CERN Open Data Group

CERN HST2018 WG5α
Angela Reisner, Anna Goryunova, Ian Watkins

opendata
CERN

CMS Open Data with Jupyter Notebooks

Teachers' Guide

Introduction

This guide accompanies the Jupyter Notebook task developed by the Open Data workgroup at CERN HST 2018. It is intended to provide teachers with the background information that they need to use the task in their classroom.

Part 1: Why should I use this task?

The world for which high school physics and mathematics students are being prepared is increasingly data driven. As data gathering technology improves, data sets in scientific research, industrial and commercial contexts are getting much larger and the skills required to manipulate and understand them are becoming increasingly important.
Most physical sciences and mathematics curriculum statements around the world make reference to training students in data handling and the scientific method. They also often emphasise the importance of relevance and applicability to real world examples. For example:

"The National Curriculum Statement Grades R-12 aims to produce learners that are able to … collect, analyse, organise and critically evaluate information." – Physical Sciences CAPS (South Africa)

"Real life problems should be incorporated into all sections whenever appropriate. Examples used should be realistic and not contrived." – Mathematics CAPS (South Africa)

"Pupils should be taught to: … construct and interpret diagrams for grouped discrete data and continuous data" – National Curriculum for Mathematics (England)

"Pupils should seek answers to questions through collecting, analysing and presenting data." – National Curriculum for Science (England)

"To know the scientific method (observe, describe, experiment, simulate, hypothesis, model, law, theory)" "Understand approach of modern physics in general" – Physics Curriculum (Switzerland)

"To use ICT competently … To make reasonable conclusions based on the given data … To read/present information in different forms" - Federal State Educational Standards (Russia)

We believe that tasks such as this, which expose students to real scientific data sets, address these requirements directly. In addition to this, students will hopefully gain a deeper understanding of the scientific method as it is applied in real-world fundamental research. We also believe that the link to real world scientific research will engage students and address common issues of motivation encountered when artificial, small and/or trivial data sets are often used to demonstrate statistical and data handling concepts.

In one sentence: This is real data from the world-famous Large Hadron Collider that has been made accessible at a high school level, what teacher or student of physics wouldn't want to do that?

Part 2: Target group and Prerequisite Knowledge

The task was created for 16-year-old students at an intermediate level in physics and mathematics, however we believe that it can be used with older or younger students with suitable differentiation. Prior knowledge of particle physics or programming is not required. The most important information about particle physics and how to use the notebook is given at the beginning of the task so that the students can read it through and gain some knowledge about particle physics and what lies behind the data they are going to use.

If you would like to discuss some further background or want to study particle physics in depth you can either do it before the use of the task or afterwards. There is a more detailed introduction to parts of particle physics provided in this document.

Even if there is no need for an educational background in particle physics the student need to have a background in some other physics aspects as listed below:

- Energy
- momentum
- Energy and momentum conservation
- Speed of light
- Collisions

Part 3: Skills Developed

Students will improve their skills in a variety of ways by using this task:

First of all Jupyter is a nice tool to present information in a different way than that usually used in schools with chalk and a blackboard. The students will learn to plot histograms and use models and programs to explore complex systems. As they do so, they extend their ability to use a variety of formats and platforms to present information.

By having a closer look on the data, they ill learn to analyse complex concepts into their constituent parts and synthesise them to create new understanding. In order to do this, they may have to revise their understanding of how the scientific method works based on the

concrete example of revealing the J/Ψ-particle out of data that was collected within a real experiment at CERN.

With the help of the reflection questions they will learn to interpret data and evaluate evidence and arguments. They will acquire the skill of reporting results and collaborating with peers and experts by discussing the outcome of the task.

Part 4: Introduction to Particle Physics

It may be interesting to you to know how the data you use in the task was collected atCERN. This section will provide some basic knowledge about accelerators, particle collisions and detectors. Further detail is provided on how the data is computed and finally it will be explained why we choose the $J/\Psi -$ data for this task but not e.g. the data of the Higgsboson.

Accelerators

There are two main types of accelerator used in modern particle physics: linear and circular accelerators. The aim of the experiment determines the type of accelerator and particle that is used. Both types of accelerator require the particles to be accelerated in a vacuum to avoid an interaction of the accelerated particles with gas molecules.

Linear accelerators (LINACs) use an oscillating electric field to accelerate the particles in a straight line. The electric field is provided by electrodes. The oscillation of the electric field is timed in such a way that the particles get accelerated each time they cross a gap between electrodes, the electrodes themselves act as a faraday cage so that the particles don't feel a force inside them. The length and design of the electrodes depends on the type of particle used (for example, electrons are lighter than protons and gain speed more quickly).

Circular accelerators (such as the LHC at CERN) use the same technique to accelerate their particles. But rather than using a straight line the particles are forced into a circle by using strong electromagnets located around the circle. The additional magnetic force causes a Lorentz force on the moving particle which forces them into circular path.

The advantage of circular accelerators over linear accelerators is the possibility of reaching higher particle energies by accelerating the particles many times as they circulate.

Collisions

Usually particles collide by smashing them against each other or against a target. In the LHC protons with opposite rotational direction are accelerated in two beams at the same time. When they have reached a predefined energy the two beams are brought together and the particles collide in bunches.

Then the particles collide, their high energy can be converted into new particles following Einstein's formula $E = mc^2$.

Detectors

After converting energy into a large number of particles different kinds of detectors are used to identify the different types of particles by measuring their different properties.

Almost every particle detector is built in the same way: shells (like the rings of an onion) around the collision point. In the first shell the charge and the momentum of the particle is detected by using a magnetic field:

- The direction of curvature of the particle reveals its charge (positively and negatively charged particles curve in opposite directions)
- The radius of the curvature reveals its momentum (the straighter the line the higher the momentum).

The line you see is called a track and the shell which reveals charge and momentum is called tracker. The other shells are used to identify one specific kind of a particle by using its unique properties. If we take muons as an example you would see the following detection:

- two tracks in opposite directions for two different charged muons (conservation of charge)
- in an almost straight line (because high energy muons have a large momentum)
- traversing all other shells (which detect other particles)
- and leaving a signal in a muon chamber (which is built to detect only the muon)

The detected properties of the different particles are saved and some of the data collected at CERN is available on an open data platform: http://opendata.cern.ch/

Invariant mass

As particles (in case of the LHC this are protons) collide, the energy is converted into mass and new particles are created. Some of these have very short lifetimes. The J/Ψ – particle for example has a lifetime of $7,2 \cdot 10^{-21} s$ and then decays into other particles with a certain probability (e.g. into two muons). Due to the short lifetime of the J/Ψ – particles you cannot detect them directly.

This is true for every short living particle – you will never be able to "see" them directly. All you can detect are the different properties of the shower of particles they produce when they decay.

Some properties that can be measured are: momentum, energy, position and electric charge of the particles. The CERN open data files use the following headings to record these properties:

Label	Meaning
E	particle's energy
px, py, pz	*x,y,z*-components of the particle's momentum respectively
pt	the combination of px and py (component of the momentum that is perpendicular to the beam's movement direction)
η, Φ	angles that describe particle's movement direction
Q	Charge

This data gives you the details of the particles which are left from the decay. But, of course, you would like to know what the original particle was before it decayed.

Because of Einstein's equation and the conservation of energy and momentum you know that the conserved properties of the particles you can detect and the particle it was before have to be the same. There is a formula which combines this information and gives you the opportunity to calculate the so called invariant mass of the original particle.

In natural units (see next passage) the equation for a one particle system is given by:

$$m_0 = \sqrt{E^2 - p^2}$$

$$m_0 = \ invariant\ mass$$
$$E = \ Energy$$
$$p = \ Momentum$$

As we mostly have a more particle system and different vectors for the properties the equation becomes:

$$m_0 = \sqrt{2 \cdot pt_1 \cdot pt_2 \cdot (\cosh(\eta_1 - \eta_2) - \cosh(\Phi_1 - \Phi_2))}$$

$$m_0 = \ invariant\ mass$$
$$rest\ as\ listed\ in\ the\ table\ above$$

Often the invariant mass of the particle systems are given in the data files. If you are interested in computing the invariant mass with Jupyter there is a notebook describing this at: https://github.com/cms-opendata-education/cms-jupyter-materials-english/tree/master/Exercises-with-open-data.

Natural units

Natural units are often used in particle physics. They are used to simplify equations by setting the speed of light equal to 1.

Using natural units you can e.g. say that the energy of an particle is equal to its mass using the formula of Einstein:

$$[m] = \frac{[E]}{[c]^2} = \frac{eV}{[c]^2} \xrightarrow{in\ natural\ units} [m] = [E] = eV$$

The mass of a proton for example can be given by following units: $m_p \approx 1\frac{GeV}{[c]^2} = 1\frac{GeV}{c^2} = 1GeV$.

What's different with the Higgs Boson?

You may ask yourself why there is no task provided where students can explore the data and find the evidence for the Higgs Boson instead of the lesser known J/Ψ – particle. In short, this is due to the extra complexity involved in processing the data to find the Higgs boson. The basic principles of the process are, however, very similar.

Part 5: Jupyter Notebooks

We chose to use Jupyter notebooks as the platform for this task for the following reasons:

- Jupyter's cell based layout provides a simple and unthreatening environment for students who may not have had much exposure to coding.
- Despite the apparent simplicity, Jupyter is a hugely powerful platform that supports many of the most important modern programming languages.
- Jupyter is widely used in the scientific community for data analysis, so students will be exposed to a real-world platform as well as real-world data.
- Python itself is an intuitive programming language that is fairly straightforward to learn as well as being highly relevant and widely used in real-world scientific computing.

The basic structure of a Jupyter notebook is a series of cells. Cells can either contain text (with html formatting supported) or code. In our notebook, the code cells can accept code in the python 3 programming language.
Further documentation on Jupyter notebooks can be found here:
And documentation for python 3 can be found here:

There are three main methods to launch the notebook for the task, choose whichever suits your context better:

1) **Google Collaboratory**
   Pros: Online and in browser, doesn't require any software installation. Will reconnect if connection is dropped for a short time.
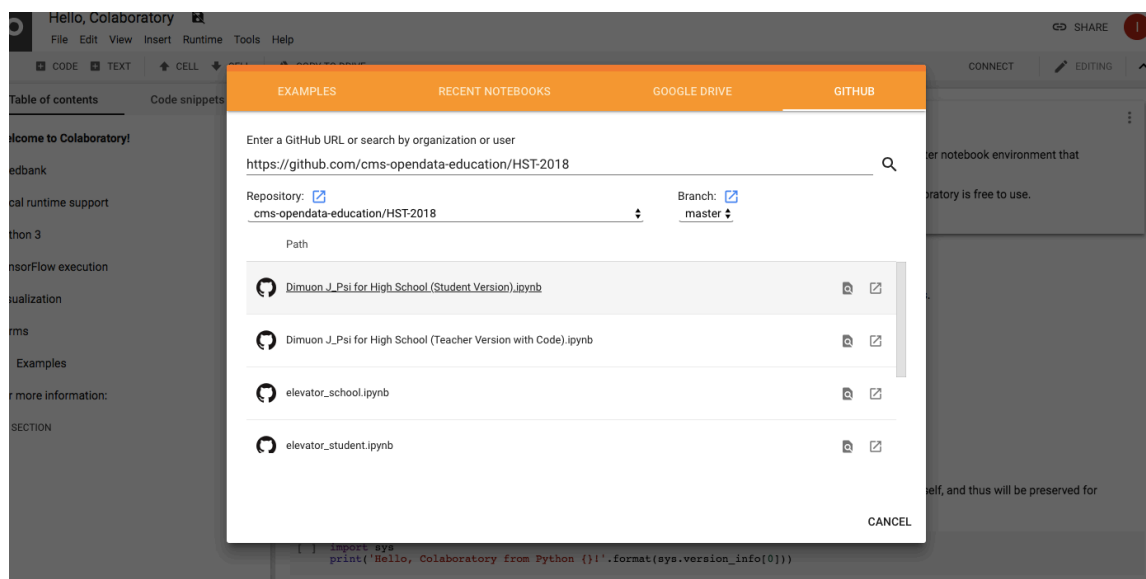   ***Cons:*** Requires a google account.
   Instructions:
   Go to https://colab.research.google.com/and sign in with a google account.
   You should be presented with a dialog box to open a notebook. Select the GitHub option at the top of the box and type this URL into the search bar:
   https://github.com/cms-opendata-education/HST-2018
   You should then be able to open the notebook from the list.

2) **My Binder**
   Pros: Online and in browser, doesn't require any software installation. No account required.
   Cons: Requires constant and stable internet connection.
   Instructions:
   Follow the links here https://github.com/cms-opendata-education/HST-2018to open the notebooks in a binder session. Ensure that your students have an active internet connection while they are working on the notebook.

3) **Install Ancaconda and run notebooks locally**
   Pros: Local solution, does not require an internet connect
   Cons: Software installation required, must be launched from terminal, may encounter issues if packages are not up to date.
   Instructions:
   Follow instructions on http://jupyter.org/installto install and run jupyter notebooks.
   The notebooks for the task can be downloaded from https://github.com/cms-opendata-education/HST-2018

Part 6: Task Description and Ideas for Differentiation

The main task provided is a Jupyter notebook for students withinformation and instructions given. It is considered as a medium-level taskto be undertaken in the classroom with discussions, brainstorming and teacher's help if needed. Students will discuss open data and it's importance, read the data collected at CMS experiment (CERN), present it using different formats, analyse the results, ending up with finding J/psi particle in the spectrum and reflecting on their work. For more specific information, consult the notebook itself.

The task is aimed at an average student and can be differentiated according to the level (or age) of the students. Here are some suggestions for both extended and lower levels.

Extended level :

If students complete previous task successfully they are encouraged to analyze another data from CMS (for instance http://opendata.cern.ch/record/304/files/dielectron.csv) without any scaffolding. It can be given as a totally open question, as in real-world science. Open question inquiry gives opportunity for unlimited and proper research.

Lower Level:

Depending on the level of the students you can skip some of the tasks provided. One of the possibilities is to give more scaffolding to the histogram construction and finish with its description. Additionally, the codecellscan be written by the teacher in advance.

At some point it may be necessary to assess the students work. A possible rubric is presented below. Each task component can be given corresponding "weight" in order to address to the skill that needs to be focused on in this specific task.

Task component/level of achievement	1	2	3
Coding	Most code cells are provided; student fills the gaps with teacher's guidance; Only some of given functions are used and filled correctly.	Some cell codes are provided; Students use all the functions given correctly.	All code cells are written completely by the student; The coding structure is precise and easy to follow; Some additional functions (except given ones) are used.
Plotting	Number of bins is chosen in a way that it is possible to detect a particle properly with teacher's guidance; The axis are labelled, the picture is named;	Number of bins is chosen in a way that it is possible to detect a particle properly; The axis are labelled, the picture is named;	Some histograms with various range limits and number of bins are presented, one is chosen as the 'best' ; The choice of the 'best' histogram is explained in cohesive, reasonable way; The axis are labelled, the picture is named;
Understanding	Some attempts to describe the histogram and outline the results are made.	Some description of the histogram is given, most significant characteristics are mentioned and linked to the theory; Most questions are answered correctly with proper argumentation; The language is usually clear and cohesive, and includes correct scientific terms.	Precise description of the histogram is given, all significant characteristics are mentioned and connected to the theory; All questions are answered correctly with proper argumentation; The language is clear and cohesive, and includes correct scientific terms.

Part 7: Using Jupyter Notebooks in General High School Physics

The Jupyter platform is helpful not only when dealing with particle physics, but as a nice tool that may be used within the curriculum in general. First of all, we can use it just for plotting data from experiments. But that doesn't reflect all the huge possibilities of the platform. For example, large sets of data can easily be plotted and analysed, including the data collected by the students themselves through smartphone applications, such as Phyphox. Except for the experimental data students can use the platform to reach and analyse some statistical data from different open data portals to investigate some questions. Some possible data sets and ideas are presented in the table below, as well as possible connections and ways to use Jupyter in different topics. If the platform is well introduced and is used often it is more possible to reach out to the high level of student's inquiry and deeper research.

Note: This workgroup also developed some more general physics tasks based on Jupyter notebooks, more information can be found at:
https://github.com/cms-opendata-education/HST-2018

Topic	Possible tasks using Jupyter platform
Acceleration	***Experiment (Phyphox):*** phone on a toy car;
Centripetal acceleration	***Experiment (Phyphox):*** rotating the phone inside a salad dryer;
Matter	***Research:*** Carbon dioxide emissions http://data.opennepal.net/keywords/co2 https://datahub.io/core/co2-ppm ***Experiment (Phyphox):*** air pressure depending on height;
Astronomie	***Research:*** solar research http://sidc.oma.be/silso/INFO/sndhemcsv.php
Light/optics	***Experiment (Phyphox):*** Intensity of light depending on the distance/angle/...;
Magnetism	***Experiment (Phyphox):*** plotting magnetic field
Magnetism+oscillations	***Experiment (Phyphox):*** influence of magnetic field on the pendulum characteristics *Note:* use a metal bob, and put a magnet under the equilibrium point;
Waves	***Experiment (Phyphox):*** analyzing the patterns in frequencies for different musical instruments; Dependence of the frequency and amplitude on the distance/...;
Particle physics	***Research:*** CMS open data ***http://opendata.cern.ch/record/545/files/Dimuon_DoubleMu.csv***
Thermodynamics	***Experiment (Phyphox):*** air pressure depending on height;

Electricity	**Research:** Power consumption https://data.world/databeats/household-power-consumption **Research:** Resistivity in geology https://gdr.openei.org/files/283/Kimberly_Elog.csv
Conservation laws	**Research:** CMS open data **Experiment (Phyphox):**i n(elastic) collision (more information you can get in the Phyphox description)