Trigger studies on 2017 CMS data for decays into VV and Vq with dijet final states

Bachelor Thesis

Marco Link

At the Department of Physics
Institut für Experimentelle Teilchenphysik (ETP)

Reviewer: PD Dr. Matthias Mozer
Second reviewer: Dr. Thorsten Chwalek

Karlsruhe, 12. March 2018
This thesis has been accepted by the first reviewer of the bachelor thesis.

Karlsruhe, 12. March 2018

..................................................  
(PD Dr. Matthias Mozer)
I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

Karlsruhe, 12. March 2018

(Marco Link)
Abstract

In this thesis the trigger efficiencies of HLT_AK8PFJet500, HLT_PFHT1050 and certain substructure triggers with the form HLT_AK8PFHT***_TrimMass50 are studied on a 2017 SingleMuon dataset recorded with the Compact Muon Solenoid (CMS) detector at $\sqrt{s} = 13$ TeV. The trigger efficiency is calculated as function of the dijet invariant mass $M_{jj}$, the transverse energy $H_T$ and the leading jet softdrop mass. The event selection matches the one used in the search for resonances in the dijet invariant mass spectrum from decays into a quark and a vector boson (Vq) or decays into two vector bosons (VV) with dijet final states.

In run period B and early run period C the HLT_PFHT1050 trigger shows an inefficiency due to a misimplementation of the Jet Energy Corrections (JEC). After that the trigger turn on shows no other suspicious behavior. The combination of the triggers has more than 99% efficiency with $M_{jj} > 1170.9$ GeV for events with single V tag and $M_{jj} > 1165.5$ GeV for events with double V tag.
1. Introduction

Subject of this bachelor thesis is the performance of the triggers used in the search for resonances in the dijet invariant mass spectrum with the 2017 data of the Compact Muon Solenoid. This specific analysis searches for massive new particles that decay into two vector bosons or into a vector boson and a quark with the vector boson decaying hadronically resulting in a dijet final state. The performance of the triggers is analyzed and monitored during the data taking to ensure that data free of errors and possible issues are fixed immediately. Also the starting point of the analysis is defined by the turn on of the trigger efficiency. An earlier turn on results in a larger range for the analysis to search for new resonances. Overall trigger studies ensure that the analysis gets qualitative data and can make more significant statements. Lastly trigger studies are performed to make sure the data recorded with the triggers is suitable for the intended analysis.

On the 2016 data these studies have already been done, but with the updated trigger system for 2017 the code needs an update to work with the new data and include the new triggers. The same studies are done by other analysis with their analysis specific selections. During the work on this thesis a monitor for the used triggers got available, which was modified slightly to fit this analysis. The quantitative effect of higher thresholds on the performance of the triggers in combination with the new substructure triggers is not clear. The new starting point of the analysis has to be defined from the data. This thesis provides the trigger studies for the dijet analysis and gives an estimate for the starting point of the analysis. Results are gained by analyzing the data of the CMS experiment and comparing the performance with previous studies. The results are used to show that the trigger efficiency is larger than 99% after a specific point, which is then used as starting point of the analysis.

Chapter 2 begins with the foundation about the experiment, the triggers and the analysis. In chapter 3 the methodology and the technical implementations are described. Then in chapter 4 the results with the full 2017 data are shown and discussed. Chapter 5 finishes with the conclusion and outlook.
2. Foundation

2.1. The Large Hadron Collider

The Large Hadron Collider (LHC) \[2\] is a particle accelerator at the European Organization for Nuclear Research (CERN) in Geneva. In its ring with 27 km circumference it accelerates protons to a maximum energy of 6.5 TeV and heavy ions to 2.56 TeV/u. The design luminosity for proton-proton collisions is \(1.2 \times 10^{34} \text{ cm}^{-1} \text{ s}^{-1}\), currently the highest luminosity at hadron colliders. In 2017 the LHC reached nearly twice its nominal value with \(2.06 \times 10^{34} \text{ cm}^{-1} \text{ s}^{-1}\) and up to 60 collisions at each bunch crossing \[3\].

Since the protons carry the same electric charge, one needs two separated beam pipes with opposing magnetic fields, that are generated by super conducting dipole magnets. The beam pipes are joined at the 4 collision points. 1232 dipole and 392 quadrupole magnets are used to guide and focus the 2808 bunches per beam. The acceleration happens in 8 radio frequency cavities at one place of the ring.

The LHC is filled by a chain of preaccelerators from the accelerator complex at CERN. A schematic view of the accelerator complex is shown in figure 2.1. The protons are generated by ionizing hydrogen gas from a small bottle of hydrogen and then injected into the Proton Synchrotron Booster (PSB) by the LINAC2 with an energy of 50 MeV. The PSB accelerates the protons to an energy of 1.4 GeV and injects them into the Proton Synchrotron (PS), where they are accelerated to 25 GeV before they are injected into the Super Proton Synchrotron (SPS). In the SPS the protons are then accelerated to 450 GeV and filled into the LHC in both directions, where they are accelerated for 20 min to a maximum energy of 6.5 TeV. Inside the LHC the beams circulates for many hours before being dumped. For heavy ions the procedure varies in the first acceleration stages.

At 4 collision points are the larger experiments: A Large Ion Collider Experiment (ALICE), A Toroidal LHC ApparatuS (ATLAS), Compact Muon Solenoid (CMS) and Large Hadron Collider beauty (LHCb). ATLAS and CMS are general purpose detectors and are used for Standard Model (SM) research at hight energy scales and for searches of Physics beyond the Standard Model (BSM). ALICE is build as heavy
2. Foundation

Figure 2.1.: The CERN accelerator complex

Schematic view of the CERN accelerator complex [1] with the chain of accelerators used to fill the Large Hadron Collider (LHC). The accelerator chain for the protons is Linac2 → PSB (Booster) → PS → SPS → LHC.
2.2. The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) [5] is a general purpose $4\pi$-detector at the LHC. Its goals are high precision SM measurements and search for Physics beyond the Standard Model (BSM). Due to its general approach, it has to be able to detect a wide range of particles and measure their properties. Since most particles decay inside the beam pipe, vertex reconstruction plays an important role in event reconstruction. With 15 m diameter CMS is approximately half as big as ATLAS [2]. To reach a comparable precision in momentum measurements, CMS has a larger magnetic field.

Figure 2.2.: The CMS detector

Wedge of the CMS-barrel with simplified structure of the detector [4] showing the tracks of detectable particles and how they are visible in the detector.

Ion detector and tries to examine quark-gluon-plasma as state of matter like present shortly after the big bang. LHCb studies the matter-antimatter asymmetry with B-mesons.
The detector is divided into a barrel and two endcap regions. Because of the cylinder symmetry it is practical to use cylindrical coordinates with the \( x \)-axis pointing inwards to the center of the LHC and the \( y \)-axis pointing upwards. The origin is the nominal collision point and \( r \) the radial coordinate. The polar angle \( \Theta \) is the angle towards the \( z \)-axis and the azimuth angle \( \phi \) is the angle to the \( x \)-axis in the \( x-y \)-plane. The transverse momentum \( p_T \) is the projection of the momentum to the \( x-y \)-plane.

The barrel tracker covers a range in pseudorapidity of \(-2.5 < \eta < 2.5\). The pseudorapidity \( \eta \), that is commonly used in detector geometries, is defined as:

\[
\eta = -\ln\left(\tan\frac{\Theta}{2}\right)
\]

with the polar angle \( \Theta \).

CMS barrel and endcaps are built in layers. Directly at the beam pipe is the pixel detector with pixel silicon sensors. It is used for primary and secondary vertex reconstruction, luminosity measurements and tracking. Around the pixel detector is the strip silicon tracker that is used to measure the tracks of charged particles. The tracker is surrounded by the calorimeter which is divided into the Electromagnetic Calorimeter (ECAL) and the Hadron Calorimeter (HCAL). The ECAL uses lead tungstate crystals (\( \text{PbWO}_4 \)) as scintillator together with avalanche photodiodes. The HCAL is built as a brass/scintillator sampling calorimeter around the ECAL. A special feature of CMS is that the calorimeter is inside a coil with a diameter of 2.95 m. This results in the HCAL being limited in its depth in hadronic interaction lengths. The superconducting coil around the tracker and calorimeter is cooled with liquid helium to 4.7 K and generates a magnetic field of 3.8 T. This allows precise measurements of the transverse momentum \( p_T \) for charged high energy particles from the tracks. Around the coil between the iron yoke are muon detectors benefiting from the outer magnetic field. Since muons are Minimum Ionizing Particles (MIPs), they can propagate through the coil and get detected by the muon system before leaving the detector. Neutrinos (and dark matter) are not detected, but vector accumulation of the transverse momenta \( p_T \) of all particles allows the reconstruction of Missing Transverse Energy (MET) \( E_T \). The endcaps close the barrel with further instrumentation.

2.3. Jets

A jet is a number of particles with nearly the same direction, originating from a single point and forming a cone. Colorcharged particles like quarks or gluons, generated in a hard scattering process, produce jets. These particles can radiate
2.3. Jets

gluons and generate new quark-antiquark-pairs out of the vacuum, which are then
collinear to the original particle. The particles then form color neutral hadrons. This
process is called fragmentation or hadronization and happens because color charged
particles cannot exist as free particles due to color confinement. Most hadrons decay
instantly, because of their short lifetime. As effect many particles with nearly the
same direction are visible in the tracker and hit the same calorimeter region.

Next the particles generate showers in the ECAL and HCAL. Since all particles
come from nearly the same direction and hit the same region in the calorimeter, the
shower in the calorimeter overlap and it is not possible to distinguish single particles.
This is not needed, because we are interested in the initial particle that generated
the jet. The particles are combined to a jet, which contains properties of the original
particle.

At CMS the anti-k\textsubscript{T} algorithm \cite{7} is used to reconstruct jets. This algorithm defines
specific distance measures \(d_{ij}\) between two objects \(i\) and \(j\) and sequentially combines
the two closest objects. If the smallest distance is \(d_{iB}\), object \(i\) is called jet and
removed from the list of objects. Then the distances are recalculated. This process
is repeated until the list is empty. The distances \(d_{ij}\) and \(d_{iB}\) are defined as:

\[
\begin{align*}
  d_{ij} &= \min\left(p_{T_i}^{-2}, p_{T_j}^{-2}\right) \left(\frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2}\right) \\
  d_{iB} &= p_{T_i}^{-2}
\end{align*}
\]  

with radius parameter \(R\), the transverse momentum \(p_{T_i}\), rapidity \(y_i\), which is equal
to pseudorapidity \(\eta\) for high energies compared to the object mass, and azimuth
angle \(\phi_i\) of object \(i\).

Two jet definitions with different radii \(R\) are used at CMS. Normal jets use a radius
\(R = 0.4\) and are called AK4 jets. Fatjets use a radius \(R = 0.8\) and are called AK8 jets.
A special feature of CMS is the Particle Flow (PF) algorithm \cite{8}. It combines the
high resolution data of the tracker with the calorimeter to measure the energy. This
compensates for the relative bad energy resolution of the CMS HCAL.

Since Quantum chromodynamics (QCD) is non perturbative in lower energy regions,
Monte Carlo simulations are used to simulate fragmentation. Monte Carlo simulations
have large uncertainties and therefore often data driven methods are used to determine
the QCD background from a non signal region instead. The jets measured in the
detector also have to be corrected because of detector effects, for example non
visible energy or non linear response. These corrections are called Jet Energy
Corrections (JEC).
2. Foundation

Figure 2.3.: The CMS trigger system

The CMS trigger system with its two stages: the L1-triggers and the High Level Triggers (HLTs) (both green). The reduction of the data rates is shown (yellow). Relations of the triggers and their datasources of the triggers are shown with arrows.

2.4. Triggers

The LHC nominal bunch crossing frequency is 40 MHz \[9\]. For a event rate of 40 MHz the bandwidth out of the detector and the available disk space is too small. Also most events contain physically uninteresting soft scattering processes. The trigger system of CMS reduces the event rate from 40 MHz to 400 Hz, so the events can be written to disk. Since the reduction of factor \(O(10^5)\) is too large to be done in one stage, its done in two stages.

The first stage are the L1-triggers \[10\] that are using the muon chamber and the calorimeter to determine if the event will be further processed. The L1-triggers are optimized to select events with hard scattering, reducing the rate by a factor \(O(400)\) to 100 kHz. If a L1-trigger fired, the event is read out from the detector and a online reconstruction is run. The online reconstruction is optimized for performance rather than accuracy.

The second stage are the High Level Triggers (HLTs) \[11\] which reduce the rates by a factor \(O(250)\) to 400 Hz. While the L1-triggers are hardware trigger, the HLTs are run on a server farm directly at the experiment. They are seeded by the L1-triggers, so only a subset of the HLTs is run. The L1-triggers used for seeding are usual strongly simplified versions of the HLTs and less specific selections. Prescaling can reduce a triggers rate by a factor to stay below a certain threshold by skipping event recording. This factor depends on the instantaneous luminosity. If a HLT is fired, the event is written to disk. Later the events are put together into datasets defined by the type of the trigger. An overview over the trigger system as just described, is given in figure 2.3.
2.5. The analysis

The analysis [12] searches for resonances from decays into Vq or VV with full hadronic final states in the dijet invariant mass spectrum of the 2017 CMS data at \( \sqrt{s} = 13 \text{ TeV} \). A new particle decays into a vector boson V (W or Z) and a quark q (Vq) for example a composite quark or two vector bosons (VV) for example a graviton. The vector boson then decays hadronically in two quarks, whose jets overlap due to the strong boost and cannot be distinguished. This results in events with two jets containing at least one fatjet from the vector boson V. Figure 2.4 shows possible Feynman diagrams for the two processes. Events with jets are collected in the JetHT dataset, which will be used for the analysis.

To find new resonances, the cross section is plotted against the dijet invariant mass. The cross section of the background processes decrease with higher energy, together with the increasing trigger efficiency this leads to a bump in the spectrum. This bump can falsely be interpreted as signal from a new particle. Therefore it is important to set the starting point when the combination of the triggers reaches full efficiency. To ensure that the triggers reach full efficiency, one monitors the trigger efficiency during the data taking.

Figure 2.4: Example Feynman diagrams of the two analysis processes

Feynman diagrams showing two production processes for new particles (for example a composite quark q\(^\ast\) or a graviton X) with decay into full hadronic final states. Figure 2.4a shows a decay in one quark and one vector boson. Figure 2.4b shows a decay into two vector bosons.
3. Trigger studies

3.1. Preselections

The events used in the analysis are preselected to ensure that only meaningful data is used. Filters are applied that for example remove events where the calorimeter did not function well due to dead cells or noise. Also the events are filtered with a luminosity mask that is provided by the experiment in a JSON file.

The selected events require the two jets with the highest $p_T > 200$ GeV and $|\eta| < 2.5$ to pass tight jet ID with muon veto [13]. The high cut on transverse momentum $p_T$ is because of the massive resonance that is searched for and the cut on $\eta$ is due to the detector geometry. Also the $\eta$-difference between the two jets is limited with $|\Delta\eta_{jj}| < 1.3$.

Events are tagged and categorized depending on the jet mass being inside the $V$ mass window between 65 GeV and 105 GeV. Events where at least one jet is tagged are included in the single $V$ tag category and are used for the search of decay into $Vq$. Events where both jets are tagged are included in the double $V$ tag category and are used for the decay into $VV$. This means events that pass the double $V$ tag selection also passed the single $V$ tag selection.

3.2. Reconstruction

The trigger efficiency is shown as a function of certain quantities. For the analysis the important quantities are the dijet invariant mass $M_{jj}$ and the softdrop mass $m_{\text{softdrop}}$. Since most of the used triggers trigger on the transverse energy $H_T$, it is interesting to look at the behavior of the trigger efficiency as function of the transverse energy $H_T$ too.

The dijet invariant mass is reconstructed from the sum of the 4-momentum of the two leading AK8 jets:

\[ M_{jj} = \sqrt{(p_1 + p_2)_\mu \cdot (p_1 + p_2)^\mu} \]  \hspace{1cm} (3.1)
where $p_1$ and $p_2$ are the 4-momenta of the two AK8 jets with the highest transverse momentum $p_T$ (leading and subleading jets).

The transverse energy $H_T$ is defined as sum over all AK4 jet transverse momenta $p_T$ with $p_T > 30$ GeV:

$$H_T = \sum_{p_T(i) > 30 \text{ GeV}} p_T(i)$$

Jet masses are reconstructed by squaring the 4-momentum of the jet:

$$m^2 = p_\mu \cdot p_\mu = E^2 - \|\vec{p}\|^2$$

with the jet energy $E$ and the jet momentum $\vec{p}$.

The softdrop mass $m_{\text{softdrop}}$ is the mass of a jet after declustering with the softdrop algorithm \cite{14}. This algorithm removes wide angled soft radiation from a jet to mitigate the contamination from underlying events including pileup and initial state radiation. The trimmed mass $m_{\text{trimmed}}$ is the mass of a jet using jet trimming \cite{15} for declustering. For the final analysis it is planned to use Pileup Per Particle Identification (PUPPI) \cite{16} for pileup mitigation before calculating the softdrop mass $m_{\text{softdrop}}$. PUPPI defines a weight for each particles pileup likeliness and removes the particles with the highest weights.

### 3.3. The triggers

The triggers used for the trigger studies are given in table 3.1. Since the analysis focuses on massive resonances, the triggers used in the analysis require jets with high transverse energy $H_T$ or high transverse momentum $p_T$ to fire. The substructure triggers (HLT\_AK8PFHT***\_TrimMass50) have an additional cut on the jets trimmed mass. This way they can reduce the threshold for the jets transverse energy $H_T$ without increasing the trigger rate.

HLT\_PFHT1050 and the substructure triggers share the same L1 seeds. The HLT\_AK8PFHT750\_TrimMass50 is deactivated for higher luminosities \cite{18}, because the trigger rate gets to high.

The two last triggers in table 3.1 trigger on muons and are used as control trigger on an orthogonal SingleMuon dataset. They were chosen because they make up a large amount of the SingleMuon dataset and are unprescaled. Because they trigger on muons and use different L1 seeds, they should be orthogonal to the other triggers. Also the HLT\_IsoMu\_27 is used in similar studies by other groups which makes it comparable.
### 3.4. Trigger efficiency

The trigger efficiency $\epsilon_{\text{trigger}}$ is calculated by dividing the number of events that passed the trigger or a combination of triggers by the number of all events, using only events that passed the selections and fired the control triggers:

$$\epsilon_{\text{trigger}} = \frac{\text{event selection} \land \& \land \text{control trigger} \land \& \land \text{studies trigger}}{\text{event selection} \land \& \land \text{control trigger}}$$  \hspace{1cm} (3.4)

The statistical uncertainties for the efficiency are approximated with Clopper–Pearson intervals [19]. This approximation is conservative, because it has never less than nominal coverage.

Events are selected with analysis specific preselections and depending on the dataset a control trigger is chosen. The control trigger should be a trigger that is quite common in the dataset to ensure high statistics. Also it should be orthogonal to the trigger under investigation, that especially means the control trigger should not share L1-seed with them.

### Table 3.1: Trigger overview

Overview over the triggers and their trigger conditions used in this studies. The quantities are the transverse momentum $p_T$ and the transverse energy $H_T$ all in GeV as defined in section 3.2.

<table>
<thead>
<tr>
<th>trigger</th>
<th>trigger condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLT_AK8PFJet500</td>
<td>AK8PFJet: $p_T &gt; 500$</td>
</tr>
<tr>
<td>HLT_PFHT1050</td>
<td>AK4PFJet: $H_T &gt; 1050$</td>
</tr>
<tr>
<td>HLT_AK8PFHT750_TrimMass50</td>
<td>AK8PFJet: $H_T &gt; 750$ and $m_{\text{trimmed}} &gt; 50$</td>
</tr>
<tr>
<td>HLT_AK8PFHT800_TrimMass50</td>
<td>AK8PFJet: $H_T &gt; 800$ and $m_{\text{trimmed}} &gt; 50$</td>
</tr>
<tr>
<td>HLT_AK8PFHT850_TrimMass50</td>
<td>AK8PFJet: $H_T &gt; 850$ and $m_{\text{trimmed}} &gt; 50$</td>
</tr>
<tr>
<td>HLT_AK8PFHT900_TrimMass50</td>
<td>AK8PFJet: $H_T &gt; 900$ and $m_{\text{trimmed}} &gt; 50$</td>
</tr>
<tr>
<td>HLT_IsoMu_27</td>
<td>isolated muon: $p_T &gt; 27$</td>
</tr>
<tr>
<td>HLT_Mu_50</td>
<td>muon: $p_T &gt; 50$</td>
</tr>
</tbody>
</table>
3. Trigger studies

3.5. Fitting the trigger turn on

The turn on is the part of the trigger efficiency curve, where the trigger efficiency raises from low efficiency to saturation. Since the task of this studies is to determine the point where the trigger reaches full efficiency, the upper part of the turn on is fitted using a modified error function with the following parameterization:

\[ y = \frac{a}{2} \left[ 1 + \text{erf} \left( b \cdot (x - c) \right) \right] \]  
\[ (3.5) \]

where \( a \) defines the efficiency at saturation, \( b \) the width and \( c \) the position of the turn on.

A similar modified logistic function can also be used to fit the turn on:

\[ y = \frac{a}{1 + \exp \left( -b \cdot (x - c) \right)} \]  
\[ (3.6) \]

where \( a, b \) and \( c \) have the similar meaning as before. However, the convergence of the modified error function is more stable.

The best parameters \( a, b \) and \( c \) are estimated by minimizing a modified \( \chi^2 \)-function:

\[ \chi^2(a, b, c) = \sum_i \left( \frac{y_i - f(x_i, a, b, c)}{\sigma_i} \right)^2 \]  
\[ (3.7) \]

where \((x_i, y_i)\) are the points of the trigger efficiency and \( f(x, a, b, c) \) is the fit function at \( x \) depending on the parameters \( a, b \) and \( c \). If \( y_i > f(x_i, a, b, c) \), \( \sigma_i \) is the lower and otherwise the upper 68\% confidence level from the Clopper-Pearson interval.

The uncertainties for the estimated parameter \( a_0, b_0, c_0 \) at the minimum \( \chi^2_0 \) are the differences between the estimated parameter and the parameter at \( \chi^2 + 1 \) with the other parameter fixed:

\[ \sigma_a = a(\chi^2_0 + 1, b_0, c_0) - a_0 \]  
\[ (3.8) \]

analog for \( \sigma_b \) and \( \sigma_c \). The two solutions are the upper and lower 68\% confidence level for the estimated parameter.

The value of \( a(\chi^2_0 + 1, b_0, c_0) \) is calculated by minimizing the function \( f(a) \):

\[ f(a) = \left( \chi^2(a, b_0, c_0) - (\chi^2_0 + 1) \right)^2 \]  
\[ (3.9) \]

analog for \( b \) and \( c \).

Because of the trigger efficiency being not completely symmetric at the turn on and the interesting part being the upper part, only points with efficiency above a certain
threshold are used for the fit. For the triggers in this thesis a threshold value of 0.8 delivered good results.

To determine the quality of fit the p-value is calculated:

$$ p\text{-value} = \int_{\chi^2_0}^{\infty} \chi^2_n(x) \, dx $$

$$ n = N_{\text{data points}} - 3 $$

where $\chi^2_n(x)$ is the $\chi^2$-distribution with $n$ degrees of freedom. The degree of freedom is the number of data points $N_{\text{data points}}$ used for the fit minus the number of parameters in the fit function.

An example for such a fitted turn on curve is shown in figure 4.5.

## 3.6. Technical implementations

Two technical implementations were realized during the work for this thesis. The first method is used in trigger studies for the analysis on an orthogonal dataset. The second method is included in the Offline Data Quality Monitoring (DQM) of CMS Software components (CMSSW) and is run over each dataset. It is used to monitor the data acquisition.

The preselection part of the code is based on the preselection code used for the analysis found on Github [20]. A number of packages [21–27] is used for the trigger studies and the monitoring.

### 3.6.1. Trigger studies with CMSSW and python

This method uses a CMSSW-EDAnalyzer to process specific MINIAOD-files on the Worldwide LHC Computing Grid. The analyzer applies the preselections and filters events where none of the given triggers fired. It also reconstructs the needed quantities. All data is saved in a csv-file. The size of the files is in the order of 1 MB to 100 MB depending on the dataset and rate of the selected triggers. Since the grid splits the dataset for parallel processing, the resulting files are then merged with a script.

The merged files are read by a python program to generate the trigger efficiency plots. During the read process the program checks for double events by their event
3. Trigger studies

id and run number Data is then split into files containing only one run number. This reduces the memory usage if the program runs only over a subset of the data.

In the python program it is possible to apply further selections on the events for example require both jets to be in the V mass window. It is possible to split different datasets and compare the performance, as shown in figure 4.2 or compare the performance of different triggers, as shown in figure 4.3. It is also possible to generate 2D-plots of the efficiency with two quantities, as shown in figure 4.4. A fit for the turn on can be applied, as shown in figure 4.5. If the program gets a list with the luminosity of the data, which can be generated with the BRIL Work Suite, it also calculates the luminosity of the subdatasets and includes it in the labeling of the plot.

The code including SingleMuon data for 2017 can be found on Github[28]. More detailed instructions on how to setup and run the code are included in the Readme (found in the repository or in appendix B). For the functions of the python program there is also a documentation available (in the repository or in appendix C).

3.6.2. Offline Data Quality Monitoring

Running the analyzer from section 3.6.1 over each dataset by hand is inconvenient and time consuming. Automating of this process is not needed, because CMSSW contains the Offline DQM, which is automatically run over all new datasets. This way the trigger efficiency plots are automatically generated and managed centrally. So other analyses using the dijet invariant mass for their searches benefit as well.

The Offline DQM already contained the code to generate trigger efficiency plots for the transverse energy $H_T$. A switch was added to this code for the dijet invariant mass $M_{jj}$ and the leading jet softdrop mass The same preselections as for the transverse energy $H_T$ are used with an additional cut on the difference of the pseudorapidity $\Delta \eta$ of the two leading jets. This means the preselections are less strict than the ones used in the analysis.

It is possible to write a quality test for the Offline DQM that compares the efficiency curve to a given efficiency curve and raises an alert if they mismatch. This can be done after the monitoring has delivered decent amounts of data.

The code for the Offline DQM was merged into the main branch[29] and will be included in the next release of CMSSW, CMSSW_9_4_X.
4. Results and Discussion

For the following results a 2017 SingleMuon dataset [30] from run period B to F was used. The SingleMuon dataset is orthogonal to the JetHT dataset, that will be used for the analysis. The combination of HLT_Mu50 and HLT_IsoMu27 is used as control trigger, because they are independent from the jet triggers [18] and quite common in SingleMuon datasets.

The JSON file used for the selection of good lumisections is /afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions17/13TeV/PromptReco/Cert_294927-306462_13TeV_PromptReco_Collisions17_JSON.txt. It does not include run period G and H since data validation was not available at the time of writing. The data from run period A is so bad, that it is not used for analysis and therefore run period A is excluded from the studies.

4.1. Comparison of run conditions

In run period B and early C the HLT_PFHT1050 trigger did not reach its full efficiency as shown in figure 4.1. The inefficiency is 2.8 % for the HLT_PFHT1050 trigger, which is determined by fitting the turn on. The HLT_AK8PFJet500 trigger does not show this inefficiency, but has a slightly shifted turn on than later run periods. The plot for this trigger can be found in appendix A as figure A.1.

The inefficiency is caused by a misimplementation in the online Jet Energy Corrections (JEC), that is fixed after run number 299 504 in early run period C. Since the HLT_AK8PFJet500 trigger still reaches full efficiency the effect on the turn on is not that strong after combining both trigger.

Figure 4.2 shows the run period comparison for the combination of HLT_PFHT1050, HLT_AK8PFJet500 and the substructure triggers for the data taken after the JEC were corrected. During run period C the substructure triggers were activated and cause an earlier turn on for the run periods C to F. Run period B remains the run period with the latest turn on, because of the missing substructure triggers.
Figure 4.1.: JEC misimplementation on the HLT_PFHT1050 trigger

The plot shows the turn on curve for the misimplementation of the Jet Energy Corrections before and after they were fixed. Before the fix there was a small inefficiency of around 2.8% in the HLT_PFHT1050 trigger, causing it to not reach full efficiency.
Figure 4.2.: Run period comparison of trigger curves as function of $M_{jj}$

Comparison of the turn on curve of the combination of all triggers for the different run periods in 2017 as function of the dijet invariant mass $M_{jj}$. The substructure triggers became active during run period C, resulting in a earlier turn on for the following run periods.
4. Results and Discussion

Because of the high trigger rates the substructure trigger HLT_AK8PFHT750_TrimMass50 is deactivated for higher luminosities in the later run periods. This is not visible in the curves, because of the small effect and low statistics of single run periods. The next unprescaled substructure trigger is HLT_AK8PFHT800_TrimMass50 with a 50 GeV higher threshold on the transverse energy $H_T$. With the tags for the analysis the statistics further decrease, but the relation stays the same and no additional insights are gained. This plots can be found in appendix A as figure A.5.

4.2. Turn on curves

Figure 4.3 shows the turn on curves of the different triggers and the combination of all against the transverse energy $H_T$ (figure 4.3a) and dijet invariant mass $M_{jj}$ (figure 4.3b). It is visible that the substructure triggers do not reach full efficiency. This is due to their additional cut on the jets trimmed mass.

The turn on curve for HLT_PFHT1050 is much steeper on the transverse energy $H_T$ than on the dijet invariant mass $M_{jj}$. The full turn on is between 900 GeV and 1300 GeV for the transverse energy $H_T$ and between 700 GeV and 1400 GeV for the dijet invariant mass $M_{jj}$. For the efficiency as function of the transverse energy $H_T$, one gets a smeared step function around the trigger threshold 1050 GeV, because of the discrepancy between online and offline reconstruction of the transverse energy $H_T$. This is shown as appendix A in figure A.7.

The turn on curve for HLT_AK8PFJet500 against the transverse energy $H_T$ begins around 800 GeV, but reaches full efficiency not before 1900 GeV. This is because the trigger triggers on a single jet with transverse momentum $p_T$ larger than the threshold of 500 GeV while the transverse energy $H_T$ is the sum over all jets transverse momentum $p_T$. Events containing multiple jets with transverse momentum $p_T$ below the threshold are not triggered, but still can reach a high transverse energy $H_T$ and therefore cause the lower efficiency at higher transverse energies. The turn on against the dijet invariant mass $M_{jj}$ is much steeper between 800 GeV and 1400 GeV because for the dijet invariant mass $M_{jj}$ only the two highest $p_T$ jets are used, most likely including a jet with $p_T > 500$ GeV.

The substructure triggers have an additional cut on the jets trimmed mass and therefore do not reach full efficiency when plotted against the dijet invariant mass $M_{jj}$. Figure 4.4a shows the efficiency against the dijet invariant mass $M_{jj}$ and the leading jet softdrop mass $m_{softdrop}$. The additional cut on the jet mass is visible as lower efficiency below a jet mass of 50 GeV. It is also possible to apply a cut on the dijet invariant mass $M_{jj}$ to remove the area before reaching full efficiency and look at the efficiency against the softdrop mass, as shown in figure A.6 in appendix A.
4.2. Turn on curves

Trigger turnon curves for all examined triggers as function of the transverse energy \( H_T \) (figure 4.3a) and the dijet invariant mass \( M_{jj} \) (figure 4.3b). For HLT_PFHT1050 the turn on in the transverse energy is much steeper, because this is the quantity it uses in its trigger condition.
4. Results and Discussion

![Graphs showing 2D trigger efficiency plots as function of $M_{jj}$ and leading jet softdrop mass.](image)

(a) substructure triggers

(b) all trigger

**Figure 4.4.** 2D trigger efficiency plots as function of $M_{jj}$ and leading jet softdrop mass

2D-plots showing the trigger efficiency as a function of the dijet invariant mass $M_{jj}$ and the leading jet softdrop mass $m_{\text{softdrop}}$ for the substructure triggers (figure 4.4a) and the combination of all triggers (figure 4.4b). The red lines show the cuts on the leading jet softdrop mass for the $V$ tagging. The cut on the jet mass from the substructure triggers results in the low efficiency below a softdrop mass of 50 GeV in figure 4.4a.
4.2. Turn on curves

Figure 4.5.: Turn on curves on 2017 data as function of $M_{jj}$

Trigger efficiency on $M_{jj}$ for the combination of all studied triggers on 2017 data for the two analysis processes with single V tag in figure 4.5a and double V tag in figure 4.5b. The blue dotted line is the turn on curve fitted with a modified error function. Only data after the JEC have been fixed is used.
The substructure triggers support an earlier turn on in the V mass window between 65 GeV and 105 GeV as shown in the combination with the other triggers in figure 4.4b. The additional cut allows to reduce the data by uninteresting events with jet masses below the V mass window and therefore maintain a low enough trigger rate despite the lower threshold on the transverse energy $H_T$.

The combination of all triggers in the dataset after the JEC have been fixed is shown in figure 4.5 for the two analysis processes with a modified error function fitted to the data. As expected the statistics for the double V tag category is lower because of both jets are inside the V mass window, resulting in fewer events. The fit describes the turn on well with low $\chi^2$-values and p-values close to 1. Also both turn on curves reach full efficiency in the saturation within parameter uncertainties.

The previous starting point of the analysis was $M_{jj} > 1050$ GeV with 35.9 fb$^{-1}$ data at $\sqrt{s} = 13$ TeV \cite{12}. Now the starting points are $M_{jj} > 1170.9$ GeV for single V tag and $M_{jj} > 1165.5$ GeV for double V tag with more than 36.3 fb$^{-1}$ data at $\sqrt{s} = 13$ TeV. This means the starting point of the analysis has to be moved by approximately 120 GeV due to the needed lowering of trigger rates for higher luminosities.
5. Conclusion

For this thesis the trigger efficiency was studied during the data taking of 2017. By studying the trigger efficiency, an inefficiency of 2.8% was found in the data of run period B and early C for the HLT_PFHT1050 trigger due to a misimplementation of the Jet Energy Corrections. After the misimplementation was fixed, the trigger reaches full efficiency. With the substructure triggers becoming active during run period C the turn on is shifted to lower dijet invariant masses. CMS collected a total of 36.3 fb$^{-1}$ of data at $\sqrt{s} = 13$ TeV relevant for this work. The starting point of the analysis, after which the combination of trigger has more than 99% efficiency, has to be moved by approximately 120 GeV higher to $M_{jj} > 1170.9$ GeV for the single V tag category and $M_{jj} > 1165.5$ GeV for the double V tag category, due to the higher thresholds of the triggers.

A program to generate trigger efficiency plots was developed. Analysis specific trigger efficiency plots were presented periodically to the Trigger Studies Group and later also at trigger meetings of the B2G group. The Offline Data Quality Monitoring was complemented with code to automate the trigger efficiency monitoring as function of the dijet invariant mass $M_{jj}$. All in all this thesis reached its intended purpose by providing the trigger efficiency studies and determining the starting point for the analysis. In the future the Offline Data Quality Monitoring will automatically provide trigger efficiency curves during data taking.

The next steps are to modify the analysis selection, once the final parameters for 2017 are public, and redo the studies with a JSON file including run period G and H with updated luminosity calculations. For 2018 the Data Quality Monitoring module has to be updated to the new triggers and validated. Also it might be worthwhile to have a look at automating the quality checks, once enough good comparison data is available. The data management of the trigger studies could be improved to work with larger datasets, presumable by streaming the data instead of loading it into memory.
## List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ALICE</strong></td>
<td>A Large Ion Collider Experiment</td>
</tr>
<tr>
<td><strong>ATLAS</strong></td>
<td>A Toroidal LHC ApparatuS</td>
</tr>
<tr>
<td><strong>BSM</strong></td>
<td>Physics beyond the Standard Model</td>
</tr>
<tr>
<td><strong>CERN</strong></td>
<td>European Organization for Nuclear Research</td>
</tr>
<tr>
<td><strong>CMS</strong></td>
<td>Compact Muon Solenoid</td>
</tr>
<tr>
<td><strong>CMSSW</strong></td>
<td>CMS Software components</td>
</tr>
<tr>
<td><strong>DQM</strong></td>
<td>Data Quality Monitoring</td>
</tr>
<tr>
<td><strong>ECAL</strong></td>
<td>Electromagnetic Calorimeter</td>
</tr>
<tr>
<td><strong>HCAL</strong></td>
<td>Hadron Calorimeter</td>
</tr>
<tr>
<td><strong>HLT</strong></td>
<td>High Level Trigger</td>
</tr>
<tr>
<td><strong>JEC</strong></td>
<td>Jet Energy Corrections</td>
</tr>
<tr>
<td><strong>LHC</strong></td>
<td>Large Hadron Collider</td>
</tr>
<tr>
<td><strong>LHCb</strong></td>
<td>Large Hadron Collider beauty</td>
</tr>
<tr>
<td><strong>MET</strong></td>
<td>Missing Transverse Energy</td>
</tr>
<tr>
<td><strong>MIP</strong></td>
<td>Minimum Ionizing Particle</td>
</tr>
<tr>
<td><strong>PF</strong></td>
<td>Particle Flow</td>
</tr>
<tr>
<td><strong>PS</strong></td>
<td>Proton Synchrotron</td>
</tr>
<tr>
<td><strong>PSB</strong></td>
<td>Proton Synchrotron Booster</td>
</tr>
</tbody>
</table>
5. Conclusion

**PUPPI** Pileup Per Particle Identification

**QCD** Quantum chromodynamics

**SM** Standard Model

**SPS** Super Proton Synchrotron
List of figures and tables

2.1. The CERN accelerator complex ........................................ 4
2.2. The CMS detector ....................................................... 5
2.3. The CMS trigger system .................................................. 8
2.4. Example Feynman diagrams of the two analysis processes ......... 9
3.1. Trigger overview .......................................................... 13
4.1. JEC misimplementation on the HLT_PFHT1050 trigger ............ 18
4.2. Run period comparison of trigger curves as function of $M_{jj}$ .... 19
4.3. Trigger efficiency as function of $H_T$ and $M_{jj}$ .................... 21
4.4. 2D trigger efficiency plots as function of $M_{jj}$ and leading jet softdrop mass ........................................ 22
4.5. Turn on curves on 2017 data as function of $M_{jj}$ ................. 23
A.1. JEC misimplementation on the HLT_AK8PFJet500 trigger ......... VIII
A.2. Run period comparison for HLT_PFHT1050 as function of $M_{jj}$ VIII
A.3. Run period comparison for HLT_AK8PFJet500 as function of $M_{jj}$ IX
A.4. Run period comparison for the substructure triggers as function of $M_{jj}$ IX
A.5. Run period comparison of trigger curves for the $V$ tag categories as function of $M_{jj}$ ................................. X
A.6. Trigger efficiency as function of leading jet softdrop mass ....... XI
A.7. Online vs. offline reconstructed transverse energy $H_T$ .......... XI
A.8. 2D Trigger efficiency as function of leading jet $\eta$ and $\phi$ ...... XII
Bibliography


[23] Python Software Foundation. python. Comp. software. Version 3.5.2. URL: https://www.python.org/


[27] W. McKinney et al. pandas. Comp. software. Version 0.22.0. URL: https://pandas.pydata.org/


A. Additional trigger plots

Additional plots not used in the main part of this thesis. Figure A.1 shows the small shifting effect of the JEC misimplementation on the HTL_AK8PFJet500 trigger. Figure A.2, figure A.3 and figure A.4 show run period comparisons for the single trigger instead of the combination, as shown in figure 4.2. Figure A.5 shows the run period comparison for the single V tag (figure A.5a) and double V tag (figure A.5b) categories of the analysis. Figure A.6 shows the turn on of the substructure trigger as function of the leading jet softdrop mass. Figure A.7 shows the discrepancy between the online and offline reconstruction of the transverse energy $H_T$. Figure A.8 shows the trigger efficiency as function of the leading jet position $\eta$ and $\phi$. 
Figure A.1.: JEC misimplementation on the HLT_AK8PFJet500 trigger

Figure A.2.: Run period comparison for HLT_PFHT1050 as function of $M_{jj}$
**Figure A.3.** Run period comparison for HLT\_AK8PFJet500 as function of $M_{jj}$

**Figure A.4.** Run period comparison for the substructure triggers as function of $M_{jj}$
Figure A.5.: Run period comparison of trigger curves for the V tag categories as function of $M_{jj}$
Figure A.6.: Trigger efficiency as function of leading jet softdrop mass

Figure A.7.: Online vs. offline reconstructed transverse energy $H_T$

The red dotted line marks the points where the online equals the offline reconstructed transverse energy $H_T$. The black lines mark the threshold of the HLT_PFHT1050 trigger at 1050 GeV.
A. Additional trigger plots

(a) without cut on $M_{jj}$

(b) with cut on $M_{jj}$

Figure A.8.: 2D Trigger efficiency as function of leading jet $\eta$ and $\phi$
B. Trigger studies setup

Instructions on how to setup the code from the Github repository [28] to generate the trigger efficiency plots used in this thesis.
trigger studies

https://github.com/marco-link/trigger-studies

Marco Link

February 22, 2018
Contents

1 Prerequisites 1

2 Setting up the analyzer 2
   2.1 Setup the CMSSW framework 2
   2.2 Setup the analyzer 2
   2.3 Send task to GRID 3
   2.4 Get the data 3

3 Generate a basic trigger report 4
   3.1 Get luminosity data 4
   3.2 Merging the data 4
   3.3 Generate trigger report 5
1 Prerequisites

- git
- CERN computing account
- GRID certificate (setup ready to use)
- write permission on any storage site
- python3 (tested with 3.5, script to setup an environment with the needed packages is included)
- doxygen (optional to generate documentation)
- \LaTeX{} (optional, but highly recommended)
2 Setting up the analyzer

2.1 Setup the CMSSW framework

Setup a new CMSSW framework with the following commands:

```bash
source /cvmfs/cms.cern.ch/cmsset_default.sh
mkdir trigger_studies
cd trigger_studies
cmsrel CMSSW_9_3_0
cd CMSSW_9_3_0/src/
cmsenv
git clone https://github.com/marco-link/trigger-studies
mv trigger-studies/analyzer/ analyzer
scram b
```

2.2 Setup the analyzer

```bash
source /cvmfs/cms.cern.ch/cmsset_default.sh
cd <your CMSSW_9_3_0 folder>/src/analyzer
cmsenv
```

In `process_data.py` you can enable and disable some preselections. For more control over the preselections you can edit `TriggerAnalyzer/plugins/TriggerAnalyzer.cc`. After editing, run a test with:

```bash
scram b
cmsRun process_data.py
```

This starts a run over 10 000 Run2017B JetHT events. This can be changed in `process_data.py`. 
2.3 Send task to GRID

```
source /cvmfs/cms.cern.ch/cmsset_default.sh
source /cvmfs/cms.cern.ch/crab3/crab.sh
voms-proxy-init -voms cms --valid 200:00
cd <your CMSSW_9_3_0 folder>/src/trigger_studies/analyzer
cmsenv
```

Now edit `crab.py` to fit your dataset, JSON-file and storageSite. Your dataset can be taken from the DAS (https://cmsweb.cern.ch/das/). For more details on the crab-config file see: https://twiki.cern.ch/twiki/bin/view/CMSPublic/CRAB3ConfigurationFile

Then submit the task with:
```
crab submit -c crab.py
```

you can check the status of your submitted task with:
```
crab status
```

after some time the task also should show up at the Task Monitoring: http://dashb-cms-job.cern.ch/dashboard/templates/task-analysis

if some jobs of your task failed, you can resubmit them:
```
crab resubmit crab_projects/<taskname>
```

kill your task with:
```
crab kill crab_projects/<taskname>
```

for more details on CRAB3 commands see:
https://twiki.cern.ch/twiki/bin/view/CMSPublic/CRAB3Commands

2.4 Get the data

After your tasks are finished, you need to get the data from your storage site. You can use a FTP-Client or gfal (https://dmc.web.cern.ch/projects/gfal-2/documentation).
3 Generate a basic trigger report

In this chapter <your report folder> corresponds to <your CMSSW_9_3_0 folder>/src/trigger-studies/report.

3.1 Get luminosity data

To calculate the luminosity of your datasets, you need to generate a file containing the luminosity for the runs defined by the JSON-file.

First we need to setup an environment like described here:

1 ssh <your lxplus username>@lxplus.cern.ch
2 export
   PATH=$HOME/.local/bin:/afs/cern.ch/cms/lumi/brilconda-1.1.7/bin:$PATH
   (bash)
3 pip install --install-option="--prefix=$HOME/.local" brilws

Then generate the luminosity file for your JSON file with:

1 brilcalc lumi -i <path to your JSON.txt> -o lumi.csv

Move lumi.csv to <your report folder>/data and remove the # second line (the line beginning with run:fill.).

3.2 Merging the data

After generating the needed data like described in chapter 2, you get a fakeroot_csv*.root-file for each job in the task. Now you can merge this files into a single one. Therefore
move the `fakeroot_csv*.root`-files into `<your report folder>/IN/crab/<task name>`. If needed rename the dataset in `packer.py`.

For merging then run:

```bash
1 python3 packer.py
```

After that place the `*.csv`-files in `<your report folder>/IN`. They are automatically read in with the next generation of a report that loads the dataset. Therefore it is important for the `*.csv`-file to be named after the dataset. If there were no conflicts with the readin process, the files are then moved to `<your report folder>/data/raw/`. Be sure to give it a unique name or otherwise it will override other files in `<your report folder>/data/raw/`, when moved.

### 3.3 Generate trigger report

Run the `install_dependencies.sh` script to generate a python environment with the needed packages. Then activate the environment.

```bash
1 ./install_dependencies.sh
2 source py_venv/bin/activate
```

To generate a trigger report you can alter the working example `SingleMuon.py` to your needs. Then run it with:

```bash
1 python3 SingleMuon.py
```

Also have a look at the documentation that can be generated by running:

```bash
1 doxygen Doxyfile
```

or look at the already generated `documentation.pdf`. 
C. Documentation of trigger-studies

Extract from the documentation of the python program used to generate the trigger efficiency plots. The full documentation is included in the Github repository [28].
Chapter 2

Namespace Documentation

2.1 dataModule Namespace Reference

module to manage the data used for trigger efficiency plots

Functions

- def loadData (dataset, limits=None)
  loads data; runs merge() before loading
- def getData (dataset, limits=None)
  loads data; use loadData() to process new files from the IN folder
- def merge (dataset)
  splits *.csv files from the IN folder into files with their runnumber in the data/dataset/ folder; after successful run the files are moved from the IN folder to data/raw
- def save (dtset, event, header, mask)
  write single event to file
- def getRunlist (data)
  generates a runlist containing all runs present in the dataset
- def printRunlist (data)
  prints runlist; the list is generated using getRunlist()
- def getLumi (data, path='data/lumi.csv')
  calculates luminosity of data

2.1.1 Detailed Description

module to manage the data used for trigger efficiency plots

Note

requires numpy and pandas

2.1.2 Function Documentation

2.1.2.1 def dataModule.getData ( dataset, limits=None )

loads data; use loadData() to process new files from the IN folder
Parameters

| dataset | name of the dataset to load; there should be a similar names folder in data |
| limits  | (optional) list containing the lower and upper limit for the runnumber to load |

Return values

| pandas_DataFrame  | loaded data |

2.1.2.2 def dataModule.getLumi ( data, path='data/lumi.csv' )
calculates luminosity of data

Parameters

| data | pandas_DataFrame to calculate luminosity of |
| path | (optional) path to lumifile e.g. generated with BRIL |

Return values

| float | luminosity in 1/fb |

2.1.2.3 def dataModule.getRunlist ( data )
generates a runlist containing all runs present in the dataset

Parameters

| data | pandas_DataFrame like loaded with this module |

Return values

| numpy_array | sorted runlist |

2.1.2.4 def dataModule.loadData ( dataset, limits=None )
loads data; runs merge() before loading

Parameters

| dataset | name of the dataset to load; there should be a similar names folder in data |
| limits  | (optional) list containing the lower and upper limit for the runnumber to load |


2.2 triggerplotModule Namespace Reference

Return values

| pandas_DataFrame | loaded data |

2.1.2.5 def dataModule.merge ( dataset )

splits *.csv files from the IN folder into files with their runnumber in the data/dataset/ folder; after succesfull run the files are moved from the IN folder to data/raw

Parameters

| dataset          | name of the dataset to merge (the file in the folder IN should contain this in their filename) |

2.1.2.6 def dataModule.printRunlist ( data )

prints runlist; the list is generated using getRunlist()

Parameters

| data            | pandas_DataFrame like loaded with this module |

2.1.2.7 def dataModule.save ( dtset, event, header, mask )

write single event to file

Parameters

<table>
<thead>
<tr>
<th>dtset</th>
<th>name of the dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>event</td>
<td>array containing data of one event</td>
</tr>
<tr>
<td>header</td>
<td>header for the file (only used if file does not exist)</td>
</tr>
<tr>
<td>mask</td>
<td>mask to write event in file, e.g. mask='[:.3f], [.3f], [.3f], [.0f], [.0f]'</td>
</tr>
</tbody>
</table>

2.2 triggerplotModule Namespace Reference

module to generate trigger efficiency plots

Functions

- def clearfile (path)
  clears a file; usually called before generating plots.
- def write2tex (txt, texpath)
  writes text to textfile
Namespace Documentation

- **def makeSlide**(name, texpath, caption='')
  writes LaTeX formatted slide with graphic to textfile

- **def getError**(k, n, gamma=0.682)
  function to calculate the asymmetric error for the trigger efficiency using Clopper-Pearson interval like defined in https://de.wikipedia.org/wiki/Konfidenzintervall_f%C3%BCr_die_Erfolgswahrscheinlichkeit_der_Binomialverteilung

- **def getEfficiency**(data, trigger, quant, denominator)
  calculates efficiency

- **def doEffPlot**(dataset, trigger, quant, texpath, fit=False, x0=[0.9, mask=''])
  generates a trigger efficiency plot for list of triggers on different subsets and writes a LaTeX formatted slide to textfile

- **def doFit**(xdata, ydata, sigma, x0, cuteff=0.99)
  fits a modified error function to data

- **def do2DPlot**(dataset, trigger, quant1, quant2, texpath, cuts=None, mask='')
  generates 2D trigger efficiency plots

Variables

- float **fitthresh** = 0.8
  y-threshold for datapoints to be considered in fitting

- string **worklabel** = 
  label shown in the top left corner of generated plots

2.2.1 Detailed Description

module to generate trigger efficiency plots

Note

requires numpy, scipy and matplotlib

2.2.2 Function Documentation

2.2.2.1 **def triggerplotModule.clearfile**( path )

clears a file; usually called before generating plots.

Parameters

<table>
<thead>
<tr>
<th>path</th>
<th>filepath of the file to clear</th>
</tr>
</thead>
</table>

2.2.2.2 **def triggerplotModule.do2DPlot**( dataset, trigger, quant1, quant2, texpath, cuts=None, mask='' )

generates 2D trigger efficiency plots

Parameters

| dataset | same as dataset in doEffPlot(), but subsets are combined and not shown separately |
2.2 triggerplotModule Namespace Reference

Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>trigger</td>
<td>label of trigger</td>
</tr>
<tr>
<td>quant1</td>
<td>dict used for x axis (same format as quant in doEffPlot())</td>
</tr>
<tr>
<td>quant2</td>
<td>dict used for y axis (same format as quant in doEffPlot())</td>
</tr>
<tr>
<td>texpath</td>
<td>path to textfile</td>
</tr>
<tr>
<td>cuts</td>
<td>(optional) shows cuts with red lines in plot; e.g. ([-1, 10000], [65, 105])</td>
</tr>
<tr>
<td>mask</td>
<td>(optional) additional mask to apply to data</td>
</tr>
</tbody>
</table>

2.2.2.3 \texttt{def triggerplotModule.doEffPlot (dataset, trigger, quant, texpath, fit = False, x0 = 0.9, mask = '')} \hfill
generates a trigger efficiency plot for list of triggers on different subsets and writes a LaTeX formatted slide to textfile

Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dataset</td>
<td>dict with entries: data: Pandas dataframe loaded with the dataModule label: label of the dataset to use for filename key: key where the subsets are stored sets: list with labels of the subsets denom: label of the denominator or combination of denominator e.g. data = {'data': data, 'label': 'SingleMuon', 'key': 'dataset', 'sets': ['SingleMuon-postfix'], 'denom': 'Mu50_OR_IsoMu27'}</td>
</tr>
<tr>
<td>trigger</td>
<td>list of triggers to generate efficiency plots of</td>
</tr>
<tr>
<td>quant</td>
<td>dict with entries: key: key of the quantity used as x-axis label: label used for x-axis label; can use LaTeX commands limits: list with lower, upper limits and stepsize for bins e.g. quant = {'key': 'Mjj', 'label': 'invariant dijet mass $M_{jj}$ in GeV', 'limits': [500, 2000, 30]}</td>
</tr>
<tr>
<td>texpath</td>
<td>path to textfile</td>
</tr>
<tr>
<td>fit</td>
<td>(optional) enables fit of modified error function</td>
</tr>
<tr>
<td>x0</td>
<td>(optional) startparameter for fit</td>
</tr>
<tr>
<td>mask</td>
<td>(optional) additional mask to apply to data</td>
</tr>
</tbody>
</table>

2.2.2.4 \texttt{def triggerplotModule.doFit (xdata, ydata, sigma, x0, cuteff = 0.99)} \hfill
fits a modified error function to data

Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>xdata</td>
<td>x value of datapoints</td>
</tr>
<tr>
<td>ydata</td>
<td>y value of datapoints</td>
</tr>
<tr>
<td>sigma</td>
<td>asymmetric uncertainty</td>
</tr>
<tr>
<td>x0</td>
<td>startparameter for fit</td>
</tr>
<tr>
<td>cuteff</td>
<td>(optional) value used to determine the plateau point (efficiency &gt; cuteff)</td>
</tr>
</tbody>
</table>
2.2.2.5 def triggerplotModule.getEfficiency( data, trigger, quant, denominator )

calculates efficiency

Parameters

data data
trigger label of trigger or trigger combination to calculate efficiency of
quant quantity used for x-axis
denominator denominator used for filtering

Return values

numpy_array x value of fit
numpy_array y value of fit
str fitlabel
bool fit succeeded
list best parameter estimation
list uncertainty of parameter estimation
float x value where fit reaches cutoff

2.2.2.6 def triggerplotModule.getError( k, n, gamma = 0.682 )

function to calculate the asymmetric error for the trigger efficiency using Clopper-Pearson interval like defined in https://de.wikipedia.org/wiki/Konfidenzintervall_f%C3%B6r_die_Erfolgswahrscheinlichkeit_der_Binomialverteilung

Parameters

k number of hits
n number of experiments
gamma (optional) confidence level for interval; default is one sigma (68.2%)

Return values

list containing the lower and upper error
2.2.2.7 def triggerplotModule.makeSlide( name, texpath, caption = '' )

writes LaTeX formatted slide with graphic to textfile

Parameters

<table>
<thead>
<tr>
<th>name</th>
<th>filename of the graphic</th>
</tr>
</thead>
<tbody>
<tr>
<td>texpath</td>
<td>filepath of textfile</td>
</tr>
<tr>
<td>caption</td>
<td>(optional) caption of the slide</td>
</tr>
</tbody>
</table>

2.2.2.8 def triggerplotModule.write2tex( txt, texpath )

writes text to textfile

Parameters

<table>
<thead>
<tr>
<th>txt</th>
<th>text to write in file</th>
</tr>
</thead>
<tbody>
<tr>
<td>texpath</td>
<td>path of textfile</td>
</tr>
</tbody>
</table>