UNIVERSIDAD SAN FRANCISCO DE QUITO

USFQ

Colegio de Ciencias e Ingenierías

Jet trigger prescale analyzer for CMS Open Data

Ricardo Paúl Escobar Gavilánez

Física

Trabajo de titulación presentado como requisito para la obtención del título de

Licenciado en Física

Quito, 18 de diciembre de 2020

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

HOJA DE CALIFICACIÓN

DE TRABAJO DE FIN DE CARRERA

Jet trigger prescale analyzer for CMS Open Data

Ricardo Paúl Escobar Gavilánez

Nombre del profesor, Título académico: Edgar Carrera, Ph. D.

Quito, 18 de diciembre de 2020

© DERECHOS DE AUTOR

Nombres y apellidos:	Ricardo Paul Escobar Gavilanez
Código:	00323575
Cédula de Identidad:	1721514766
Lugar y fecha:	Quito, 18 de diciembre de 2020

ACLARACIÓN PARA PUBLICACIÓN

UNPUBLISHED DOCUMENT

Resumen

Usamos datos públicos del experimento CMS para desarrollar un ejemplo sobre cómo extraer la información de los valores de las preescalas por evento para jets en dicho evento sabiendo si el trigger se disparó o no. Centramos nuestra atención en jets con momento transversal mayor que los valores umbral para los triggers en el rango de 60 GeV a 150 GeV. Como punto de partida se utilizaron guías anteriores del portal de datos abiertos de CMS. Además, agregamos métodos de análisis nuevos: uso de comodines para las diferentes versiones de los trigger, pareo entre los objetos del trigger y los objetos de jets y verificación de la calidad de los datos con archivos JSON. Este proyecto se puede encontrar en la carpeta TriggerPrescalesAnalyzer en el repositorio de GitHub de datos abiertos de CMS.

Palabras clave: Datos abiertos de CMS, Jets, Trigger, valores de preescala

Abstract

We use public data from the CMS experiment to develop an example on how to extract the information of the prescales values per event for jets in that event knowing if the trigger was fired or not. We focus our attention in jets with a transverse momentum greater than the threshold values for the triggers in the range from 60 GeV to 150 GeV. Previous guides from the CMS Open Data Portal were used as a starting point. Moreover, we added some new analysis methods: use of wild-cards for the different versions of the triggers, matching between the trigger objects and jet objects, and data quality verification with JSON files. This project can be found under the folder TriggerPrescalesAnalyzer at the CMS open data GitHub repository.

Keywords: CMS Open Data, Jets, Trigger, Prescale values

Contents

List of Figures

1 Introduction

The description and interaction of fundamental particles have been precisely predicted by the Standard Model of particle physics [Griffiths, 2008]. This theory, developed in the early 1970's has explained may of the experimental results in particle physics. This model can describe three of the four fundamental forces in the universe: electromagnetic, weak, and strong interactions. Also it classifies the elementary particles as shown in Figure 1.

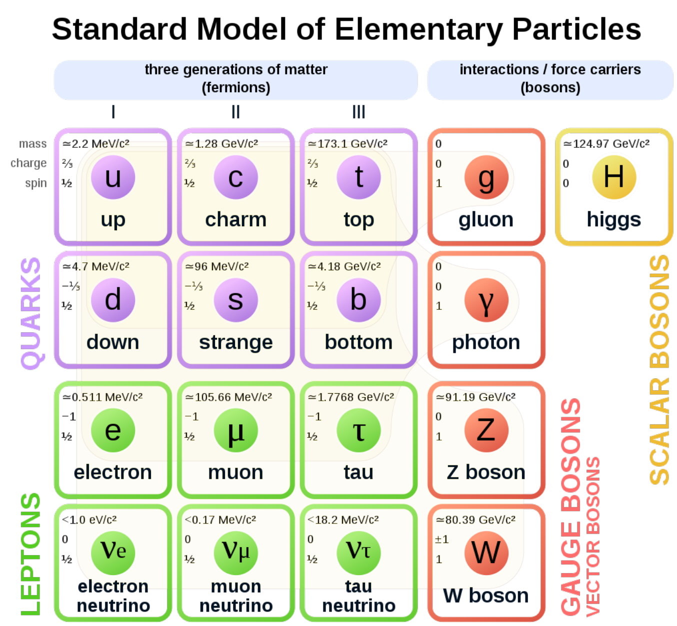

Figure 1: Standard Model of Elementary Particles [Oerter, 2006].

The elementary particles are classified into two groups, fermions, which have half-integer spin and obey the Fermi-Dirac statistics, and bosons, which have integer spin and obey the Bose-Einstein statistics. Within the Fermions we have quarks, which participate in strong interactions and have color charge, and leptons, which participate in electroweak interactions and have no color charge. On the other hand, within the bosons we have gauge bosons, which are the force carriers and have spin = 1, and the scalar bosons which have spin = 0 [Griffiths, 2008].

The Standard Model is theoretically self-consistent, but it is incomplete because we still have many open questions about the origin of mass, dark matter, dark energy, among others [Mann, 2011]. All these questions are closely related to the beginning of the Universe. How can we understand something that happened 14 billion years ago? The simplest approach to understand the fundamental pieces of our Universe is to smash things together, take "pictures", and see what happens.

The Large Hadron Collider (LHC) is the largest and most powerful particle accelerator. With a circumference of 27 km and dipole magnet operating temperature of 1.9 K (-271.3 $^{\circ}C$), it can accelerate protons to 0.999999991 times the speed of light, which is only 3 meters per second slower than the speed of light. Acceleration is necessary because the energy of the particles will be higher if their speed is higher. There are seven experiments at the LHC and four detectors. The two largest ones are the Compact Muon Solenoid (CMS) and the A Toroidal LHC Apparatus (ATLAS). After the collision of particles, we try to measure all the produced particles and reconstruct what was created in the initial interaction. After repeating millions of collisions, we could obtain statistically significant results about the existence of new particles.

In this work, we focus our attention in the product of the decay of quarks and gluons which are particles carrying a color charge and cannot exist in free form because of QCD confinement. The decay products is called Jet, which is a narrow cone of hadrons and other particles produced by the hadronization of a quark or gluon. When there is an object containing color charge fragments, it creates other colored objects around them. However, the lack of understanding of this process is a serious barrier to further progress in jet physics. Jets are built from clusters of energy depositions in the detector calorimeter and they help us to determine the properties of the original quarks or gluons. These objects are important because analyzing them is the only way that scientists have to study quarks and gluons [Rafelski, 2015].

Usually, the data from the collisions in official experimental collaborations is not publicly available. This is not the case at CERN, the CMS Open Data project, for example, was announced on November 2014 and it was the first time that research-grade collision data have been made publicly available for use outside of the LHC collaborations. The CMS data preservation, re-use and open access policy states that 50% of the data will be released after 6 years and 100% within 10 years, or when the main analysis work on these data in CMS has ended [CMS_Collaboration, 2020a]. Following this policy, the 2011 dataset A was released on 2016 and the 2011 dataset B was released on 2020 [CMS_Collaboration, 2016b, CMS_Collaboration, 2020b]. This is the data that we are using in this project. It is important to point out that the CMS Open Data is a great opportunity for physicists, specially for theorists who are not members of experimental collaborations, to use real proton collisions data in their studies.

There are different examples and guides including usage of trigger information at the Open Data Portal, but it would be easier for the users to have more specific guidance about how to get certain information of the trigger. This work combines and completes these guides and examples to create a tool to extract information about the trigger, a concept that not all users of CMS Open Data are necessarily familiar with. In simulations, a trigger and data acquisition is not necessarily needed. In real collisions at the LHC, however, it is mandatory due to the huge amount of data that it would need to be stored without online trigger selection. Here is where the concept of the prescale value is important because it gives us information about how many events, with similar decay topology, passed before we save one of them.

There are two independent values: at Level 1 Trigger and at High Level Trigger. We are interested in the total value of the prescale that is obtained multiplying this two values and that will be the "weight" that each event will have in order to be included in the analysis. In this project, we use the 2011 Jet dataset A and B to show the importance of the correct use of the prescale values per event for maximum precision of the results instead of taking the average of all the prescale values of the events [Tripathee et al., 2017].

This work is a guide to get the prescale values per event. We provide the tools to use "wild-cards" to run over a list of files where different versions of the trigger are being used. Also, we use JSON files to verify the quality of the lumisections that are useful for physics analysis and we show how to add this information in any configuration file [Carrera, 2017]. Finally, the trigger matching tool is implemented to know if the object measured by the trigger corresponds to the reconstructed jet object. In brief, this work takes and expands the information available at the CMS Open Data Portal about prescale values, how to deal with different versions of the trigger, useful lumisections for physics analysis, and an application of trigger matching with jets, providing a better way to understand how to use the CMS open data.

2 LHC trigger

Due to real-world limitations in data storage capacity and rates, we need a system to decide which events in a particle detector to keep. This is because we are only able to save certain amounts of data. The trigger system will decide to save the "interesting" events that occur at relatively low rate for example, decays of rare particles. Physics processes like QCD events or even vector boson production have large cross sections at the LHC, therefore, the number of events is large. Meanwhile, boosted Higgs will have smaller probability of being produced. It is then logical to prescale the former but not the latter.

In CMS data taking, the type of triggers that are used provide a way to group the primary dataset. For this work, the 2011 Jet Primary Datasets A and B are used because they have the information for the single-jet triggers. Each trigger has an associated prescale factor that keeps record of how often an event passing this trigger is saved. For example, a prescale value equal to 1 means that all the events are being saved. On the other hand, a prescale value of 100 means that 1 out of 100 events are being saved. This value is equal to the product of two independent prescale values: one at Level 1 (L1) Trigger and other at Hight Level Trigger (HLT) [CMS_Collaboration, 2016a]. There are different versions for the triggers, going usually from 1 to 9, which reflect updates and improvements to the underlying code.

The trigger system has a list of paths that will accept or not an event. For each trigger we can know if it was accepted (we saved the event due to this trigger), if it was run (the event passed at least the first threshold) or if there was an error. We will focus our attention on the hardest jet of each event. Generally, the information of the jets should be in a descendent order, but to be sure, we analyze all of them and we take the one with higher transverse momentum. The trigger system is complex and in this work it was considered more important and useful to know where to find information and how to use it instead of trying to understand the full complexity.

The CMS single-jet triggers will be "fired" whenever any jet transverse momentum [1] is above the threshold value for the trigger. In other words, the trigger will send a signal each time it detects a jet with more transverse momentum than a value defined in it. The transverse momentum measured at the trigger level is not equal to the transverse momentum measured offline. As a consequence, we have a turn-on curve for the triggers that we use. Ideally, this turn-on curve should be a step function, but in reality there will be jets with transverse momentum less than the trigger threshold that were saved and jets with transverse momentum greater than the trigger threshold that were not saved. Figure 2 shows one example of the turn-on curve for the Run 1 of the LHC where three L1 threshold $(E_T > 16, 26, 92)$ are shown.

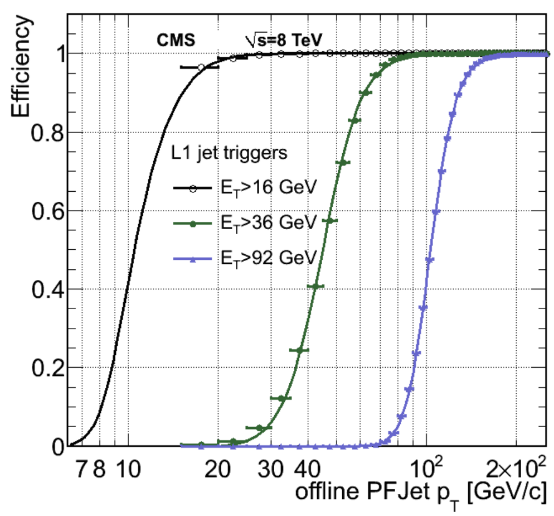

Figure 2: Turn-on curves using L1 jet trigger efficiencies as a function of the Particle Flow jet transverse momentum during the Run 1 of the LHC [Khachatryan et al., 2017].

[1]In hadron collider physics, the pseudorapidity, $\eta = -\ln[tan(\theta/2)]$ and the transversal momentum, pT, are preferred over the traditional spherical coordinates centered at the collision point: the momentum's magnitude $\mid \vec{p} \mid$ and the polar angle θ respectively. In particular, the transversal momentum is calculated from transversal energy delivered on calorimeters. The linear momentum is represented by $\vec{p} = (pT, \eta, \phi)$ with ϕ as the azimuthal angle

The Open Data from the CMS experiment are available from the CERN Open Data Portal. The information that we will analyze in this project are the primary datasets that are in the form of Analysis Object Data (AOD) files [Tripathee et al., 2017]. This format is used internally within CMS and it is based on the ROOT framework [Brun and Rademakers, 1997]. As a first steep, one has to install the Virtual Machine (VM) for 2011 that runs Scientific Linux CERN 6. The steps to install the VM and how to use it are in the guides provided on the CERN Open Data Portal. Inside this VM, one can run the CMS Software framework (CMSSW) and it provides access to the complete analysis tools needed to parse the AOD files.

The CMSSW is a hybrid Python/C++ analysis framework where event processing takes place using user-defined modules. One can think of these modules as blocks with which a complete analysis process can be built. One advantage of the modularity is that we can separate the analysis processes and reuse them in other applications. In this project, we create and then modify an EDAnalyzer module and a configuration file that will run it.

3 Trigger Prescale Analyzer

As a result of the project, we provide an example that extracts the information about:

- How to know what trigger was fired
- How to deal with different versions of the same trigger
- How to take the value of the hardest jet transverse momentum per event
- How to extract the prescale information and how to fill histograms with this value
- How to match the trigger object with the jet object
- How to use a JSON file to take lumisections that are useful for physics analysis in this context

For this project, we use an EDAnalyzer template and modify it as needed. It consists of a python configuration file and a C++ code that will run over all the events and that will perform the analysis. We are running our analysis over 2043 ROOT files with a total of 44051818 events using the offline reconstruction information of Particle Flow (PF).

3.1 Fired triggers and hardest Jet transverse momentum

The trigger system works in a way that if an event is accepted in any trigger path, it will have the information about the other paths. This information will show the paths that were run but not accepted and also the last module (or decision) that decided not to save that event in a certain path. For this project, we will focus our attention just in knowing if the events passed a specific trigger path, and if they did, we will extract the information that we need. In order to do this, we need to import libraries and store the information about the triggers. We pass the information about the event and the trigger name to a Boolean method that will return the information if the trigger was accepted or not. After we know that the trigger we are interested in was fired and accepted, we pass this event to the methods that will extract the information and that will fill the histogram that we need. At this stage, we need to know if the trigger that we are studying is actually taking the data that is supposed to take. To verify this, we can add a *filter name* to the configuration file and fill a histogram with the values that the trigger system is taking. Moreover, we are matching the trigger objects (what the trigger "sees") and the jet objects (the actual objects). To perform this matching we use the angular information of the events to calculate the quantity $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. We say that we have the same object in the trigger collection and the jet collection if $\Delta R < 0.1$. In Figure 3, histograms with the events fired by the single jet triggers in the range from 60 GeV to 150 GeV are shown. We can see that there are no events below the threshold value for each trigger, as expected.

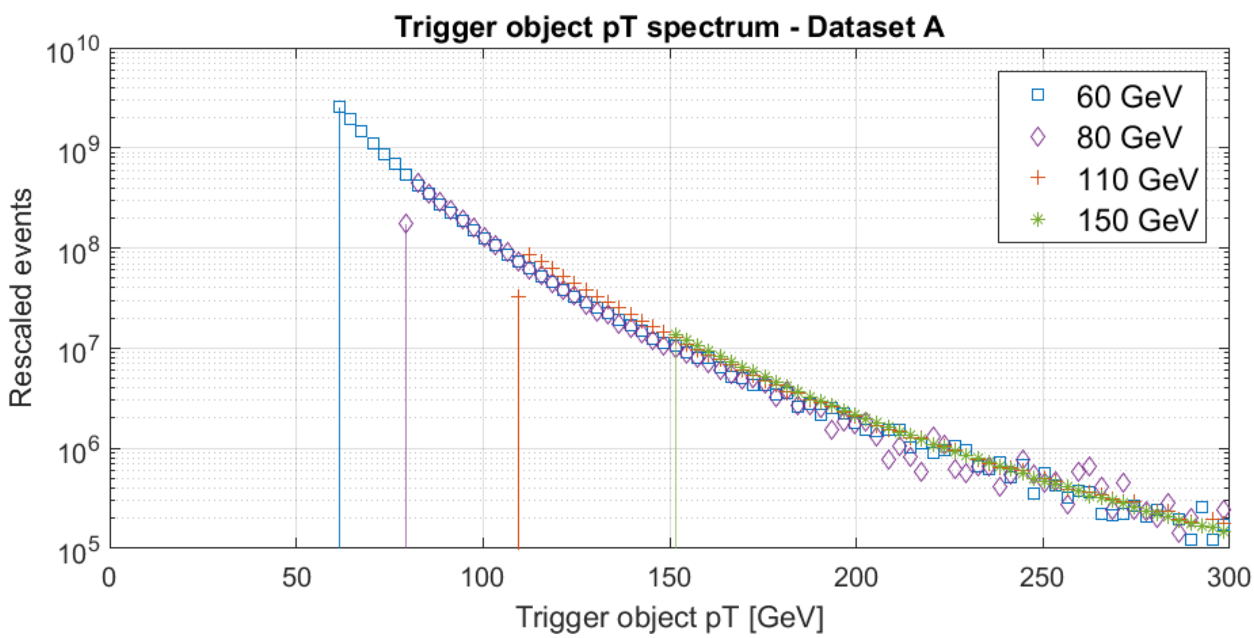

Figure 3: Rescaled events that were fired by the different single jet triggers.

3.2 Prescale values extraction and implementation

It is important to know how to extract the information about the prescales per value and how to use it in the functions that are analyzing the events. One could take all these prescale values and use an average. In this work we will take the values event per event. In brief, we define the information about the trigger and we pass it to the methods that are analyzing the trigger. In the analyzer we take the information of the transverse momentum of the jets and then we apply the prescale weight when we fill the histogram. The prescale values can change in the same run but they do not change in the same lumisection. We conclude that doing the analysis event per event is not the same as to just take an average of the prescale values because the shape of the histogram is different. It is shown, in Figure 4, a plot with example histograms of jet events when the prescale values per event are ignored. Taking the average of the prescale values is equivalent to have the same scaling factor for every bin of the histogram. As a consequence we will not have the correct shape of the histogram and within different triggers, they will not agree in the fall-off.

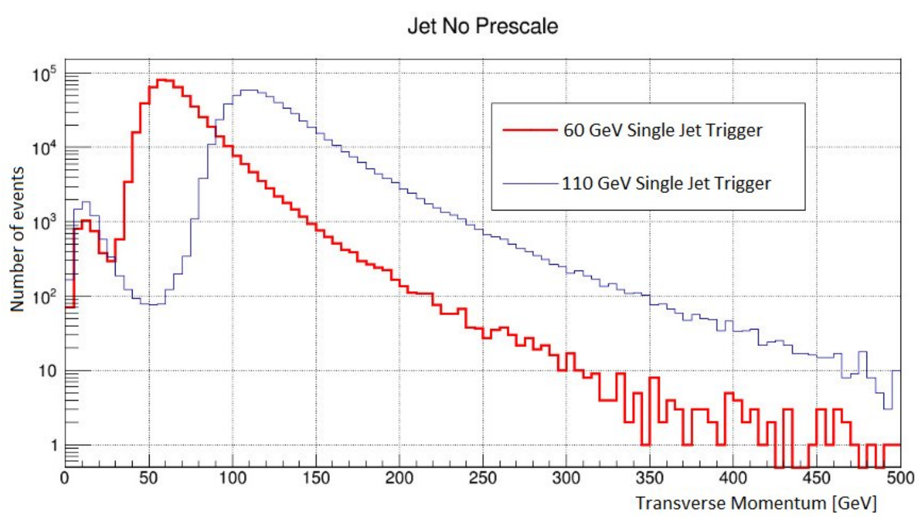

Figure 4: Single jet events in function of the transverse momentum that fired the trigger at 60 GeV and 110 GeV without taking into account the prescale value [Escobar et al., 2019].

In this work we analyze triggers in the range from 60 GeV to 150 GeV in the datasets A and B. We found that the triggers with threshold values at 80 GeV and 150 GeV were not fired in the data from dataset B. The following figures show the histograms of the jet events when we take into account the prescale values per event and the trigger matching between the jet objects and the trigger objects. The uncertainty errors are not displayed for any of the pT distributions. However, it can be seen that the statistical errors are small for lower pT and the dispersion starts at around 230 GeV. Triggered events corresponding to dataset A are shown in Figure 5; to dataset B in Figure 6; and to datasets A and B in Figure 7.

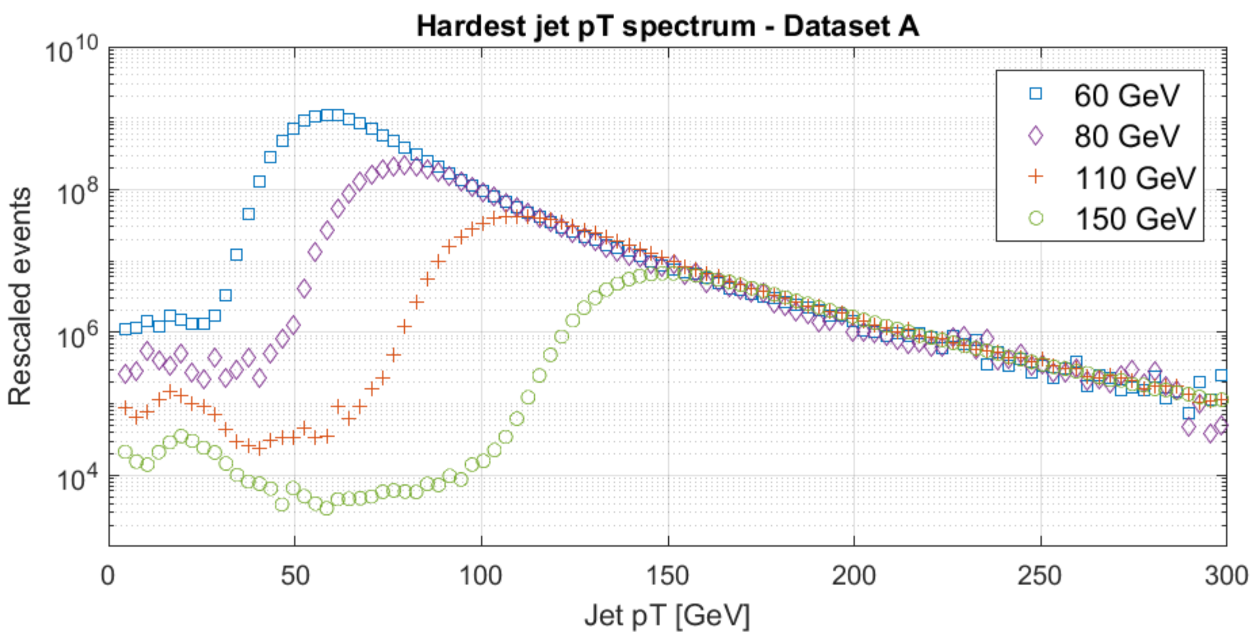

Figure 5: Single jet events in function of the transverse momentum that fired the triggers from 60 GeV to 150 GeV in the dataset A taking into account the prescale value.

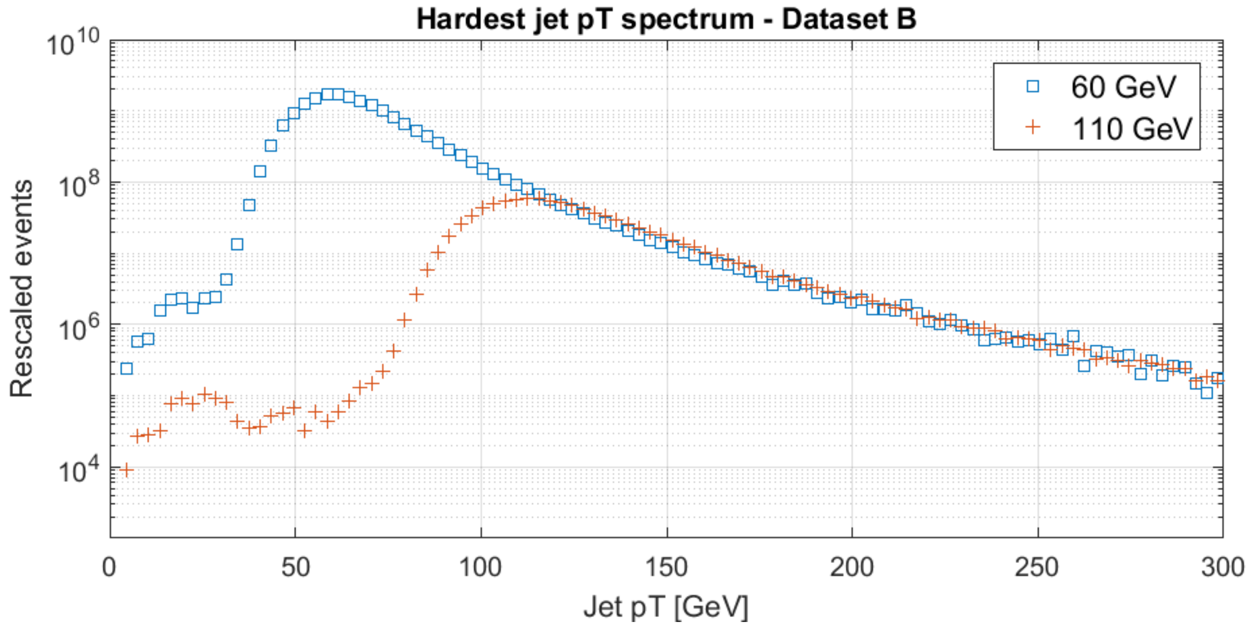

Figure 6: Single jet events in function of the transverse momentum that fired the triggers 60 GeV and 110 GeV in the dataset B taking into account the prescale value.

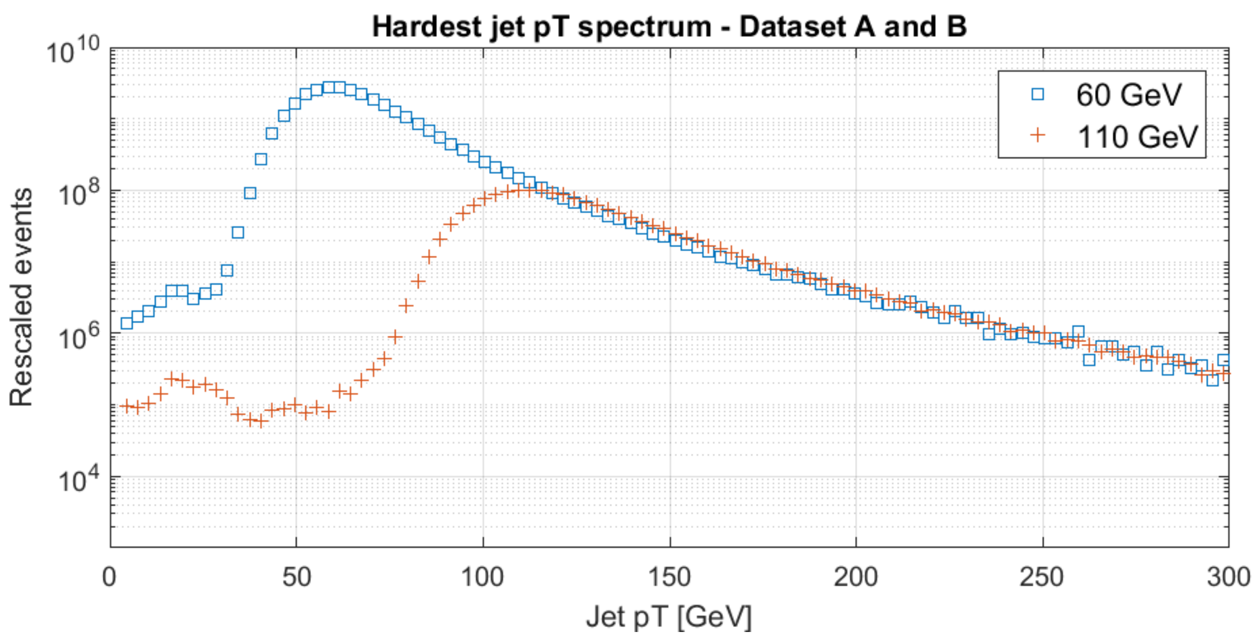

Figure 7: Single jet events in function of the transverse momentum that fired the triggers 60 GeV and 110 GeV in the dataset A and B taking into account the prescale value.

3.3 Data quality check with JSON files

Some of the data available in the datasets may not be good for physics analysis but one can use that data for other purposes [CMS_Collaboration, 2017]. As instructed for the data usage on the CERN Open Data portal, we used a JSON file with the information of the lumisections that are "good" to perform physics analysis. In this work, we implemented this data quality check method to filter events in the configuration file in a way that the user can change this information without the need of compiling the entire code every time there is a change. Applying this data quality check, we get rid of many events that are not useful resulting in a histogram similar to the ones that are shown by other works that used these data. However, there are still events at low transverse momentum, the origin of which needs to be studied more in detail. The low momentum values could arise when the online transverse momentum computation is not exact and it is improved in the offline reconstruction. This may be true for events with transverse momentum close to the trigger threshold, but we can see events with very low transverse momentum. We may need to impose more filters to obtain a better pT distribution.

3.4 GitHub repository

The directory that contains all the programs and necessary files can be found at the CMS Open Data GitHub repository. It has full instructions on how to run this example with all the files needed. The final version of this work is still under development but it can be found under TriggerInfoTool in the directory named as TriggerPrescalesAnalyzer here:

https://github.com/cms-opendata-analyses/TriggerInfoTool/tree/dev

4 Conclusions

In high energy physics processes, quarks are fundamental constituent objects of the theory and gluons are fundamental gauge bosons. Each of these particles are colored, and in the QCD sense, they do not survive long enough to be detected directly. Instead, we detect hadrons as clusters of energy in the calorimeter. We call these objects jets and they help us determine the properties of the original quarks and gluons. This work is about the extraction of the trigger information and the main goal was to get the prescale values per event and plot a histogram as a function of the transverse momentum of single jets. The example can be found in a GitHub repository, and it is still under development. We analyzed a total of four triggers in the range of 60 GeV to 150 GeV. We used previous examples on how to extract the information from the trigger and after this work, this example can be used in the same way, to develop new examples and guides for scientists outside the CMS Collaboration. During the implementation of the different methods, we found more challenges that we expected. For example, the different versions of the triggers, the unmatched trigger and jet objects, and data that are not useful for physics analysis. To start with the project was not easy because the trigger system is complex and it takes some time to understand how it works and how will our project be developed. The personal goal was not to understand the code that is run online (during the run) to save the interesting events, but to understand how and where to look for information, understand how to use different methods in our analysis, have a good idea of the general structure of the trigger system, and with all the machinery available, develop our own example.

The CMS Open Portal is continuously developing and providing new tools, examples, and guides to help external people in education and research. This work started during a Summer Program at CERN and it is an important contribution for the CMS Open Data. There are some improvements that can be envisaged related with this project and in general with the CMS Open Data specially related with the trigger system. For example, calculate the trigger efficiency and run the analyses on the cloud.

References

[Brun and Rademakers, 1997] Brun, R. and Rademakers, F. (1997). Root—an object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 389(1-2):81–86.

[Carrera, 2017] Carrera, E. (2017). Analysis code for extracting the trigger information from the cms 2011 data. cern open data portal. doi:10.7483/opendata.cms.te11.l03e.

[CMS_Collaboration, 2016a] CMS_Collaboration (2016a). The cms trigger system. *arXiv preprint arXiv:1609.02366*.

[CMS_Collaboration, 2016b] CMS_Collaboration (2016b). Jet primary dataset in aod format from runa of 2011 (/jet/run2011a-12oct2013-v1/aod). doi:10.7483/opendata.cms.up77.p6pq. *CERN Open Data Portal*.

[CMS_Collaboration, 2017] CMS_Collaboration (2017). Cms list of validated runs for primary datasets of 2012 data taking.

[CMS_Collaboration, 2020a] CMS_Collaboration (2020a). Cms data preservation, re-use and open access policy.

[CMS_Collaboration, 2020b] CMS_Collaboration (2020b). Jet primary dataset in aod format from runb of 2011 (/jet/run2011b-12oct2013-v1/aod). doi:10.7483/opendata.cms.5jya.hksg. *CERN Open Data Portal*.

[Escobar et al., 2019] Escobar, R., Lassila-Perini, K., and Carrera, E. (2019). Jet trigger prescale analyzer for cms open data. Technical report.

[Griffiths, 2008] Griffiths, D. (2008). *Introduction to elementary particles*. John Wiley & Sons.

[Khachatryan et al., 2017] Khachatryan, V., Anderson, D., Apresyan, A., Bornheim, A., Bunn, J., Chen, Y., Duarte, J., Mott, A., Newman, H., Pena, C., et al. (2017). The cms trigger system. *Journal of Instrumentation*, 12(1):Art–No.

[Mann, 2011] Mann, R. (2011). *An introduction to particle physics and the standard model*. CRC press.

[Oerter, 2006] Oerter, R. (2006). *The theory of almost everything: The standard model, the unsung triumph of modern physics*. Penguin.

[Rafelski, 2015] Rafelski, J. (2015). Melting hadrons, boiling quarks. *Eur. Phys. J. A*, 51(9):114.

[Tripathee et al., 2017] Tripathee, A., Xue, W., Larkoski, A., Marzani, S., and Thaler, J. (2017). Jet substructure studies with cms open data. *Physical Review D*, 96(7):074003.