. This choice was made because in the expression of $OO_{hel.avg.}$ only integer constants are present. PySR should have the possibility to evolve in the direction of those constants before replacing the oldest individual, if the overall form is correct. 'parsimony' was set for a lower than default value because in case the frequency is not used, PySR should not be discouraged much from using higher complexities since the target is already a complex expression. Lower parsimony should also prevent having too simple models with highly optimized constants because constant optimization during mutation is allowed. 'adaptive_parsimony_scaling' is used to achieve the same thing when using frequency but here a higher than default value was chosen, because lower complexity models will most likely be more abundant than ones with higher complexity.

option name	set values
maxsize	40
maxdepth	9
weight_optimize	0.001
parsimony	0.0001
adaptive_parsimony_scaling	30

Table 7.2: Lesser parameters of PySR and their set values for the remainder of this thesis.

The models will be trained on the SM sample. This makes the use of weights in training not necessary since they are the same for all events. Testing is carried out on the BSM sample, reweighted to the SM sample of $a_\tau = 0.0$.

7.1.2 Comparing symbolic models

For comparing the 486 models from symbolic regression, three methods have been chosen. For all three methods, the four best models are presented and compared by performance with respect to the method. Their hyperparameter combination will be investigated for any similarities or differences. The first method will compare models by mean loss. The distribution of the normalised difference between prediction and true value will be shown and models will be sorted by the mean of this distribution. The second method is to consider the correlation between prediction and true expression for each event, which results in a scatterplot for each model. Finally, the Confidence intervals will be considered and compared by using the ΔNLL fits on model predictions.

7.1.2.1 Comparing by loss

The models have been sorted by the lowest mean of the normalised difference distribution. The normalised difference is defined as $\frac{True-prediction}{True}$ for each event. The four models with the lowest mean and their hyperparameter combination can be seen in table 7.3.

option name	1 _{st} model	2 _{nd} model	3 _{rd} model	4 _{th} model
events_used	5000	5000	20000	20000
model_selection	best	best	best	best
frequency_use	False	False	False	True
iterations	30	30	50	30
populations	15	5	10	5
population_size	40	40	40	20
mean	-4.091×10^{-7}	2.125×10^{-6}	3.590×10^{-6}	-3.777×10^{-6}

Table 7.3: The best 4 models sorted by lowest mean loss, and their hyperparameter combination.

Figure 7.2 shows the loss distributions for the four best models. They are restricted to a range of $[-0.025, 0.05]$ and have small variations. Models 1 in 7.2a and 3 in 7.2c have a double peak structure, such that the most probable values do not correspond to the mean. For model 3 in 7.2c this is most pronounced, where almost 14% of all events lie in the bin around -0.004 , whereas its mean is within 10^{-7} of 0. In contrast to this, model 2 in 7.2b has a particularly small range of $[-0.016, 0.1]$ which manifests in its standard deviation of $\sigma = 0.0028$, and with a peak that represents the mean more accurately. The same holds for model 4 in 7.2d, although its standard deviation is larger at $\sigma = 0.0081$.

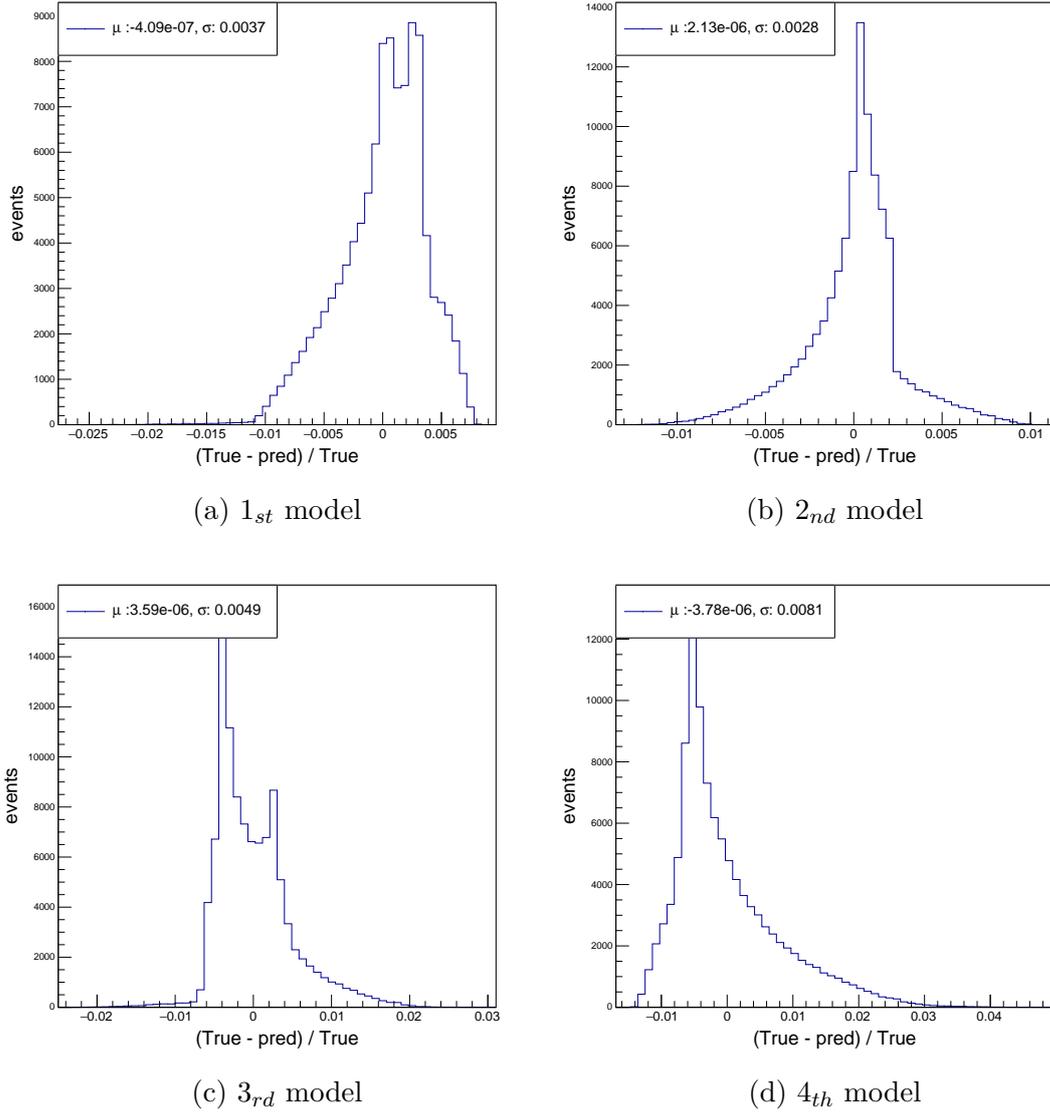


Figure 7.2: The four best models for $OO_{hel.avg.}$, sorted by lowest mean loss.

In equations 7.1, 7.2, 7.3, 7.4 the expressions for the four models are shown. All of the expressions have higher complexity than 20, some higher than 30. The expressions have high powers in β and/or $\cos \theta$. Two reasons for this are possible. Either when the multiplication operator is chosen, there is a higher chance that both operands will be the same variable since there are only two available or high powers bring a high fidelity to small changes in both variables because they are always smaller than 1, resulting in better performance. It is possible that both effects are simultaneously influencing the outcome.

$$model_1 = \frac{1.36023516005649 \left(\beta^2 - \frac{0.8574188 \cos^2 \theta}{\beta} \right)^2 (\beta^4 + 0.20372193)^2 + 1 + \frac{\beta + 0.019046964}{\beta^2}}{\beta} \quad (7.1)$$

$$model_2 = \frac{0.6032367}{0.0670197060177826\beta^4(-0.22943472194416\beta^8 - 0.89388067\beta^4 \dots} \dots + 2.30722435196801 \cos^2 \theta) + 0.22556765 \quad (7.2)$$

$$model_3 = -\beta^2(-2.856837\beta + 2.60865610534389 \cos^2 \theta) + \beta + 1.26383926247537(-0.20558754145143\beta^8 + \frac{\cos^2 \theta}{\beta^2}) \quad (7.3)$$

$$model_4 = 4\left(\beta - 0.37210544 \cos^2 \theta + \frac{0.0959514315788578 \cos^4 \theta}{\beta^4} - \frac{0.034527775}{\beta^6(\beta^2 + 0.48134148)^2}\right)^2 \quad (7.4)$$

Most models, that have a good mean loss also share some similarity in their hyperparameter combination. All four models with the lowest mean loss use the model selection method "best" and three of them did not use frequency. For both 'iterations' and 'population_size' higher numbers seem to contribute to a better performance. For 'populations' and the number of events that characterise these models, values are more scattered over the full range. Both models 2 and 4 used 5 populations, they were the ones with the 1 defined peak in their distribution but this is most likely coincidence because the shape of this distribution only depends on the form of the expressions. It is possible that this distribution form emerges from an expression type that is generally good at the problem at hand and therefore more abundant but this can not be determined.

7.1.2.2 Comparing by correlation

For all models obtained from symbolic regression the correlation between prediction and true value of $OO_{hel.avg.}$ has been determined. In table 7.4 the four chosen models with the highest correlation are presented with their hyperparameter combination. The values are all within 0.01% of perfect correlation of 1.

option name	1 _{st} model	2 _{nd} model	3 _{rd} model	4 _{th} model
events_used	5000	5000	20000	10000
model_selection	score	accuracy	accuracy	best
frequency_use	False	False	False	False
iterations	30	30	30	50
populations	15	15	15	15
population_size	20	20	40	30
correlation	0.99999	0.99999	0.99998	0.99997

Table 7.4: The best 4 models sorted by highest correlation, and their hyperparameter combination.

The scatterplots for all four models can be seen in figure 7.3 with their calculated correlation. They are mostly indistinguishable by bare eye except for model 3 in 7.3c having a slight asymmetry. The reason these are not the same models found in 7.1.2.1 is assumed

to be due to small losses being tolerant to a few large outliers. The correlation is not affected much by a slight shift as long as the variance is very small.

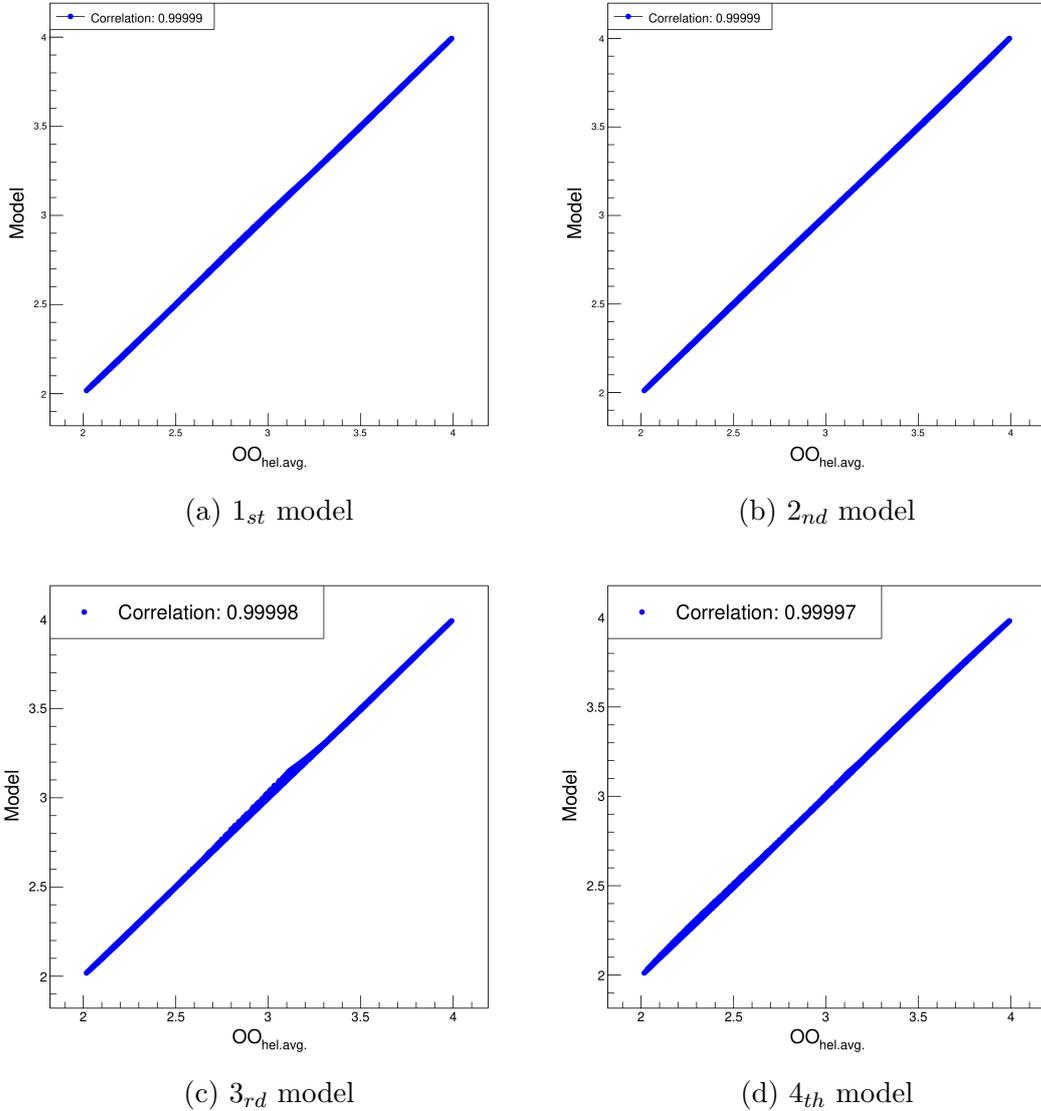


Figure 7.3: The four best models for $OO_{hel.avg.}$, sorted by the highest correlation to the true values.

The expression for the models shown in equations 7.5, 7.6, 7.7, 7.8 all show high powers in one or both variables. They seem similar in shape and complexity to the expressions gained from comparing the loss.

$$model_1 = \frac{\beta(\beta^{12}(\beta + 0.25473875) + 2.7468102)}{\beta^7 \cos^2 \theta + \beta} \quad (7.5)$$

$$model_2 = \frac{\beta^5 + 2.6223605\beta}{\beta(0.11678035\beta^8(-2.0130465\beta^4 + \frac{7.82168436844201 \cos^2 \theta}{\beta^4}) + 1.138608)} \quad (7.6)$$

$$model_3 = \frac{(\beta^8 + 1)^2}{0.998705329663195\beta^{10} \cos^2 \theta + \beta} \quad (7.7)$$

$$\begin{aligned}
model_4 = & \beta^2 - 0.41321951059433 \cos^4 \theta + (1.5996275 - \cos^2 \theta)(\beta^5 - 0.835632417757799\beta \cos^2 \theta)^2 \\
& + \frac{1.3900903}{\beta^2}
\end{aligned}
\tag{7.8}$$

The comparison by correlation reinforces the notion that using frequency does not correlate to performance, as all four models do not use the frequency option. It also shares the higher 'iteration' values with the previous comparison. For 'populations' all models use the highest possible value of 15, in contrast to the comparison from before. Again for events_used', a similar random choice can be seen. This time for 'population_size' there does not seem to be the same influence on performance as before.

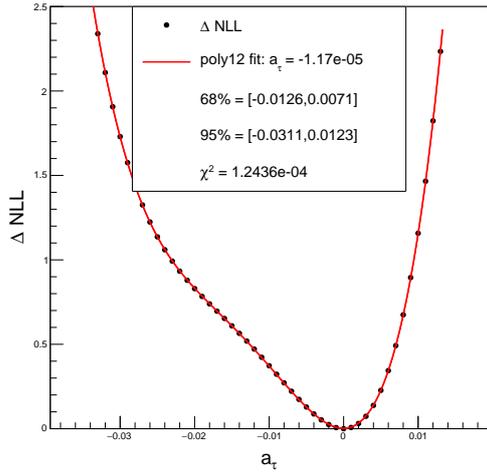
7.1.2.3 Comparing by confidence intervals

For each of the 486 models obtained from symbolic regression, the best-fit value for a_τ and its confidence intervals at 68% and 95% have been determined using the maximum likelihood method. In table 7.5 the four models with the smallest length of the interval on their 95% confidence level are listed with their hyperparameter combination. The lengths of their confidence intervals agree within 2%.

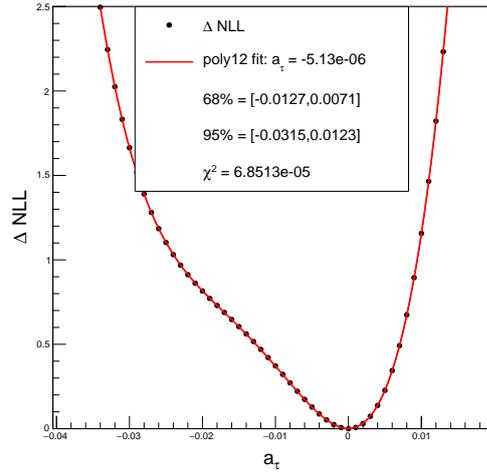
option name	1 _{st} model	2 _{nd} model	3 _{rd} model	4 _{th} model
events_used	5000	5000	20000	5000
model_selection	score	score	score	score
frequency_use	False	True	False	False
iterations	50	50	10	10
populations	10	5	5	15
population_size	20	30	30	40
95% range	0.0434	0.0438	0.0438	0.0443

Table 7.5: The best 4 models sorted by smallest 95% confidence interval, and their hyperparameter combination.

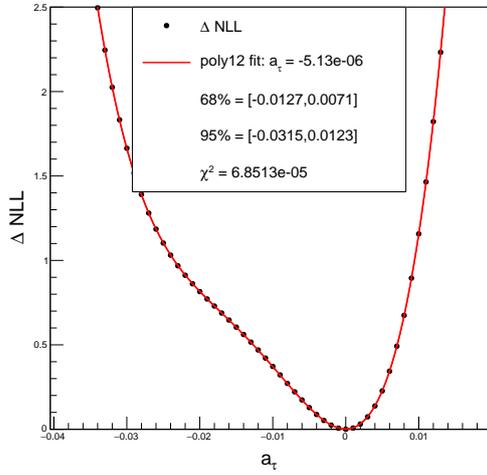
The ΔNLL fits for all four models can be seen In figure 7.4. They all have the same general shape and share similar estimates for a_τ as well as similar confidence intervals. The fits all have $\chi^2 > 10^{-6}$, which is a sufficient description of the given ΔNLL points. The confidence intervals are better than for the analytic form of the targeted $OO_{hel.avg}$. as seen in equations 5.11. Corresponding example histograms can be seen in figure 7.5.



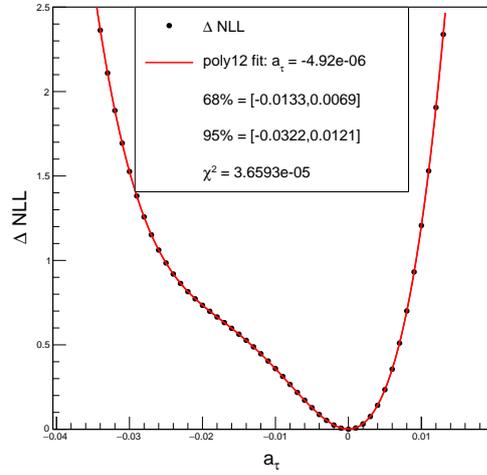
(a) 1_{st} model



(b) 2_{nd} model



(c) 3_{rd} model



(d) 4_{th} model

Figure 7.4: ΔNLL curves for the four best models for $OO_{hel.avg.}$, sorted by smallest 95% confidence interval.

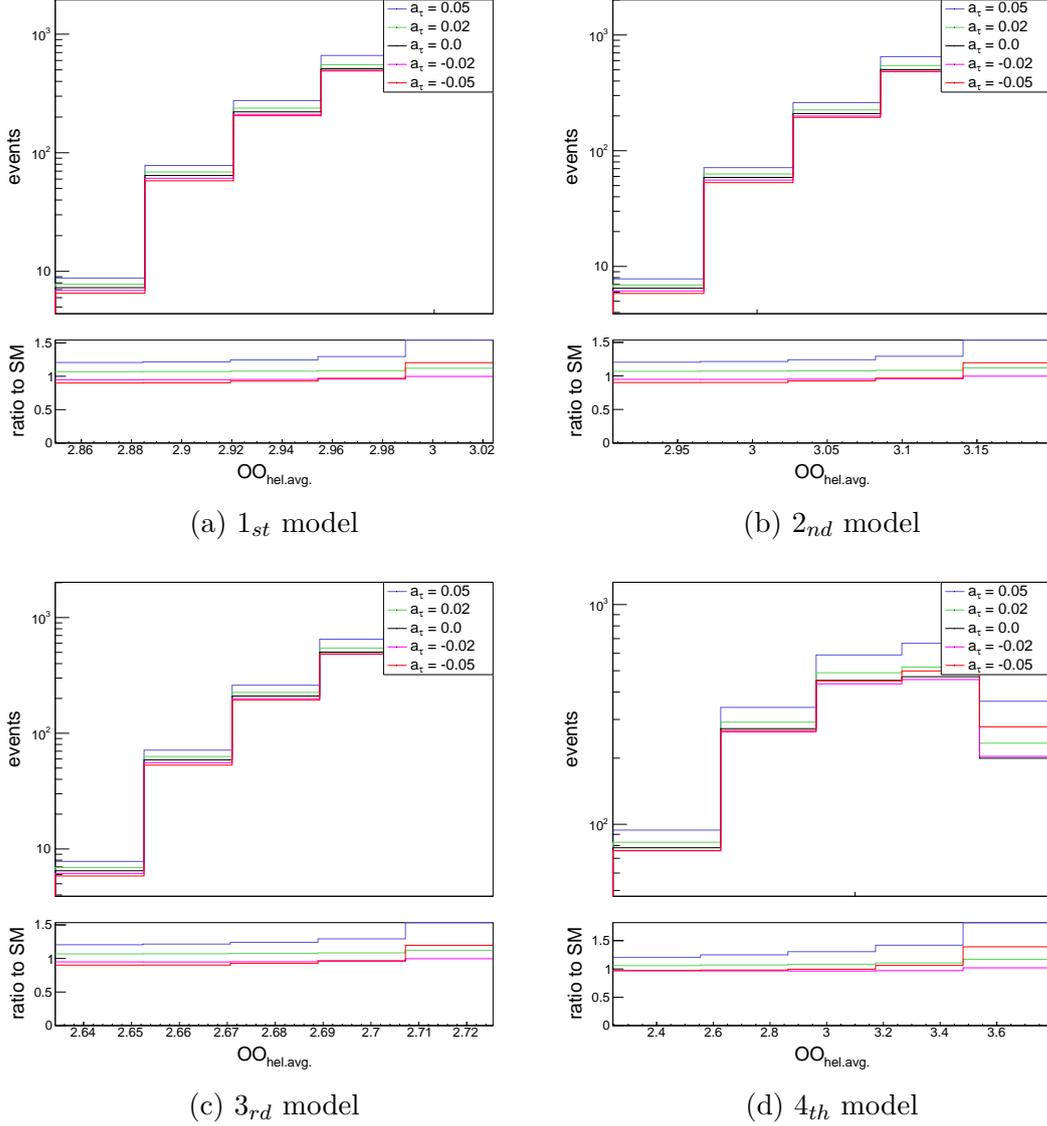


Figure 7.5: distributions at various a_τ values for the four best models for $OO_{hel.avg.}$, sorted by smallest length of the confidence interval at 95% confidence level.

The expressions for the models are presented in equations 7.9, 7.10, 7.11, 7.12. Three out of four models only use β with lower power than before and only one or two constants. Their shape and complexity are simple compared to the earlier models.

$$model_1 = \beta^2 + 2.0243666 \quad (7.9)$$

$$model_2 = 3.1997702\beta \quad (7.10)$$

$$model_3 = \beta + 1.7256507 \quad (7.11)$$

$$model_4 = -1.5127169 \cos^2 \theta + (\beta^4 + 0.94814818118951)^2 \quad (7.12)$$

The comparison of the hyperparameters for these models yields a slightly different outcome than the previous ones. The models with the best confidence intervals at 95% confidence level indicate the use of 'score' as the model selection method. Three of them were trained on 5000 events which is the smallest possible value. For 'iterations' the two best models used the highest value of 50 whereas the 3_{rd} and 4_{th} model used the smallest value of 10. The other hyperparameters also seem contradictory to what was observed earlier, where no real pattern can be seen.

7.1.3 Interim conclusion

PySR can achieve great precision in approximating a function via an evolved analytic expression as was presented in this chapter. It uses the variables to get general dependencies and finds an evolved model that maximises performance by combining operators and optimised constants. The hyperparameters that were included in the grid search showed different magnitudes of influence on performance. Most notably, frequency was used in 2/12 cases in contrast to the 50% average that would be expected, if the two possible values did not matter. The expectation of larger values corresponding to better performance for the options like events used, iterations, populations and population size, can be confirmed in many cases. However, 'events_used' did not seem to have a strict influence on performance, since often models that were trained on only 5000 events have better performance than models with higher values.

For the selection method of the final model, a discrepancy can be seen between the performance quantified by loss or correlation and the performance by confidence intervals. While the first two exclusively featured the method 'best' or were random, the latter exclusively featured models selected using 'score'. On one hand, the first two comparison methods yielded models with higher complexity and bigger powers in the given variables. On the other hand, the 'confidence interval' method yields simpler models, with lower powers and in 3 of 4 cases not using $\cos \theta$ at all.

This can be understood, by inspecting the cross-check of the two performance types correlation and confidence intervals. The correlation scatterplot for the model with the smallest 95% confidence interval can be seen in figure 7.6a. This model describes the true values of $OO_{hel.avg.}$ poorly with a correlation of -0.1442 . The ΔNLL fit for the model with the highest correlation, seen in figure 7.6b, shows that its confidence intervals are worse than the ones presented in table 7.5. The length of its confidence interval is 0.0509 which is more similar to the confidence from the analytic form of $OO_{hel.avg.}$ in equation 5.11.

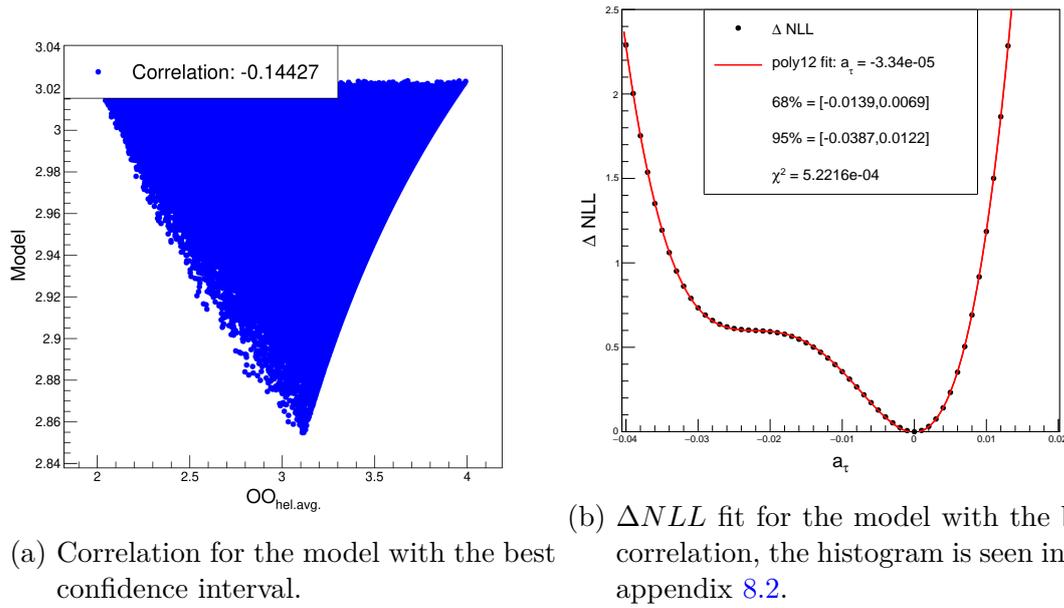


Figure 7.6: Cross-checking the performances by correlation and confidence intervals.

It can be concluded, that PySR accidentally learned observables with higher sensitivity than the original optimal observables. The most likely scenario for this is, that the model selection 'score', which is prevalent in the models with the best confidence intervals, leads to relatively simple equations being chosen if all the models in the HOF of PySR have similar accuracy. These expressions had better confidence intervals by chance.

In figure 7.7 the confidence intervals of the best models from all three comparison methods are shown together with the known optimal observables and the current experimental benchmark of p_T^μ . It can be observed, that models 1 and 2, corresponding to rating by loss and correlation have very similar confidence intervals to $OO_{hel.avg.}$ which is desirable because they describe its distribution relatively well. Model 3 from comparing by confidence intervals, shows an interval length more similar to p_T^μ but since this behaviour was declared a random effect via the selection method 'score', it is not of further interest.

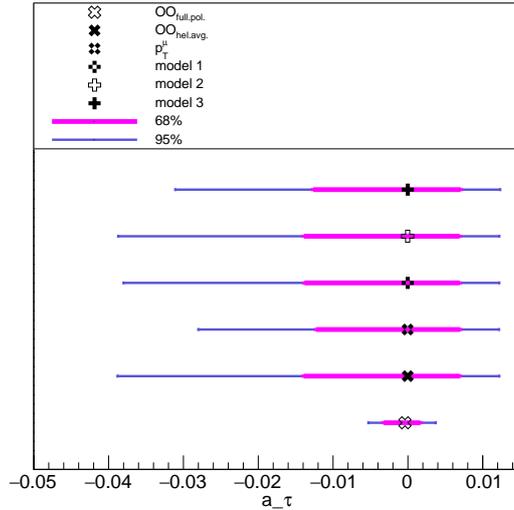


Figure 7.7: Comparing the confidence intervals of the best models from the three comparison methods loss (model 1), correlation (model 2) and confidence intervals (model 3) with the reference intervals from $OO_{hel.avg.}$, $OO_{full.pol}$ and p_T^μ .

The values corresponding to the above intervals are listed in table 7.6. The best estimates \hat{a}_τ are all constructed with the SM value of $a_\tau = 0.0$ as expectation, with which they agree. The ranges of the confidence interval length vary. Model 1 has a slightly better performance than model 2 and model 3, but the benchmark of p_T^μ is still not reached.

model	95% CI range
$OO_{hel.avg.}$	0.0510
$OO_{full.pol.}$	0.009
p_T^μ	0.0402
model 3 (CI)	0.0434
model 2 (corr)	0.0509
model 1 (loss)	0.0502

Table 7.6: Best estimates and ranges on the 95% confidence interval for the observables shown in figure 7.7

When counting the model that was selected by loss as the best approximation for $OO_{hel.avg.}$, the result for estimating a_τ with a learned model from parton variables is:

$$-0.0380 < \hat{a}_\tau < 0.0122 \quad (7.13)$$

at 95% confidence level. From this hyperparameter optimisation, a general trend for the hyperparameters can be gathered from the methods of comparing by loss and correlation, which indicated good prediction performance. The selection method 'best' is a good compromise between accuracy and simplicity as it selects the models similar to the most accurate but with respect to complexity. For the numerical options 'iterations', 'populations' and 'population_size' there is a small tendency towards higher values corresponding to better performance when compared to the true expression. This was expected since

these options increase the algorithm's resources to approximate the targeted expression. Frequency is an option that is not used much. Two possibilities are most convincing. The first possibility is that the 'adaptive_parsimony_scaling' hyperparameter was set too high such that complex enough expressions did not arise to achieve the same performance as without frequency. The second possibility is that the 'parsimony' parameter was set low enough to mostly evolve high complexity models. The latter possibility could be interpreted as overfitting, even though the example expressions shown for the first two comparison methods are not at the limit of allowed complexity. More complicated interactions between the parameters of the algorithm could likely cause such behaviour, but they are not determinable due to the randomness of this process.

In this part of the analysis, the analytical form of $OO_{hel.avg.}$ was not found as a model, this is not surprising, since the number of possible equations from PySR is infinite. However, it was observed that the best models had some semblance of the true expression, often combining β and $\cos\theta$ in divisions of sums albeit with high powers. The fourth model from comparing by correlation in equation 7.8 showed a term $(1.5996275 - \cos^2\theta)$ similar to the trigonometric relation $\sin^2\theta = 1 - \cos^2\theta$ which is present in the analytical form.

7.2 Regressing the fully polarised optimal observable

Learning the fully polarised optimal observable $OO_{full.pol.}$ is an important test for the symbolic regression method. A general analytic expression is not known for $OO_{full.pol.}$ but an approximation by PySR could be worthwhile if it achieves similar precision as the best models on $OO_{hel.avg.}$. The advantage of such an observable would be its simplicity compared to calculating the matrix elements. By training on the kinematics of final state particles, it is ensured that this is applicable in experiments.

7.2.1 Special phase-space and helicity groups of $OO_{full.pol.}$

The fully polarised optimal observable $OO_{full.pol.}$ can be calculated on an event-by-event basis by equation 3.8, using the parton helicities and momenta and the corresponding matrix elements from equation 1.10. There is no analytic expression that generalises this OO. It can nonetheless be used for training ML algorithms, especially symbolic regression could yield an interpretable approximation. However the distribution of $OO_{full.pol.}$ is complex. The helicity combinations can be sorted into groups of identical matrix elements, which appear in different forms in the distribution of the OO. These groups, whose contribution to the total distribution at $a_\tau = 0.0$ can be seen in figure 7.8, and their contributions to the distribution are listed in table 7.7. The groups are filled with helicity combinations $\{\lambda_1; \lambda_2; \lambda_3; \lambda_4\}$ for the partons $\{\gamma_1, \gamma_2, \tau_1, \tau_2\}$, where only the sign is of interest. For the missing combinations $\{++-+; +++-; --+-; ---+\}$ the matrix elements are calculated to be 0, therefore they do not contribute to the $\gamma\gamma \rightarrow \tau\tau$ process at all.

Group index	associated helicity combinations	event portion [%]
0	$+-+-; -+ -+; + - -+; - + +- $	90
1	$+- ++; - + ++; + - --; - + -- $	2.69
2	$++ --; -- ++$	0.0028
3	$+++ +; - - - -$	8.5

Table 7.7: Helicity groups defined by identical matrix elements and their probability.

It can be gathered, that groups 0 and 3 dominate the events by orders of magnitudes. Groups 1 and 2 only make up a small fraction of all events. Figure 7.8 shows a small gap in the phase space of $OO_{full.pol.}$ between values of 2 and 10. For group 0 the optimal observable becomes exactly 2 because of the matrix elements $M_1 = M_2$. Therefore the OO displays a constant in this region manifesting as a single bin peak in the distribution, yielding little to no information. This phenomenon is explained in detail in [20]. The important distribution part, where sensitivity is high, only consists of groups 1 and 2. For the SM value at $a_\tau = 0.0$ as seen in table 7.7 2.7% fall in this category. This fraction changes depending on a_τ for example at $a_\tau = 0.06$ it is 18%.

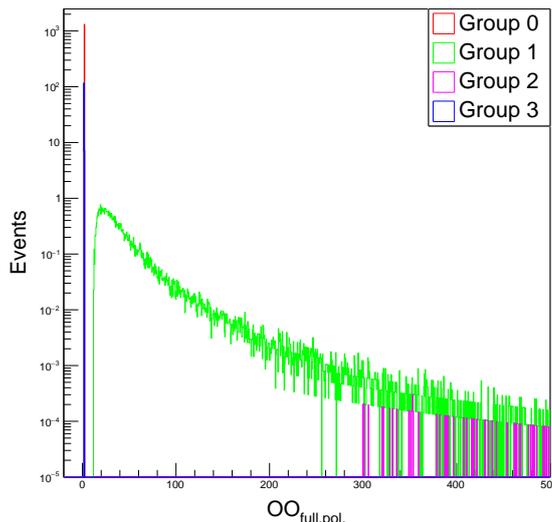


Figure 7.8: Distribution of $OO_{full.pol.}$, separated into the helicity groups 0-3.

The shape of this distribution will most likely influence the ML process. In the first part of section 7.2.4 it will be investigated if regression on the entire distribution of $OO_{full.pol.}$ is viable and if the gap structure will be approximated by PySR. In the second part, PySR is trained only on the events that fall into the helicity groups 1 and 2 for approximating the important distribution fraction for estimating a_τ .

7.2.2 Selection of observables for regression

The observables constructed from final state particles that are selected for training should have some connection to the parton kinematics, the OO itself or the helicities that influence it. 11 different observables from final state particle kinematics were investigated

towards their correlation to the significant distribution part of the helicity groups 1 and 2 in $OO_{full.pol.}$. For calculating them, the four-momenta of the final state particles in the laboratory frame were used. The final state particles are the muon and all existing pions for each event. The combined four-momentum of all pions is used as the 'visible hadronic vector' abbreviated by 'vishad'. As observables mostly the transverse momentum p_T and the pseudorapidity η are used as well as the polar and azimuth angles Φ and θ . The missing transverse energy has been calculated from the four-momenta of the three outgoing neutrinos. The full set is listed in table 8.1. An example of the scatterplots used to calculate the correlations can be seen in figure 7.9.

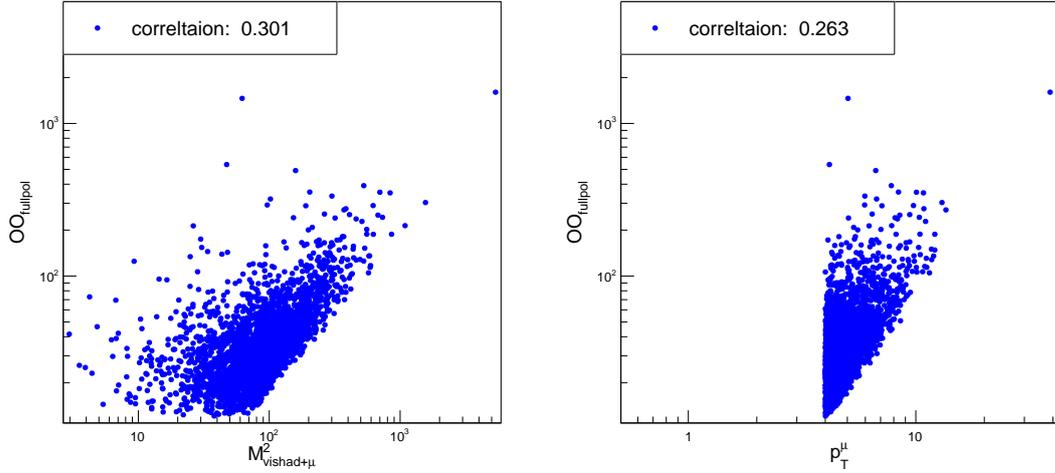


Figure 7.9: Correlation to $OO_{full.pol.}$ for the two variables $M_{vishad+\mu}^2$ and p_T^μ .

The six observables with the highest correlation have been chosen for training. The higher dimensionality of this observable space in comparison to the training on $OO_{hel.avg.}$ has been chosen to grant PySR a wider diversity in possible equations. The chosen variables can be seen in table 7.8 together with their correlation to $OO_{full.pol.}$.

variable	correlation
$M_{vishad+\mu}^2$	0.301
p_T^{vishad}	0.165
η^{vishad}	-0.020
p_T^μ	0.263
η^μ	-0.022
MET	0.186

Table 7.8: Chosen observables for learning $OO_{full.pol.}$ and their correlation to the distribution restricted to helicity groups 1 and 2.

7.2.3 Choosing hyperparameters from earlier experience

The hyperparameters for learning $OO_{full.pol.}$ are shown in table 7.9. They have been chosen, based on the results from learning $OO_{hel.avg.}$.

name	values chosen
events_used	[3000, 100000]
model_selection	['best']
frequency_use	[False]
iterations	[50]
populations	[15]
population_size	[40]

Table 7.9: Hyperparameters and their chosen values for learning $OO_{full.pol.}$.

The range of 'events_used' corresponds to training once with applying a condition of $OO_{full.pol.} > 3$ on the events, which ensures training will only be carried out on the important groups 1 and 2. The second value corresponds to training with all the events, where groups 1 and 2 only account for $\sim 2.7\%$ in the training sample. The model selection method 'best' has been chosen, because a good approximation for $OO_{full.pol.}$ is preferred over simpler expressions. In this case, simpler expressions are not likely to yield a better performance in confidence intervals than the OO, because it was precisely calculated. Frequency is not used for reasons explained earlier. For the rest, the highest values have been chosen.

For the less important hyperparameters, it was chosen not to change them, because of comparability. The properties of maximum depth and maximum number of nodes are not known for $OO_{full.pol.}$ since it is not a single analytic expression for all events. Potential approximations may look very different based on the observables chosen such that room for potential exploration should be given to the algorithm.

7.2.4 Comparing symbolic models

For examining the performance of the two models trained on $OO_{full.pol.}$, the same methods will be applied as before. This time, the model that was trained without applying a condition on the phase space of the observable will be presented first. The model trained on the subsample defined by the helicity groups 1 and 2 will be discussed afterwards. Finally, the possibility of application to experimental data is discussed.

7.2.4.1 Predicting $OO_{full.pol.}$ from the entire distribution

The expression of the model that was trained on the entire distribution of $OO_{full.pol.}$ is shown in figure 7.14. It consists of 3 of the 6 given variables and shows large powers in p_T^μ up to a power of 12. There is also a small constant in the range of 10^{-18} , which was not observed before.

$$\begin{aligned}
model_{full\ distribution} = & 3.01339931077764 \times 10^{-18} p_T^{\mu 8} (p_T^\mu + p_T^{vishad})^4 \\
& (0.733440992691847\eta^{vishad} - 1)^4 + 3.1669834
\end{aligned} \tag{7.14}$$

The loss distribution and Correlation to $OO_{full.pol.}$ of the prediction from this model can be seen in figure 7.10. The loss distribution is mainly populated around -1 and 1, where there is only one event between -0.6 and 0.7. From this distribution, a few extreme outliers up to -160 have been removed for better discernibility of the main distribution. This is most likely from the prediction not correctly assigning any of the events associated with the helicity groups 1 and 2, that populate the higher ranges of $OO_{full.pol.}$, leading to large negative values corresponding to the definition of this loss. For the correlation, a similar effect is seen. Almost all events are predicted to lie around 3, especially the ones corresponding to the higher values in $OO_{full.pol.}$ only one of which was predicted near correctly. The predictions at large values stem from the original part of $OO_{full.pol.}$ where the value 2 was dominant.

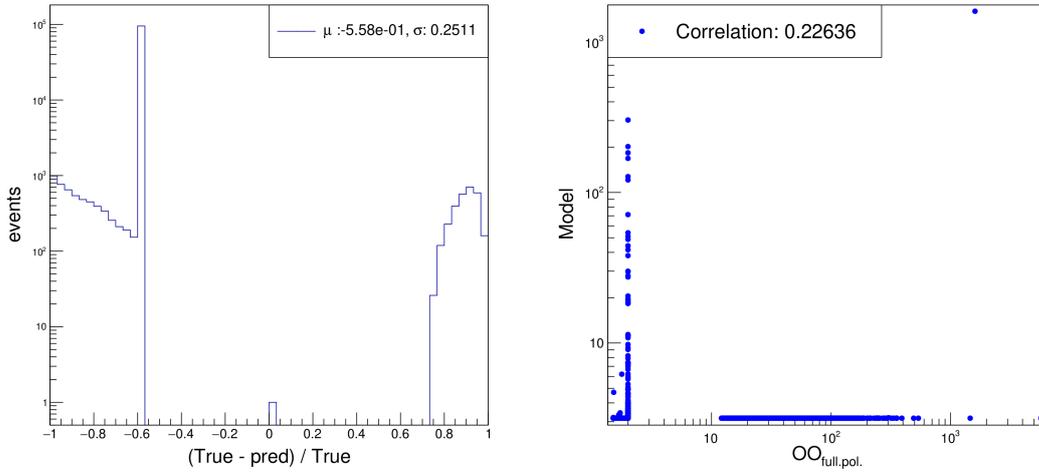


Figure 7.10: Loss distribution and correlation for the model on the full distribution of $OO_{full.pol.}$.

In figure 7.11 the ΔNLL fit for the model is shown as well as an exemplary distribution at various a_τ values. The ΔNLL curve shows two local minima. Due to this double-peak structure, the confidence intervals extracted from this model are significantly lower than any observable considered so far.

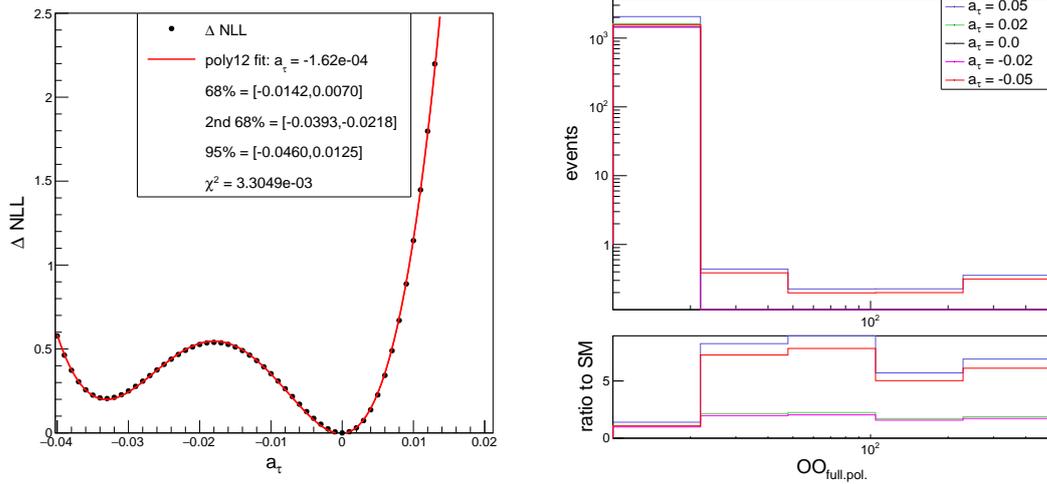


Figure 7.11: ΔNLL fit and distribution at different a_τ values for the model on the full distribution of $OO_{full.pol.}$.

The final estimation of this model at 95% confidence level comes to

$$-0.046043 < \hat{a}_\tau^{\text{full distribution}} < 0.012473. \quad (7.15)$$

As expected, the helicity groups 0 and 3 dominate the training of PySR and the model is not able to discern the gap in the phase space as important due to the event ratio of the smaller parts. This yields poor overall performance compared to the models on $OO_{hel.avg.}$, making this method not viable for further investigation, even though it is transferable to experiment i.e. can be applied to momenta of final state particles.

7.2.4.2 Predicting $OO_{full.pol.}$ from helicity groups 1 and 2

For training the PySR model only helicity groups 1 and 2, a condition on the full sample is applied of $OO_{full.pol.} > 3$. This yields a small sample of ~ 3000 events on which training can be carried out to predict $OO_{full.pol.}$ where it is most sensitive to a_τ .

The expression of this model can be seen in figure 7.16. This formula makes use of the most correlated observables MET , p_T^μ , $M2_{vishad+\mu}$, p_T^{vishad} , combining them in a mostly linear way apart from the fraction. All variables are not raised to a higher power in contrast to the models on $OO_{hel.avg.}$, where only 2 variables were available.

$$\begin{aligned} model_{hel. \text{ groups } 1+2} = & MET p_T^\mu + 0.22548035 M_{vishad+\mu}^2 + \\ & 1.7784543 MET - 1.5180968085157 p_T^\mu - 0.338300443942897 \\ & \frac{p_T^{vishad} - 0.54388505}{\phantom{MET p_T^\mu + 0.22548035 M_{vishad+\mu}^2 +}} \end{aligned} \quad (7.16)$$

The loss distribution and the correlation of the model can be seen in figure 7.12. The peak of the loss distribution is strongly pronounced, although there are some outliers in extreme ranges. This leads to a mean loss in the order of 10^{-2} which is not comparable to the values for $OO_{hel.avg.}$ that were in the order of 10^{-6} . The correlation of this model to the true OO shows two separated concentrations that can be identified as the helicity groups 1 and 2 when comparing with figure 7.8. This is due to the testing sample having

significantly more events in both helicity groups than the training sample. The correlation for the prediction on the training sample can be seen in figure 7.13 which is within 6% of perfect correlation and shows almost no events for helicity group 2 starting at higher values of $OO_{full.pol.}$. The comparably low performance of this model is expected, since the OO still depends on helicities and the variables chosen for this model do not represent the actual form that was calculated from matrix elements. The model yields a higher correlation than the individual observables on their own, which displays the capability of PySR to combine features into a viable model. The ΔNLL fit and corresponding distributions at various a_τ values for this model are shown in figure 7.14. It displays a small confidence interval compared to the previous model when tested on the same subsample of helicity groups 1 and 2 of

$$-0.0099 < a_\tau^{\text{hel. groups 1+2}} < 0.0050 \quad (7.17)$$

at 95% confidence level. This result is almost comparable to the true $OO_{full.pol.}$.

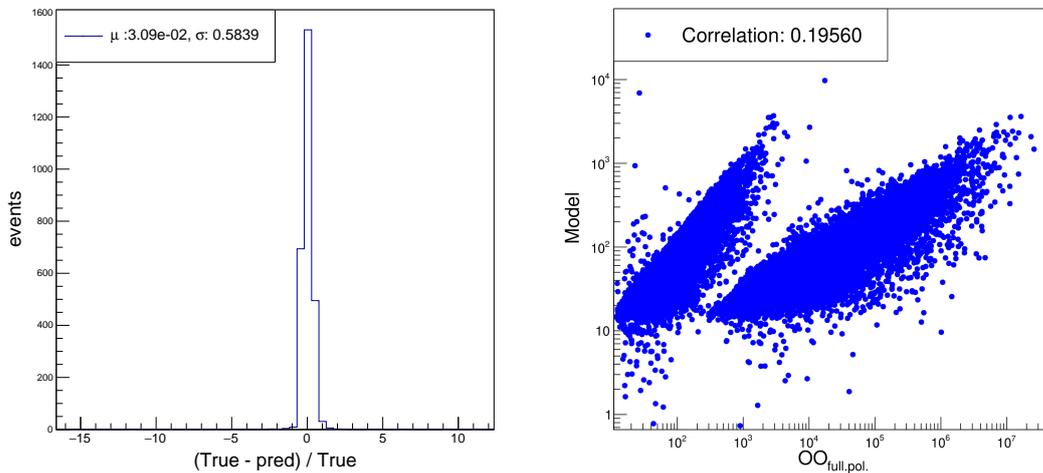


Figure 7.12: Loss distribution and correlation for the model trained on helicity groups 1 and 2.

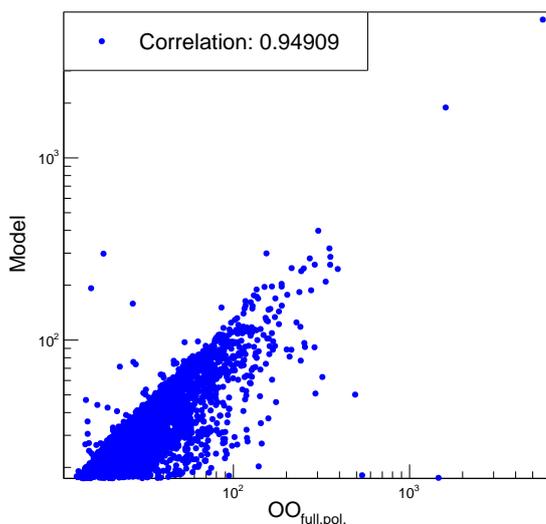


Figure 7.13: Scatterplot for the model trained on helicity groups 1 and 2 only predicting helicity group 1.

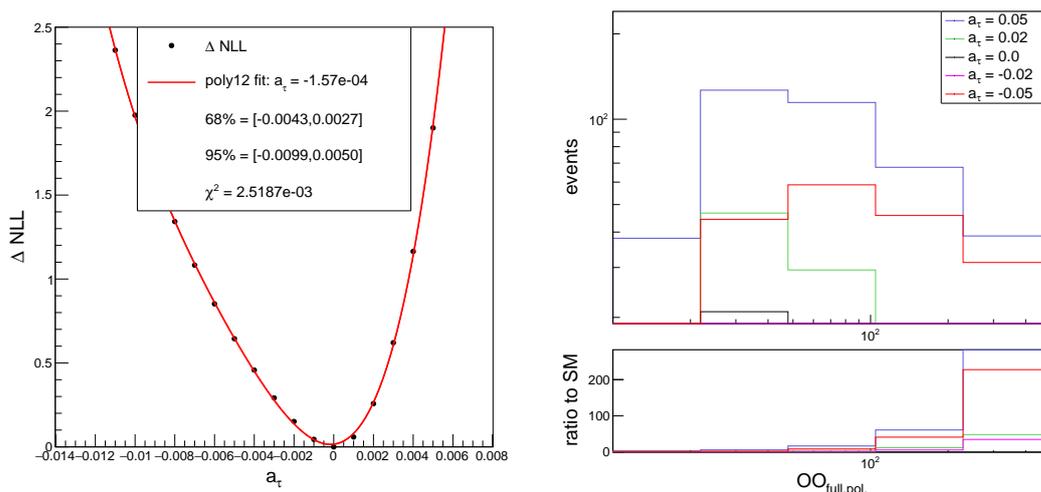


Figure 7.14: ΔNLL fit and corresponding distributions at various a_τ values for the model trained on helicity groups 1 and 2.

The problem with this evaluation is, that it is not repeatable in experiments, because one cannot select only helicity groups 1 and 2 in data. The comparison of confidence intervals in different evaluations of the model trained on helicity groups 1 and 2 can be seen in figure 7.17. This model has been trained on a fraction of the entire sample as explained above. It has been evaluated for confidence intervals from the entire BSM sample, being the equivalent of using this observable on experimental data, which is represented as model 1 in the comparison. The performance from above, where a condition was applied to the testing sample is included as model 2. Then to represent a hybrid method, model 3 displays the performance, where a condition was applied to the prediction itself, before computing the ΔNLL fit. ΔNLL curves and corresponding distributions for the PySR model for $OO_{full.pol.}$, trained on the helicity groups 1 and 2. The model predicting the entire testing sample can be seen in 7.15a and 7.15c. The model where a condition of

$OO_{full.pol.}^{pred} > 3$ was applied is shown in 7.15b and 7.15d.

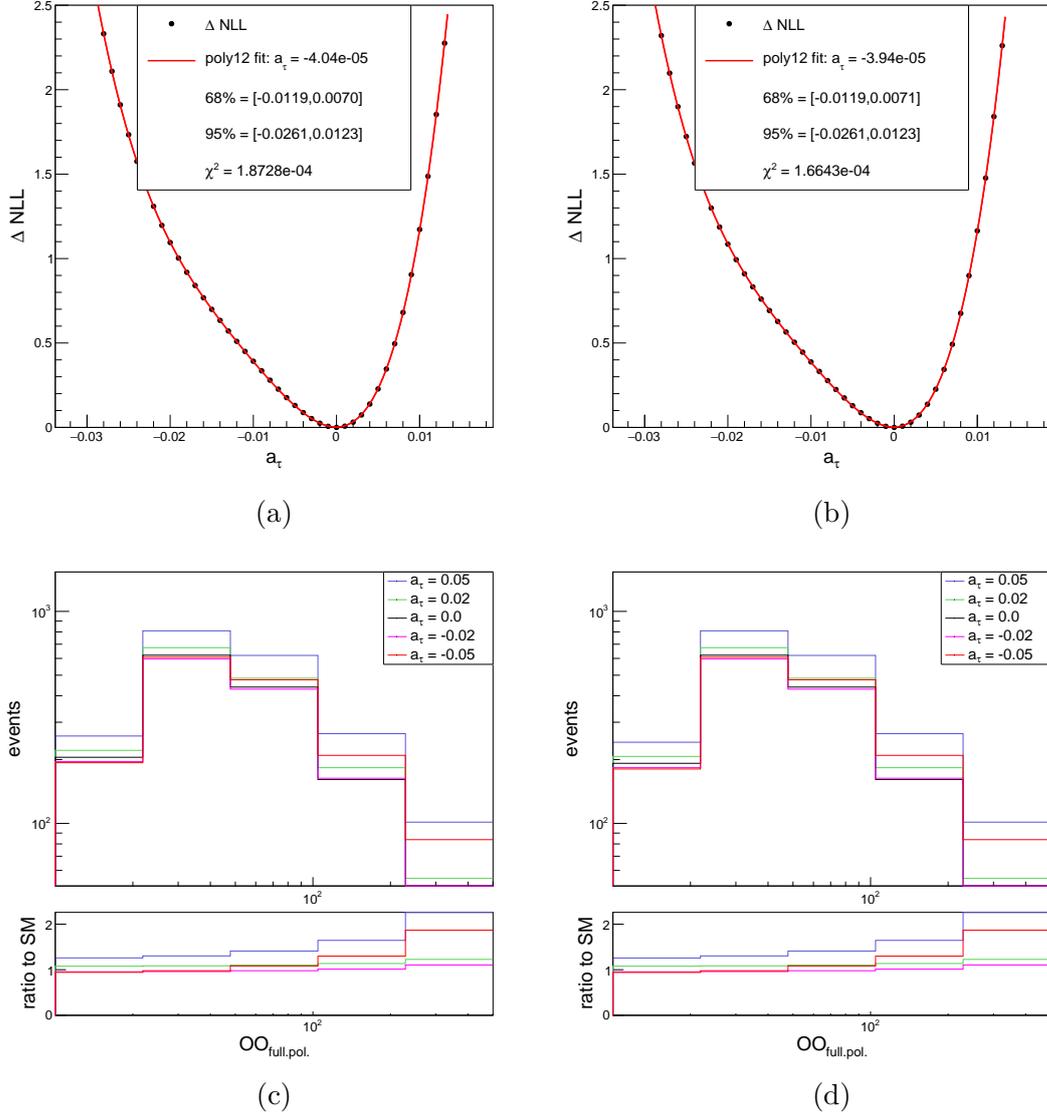


Figure 7.15: ΔNLL curves and corresponding distributions for the PySR model for $OO_{full.pol.}$, trained on the helicity groups 1 and 2. The model predicting the entire testing sample can be seen in 7.15a and 7.15c. The model where a condition of $OO_{full.pol.}^{pred} > 3$ was applied is shown in 7.15b and 7.15d.

A correlation scatterplot can be seen in figure 7.16. It shows that knowing the prediction does not yield further information. All predicted values are larger than 2, because as was the regressed distribution. However one can see good separation of the helicity groups 1 and 2 as seen before. The behaviour that events are not separated in the distribution of the prediction is expected because the events not contained in the training sample, i.e. helicity groups 0 and 3 could not be predicted correctly.

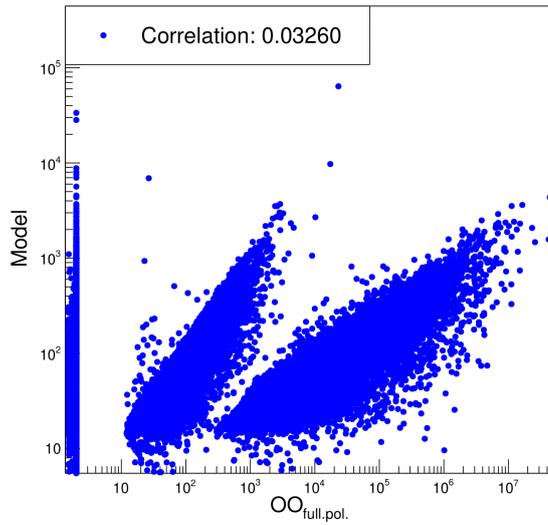


Figure 7.16: The correlation scatterplot for the model trained on helicity groups 1 and 2, predicting the entire testing set.

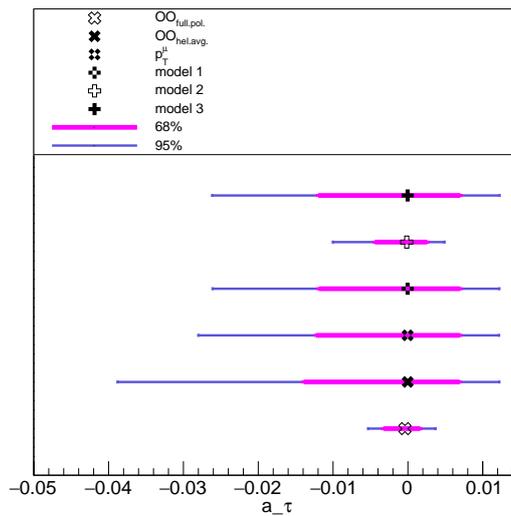


Figure 7.17: Comparing confidence intervals for the viable models for $OO_{full.pol.}$.

The values corresponding to the comparison of the different evaluation methods are listed in table 7.10. Models 1 and 3 show similar performance. This was expected because the prediction cannot separate the phase space but predict the shape of the helicity groups 1 and 2 for all events. Therefore the condition on the prediction itself will remove an equal portion of events that do not lie in the desired part and events that do. Model 2 is not yet experimentally achievable, because it requires selecting helicity groups 1 and 2 in data, however, it has the most similar performance to the true optimal observable.

model	\hat{a}_τ	95% CI range
$OO_{hel.avg.}$	-2.74×10^{-5}	0.051019
$OO_{full.pol.}$	-3.93×10^{-4}	0.008987
p_T^μ	-3.96×10^{-5}	0.040196
model 1 (no cond.)	-3.9×10^{-5}	0.038328
model 2 (cond)	-1.57×10^{-4}	0.017427
model 3 (self cond.)	-4.0×10^{-5}	0.038416

Table 7.10: Best estimates and ranges on the 95% confidence interval for the observables shown in figure 7.17

Models 1 and 3 are experimentally repeatable and are almost identical in their best estimate \hat{a}_τ and their range on the confidence interval. They gain 4.6% and 4.4% in sensitivity respectively compared to using the simple observable p_T^μ . The final result for estimating a_τ with an observable obtained from symbolic regression on experimentally available data is from applying model 1 to all events is

$$-0.026065 < \hat{a}_\tau^{\text{small sample}} < 0.012263 \quad (7.18)$$

at a 95% confidence level.

7.3 Final comparison of observables considering experimental viability

There are 3 models that are useful for the final comparison and for the viability of PySR as a tool. In the following model 1 will describe the observable that was learned from the entire distribution of $OO_{full.pol.}$. Model 2 will denote the observable obtained from only the helicity groups 1 and 2 and evaluated using the entire testing sample for experimental applicability. Model 3 describes the observable learned from $OO_{hel.avg.}$ that was selected via the 'loss' method. Their confidence intervals are presented in figure 7.18 together with the confidence intervals of the two target observables $OO_{hel.avg.}$ and $OO_{full.pol.}$ and the observable p_T^μ currently used in experiments.

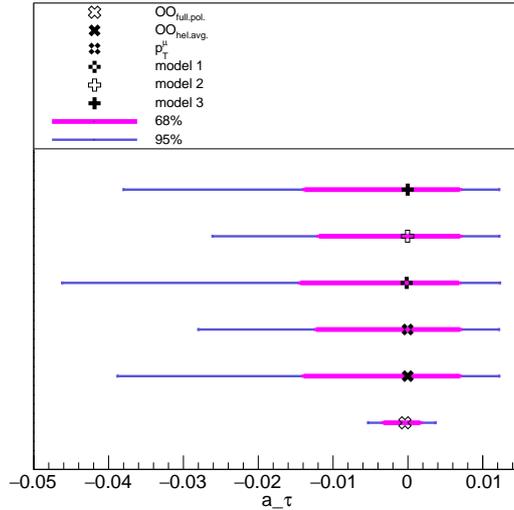


Figure 7.18: Comparing confidence intervals for the known observables $OO_{hel.avg.}$, $OO_{full.pol.}$ and p_T^μ to the 3 models (see text for explanation) learned by PySR.

$OO_{full.pol.}$ itself is experimentally not viable, because of the dependency on helicity. Model 3 is mostly a proof of concept for using symbolic regression as a tool for extracting observables. The confidence intervals from model 3 and $OO_{hel.avg.}$ are identical within 1.8%. Model 1 was chosen for its low loss in predicting the values of $OO_{hel.avg.}$ and even though it is not the same expression, its approximation is satisfying. Model 1 has the highest range of all if tested on the entire testing sample. Its regression process was dominated by the helicity groups 0 and 3 and therefore has not acquired much of the sensitivity of the helicity groups 1 and 2. Model 2 describes the helicity group 1 and 2 relatively well but if this is projected onto the entire testing sample it yields only a 4.6% improvement in sensitivity compared to p_T^μ .

model	95% CI range
$OO_{hel.avg.}$	0.051019
$OO_{full.pol.}$	0.008987
p_T^μ	0.040196
model 1 ($OO_{full.pol.}$)	0.058516
model 2 ($OO_{full.pol.}$)	0.038328
model 3 ($OO_{hel.avg.}$)	0.050928

Table 7.11: Best estimates and ranges on the 95% confidence interval for the observables shown in figure 7.18

An effect of the chosen histogram limit when calculating the ΔNLL values as discussed in section 5.2 on the confidence intervals was observed. Model 1 resulted in a larger confidence interval than p_T^μ when using a lower histogram limit. This is explainable by the fact that the helicity groups 1 and 2 which hold most of the sensitivity lie at the larger ranges of the distribution. Model 1 predicted them well but also blended some predictions from the other helicity groups into this distribution. Therefore when a smaller histogram limit is chosen the sensitivity from the additional bins is lost. It is also to note, that the

distribution of $OO_{full.pol.}$ in the training sample had a smaller range than the distribution in the training sample because of the discussed ratios of the helicity groups in dependence of a_τ for which the samples have been created. Therefore model 1 will also not predict the sensitivity of the outer ranges of the testing sample because the training was done on the more limited distribution.

In the end, PySR gave model 1 which is not the optimal observable, but it is a "semi-optimal" observable that is usable experimentally and fulfils its purpose of yielding an improvement in sensitivity. This "semi-optimal" observable is a relatively simple function of variables that can be computed using reconstructed final state particle kinematics. Symbolic regression was, therefore, able to yield an analytic function with similar sensitivity than the currently used observable p_T^μ .

7.4 Recommendations for further studies of a_τ using symbolic regression

It has been shown, that for increasing the sensitivity to determine a_τ , modelling the specific distribution of $OO_{full.pol.}$ is crucial for the performance of any observable obtained from symbolic regression. In this thesis, learning this distribution was not possible due to the statistical difference in events from more probable helicity groups. As seen in section 7.2.1 the helicity groups 0 and 3 make up more than 90% of all events, dominating the symbolic regression in section 7.2.4.1. Whereas the helicity groups 1 and 2 hold the most sensitivity. Rather than trying to regress the two parts in the distribution of $OO_{full.pol.}$ separately, it should be considered manually normalising the event numbers in both regions to the same number. In this way, the symbolic regression by PySR could treat both equally and learn the gap in the phase space of the observable.

In the future, studies should consider comparing observables from symbolic regression with predictions from other ML methods such as regression neural networks. If symbolic regression methods yield similar results to these other methods, they would be preferable for their interpretability. Concerning PySR in particular, a more in-depth hyperparameter optimisation needs to be carried out on more than the basic options used in this thesis, to effectively chose them for serious learning. Simulated annealing is an option not used in this thesis but could have an interesting influence on model evolution. Other investigations could include searching for optimal parsimony or dynamic parsimony scaling, varying the possibilities for different events within the algorithm and increasing the number of operators to include common functions that characterise many physical theories like sin, cos or exp. In this thesis, not many constraints were put on the learning process in terms of the combination of operators or constants. Since constants are an infinite space of mutation possibility for the algorithm, constraining them to a sensible set or even allowing PySR to use physical constants could be an interesting method.

8 Conclusion

The goal of this thesis was to study whether observables constructed with symbolic regression can increase the sensitivity to measure a_τ in ultraperipheral PbPb collisions at the LHC. The sensitivities were estimated assuming an integrated luminosity of $\mathcal{L} = 2.0\text{nb}^{-1}$ corresponding to the 2015 and 2018 data sets collected by the ATLAS detector. The objective was to learn expressions that approximate the optimal observables from particle-level variables via this method.

Optimal observables, which are able to maximise sensitivity for a_τ , were calculated using matrix elements of the $\gamma\gamma \rightarrow \tau\tau$ process. The distinction was made between the true OO, which needs helicities of the particles to compute the matrix elements for each event, called the fully polarised optimal observable $OO_{full.pol}$, and the OO where the average matrix elements from the possible helicity combinations were used to calculate the helicity averaged optimal observable $OO_{hel.avg.}$. An analytic form was calculated for this observable, which is valid for all events and depends on only two parton kinematic observables θ_τ and β_{cm} .

The symbolic regression tool PySR was used to train models based on simulated data, to learn the two optimal observables mentioned above. A hyperparameter optimisation was carried out on $OO_{hel.avg.}$. The parameters considered are generations of the genetic algorithm, populations and population size to simulate evolution, the use of frequency to punish the complexity of models, measured by the number of nodes in the expression tree and the method by which PySR selects the final model. 486 different models were compared via three figures of merit, the normalised loss distribution, correlation of prediction to true OO and the length of resulting confidence intervals on a_τ . It was found, that PySR was able to model $OO_{hel.avg.}$ effectively, with a correlation of 0.99999 and yields a constraint on a_τ of $[-0.037977, 0.012235]$ at 95% confidence level, which is within 1.8% of the confidence extracted from the true $OO_{hel.avg.}$. A random effect was encountered for the method of selecting the final model 'score', where some of these models were represented by a small function, not describing $OO_{hel.avg.}$ but getting a better confidence interval than the OO itself. This was discarded as a fluctuation in the learning process due to the selection method and the randomness of the algorithm.

For learning $OO_{full.pol}$, the optimised hyperparameters previously studied were adapted. As variables for the training process, the correlations of 11 final state observables to $OO_{full.pol}$ were investigated, of which 6 were chosen. $OO_{full.pol}$ showed two benches of values stemming from different helicity groups, where 98% of the testing sample were located in an interval of $[0, 2]$, consisting of a single sharp peak, and 2% of the events sowed sensitive distribution contributions in the range of $[10, \infty]$. This effect was due to different helicity combinations in the $\gamma\gamma \rightarrow \tau\tau$ process contributing at different probabilities to the events and manifesting at different ranges in the $OO_{full.pol}$ distribution. One model was trained on the entire sample. It was not able to reconstruct the shape of the distribution of $OO_{full.pol.}$, as the learning process was dominated by events from the helicity groups 0 and 3, treating most other events as outliers. This yielded a constraint

on a_τ of $[-0.046, 0.013]$ with 95% confidence level.

The second model for $OO_{full.pol}$ was only trained on the subsample from the helicity groups 1 and 2. Testing it on the same distribution part yielded a good fit for $OO_{full.pol}$ in this range with a correlation of 0.95 and a constraint on a_τ of $[-0.0099, 0.0050]$ at 95% confidence level. The extraction of this constraint from the helicity groups 1 and 2 was not experimentally applicable, therefore the model was tested on the entire testing sample. This yielded a correlation of 0.033 and a constraint on a_τ of $[-0.026, 0.012]$. This represents an improvement of 4.6% compared to using the observable p_T^μ which is currently used in experimental analysis. It was concluded, that symbolic regression is an interesting tool and could yield significant improvements to current methods with further investigation. Future studies should account for the unequal distribution of helicity groups in $OO_{full.pol}$. and consider larger hyperparameter investigations.

8 References

- [1] Sheldon L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2). URL: <https://www.sciencedirect.com/science/article/pii/0029558261904692>.
- [2] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [3] Abdus Salam. “Weak and Electromagnetic Interactions”. In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: [10.1142/9789812795915_0034](https://doi.org/10.1142/9789812795915_0034).
- [4] 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”. In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020). arXiv: [1207.7214](https://arxiv.org/abs/1207.7214) [[hep-ex](#)].
- [5] Serguei Chatrchyan et al. “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021). arXiv: [1207.7235](https://arxiv.org/abs/1207.7235) [[hep-ex](#)].
- [6] P. J. E. Peebles and Bharat Ratra. “The cosmological constant and dark energy”. In: *Rev. Mod. Phys.* 75 (2 Apr. 2003), pp. 559–606. DOI: [10.1103/RevModPhys.75.559](https://doi.org/10.1103/RevModPhys.75.559). URL: <https://link.aps.org/doi/10.1103/RevModPhys.75.559>.
- [7] Planck Collaboration. “Planck 2018 results - VI. Cosmological parameters”. In: *A&A* 641 (2020), A6. DOI: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910). URL: <https://doi.org/10.1051/0004-6361/201833910>.
- [8] Stephen P. Martin and James D. Wells. “Muon Anomalous Magnetic Dipole Moment in Supersymmetric Theories”. In: *Phys. Rev. D* 64 (2001), p. 035003. DOI: [10.1103/PhysRevD.64.035003](https://doi.org/10.1103/PhysRevD.64.035003). arXiv: [hep-ph/0103067](https://arxiv.org/abs/hep-ph/0103067).
- [9] D. Hanneke, S. Fogwell and G. Gabrielse. “New Measurement of the Electron Magnetic Moment and the Fine Structure Constant”. In: *Phys. Rev. Lett.* 100 (12 Mar. 2008), p. 120801. DOI: [10.1103/PhysRevLett.100.120801](https://doi.org/10.1103/PhysRevLett.100.120801). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.100.120801>.
- [10] B. Abi et al. “Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm”. In: *Phys. Rev. Lett.* 126 (14 Apr. 2021), p. 141801. DOI: [10.1103/PhysRevLett.126.141801](https://doi.org/10.1103/PhysRevLett.126.141801). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.126.141801>.
- [11] T. Aoyama et al. “The anomalous magnetic moment of the muon in the Standard Model”. In: *Physics Reports* 887 (2020). The anomalous magnetic moment of the muon in the Standard Model, pp. 1–166. ISSN: 0370-1573. DOI: <https://doi.org/10.1016/j.physrep.2020.07.006>. URL: <https://www.sciencedirect.com/science/article/pii/S0370157320302556>.
- [12] S. Eidelmann and M. Passera. “Theory of the Lepton Anomalous Magnetic Moment”. In: *Modern Physics Letters A* 22.03 (Jan. 2007), pp. 159–179. DOI: [10.1142/S0217732307022694](https://doi.org/10.1142/S0217732307022694). URL: <https://doi.org/10.1142/S0217732307022694>.

- [13] J. Abdallah et al. “Study of tau-pair production in photon-photon collisions at LEP and limits on the anomalous electromagnetic moments of the tau lepton”. In: *Eur. Phys. J. C* 35 (2004), pp. 159–170. DOI: [10.1140/epjc/s2004-01852-y](https://doi.org/10.1140/epjc/s2004-01852-y). arXiv: [hep-ex/0406010](https://arxiv.org/abs/hep-ex/0406010).
- [14] The ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: 3.08 (Aug. 2008), S08003. DOI: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003). URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08003>.
- [15] *2018 5.02 TeV lead-lead final January 2022*. URL: https://twiki.cern.ch/twiki/bin/viewauth/Atlas/LuminosityForPhysics#2018_5_02_TeV_lead_lead_final_Ja. last accessed Jul 08 2023.
- [16] Morad Aaboud et al. “Evidence for light-by-light scattering in heavy-ion collisions with the ATLAS detector at the LHC”. In: *Nature Phys.* 13.9 (2017), pp. 852–858. DOI: [10.1038/nphys4208](https://doi.org/10.1038/nphys4208). arXiv: [1702.01625](https://arxiv.org/abs/1702.01625) [hep-ex].
- [17] M. Diehl and O. Nachtmann. “Optimal observables for the measurement of three gauge boson couplings in $e^+ e^- \rightarrow W^+ W^-$ ”. In: *Z. Phys. C* 62 (1994), pp. 397–412. DOI: [10.1007/BF01555899](https://doi.org/10.1007/BF01555899).
- [18] D. Atwood and A. Soni. “Analysis for magnetic moment and electric dipole moment form-factors of the top quark via $e^+ e^- \rightarrow t \text{ anti-}t$ ”. In: *Phys. Rev. D* 45 (1992), pp. 2405–2413. DOI: [10.1103/PhysRevD.45.2405](https://doi.org/10.1103/PhysRevD.45.2405).
- [19] D. Buskulic et al. “Measurement of the tau polarization at the Z resonance”. In: *Z. Phys. C* 59 (1993), pp. 369–386. DOI: [10.1007/BF01498618](https://doi.org/10.1007/BF01498618).
- [20] Kartik Bhide. “Probing optimal measurements of the electromagnetic dipole moments of the tau lepton”. In: *to be published* ().
- [21] Miles Cranmer et al. *PySR*. URL: <https://astroautomata.com/PySR/>. last accessed Jun 24 2023.
- [22] *Miles Cranmer Github*. URL: <https://github.com/MilesCranmer>. last accessed Jul 07 2023.
- [23] G. Arnison et al. “Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV”. In: *Phys. Lett. B* 122 (1983), pp. 103–116. DOI: [10.1016/0370-2693\(83\)91177-2](https://doi.org/10.1016/0370-2693(83)91177-2).
- [24] P. Bagnaia et al. “Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN $\bar{p}p$ Collider”. In: *Phys. Lett. B* 129 (1983), pp. 130–140. DOI: [10.1016/0370-2693\(83\)90744-X](https://doi.org/10.1016/0370-2693(83)90744-X).
- [25] S. Abachi and et.al. Abbott. “Observation of the Top Quark”. In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2632–2637. DOI: [10.1103/PhysRevLett.74.2632](https://doi.org/10.1103/PhysRevLett.74.2632). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.74.2632>.
- [26] M.E. Peskin and D.V. Schroeder. *An Introduction to Quantum Field Theory*. Advanced book classics. Avalon Publishing, 1995. ISBN: 9780201503975. URL: <https://books.google.de/books?id=i35LALN0GosC>.
- [27] G. SHAW B.R. MARTIN. *Particle Physics*. The Manchester Physics Series. John Wiley Sons, Ltd, 2017. ISBN: 9781118911907.
- [28] Various. *The Standard Model infographic*. URL: https://cds.cern.ch/record/1473657/files/SMinfographic_image.png?subformat=. last accessed Jun 17 2023.

- [29] F. Halzen and Alan D. Martin. *Quarks and Leptons: an introductory Course in modern Particle Physics*. 1984. ISBN: 978-0-471-88741-6.
- [30] R. L. Workman et al. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. DOI: [10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097).
- [31] Rabindra N. Mohapatra and Goran Senjanovi ć. “Neutrino Mass and Spontaneous Parity Nonconservation”. In: *Phys. Rev. Lett.* 44 (14 Apr. 1980), pp. 912–915. DOI: [10.1103/PhysRevLett.44.912](https://doi.org/10.1103/PhysRevLett.44.912). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.44.912>.
- [32] Nima Arkani-Hamed et al. “The String landscape, black holes and gravity as the weakest force”. In: *JHEP* 06 (2007), p. 060. DOI: [10.1088/1126-6708/2007/06/060](https://doi.org/10.1088/1126-6708/2007/06/060). arXiv: [hep-th/0601001](https://arxiv.org/abs/hep-th/0601001).
- [33] Vincenzo Cavasinni. “Observation of the $\gamma\gamma \rightarrow \pi$ process in Pb+Pb collisions and constraints on the τ -lepton anomalous magnetic moment with the ATLAS detector at LHC”. In: *EPJ Web Conf.* 282 (2023), p. 01009. DOI: [10.1051/epjconf/202328201009](https://doi.org/10.1051/epjconf/202328201009).
- [34] L3 Collaboration. “Measurement of the anomalous magnetic and electric dipole moments of the tau lepton”. In: *Physics Letters B* 434.1 (1998), pp. 169–179. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/S0370-2693\(98\)00736-9](https://doi.org/10.1016/S0370-2693(98)00736-9). URL: <https://www.sciencedirect.com/science/article/pii/S0370269398007369>.
- [35] OPAL Collaboration. “An upper limit on the anomalous magnetic moment of the lepton”. In: *Physics Letters B* 431.1 (1998), pp. 188–198. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/S0370-2693\(98\)00520-6](https://doi.org/10.1016/S0370-2693(98)00520-6). URL: <https://www.sciencedirect.com/science/article/pii/S0370269398005206>.
- [36] Mariola Klusek-Gawenda et al. “From the $\gamma\gamma \rightarrow p\bar{p}$ reaction to the production of $p\bar{p}$ pairs in ultraperipheral ultrarelativistic heavy-ion collisions at the LHC”. In: *Phys. Rev. D* 96 (9 Nov. 2017), p. 094029. DOI: [10.1103/PhysRevD.96.094029](https://doi.org/10.1103/PhysRevD.96.094029). URL: <https://link.aps.org/doi/10.1103/PhysRevD.96.094029>.
- [37] Mateusz Dyndał et al. “Anomalous electromagnetic moments of lepton in $\gamma\gamma \rightarrow \tau\bar{\tau}$ reaction in Pb+Pb collisions at the LHC”. In: *Physics Letters B* 809 (2020), p. 135682. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2020.135682>. URL: <https://www.sciencedirect.com/science/article/pii/S0370269320304858>.
- [38] *About CERN - Who we are*. URL: <https://home.cern/about/who-we-are>. last accessed Jun 21 2023.
- [39] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001. DOI: [10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001). URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08001>.
- [40] V. Shiltsev and F. Zimmermann. “Modern and future colliders”. In: *Rev. Mod. Phys.* 93 (1 Mar. 2021), p. 015006. DOI: [10.1103/RevModPhys.93.015006](https://doi.org/10.1103/RevModPhys.93.015006). URL: <https://link.aps.org/doi/10.1103/RevModPhys.93.015006>.
- [41] *ATLAS Experiment*. URL: <https://atlas.cern/>. last accessed Jul 07 2023.
- [42] *The CMS detector*. URL: <https://cms.cern/>. last accessed Jul 07 2023.
- [43] *LHCb Collaboration*. URL: <https://lhcb.web.cern.ch/>. last accessed Jul 07 2023.

- [44] *ALICE Collaboration*. URL: <https://alice-collaboration.web.cern.ch/>. last accessed Jul 07 2023.
- [45] *Public ATLAS Luminosity Results for Run-3 of the LHC*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3>. last accessed Jul 08 2023.
- [46] *Public ATLAS Luminosity Results for Run-2 of the LHC*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>. last accessed Jul 08 2023.
- [47] Matthias Schott and Monica Dunford. “Review of single vector boson production in pp collisions at $\sqrt{s} = 7$ TeV”. In: *Eur. Phys. J. C* 74 (2014), p. 2916. DOI: [10.1140/epjc/s10052-014-2916-1](https://doi.org/10.1140/epjc/s10052-014-2916-1). arXiv: [1405.1160](https://arxiv.org/abs/1405.1160) [hep-ex].
- [48] Lydia Beresford and Jesse Liu. “New physics and tau $g-2$ using LHC heavy ion collisions”. In: *Phys. Rev. D* 102.11 (2020). [Erratum: *Phys.Rev.D* 106, 039902 (2022)], p. 113008. DOI: [10.1103/PhysRevD.102.113008](https://doi.org/10.1103/PhysRevD.102.113008). arXiv: [1908.05180](https://arxiv.org/abs/1908.05180) [hep-ph].
- [49] M. Davier et al. “The Optimal method for the measurement of tau polarization”. In: *Phys. Lett. B* 306 (1993), pp. 411–417. DOI: [10.1016/0370-2693\(93\)90101-M](https://doi.org/10.1016/0370-2693(93)90101-M).
- [50] G. Aad et al. “Test of CP invariance in vector-boson fusion production of the Higgs boson in the $H \rightarrow \tau\tau$ channel in proton-proton collisions at $s=13$ TeV with the ATLAS detector”. In: *Physics Letters B* 805 (2020), p. 135426. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2020.135426>. URL: <https://www.sciencedirect.com/science/article/pii/S0370269320302306>.
- [51] Davide Castelvecchi. “The Stern–Gerlach experiment at 100”. In: *Nature Reviews Physics* 4 (2022), pp. 140–142.
- [52] Deaglan J. Bartlett, Harry Desmond and Pedro G. Ferreira. “Exhaustive Symbolic Regression”. In: (Nov. 2022). DOI: [10.1109/TEVC.2023.3280250](https://doi.org/10.1109/TEVC.2023.3280250). arXiv: [2211.11461](https://arxiv.org/abs/2211.11461) [astro-ph.CO].
- [53] Matiur Rahman Minar and Jibon Naher. *Recent Advances in Deep Learning: An Overview*. 2018. DOI: [10.13140/RG.2.2.24831.10403](https://doi.org/10.13140/RG.2.2.24831.10403). URL: <http://rgdoi.net/10.13140/RG.2.2.24831.10403>.
- [54] Zhongtian Dong et al. “Is the machine smarter than the theorist: Deriving formulas for particle kinematics with symbolic regression”. In: *Phys. Rev. D* 107.5 (2023), p. 055018. DOI: [10.1103/PhysRevD.107.055018](https://doi.org/10.1103/PhysRevD.107.055018). arXiv: [2211.08420](https://arxiv.org/abs/2211.08420) [hep-ph].
- [55] Suyong Choi. “Construction of a Kinematic Variable Sensitive to the Mass of the Standard Model Higgs Boson in $H \rightarrow WW^* \rightarrow l^+\nu l^-\bar{\nu}$ using Symbolic Regression”. In: *JHEP* 08 (2011), p. 110. DOI: [10.1007/JHEP08\(2011\)110](https://doi.org/10.1007/JHEP08(2011)110). arXiv: [1006.4998](https://arxiv.org/abs/1006.4998) [hep-ph].
- [56] Pablo Lemos et al. *Rediscovering orbital mechanics with machine learning*. 2022. arXiv: [2202.02306](https://arxiv.org/abs/2202.02306) [astro-ph.EP].
- [57] Miles D. Cranmer et al. *Learning Symbolic Physics with Graph Networks*. 2019. arXiv: [1909.05862](https://arxiv.org/abs/1909.05862) [cs.LG].
- [58] Silviu-Marian Udrescu and Max Tegmark. “AI Feynman: a Physics-Inspired Method for Symbolic Regression”. In: *Sci. Adv.* 6.16 (2020), eaay2631. DOI: [10.1126/sciadv.aay2631](https://doi.org/10.1126/sciadv.aay2631). arXiv: [1905.11481](https://arxiv.org/abs/1905.11481) [physics.comp-ph].

- [59] Miles Cranmer et al. “Discovering Symbolic Models from Deep Learning with Inductive Biases”. In: (June 2020). arXiv: [2006.11287](https://arxiv.org/abs/2006.11287) [cs.LG].
- [60] Nour Makke and Sanjay Chawla. “Interpretable Scientific Discovery with Symbolic Regression: A Review”. In: (Nov. 2022). arXiv: [2211.10873](https://arxiv.org/abs/2211.10873) [cs.LG].
- [61] Miles Cranmer. *PySR Github*. URL: <https://github.com/MilesCranmer/PySR>. last accessed Jun 24 2023.
- [62] Python Software Foundation. *Python Language Reference, version 3.11.4*. URL: <https://www.python.org/downloads/release/python-3114/>. last accessed Jul 13 2023.
- [63] Miles Cranmer. *SymbolicRegression.jl Github*. URL: <https://github.com/MilesCranmer/SymbolicRegression.jl>. last accessed Jun 24 2023.
- [64] Digvijay Wadekar et al. “Modeling assembly bias with machine learning and symbolic regression”. In: (Nov. 2020). arXiv: [2012.00111](https://arxiv.org/abs/2012.00111) [astro-ph.CO].
- [65] Jessica Craven, Vishnu Jejjala and Arjun Kar. “Disentangling a deep learned volume formula”. In: *JHEP* 06 (2021), p. 040. DOI: [10.1007/JHEP06\(2021\)040](https://doi.org/10.1007/JHEP06(2021)040). arXiv: [2012.03955](https://arxiv.org/abs/2012.03955) [hep-th].
- [66] Digvijay Wadekar et al. “The SZ flux-mass (Y–M) relation at low-halo masses: improvements with symbolic regression and strong constraints on baryonic feedback”. In: *Mon. Not. Roy. Astron. Soc.* 522.2 (2023), pp. 2628–2643. DOI: [10.1093/mnras/stad1128](https://doi.org/10.1093/mnras/stad1128). arXiv: [2209.02075](https://arxiv.org/abs/2209.02075) [astro-ph.CO].
- [67] Patrick Kidger. *On Neural Differential Equations*. 2022. arXiv: [2202.02435](https://arxiv.org/abs/2202.02435) [cs.LG].
- [68] Tilman Plehn et al. “Modern Machine Learning for LHC Physicists”. In: (Nov. 2022). arXiv: [2211.01421](https://arxiv.org/abs/2211.01421) [hep-ph].
- [69] Anja Butter et al. “Back to the Formula – LHC Edition”. In: (Sept. 2021). arXiv: [2109.10414](https://arxiv.org/abs/2109.10414) [hep-ph].
- [70] Miles Cranmer. *Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl*. 2023. arXiv: [2305.01582](https://arxiv.org/abs/2305.01582) [astro-ph.IM].
- [71] Glen Cowan. *Statistical Data Analysis*. Oxford University Press, 2002. ISBN: 0198501552.
- [72] *ROOT Data Analysis Framework*. URL: <https://root.cern/>. last accessed Jun 27 2023.
- [73] *ROOT TGraph Class Reference*. URL: <https://root.cern/doc/master/classTGraph.html#aa978c8ee0162e661eae795f6f3a35589>. last accessed Jun 27 2023.
- [74] Christophe Andrieu et al. “An Introduction to MCMC for Machine Learning”. In: *Machine Learning* 50 (Jan. 2003), pp. 5–43. DOI: [10.1023/A:1020281327116](https://doi.org/10.1023/A:1020281327116).
- [75] *gamma-UPC*. URL: <http://hshao.web.cern.ch/hshao/gammaupc.html>. last accessed Jun 27 2023.
- [76] *MadGraph5*. URL: <https://cp3.irmp.ucl.ac.be/projects/madgraph/>. last accessed Jun 27 2023.
- [77] *PYTHIA*. URL: <https://pythia.org/>. last accessed Jun 27 2023.
- [78] *PHOTOS*. URL: http://photospp.web.cern.ch/photospp/resources/PHOTOS.3.64/Photos_interface_design.3.64.pdf. last accessed Jun 27 2023.
- [79] *Rivet — the particle-physics MC analysis toolkit*. URL: <https://rivet.hepforge.org/>. last accessed Jun 27 2023.

- [80] Melike Akbiyik et al. *Search for anomalous magnetic moment of τ lepton in ultra-peripheral Pb+Pb collisions with ATLAS*. Tech. rep. Geneva: CERN, 2021. URL: <https://cds.cern.ch/record/2798409>.
- [81] Olivier Mattelaer. “On the maximal use of Monte Carlo samples: re-weighting events at NLO accuracy”. In: *The European Physical Journal C* 76 (2016). DOI: [10.1140/epjc/s10052-016-4533-7](https://doi.org/10.1140/epjc/s10052-016-4533-7). URL: <https://doi.org/10.1140/epjc/s10052-016-4533-7>.
- [82] *Vistar*. URL: <https://op-webtools.web.cern.ch/vistar/vistars.php?usr=LHCATLAS>. last accessed Jul 06 2023.

8 Acknowledgements

This thesis will mark my first academic degree and it is a gratifying experience to learn what scientific work means and how fun and stressful it can be.

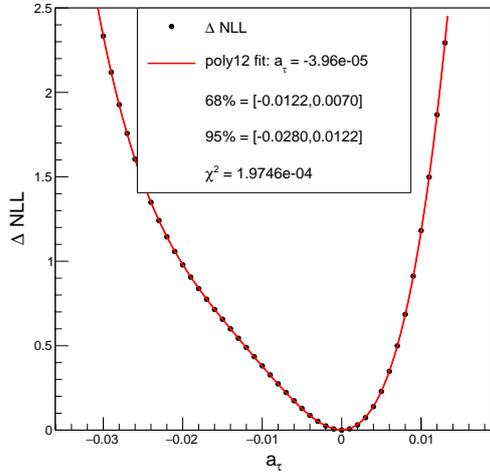
Many thanks go to Prof. Dr. Markus Schumacher for trusting me with this very interesting, experimental study and for his continuous help and guidance throughout my time in his group.

I want to thank my supervisor Dr. Valerie Lang who was always available for questions and helped immensely with overcoming my concerns about my performance. Also for showing me the wonderful Vistars website [82] to keep up with my favourite collider.

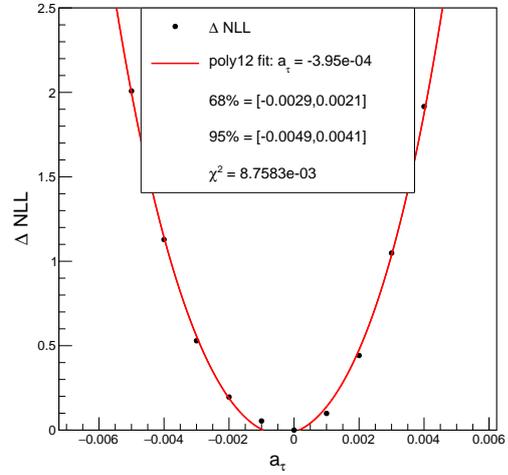
Additional thanks go to Kartik Bhide, with whom I shared an office and who therefore provided me with many explanations and support for both my theoretical and technical understanding. His willingness to share his extended knowledge on the topic was essential for this thesis.

Special thanks go to Lena Schönstein. Without her never-ending support and love, I would not have managed this thesis let alone all the other things around as life continues happening despite work being a point of focus.

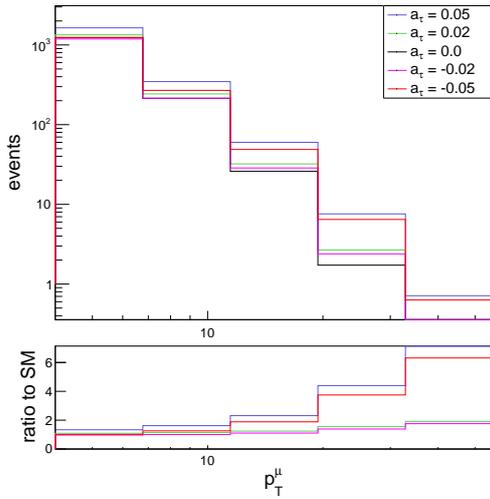
8 Appendix



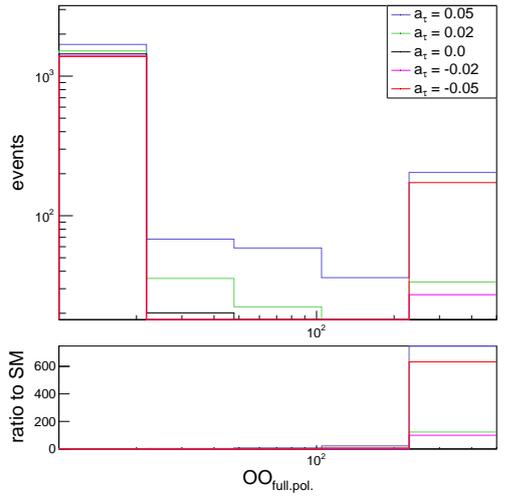
(a) ΔNLL curve for p_T^μ .



(b) ΔNLL curve for $OO_{full.pol.}$.



(c) Distribution of p_T^μ for various a_τ values.



(d) Distribution of $OO_{full.pol.}$ for various a_τ values.

Figure 8.1: ΔNLL curves and distributions for the observables p_T^μ and $OO_{full.pol.}$.

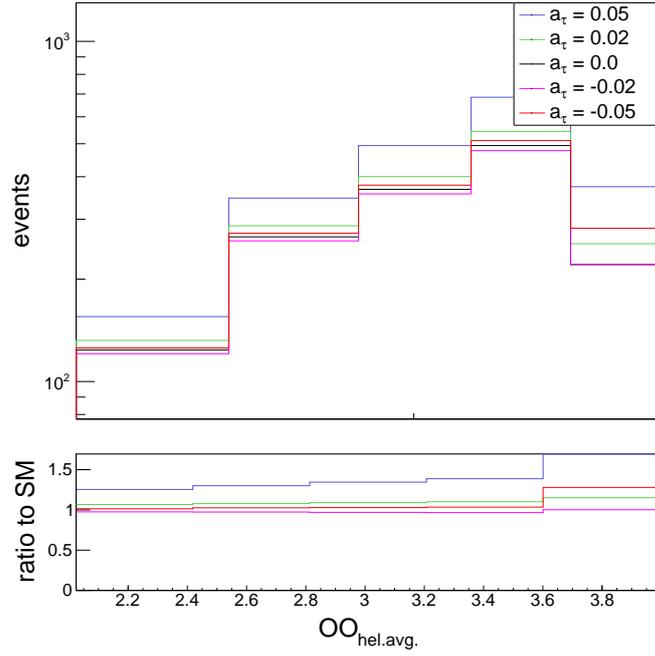


Figure 8.2: Distribution of the model with the highest correlation for various a_τ values.

variable	correlation
p_T^μ	0.262936
η^μ	-0.022407
$p_T^{\pi^\pm}$	0.093358
η^{π^\pm}	-0.019768
$\Delta\eta(\pi^\pm, \mu)$	0.058850
$\Delta\Phi(\pi^\pm, \mu)$	-0.032330
MET	0.186475
$M_{vishad+\mu}^2$	0.300660
p_T^{vishad}	0.165181
η^{vishad}	-0.020017
$\Delta\theta(vishad, \mu)$	-0.002869

Table 8.1: Investigated particle level variables and their correlation to the fully polarised optima observable restricted to helicity groups 1 and 2.