Dissertation

Automated optimization of sensitivity in a search for boosted VBF Higgs pair production in the $b\bar{b}b\bar{b}$ quark final state with the ATLAS detector

For the attainment of the academic degree doctor rerum naturalium

(Dr. rer. nat.) in the subject: Physics

Frederic Renner

Berlin, 15.12.2023

Faculty of Mathematics and Natural Sciences of the Humboldt University of Berlin

1st Supervisor: Dr. Clara Elisabeth Leitgeb

2nd Supervisor: Prof. Dr. Cigdem Issever

(Only after the disputation for publication in the university library according to § 15 of the doctoral regulations enter the names and the date):

Reviewers:

1st:

2nd:

3rd:

Date of the oral examination:

Abstract

I am an abstract.

Contents

Chapter 1

Statistics

Every scientific investigation starts with a hypothesis that is to be tested empirically. The main objective is to evaluate if the proposed hypothesis agrees or disagrees with observed data, to either accept or reject it against the null-hypothesis which represents a baseline scenario where only known phenomena are presumed to occur.

A key metric that quantifies this is the p-value that arises within hypothesis testing. Test results of an experiment follow some probability density function. Assuming some hypothesis, the p-value is the integrated probability for test results compatible with this hypothesis and ergo measuring the compatibility of the observation to the assumption. In other words if the experiment where to be repeated it gives the probability that the result favors the proposed hypothesis.

In the field of high energy physics a framework has been developed specifically for this task. This section begins to lay out the mathematical fundamentals of the approach and explains its implementation in PYHF [1, 2]. The following is based on [1, 3, 4].

1.1 Profile Likelihood Ratio

The statistical model needs to reflect the compatibility of predictions with the observed collision events. This can be quantified by a likelihood $L(\boldsymbol{x}|\boldsymbol{\phi})$ which is a probability for an observation $\boldsymbol{x}$ under a given set of parameters $\boldsymbol{\phi}$ that govern

the predictions. Given that this is a counting experiment bins of a histogram $\boldsymbol{h} = (h_1, ..., h_N)$ are the main tool of analysis.

The observation can be subdivided $\boldsymbol{x} = (\boldsymbol{n}, \boldsymbol{a})$ into observable histograms $\boldsymbol{n}$ and auxiliary measurements. Observable histograms could be the invariant mass of a particle on the other hand auxiliary measurement histograms $\boldsymbol{a}$ can be any additional observable that assist in constraining the model. For instance they can be a measurement of a kinematic variable in a phase space region where only background is expected. It should be noted that these auxiliary measurements are not equivalent to the inclusion of uncertainties into the model. The treatment of uncertainties is addressed separately in section 1.4.

Another useful splitting for the set of parameters $\boldsymbol{\phi} = (\boldsymbol{\psi}, \boldsymbol{\Theta})$ into so-called parameters of interest $\boldsymbol{\psi}$ and nuisance parameters $\boldsymbol{\Theta}$. For this section only one parameter of interest is considered, the signal strength μ.

The bin contents can then be expressed in terms of the amount of signal $s_i(\boldsymbol{\Theta})$ and background $b_i(\boldsymbol{\Theta})$ in bin i that depend on the nuisance parameters. The prediction (expectation value) of the histogram bins of the observable n_i can then be expressed as

$$\langle n_i(\mu, \boldsymbol{\Theta}) \rangle = \mu s_i(\boldsymbol{\Theta}) + b_i(\boldsymbol{\Theta}). \tag{1.1.1}$$

Similarly for auxiliary measurement bins a_i their expectation value are calculable from some function $u_i(\boldsymbol{\Theta})$ modeling some observable and is also dependent on the nuisance parameters

$$\langle a_i(\boldsymbol{\Theta}) \rangle = u_i(\boldsymbol{\Theta}). \tag{1.1.2}$$

Since this is a counting experiment in which events occur at a constant mean rate and independently of time each bin follows a Poisson distribution

$$P(r, k) = \frac{r^k e^{-r}}{k!}. \tag{1.1.3}$$

r is the expected rate of occurrences which translates as the prediction whereas k are the actual measured occurrences. A likelihood can then be constructed from a

product of Poisson probabilities

$$L(\mu, \boldsymbol{\Theta}) = \prod_{j=1}^{N} \frac{(\mu s_j(\boldsymbol{\Theta}) + b_j(\boldsymbol{\Theta}))^{n_j}}{n_j!} e^{-(\mu s_j(\boldsymbol{\Theta}) + b_j(\boldsymbol{\Theta}))} \prod_{k=1}^{M} \frac{u_k(\boldsymbol{\Theta})^{a_k}}{a_k!} e^{-u_k(\boldsymbol{\Theta})}. \quad (1.1.4)$$

The last product can also be thought of penalizing the likelihood if e.g. an auxiliary measurement displays a very improbable value for some quantity. To test for a hypothesized value of μ, the best choice according to the Neyman-Pearson lemma [4], is the profile likelihood ratio that reduces the dependence to the parameter(s) of interest

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\Theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\Theta}})}. \quad (1.1.5)$$

The denominator is the unconditional maximum likelihood estimate so that $\hat{\mu}$ and $\hat{\boldsymbol{\Theta}}$ both are free to vary to maximize L, whereas the numerator is the found maximum likelihood conditioned on some chosen μ and the set of nuisance parameters $\hat{\hat{\boldsymbol{\Theta}}}$ that maximize the likelihood for that particular μ. This definition gives $0 \leq \lambda \leq 1$ where $\lambda = 1$ corresponds to perfect agreement of the hypothesized value of μ to the model.

1.2 Test Statistic and p-value

To test for alternative hypotheses it is useful to transform the profile likelihood into a test statistic

$$t(\mu) = -2 \ln \lambda(\mu). \quad (1.2.1)$$

This translates to $t \to 0$ as increasing agreement and $t \to \infty$ as decreasing agreement to the model. A right-tail p-value can then be calculated from the probability density function of the test statistic: Parton Density Function (PDF)$(t) = f(t|\mu)$

$$p = \int_{t_{\text{obs}}}^{\infty} f(t|\mu) \mathrm{d}t \quad (1.2.2)$$

t_{obs} is the test statistic t evaluated at the observed data which means replacing the predictions in the likelihood of the numerator of equation 1.1.5 with the values observed in data. Similar to the PDF of a standard normal distribution the PDF

in this context quantifies how probable a particular value of the test statistic t is under a fixed value of the signal strength. This essentially measures how frequently a particular value of t occurs in comparison to all other possible values that t can take.

To calculate p-values the integral of equation 1.2.2 must be solved. The test statistic's specific form is useful because of existing approximations for $f(t|\mu)$ [3]. Let $f(t|\mu')$ be the probability distribution for the true strength parameter μ'. Wald [5] demonstrated that for a single parameter of interest the test statistic is equivalent to a normalized sum of squared distances between the tested parameter μ and its maximum likelihood estimate $\hat{\mu}$

$$t(\mu) = -2\ln\lambda(\mu) = \left(\frac{\mu - \hat{\mu}}{\sigma_{\hat{\mu}}}\right)^2 + \mathcal{O}(\frac{1}{\sqrt{N}}). \tag{1.2.3}$$

The maximum likelihood estimate $\hat{\mu}$ is in the large sample limit normally distributed around their true value μ' with standard deviation $\sigma_{\hat{\mu}}$. This is the definition of a χ-squared distribution with one degree of freedom. It can be shown that [3] the PDF of t is asymptotically follows

$$f(t|\mu) = \frac{1}{2\sqrt{t}}\frac{1}{\sqrt{2\pi}}\left[\exp\left(-\frac{1}{2}\left(\sqrt{t} + \sqrt{\Lambda_\mu}\right)\right) + \exp\left(-\frac{1}{2}\left(\sqrt{t} - \sqrt{\Lambda_\mu}\right)\right)\right], \tag{1.2.4}$$

with the non-centrality parameter as the normalized distance between the tested μ and true parameter of interest μ'

$$\Lambda_\mu = \frac{(\mu - \mu')^2}{\sigma^2}. \tag{1.2.5}$$

Figure 1.1 illustrates these steps. Being able to calculate p-values allows to state how likely it is that the proposed hypothesis is reflected by the observed data. In other words, the p-value represents the probability, how incompatible the proposed hypothesis or prediction is with the observation.

In the scientific community a p-value of 0.05 is commonly accepted as significant. Though particle physicists only claim discovery of a new phenomenon for $p < 2.87 \times 10^{-7}$ corresponding to 5 standard deviations of the standard normal distribution and exclude hypotheses if the p-value is not below 2 standard deviations of the

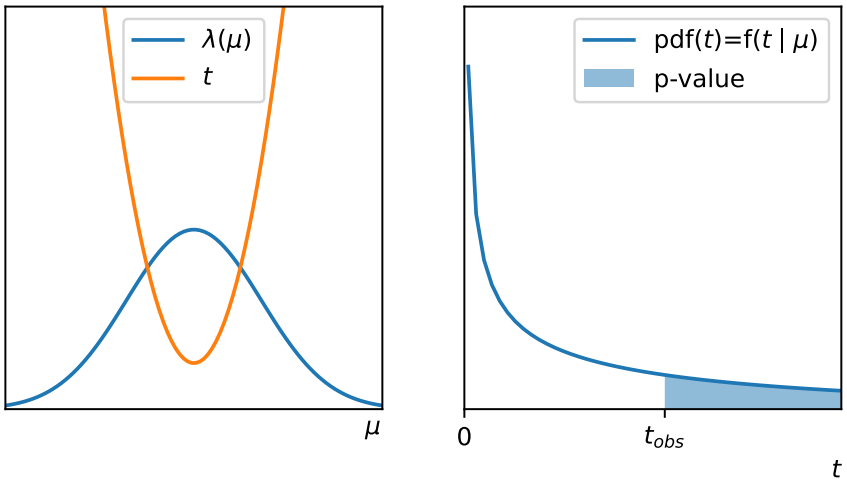


Figure 1.1: A sketch to follow the steps to calculate p-values. (**left**) The profile likelihood (／) has essentially some hill-like form with a maximum at $\lambda(\hat{\mu}, \hat{\mathbf{\Theta}})$. The test statistic t (／) is calculated as $-2\ln(\lambda)$. (**right**) For one parameter of interest in the large sample limit $f(t|\mu)$ from equation 1.2.4 follows a non-central chi-squared distribution with one degree of freedom. The blue shaded area under the PDFs is a right hand sided p-value.

standard normal distribution $p \lesssim 0.05$. A key consideration is that t can assume negative values for μ which might be non-physical depending on the context. This is handled by cutting off the test statistic for undesired behavior. An example of an adjusted test statistic for setting upper limits is

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}. \tag{1.2.6}$$

Here if a tested signal strength μ is not larger than the maximum estimate it would not be regarded as less compatible and is therefore set to zero. [3] covers other cases and pdf approximations for various scenarios.

1.3 The CL_s value

Particle physicists typically concentrate on two key aspects when performing statistical tests to discover new phenomena: the accurate modeling of known backgrounds and whether there is evidence in the observations for a new phenomenon. This involves assessing two distinct hypotheses: a background only (b) and one that involves signal and background ($s + b$). Each will result in a p-value of their own.

For example, a p-value of $p_b = 0$ would indicate a perfect modeling of the background model reflected in observational data, meaning that known phenomena have been accurately accounted for. Conversely, a p-value of $p_{s+b} < 0.05$ might signal the presence of new phenomena, such as previously undiscovered physics.

To synthesize these two aspects into a unified metric particle physicists developed a pseudo Confidence Level/p-value referred to as CL_s. This measure not only considers the potential presence of new phenomena but also the accuracy of the background modeling.

$$\mathrm{CL}_s = \frac{p_{s+b}}{1 - p_b} = \frac{\int_{t_{\mathrm{obs}}}^{\infty} f(t_{s+b}|\mu)\mathrm{d}t}{1 - \int_{t_{\mathrm{obs}}}^{\infty} f(t_b|\mu)\mathrm{d}t}. \tag{1.3.1}$$

The numerator represents the p-value for the alternative hypothesis while the denominator p-value penalizes the CL_s based on how compatible the background model is with the observational data. This concept can also be understood visually

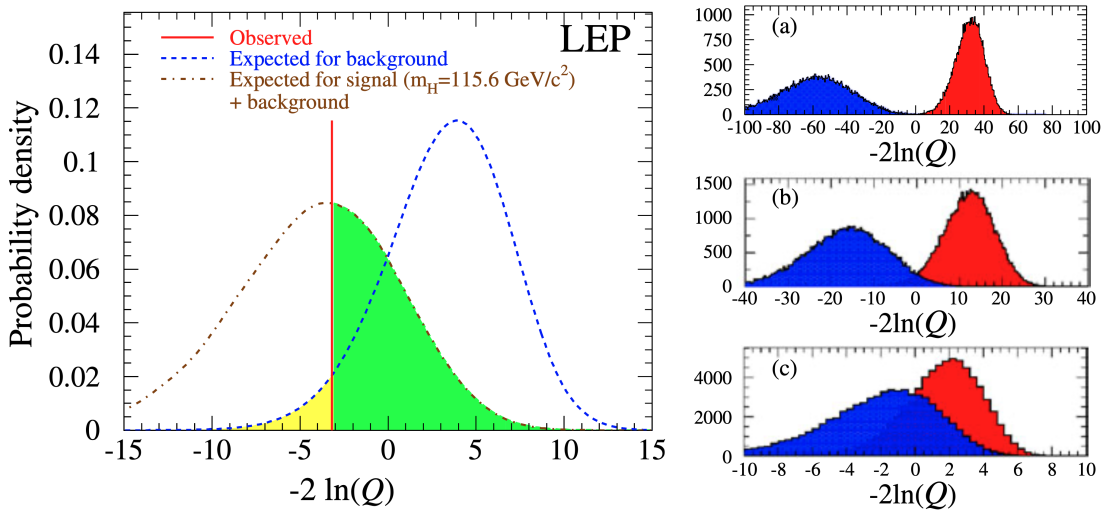


Figure 1.2: Probability density functions of test statistics from a Higgs search at LEP illustrating the calculation of p-values (λ becomes Q). (**left**) The PDFs's of the test statistic $f(t|\mu)$ of the signal + background (/) and background (/) only hypotheses. The p-value is calculated by integration from t_{obs} (the red observed line (/)) to infinity (see eq. 1.2.2). The green shaded area (■) corresponds to p_{s+b} whereas the yellow area (■) corresponds to $1 - p_b$ since the integral over one whole PDFs is 1. (**right**) Degradation of search sensitivity from (a) to (c). Note that the colors of the PDFs's change here to signal + background (■) and background only (■). For example putting the observation (t_{obs}) on the x-axis at 0 in these plots, one would get for plot (a) $p_b \approx 1$ and $p_{s+b} \approx 0$ resulting in a $\text{CL}_s \approx 0$, whereas with increasing overlap the CL_s value increases and the sensitivity decreases. Adopted from [6].

from the first figure of the paper that introduced the CL_s quantity [6] and is explained here in figure 1.2.

1.4 HistFactory

A widely-used model for constructing likelihoods as discussed in section 1.1 is known as HistFactory [7]. This model is implemented in the PYHF toolkit [1] and the following section is primarily based on the introduction to HistFactory found in the PYHF documentation. HistFactory simplifies the process of building a likelihood by breaking it down into several fundamental components. To understand this it is

helpful to consider a different categorization of the model parameters $\boldsymbol{\phi}$

$$L(\boldsymbol{x}|\boldsymbol{\phi}) = L(\boldsymbol{x}|\overbrace{\boldsymbol{\psi}}^{\text{parameters of interest}}, \underbrace{\boldsymbol{\theta}}_{\text{nuisance parameters}}) = L(\boldsymbol{x}|\overbrace{\boldsymbol{\eta}}^{\text{free}}, \underbrace{\boldsymbol{\chi}}_{\text{constrained}}), \tag{1.4.1}$$

Free parameters $\boldsymbol{\eta}$ are free to choose in the model and can be for example a cross-section of a process. Constrained parameters $\boldsymbol{\chi}$ are used to incorporate uncertainties into the likelihood to constrain it. Further there might be several histograms of an observable, for example measured in orthogonal kinematic regions, that are called channels c. Bins have the index b here and constraint terms are denoted c_χ. The likelihood can thus be described by

$$L(\boldsymbol{n}, \boldsymbol{a} \,|\, \boldsymbol{\eta}, \boldsymbol{\chi}) = \underbrace{\prod_{c \in \text{ channels}} \prod_{b \in \text{ bins}_c} \text{Pois}(\overbrace{n_{cb}}^{\text{observed}} \,|\, \overbrace{\nu_{cb}(\boldsymbol{\eta}, \boldsymbol{\chi})}^{\text{predicted}})}_{\substack{\text{Simultaneous measurement} \\ \text{of multiple channels}}} \underbrace{\prod_{\chi \in \boldsymbol{\chi}} c_\chi(a_\chi | \, \chi)}_{\substack{\text{constraint terms} \\ \text{for auxiliary measurements}}} . \tag{1.4.2}$$

$\boldsymbol{n}$ and $\boldsymbol{a}$ are the observations auxiliary measurement hisograms as defined in section 1.1. The n_{cb} is the observed bin content and $\nu_{cb}(\boldsymbol{\eta}, \boldsymbol{\chi})$ the predicted bin content. $c_\chi(a_\chi | \, \chi)$ are functions that calculate a probabilistic impact to the likelihood L of uncertainties a_χ to constrain the parameter χ and are discussed in detail section 1.4.2.

The prediction is a sum of nominal bin counts[1] ν^0_{scb} over all samples s (e.g. $t\bar{t}$, multijet-background, etc.). These nominal bin counts are subject to uncertainties. Therefore the bin content can be varied within the bounds of these uncertainties. However the effect of this modification to the likelihood must be taken into account which is through the constraint terms. These penalize the likelihood proportional to the modification. Modifiers are discussed in detail in section 1.4.1. They enter the likelihood through linear modeling of the nominal bin content ν^0_{scb} with

[1] also called rates, like in the definition of a Poisson distribution

multiplicative κ_{scb} and additive modifiers Δ_{scb}

$$\nu_{cb}\left(\boldsymbol{\eta}, \boldsymbol{\chi}\right) = \sum_{s \in \text{ samples}} \nu_{scb}\left(\boldsymbol{\eta}, \boldsymbol{\chi}\right) \tag{1.4.3}$$

$$= \sum_{s \in \text{ samples}} \underbrace{\left(\prod_{\kappa \in \boldsymbol{\kappa}} \kappa_{scb}\left(\boldsymbol{\eta}, \boldsymbol{\chi}\right)\right)}_{\text{multiplicative modifiers}} \left(\nu^0_{scb} + \underbrace{\sum_{\Delta \in \boldsymbol{\Delta}} \Delta_{scb}\left(\boldsymbol{\eta}, \boldsymbol{\chi}\right)}_{\text{additive modifiers}}\right). \tag{1.4.4}$$

The usefulness of this approach becomes clear when considering one uncertainty a_χ on a nominal bin count estimate ν^0_{scb}. The main goal remains to maximize the overall likelihood L. This can be achieved via maximizing the Poisson probability (blue part in equation 1.4.2) and at the same time keeping the constraint term (red part in equation 1.4.2) large. It is illustrative to consider one nuisance parameter χ as a multiplicative modifier $\kappa_{scb} = \chi$ on ν^0_{scb}. An optimum can be found by modifying the prediction to move closer to the observed value (blue part in equation 1.4.2) while at the same time keeping the constraint term $c_\kappa(\kappa)$ (red part in equation 1.4.2) controlled by the same χ at values where the penalization of the likelihood stays insignificant. This works well for example when κ has a very large uncertainty rendering the constraint term's influence on the likelihood relatively small.

1.4.1 The Modifiers

In HistFactory there are by convention four types $\{\lambda, \mu, \gamma, \alpha\}$ of such multiplicative rate modifiers that are explained in this section. There are **free rate modifiers** for the luminosity λ and signal strength μ that affect all bins equally

$$\nu_{scb}(\mu) = \mu \nu^0_{scb}. \tag{1.4.5}$$

These are bin-independent normalization factors and preserve the shape of the histogram. Further there are **bin-wise modifiers** γ_b (uncorrelated shape)

$$\nu_{scb}(\gamma_b) = \gamma_b \nu^0_{scb}. \tag{1.4.6}$$

These are useful for example to include uncertainties of a per bin data-driven background estimate. This type without a constraint term is not of much use as if

there is only one sample or channel, the fit would always match the data perfectly. In addition there exist **interpolation parameters** α (shape factors) that enter the modeling through an interpolation function η instead of being the factor itself. They exist in multiplicative versions

$$\nu_{scb}(\alpha) = \eta(\alpha)\nu^0_{scb}, \tag{1.4.7}$$

and additive versions

$$\nu_{scb}(\alpha) = \nu^0_{scb} + \eta(\alpha). \tag{1.4.8}$$

This is useful to include systematic uncertainties. Typically they are known for one standard deviation of a bin count $\eta_{-1} = \nu^{\text{1down}}_{scb}$ and $\eta_1 = \nu^{\text{1up}}_{scb}$ to the nominal value ν^0_{scb}. These are then used to construct interpolation functions that modify the nominal value controlled by the nuisance parameter that also controls the constraint term c_α which is further explained in section 1.4.2. Thus one parameter controls the modification and constraint at a time.

In HistFactory there exists four of such interpolation functions. For those exist an identity operator

$$\eta_0 = \eta(\alpha = 0) = \begin{cases} 1, & \text{multiplicative modifier, } (\kappa) \\ 0, & \text{additive modifier, } (\lambda). \end{cases} \tag{1.4.9}$$

One example of these interpolation functions that scales the bin count linearly over the known deviations $\eta_{-1} = \nu^{\text{1down}}_{scb}$ and $\eta_1 = \nu^{\text{1up}}_{scb}$ is

$$\eta_{\text{linear}}(\alpha) = \begin{cases} \alpha(\eta_0 - \eta_1), & \alpha > 0 \\ \alpha(\eta_0 - \eta_{-1}), & \alpha < 0 \end{cases} \tag{1.4.10}$$

This is illustrated in fig. 1.3(a). For the other ones see e.g. [8]. It is noted that α is the nuisance parameter and not the function $\eta(\alpha)$ and there is an associated constraint term c_α to each α.

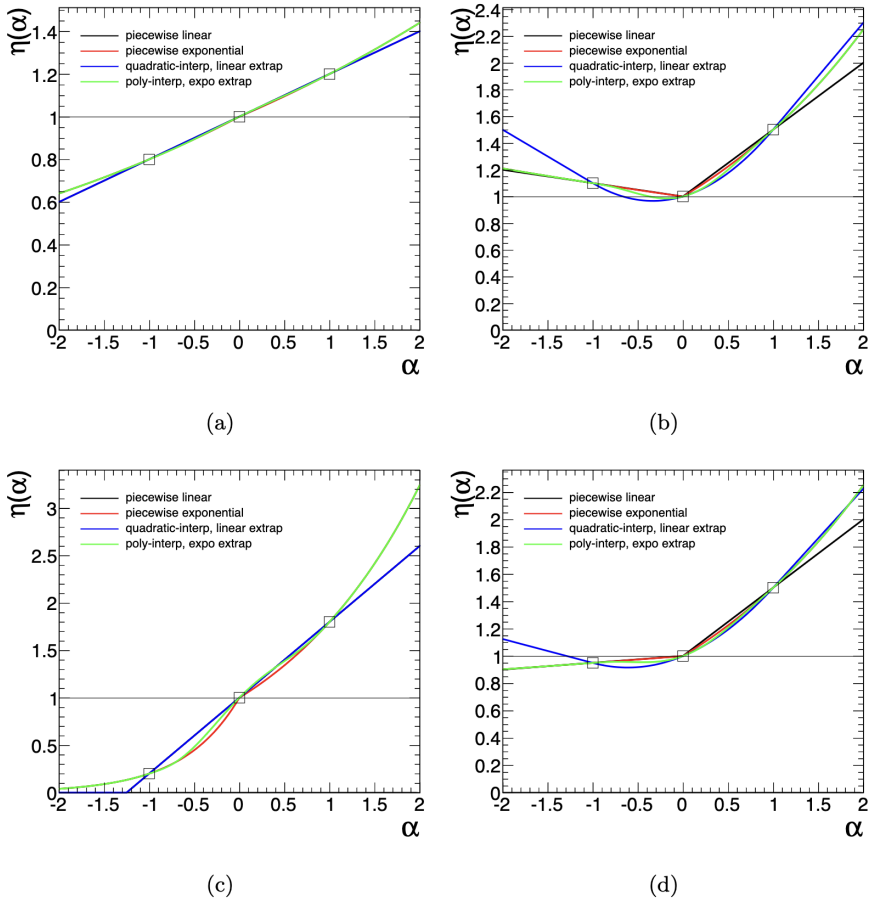


Figure 1.3: The four interpolation functions $\eta(\alpha)$ for different up and down standard deviation values. For example in (a) the bin count will be scaled with a factor of 0.8 for an $\alpha = -1$ (1.2 for an $\alpha = 1$). From [7].

1.4.2 The constraint terms

Uncertainties are modeled either Gaussian or Poissonian. The Gaussian uncertainty implementation is straightforward as the standard deviation σ appears in the definition of the Gaussian with mean μ

$$\text{Gaus}(\mu|x,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \tag{1.4.11}$$

So that the likelihood for a Gaussian uncertainty is constrained by a Gaussian scaled to one standard deviation controlled by the nuisance parameter α: $\text{Gauss}(\alpha|a,\sigma=1)$.

Similar to the Gaussian uncertainty for an uncertainty that is Poissonian distributed

$$\text{Pois}(k|r) = \frac{r^k e^{-r}}{k!}. \tag{1.4.12}$$

the nuisance parameter for a multiplicative factor $\kappa_{scb} = \gamma$ in equation 1.4.4 should control the Poisson constraint term such that the uncertainty σ is reflected by the variance of the Poisson $\text{Var}(\text{Pois}) = r = \sigma^2$. This is achieved by scaling the distribution with a factor f which is then solved for the one with the desired uncertainty by evaluating it at the nominal value of the multiplicative modifier $\gamma_0 = 1$

$$\text{Var}\left[\text{Pois}(k = f\gamma_0|r = f\gamma)\right] = r = f\gamma \overset{\gamma=\gamma_0}{=} f\gamma_0 = (f\sigma)^2 \quad \rightarrow \quad f = (1/\sigma^2). \tag{1.4.13}$$

Thus a Poissonian constraint term for a multiplicative modifier γ with uncertainty σ reads $\text{Pois}(k = \sigma^{-2}|r = \sigma^{-2}\gamma)$. This completes the necessities for the HistFactory model. The different types of modifiers and their constraint terms are summarized in table 1.1.

Table 1.1: Modifiers and constraint terms used in HistFactory implemented by PYHF. Note that the interpolation functions are called f_p and g_p here instead of η as chosen in the full text. Input for the constraint terms are the corresponding uncertainties. Adapted from [1].

Description	Modification	Constraint Term c_χ	c_χ input
Uncorrelated Shape	$\kappa_{scb}(\gamma_b) = \gamma_b$	$\prod_b \text{Pois}\left(k_b = \sigma_b^{-2} \middle\vert r_b = \sigma_b^{-2}\gamma_b\right)$	σ_b
Correlated Shape	$\Delta_{scb}(\alpha) = f_p\left(\alpha \middle\vert \Delta_{scb,\alpha=-1}, \Delta_{scb,\alpha=1}\right)$	$\text{Gaus}\left(a = 0 \middle\vert \alpha, \sigma = 1\right)$	$\Delta_{scb,\alpha=\pm1}$
Normalisation Unc.	$\kappa_{scb}(\alpha) = g_p\left(\alpha \middle\vert \kappa_{scb,\alpha=-1}, \kappa_{scb,\alpha=1}\right)$	$\text{Gaus}\left(a = 0 \middle\vert \alpha, \sigma = 1\right)$	$\kappa_{scb,\alpha=\pm1}$
MC Stat. Uncertainty	$\kappa_{scb}(\gamma_b) = \gamma_b$	$\prod_b \text{Gaus}\left(a_{\gamma_b} = 1 \middle\vert \gamma_b, \delta_b\right)$	$\delta_b^2 = \sum_s \delta_{sb}^2$
Luminosity	$\kappa_{scb}(\lambda) = \lambda$	$\text{Gaus}\left(l = \lambda_0 \middle\vert \lambda, \sigma_\lambda\right)$	$\lambda_0, \sigma_\lambda$
Normalisation	$\kappa_{scb}(\mu_b) = \mu_b$		
Data-driven Shape	$\kappa_{scb}(\gamma_b) = \gamma_b$		

Chapter 2

Studies with the Soft Muon Tagger

Analyses that involve final states with b-quarks heavily rely on the accurate identification of these quarks. Ongoing efforts are dedicated to refine this identification process. This chapter presents a study on how muons can contribute to the identification of b-quarks in current b-tagging algorithms applied to small-R jets.

2.1 Soft Muon tagging

Algorithms currently in use for b-tagging described in section **??**, primarily exploit the displaced secondary vertex that is characteristic of the long lifetime of b-quarks. Besides vertex finding algorithms muons can be used as indication for the presence of a b-hadron.This is because approximately 20 % of the b-jets undergo semi-leptonic decays that involve a muon as exemplified in figure 2.1. The branching fractions for these decays are roughly 11 % for direct b to muon decays $BR(b \to \mu\nu X)$ and about 10 % for cascade decays $BR(b \to c \to \mu\nu X)$ [9].

Despite their limited branching ratio, muons from semi-leptonic heavy-flavor decays are valuable in complementing impact parameter- and vertex-based b-tagging methods. These muons typically exhibit higher transverse momentum compared to those from lighter hadron decays, as shown in figure 2.2(a). This relates also to the fact that muons from b-decays tend to be more boosted transverse to the jet

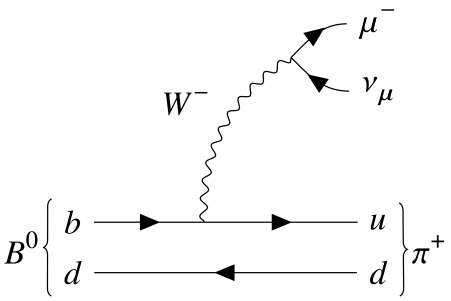


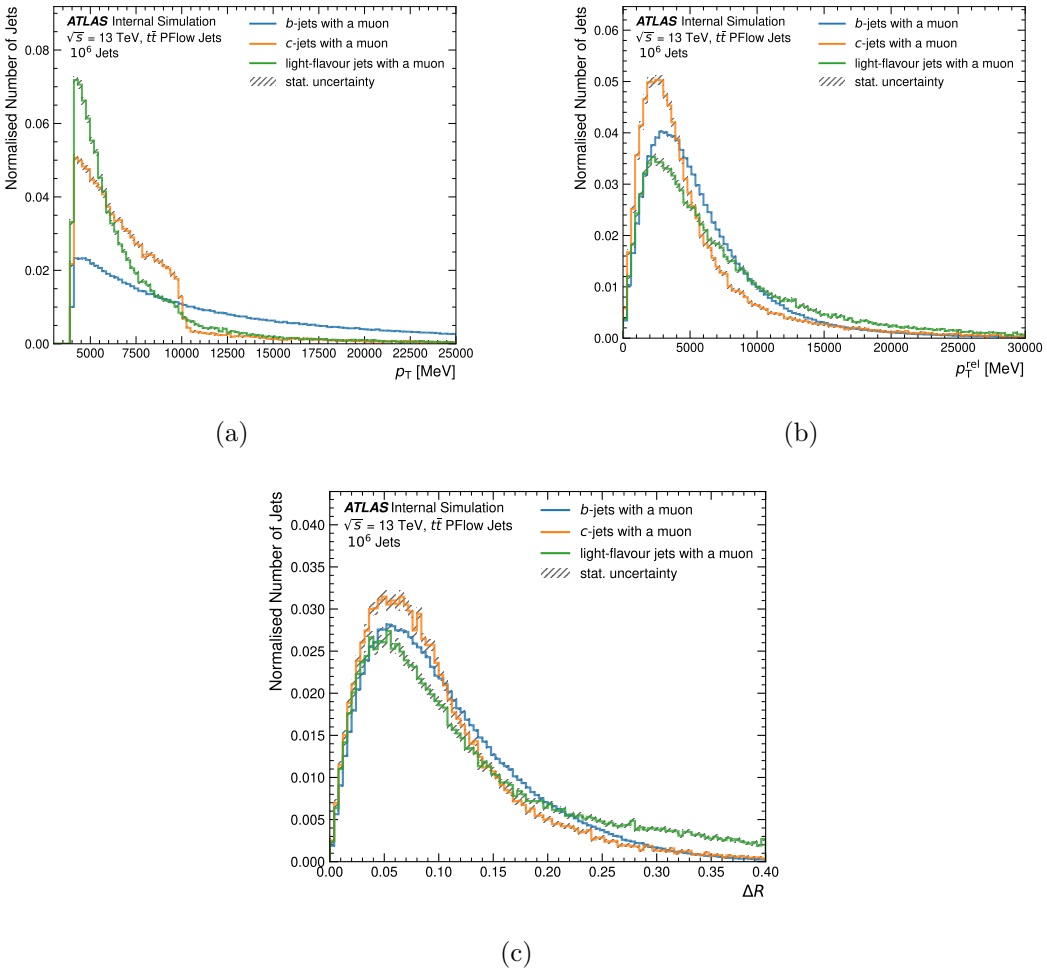
Figure 2.1: Feynman Diagram depicting a typical semi-leptonic decay of an Anti-B^0-meson containing a muon in the final state.

axis and therefore have a larger orthogonal projection $p_{\mathrm{T}}^{\mathrm{rel}}$ of the muon-p_{T} onto the jet axis. This is calculated by taking the perpendicular part of the three vector muon-momentum to the three-vector jet momentum. The term "soft" is derived from the fact that their p_{T} is smaller than that typically possessed by muons from electroweak boson decays, since they originate from a secondary process [10].

2.2 Muon Selection

The particle flow algorithm used to reconstruct small-R jets described in section **??** does not include muons. To address this muons are associated to jets using the shrinking association ΔR cone as depicted in figure 2.3. Selected muons must be combined muons meaning they are reconstructed using both the inner detector and the muon spectrometer. The closest muon within $\Delta R_{\mu,\mathrm{jet}} < 0.4$ to the jet axis is chosen provided it has a minimum p_{T} of 4 GeV. This threshold is set as muon reconstruction below 3 GeV is generally unreliable due to minimally ionizing particles losing approximately 3 GeV in the ATLAS calorimeters [9].

Truth studies revealed that most muons associated to jets result from secondary (non-prompt) decays of c- and b-hadrons as as evidenced in figure 2.4(a). However this figure also reveals the introduction of b-tagging backgrounds since muons are also associated to light jets. These backgrounds predominantly arise from in flight decays of pions and kaons but also prompt muons from nearby W-boson decays and muons from light and strange mesons as shown in figure 2.4(b). Table 2.1 shows the proportion of jets with an associated muon for each flavor. Approximately 15 % of



(a) (b)

(c)

Figure 2.2: Kinematics of soft muons: **(a)** transverse momentum distribution, **(b)** relative transverse momentum ($p_{\mathrm{T}}^{\mathrm{rel}}$) indicative of muons from direct b-jet decays being more boosted transversely to the jet axis, and **(c)** ΔR distribution between soft muons and jets, normalized per flavor.

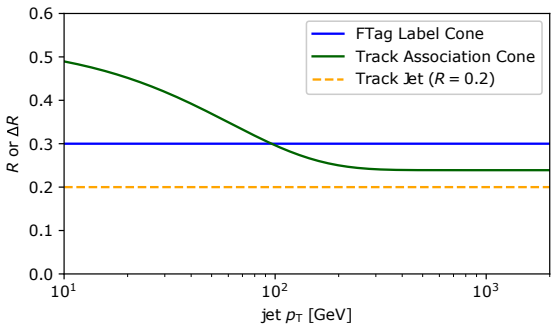
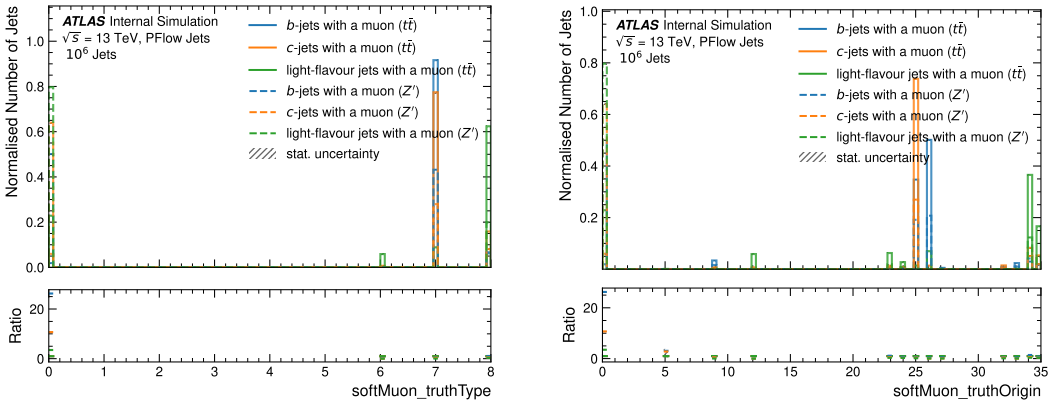


Figure 2.3: The shrinking particle association cone (／) used to associate muons with a $\Delta R_{\mu,\text{jet}}$ smaller than the cone depending on the transverse momentum of the jet. Adopted from [11].



(a) (b)

Figure 2.4: Jet truth type (a) and truth origin (b) for jets with an associated muon normalized per flavor from $t\bar{t}$ and Z' samples. Truth types: unclassified (0), prompt (6), non-prompt (7) and background muon (8). Truth origins: tau lepton decays (9), prompt muons from nearby W decays (12), light (23), strange (24), charm (25), bottom (26) mesons and in flight decays of pions (34) and kaons (35) according to the Monte Carlo truth classification [12].

b-jets and around 1 % of both light and tau jets have an associated muon. Figure 2.5

Table 2.1: Fraction of associated muons per flavor. The large fraction for b-jets compared to the other flavors is the reason why muons are a useful discriminator for b-tagging.

Jet flavor	Fraction of jets with an associated muon
light	0.0134
c	0.0472
b	0.1490
tau	0.0138

visualizes the amount of fake muons in b-jets with an associated muons depending on the jet- and muon-p_{T}. A notable trend is that b-jets with low momentum tend to have low momentum muons. Falsely associated muons become more common at high jet momentum and low muon momentum.

2.3 Soft Muon Variables

The soft muon variables are selected based on their b-tagging discrimination power. The impact parameters of jets as described in section **??** are determined with the three dimensional impact parameter algorithm (IP3D) detailed in [10] and are shown in figure 2.6(a)-2.6(d). In addition variables are used to reject muons from in flight decays of Kaons and Pions [13]. The Scattering Neighbor Significance measures if the track of a particle has kinks that could be a sign of an additional decay. By connecting neighboring detector hits along the track with straight lines and measuring the angles between adjacent lines the Scattering Neighbor Significance is obtained by taking the maximum of the angle divided by its uncertainty. In figure 2.7(a) light jets therefore tend to have larger values as they have a smaller p_{T} (cf. figure 2.2(a)). Another variable to identify possible background sources is the Momentum Balance Significance. It calculates the difference of the muon-p_{T} determined with the inner detector with the one from the muon spectrometer and corrects it for energy deposits in the calorimeters. This would be zero in figure 2.7(b) for a perfect energy loss correction and if the muon did not arise from an in flight decay. Furthermore there is the curvature comparison between inner

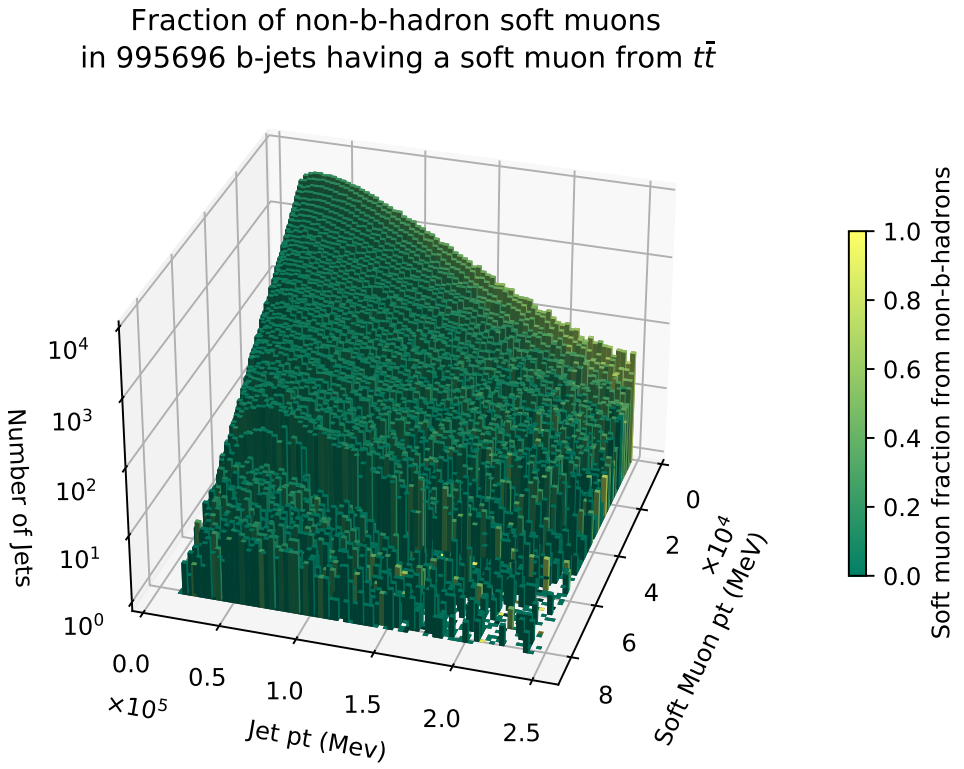
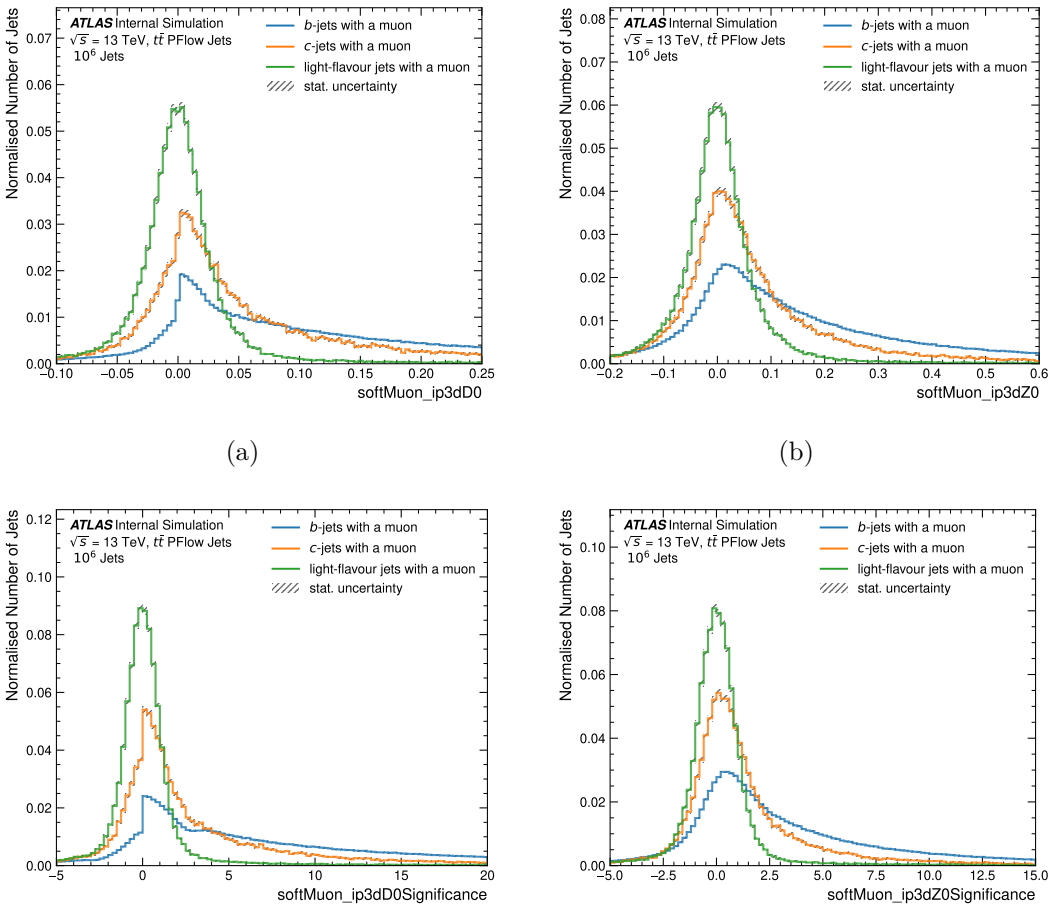


Figure 2.5: 2D histograms of $p_{\mathrm{T}}^{\mathrm{jet}}$ and $p_{\mathrm{T}}^{\mathrm{muon}}$ for $t\bar{t}$ selected on b-jets only. The colorbar is the fraction of muons that do not originate from b-hadrons. The distributions are actually smooth although some irregularities are visible which is due to a bug in the plotting library.

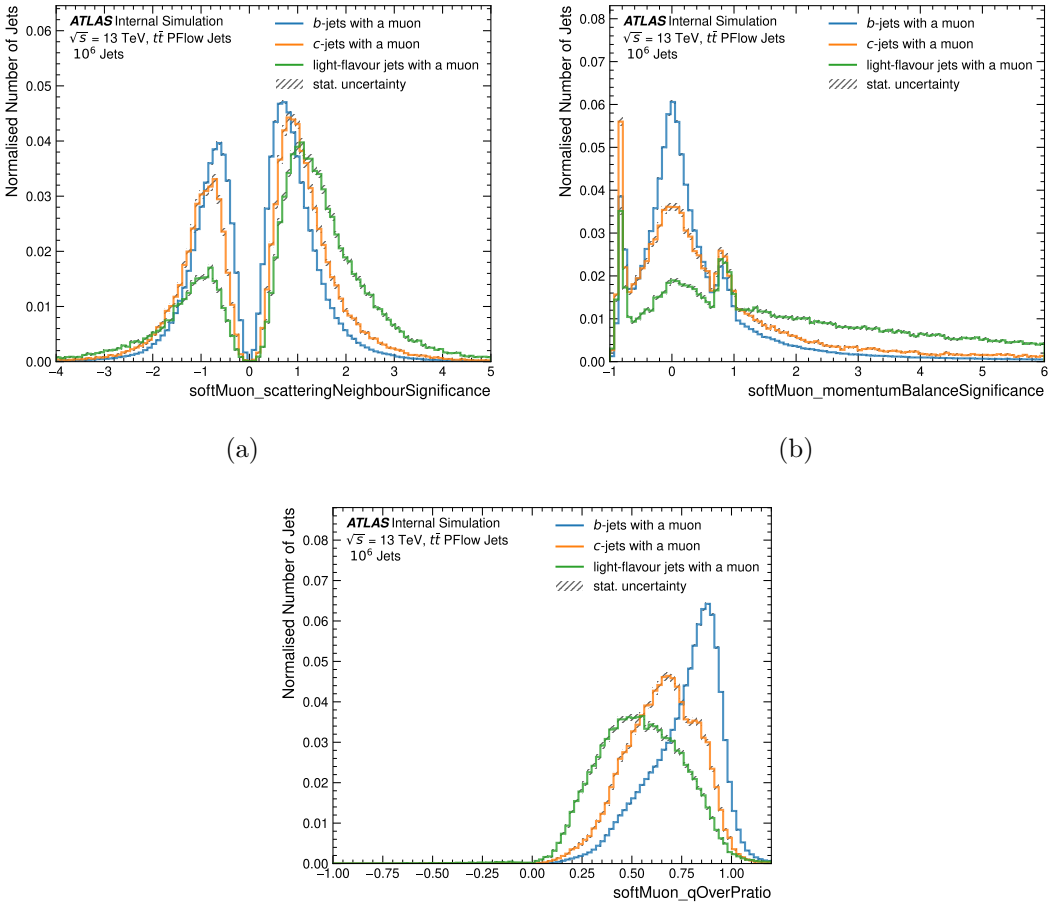


(a) (b)

(c) (d)

Figure 2.6: Impact parameters of the soft muon retrieved with the IP3D algorithm [10], normalized per flavor.

detector (ID) and the one from the muon spectrometer (MS) termed q over p ratio: $(q/p)_{\mathrm{ID}}$ / $(q/p)_{\mathrm{MS}}$ in figure 2.7(c).



(a) (b)

(c)

Figure 2.7: Variables to reject light jet background. Description in section 2.3.

2.3.1 Adding muons to *b*-tagging algorithms

Current *b*-tagging algorithms in use are deep feed-forward neural networks with the prefix DL1 that assign a score $(p_{\mathrm{b}}, p_{\mathrm{c}}, p_{\mathrm{light}})$ to a jet depending on its flavor content. Depending on the strategy used to introduce additional tracking information associated with the jet it can become DL1r if the outputs of the RNNIP (Recurrent Neural Network Impact Parameter) [14] tagger or DL1d if the outputs of the DIPS

(Deep Impact Parameter Sets) [15] tagger are added as inputs to DL1. The ouputs of these taggers are then combined into a signal to background ratio variable, the discriminant

$$\mathrm{D} = \ln\left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{\mathrm{light}}}\right). \tag{2.3.1}$$

f_c is the fraction of charm jets in the background and can be used to tune the importance of the different background classes ($\sum f_{\mathrm{bkg}} = 1$).

Information of the selected muon variables are added to these tracking information inclusive taggers as additional inputs. Two ways of introducing the muon information were explored. The first refers to an approach where the muon variables itself were used to train an additional neural network called Soft Muon Tagger (SMT) with architecture: (soft muon input variables) $\rightarrow$ 3 hidden layers (100, 20, 10) $\rightarrow$ 3 outputs nodes for classification of the jet flavor ($p_{\mathrm{b}}, p_{\mathrm{c}}, p_{\mathrm{light}}$). The output scores of the SMT are then added as inputs to the DL1 training. The tagger resulting from this method is called DL1r_SMT. In a second approach the muon variables were added directly as inputs to the DL1r training and are referred to here as DL1rmu. The resulting discriminant comparing DL1d and DL1dmu is depicted in Fig. 2.8

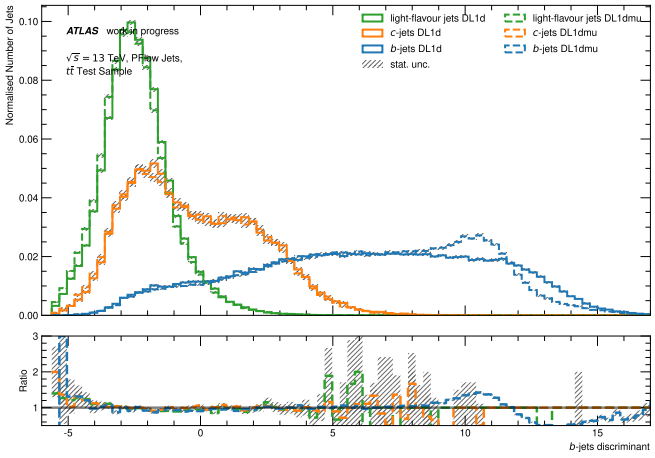
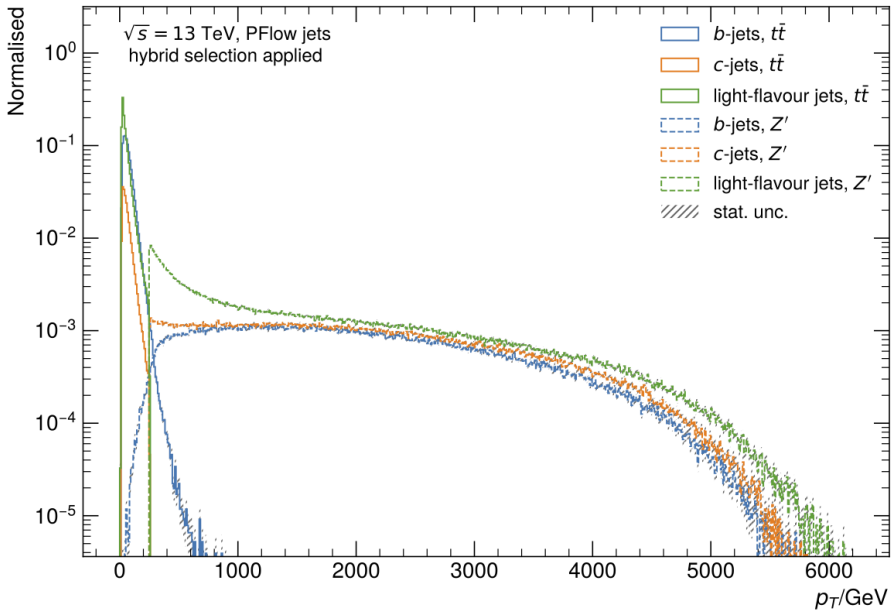
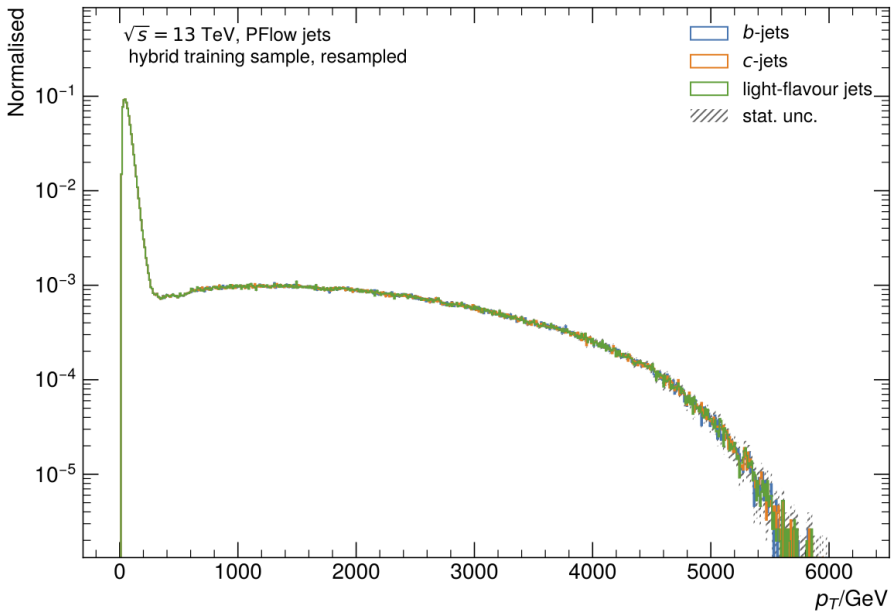


Figure 2.8: Discriminant comparison between DL1d (／) and DL1dmu (- -). The performance of the model increases if the different classes are visually more separated.

For all trainings the Umami neural network training framework [16] is used with $t\bar{t}$ and Z' Monte Carlo samples described in [10]. Z' is a hypothetical particle used here to enhance the statistics at larger jet-p_{T} >250 GeV. For the training both $t\bar{t}$ and Z' samples are merged and resampled to make the p_{T} and η distributions appear the same for all flavors. This is done to allow for a fair training between flavor categories, as the kinematic regimes are neither over- nor under-represented between flavors when the neural network is trained. The flavor composition of jets can be seen before and after applying the resampling in figure 2.9.



(a)



(b)

Figure 2.9: Jet-p_{T} per flavor normalized to unity for $t\bar{t}$ and Z' samples **(a)** before and **(b)** after resampling. Adopted from [17].

Appendices

Appendix A

Acronyms

CERN Organisation européenne pour la recherche nucléaire

ATLAS A Toroidal LHC Apparatus

SM Standard Model

QFT Quantum Field Theory

QCD Quantum Chromodynamics

QED Quantum Electrodynamics

EW Electroweak

EWSB Electroweak Symmetry Breaking

VEV Vacuum Expectation Value

CKM Cabibbo-Kobayashi-Maskawa

EM electromagnetic

IP impact parameter of tracks

ML Machine Learning

neos neural end-to-end-optimized summary statistics

HEP High Energy Physics

LHC Large Hadron Collider

HL-LHC High Luminosity LHC

ID Inner Detector

SCT semiconductor tracker

TRT transition radiation tracker

IBL insertable b-layer

HLT high level trigger

L1 Level-1

PFO Particle Flow Object

TCC Track CaloCluster

UFO Unified Flow Object

JES Jet Energy Scale

JER Jet Energy Resolution

JMR Jet Mass Resolution

GGF gluon-gluon fusion

VBF vector-boson fusion

NNLO next-to-next-to-leading order

N^{3}LO next-to-next-to-next-to-leading order

SR Signal Region

VR Validation Region

CR Control Region

KDE Kernel Density Estimation

bKDE binned Kernel Density Estimation

MC Monte Carlo

PDF Parton Density Function

PV primary vertex

JVT jet vertex tagger

NN Neural Network

ANN Artificial Neural Network

WP working point

VR variable radius

SMT Soft Muon Tagger

Appendix B

Cutflow

TODO, also fine like that?

Selection	Event	Fraction [%]	Total Fraction [%]
Initial	16854036422.000		
Preselections (MNT + Jet Cleaning)	670573995.000	100.000	100.000
PassTrigBoosted	63944638.000	9.536	9.536
PassTwoFatJets	57510800.000	89.938	8.576
PassTwoHbbJets	12875.000	0.0223	<0.001
PassVBFJets	5762.000	44.753	<0.001
PassFatJetPt	3902.000	67.720	<0.001
PassVBFCut	314.000	8.047	<0.001

Table B.1: Cut-flow table for data before signal region cut

Selection	Event	Fraction [%]	Total Fraction [%]
Initial	1475.226		
Preselections (MNT + Jet Cleaning)	547.960	100.000	100.000
PassTrigBoosted	20.926	3.819	3.819
PassTwoFatJets	14.141	67.576	2.581
PassTwoHbbJets	5.353	37.852	0.977
PassVBFJets	2.243	41.903	0.409
PassFatJetPt	1.408	62.793	0.257
PassVBFCut	0.148	10.539	0.027
PassSR	0.097	65.484	0.018
OverlapRemoval	0.059	61.200	0.011

Table B.2: Cut-flow table for DSID = 600463

Bibliography

[1] Lukas Heinrich, Matthew Feickert, and Giordon Stark. pyhf: v0.7.2. URL https://doi.org/10.5281/zenodo.1169739. https://github.com/scikit-hep/pyhf/releases/tag/v0.7.2.

[2] Lukas Heinrich, Matthew Feickert, Giordon Stark, and Kyle Cranmer. pyhf: pure-python implementation of histfactory statistical models. *Journal of Open Source Software*, 6(58):2823, 2021. doi:10.21105/joss.02823. URL https://doi.org/10.21105/joss.02823.

[3] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71:1–19, 2011.

[4] Olaf Behnke, Kevin Kröninger, Grégory Schott, and Thomas Schörner-Sadenius. *Data analysis in high energy physics: a practical guide to statistical methods*. John Wiley & Sons, 2013.

[5] Abraham Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, 54(3):426–482, 1943.

[6] Alexander L Read. Presentation of search results: the cls technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693, 2002.

[7] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. Technical report, New York U., New York, 2012. URL https://cds.cern.ch/record/1456844.

[8] Lukas Heinrich. *Searches for Supersymmetry, RECAST, and Contributions to Computational High Energy Physics*. PhD thesis, New York University, 2019.

[9] The ATLAS Collaboration. *Expected performance of the ATLAS experiment: detector, trigger and physics*. CERN, Geneva, 2009. URL `http://cds.cern.ch/record/1125884`.

[10] ATLAS Collaboration. Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run. ATL-PHYS-PUB-2017-013, 2017. URL `https://cds.cern.ch/record/2273281`.

[11] Ruth Magdalena Jacobs. Probing new physics with boosted $H \to b\bar{b}$ decays with the ATLAS detector at 13 TeV, 2019. URL `https://cds.cern.ch/record/2697316`. Presented 08 Oct 2019.

[12] The ATLAS Collaboration. MC truth classification, . URL `https://gitlab.cern.ch/atlas/athena/-/blob/005da0491b3f8bd457f4c00bcadc3a92b9ef6f14/PhysicsAnalysis/MCTruthClassifier/MCTruthClassifier/MCTruthClassifierDefs.h`.

[13] Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2020. URL `https://cds.cern.ch/record/2725736`. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2020-030.

[14] Geoffrey Gilles, Manuel Guth, Philipp Gadow, Chris Pollard, Sanmay Ganguly, Rafael Teixeira De Lima, Michael Aaron Kagan, Binbin Dong, Nicole Michelle Hartman, Ning Zhou, Dan Guest, Andrea Coccaro, Francesco Armando Di Bello, Martino Tanasini, Marco Cristoforetti, Andrea Helen Knue, Arnaud Duperrin, and Andrea Di Luca. ATLAS b-tagging algorithms for the LHC Run 2 dataset. Technical report, CERN, Geneva, 2022. URL `https://cds.cern.ch/record/2806947`.

[15] Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS. Technical report, CERN, Geneva, 2020. URL https://cds.cern.ch/record/2718948. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-014.

[16] Alexander Froch, Manuel Guth, Joschka Birk, Philipp Gadow, Maxence Draguet, Tomke Schroer, Samuel Van Stroud, Victor Hugo Ruelas Rivera, Frederic Renner, Jackson Barr, Janik Von Ahnen, Stefano Franchellucci, Sebastien Rettie, Nikita Ivvan Pond, Martino Tanasini, Ivan Oleksiyuk, Osama Karkout, and Dmitrii Kobylianskii. Umami: A Python toolkit for jet flavour tagging in the ATLAS experiment. Technical report, CERN, Geneva, 2023. URL https://cds.cern.ch/record/2857164.

[17] The ATLAS Collaboration. Documentation for the umami framework, . URL https://umami.docs.cern.ch.

Statutory Declaration - Eidesstattliche Erklärung

I declare that I have authored this thesis independently, that I have not used other than the declared sources/ resources and that I have explicitly marked all materials which has been quoted either literally or by content form the used sources.

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Berlin, 15.12.2023

Frederic Renner