

---

# ATLAS *Open Data*: BUSCANDO EL HIGGS CON ROOT

---

Jennifer Grisales Casadiegos

Curso introducción a la ciencia de datos  
Universidad Industrial de Santander  
LA-CoNGA Physics  
Colombia

## 1. Introducción

ROOT es una poderosa herramienta computacional para el análisis de grandes cúmulos de datos creado y usado en el CERN. Hoy día es de gran utilidad no sólo para los experimentos de física de altas energías; este framework ha sido de utilidad en otras áreas del conocimiento por su gran versatilidad.

En el presente reporte, se usa ROOT de forma propedéutica sobre datos simulados a 13TeV, de la producción de bosones de Higgs en el LHC. Usando histogramas, se busca reproducir y comparar los resultados mostrados en el *Histogram Analyser* [1], una aplicación interactiva para manipular y comprender los datos abiertos del experimento ATLAS.

## 2. Obtención de los datos

Los datos que se usarán corresponden con simulaciones de la producción del Higgs como un decaimiento en bosones  $WW$  a 13TeV.

Para hacer el análisis de datos con ROOT es necesario convertir este archivo a formato ROOT para ello se utilizó un script en c++ disponible en la máquina virtual de ATLAS *Open Data* [2]. La notebook en donde se hace la transformación se ha añadido como documento anexo a este trabajo.

Un archivo ROOT es un objeto "TFile" que puede contener otros objetos organizados en niveles. El archivo es almacenado en un formato de máquina independiente, lo que reduce considerablemente la memoria que puede ocupar. Un archivo ROOT puede hacer que los objetos creados sean "persistentes".<sup>es</sup> decir, al volver a leer el archivo ROOT, el objeto se reconstruye en la memoria [3]. Estos archivos suelen contener árboles (*Trees*), una clase que permite almacenar grandes cantidades de objetos de la misma clase. Básicamente es un contenedor de datos que alberga columnas independientes (ramas) representado por la clase TBranch, cada una de estas ramas puede contener todo tipo de datos.

Para crear y posteriormente escribir en el archivo root se usa:

```
TFile *f = new TFile("outreach.root", "RECREATE");
```

Luego se crea el TTree de datos y se le asigna un nombre, en este caso "ntuple" (En root existe Ntuple que es un tipo de Tree que contiene sólo números del tipo flotante), luego se asignan los identificadores de cada columna de datos (los de la primera fila del dataset) y se escribe:

```
TTree *tree = new TTree("ntuple", "data from csv file"); // Se crea un tree de ROOT

tree->ReadFile("outreach.csv", "type/I:Channel/I:NJets/I:MET/F:M11/F:LepDeltaPhi
/F:METLLDeltaPhi/F:SumLepPt/F:BTags/F:weight/F", ',', '');

f->Write();
```

File	Edit	View	Language
1	type,Channel,NJets,MET,Mll,LepDeltaPhi,METLLDeltaPhi,SumLepPt,BTags,weight		
2	3,0,1,7.57,11.27,0.03,2.61,66.87,0,0.04678		
3	3,0,1,12.99,12.17,0.31,-2.91,81.43,0,0.11052		
4	3,0,1,34.26,14.36,0.32,2.42,87.85,0,0.06163		
5	3,0,0,17.20,11.49,0.30,0.91,73.51,0,0.02291		
6	3,0,1,42.58,12.16,0.39,0.81,57.07,0,0.07487		
7	3,0,1,14.07,10.74,0.17,-0.25,52.11,0,0.09142		
8	3,0,1,30.56,8.85,0.32,-0.89,52.17,0,0.09957		
9	3,0,3,6.00,14.18,0.40,0.71,70.77,0,0.05642		
10	3,0,3,28.34,13.90,0.13,-2.83,72.63,0,0.10075		
11	3,0,1,23.13,14.09,0.23,-2.52,65.19,0,0.08590		
12	3,0,1,27.18,11.30,0.04,-0.36,51.03,0,0.06117		
13	3,0,1,16.41,9.91,0.25,1.19,60.95,1,0.11450		
14	3,0,1,25.99,10.76,0.31,-2.52,63.50,0,0.10902		
15	3,0,1,11.52,9.15,0.25,0.01,71.73,0,0.00289		
16	3,0,1,22.23,8.66,0.32,3.09,53.56,0,0.06102		
17	3,0,1,30.56,13.33,0.33,2.10,71.05,0,0.09388		
18	3,0,1,20.97,11.79,0.28,1.86,68.88,1,0.09318		
19	3,0,3,63.04,10.15,0.32,-1.96,61.86,0,0.07546		
20	3,0,4,36.78,13.95,0.10,3.13,90.19,0,0.09400		

Figura 1: Cabecera del archivo outreach.csv [2] que corresponden a datos simulados de la producción del Bosón de Higgs como decaimiento de bosones  $WW$  a  $13TeV$ .

### 3. Exploración de los datos

Los datos corresponden a 196491 eventos caracterizados por:

- Tipo de dato → Type
- Canales: 0, es el canal  $ee$ , 1 es  $\mu\mu$  y 2 corresponde a  $e\mu$  → Channel
- Cantidad de Jets o chorros de hadrones en el evento → NJets
- Masa reconstruída del dileptón → M(ll)
- Jets de quarks b (bottom) → BTags
- Momento transverso perdido (evidencia de neutrinos) → MET
- Momento transverso total del leptón → PT(ll)
- Ángulo de apertura  $\phi$  entre leptones → LepDeltaPhi
- Ángulo de apertura entre el momento transverso perdido MET y los leptones → METLLDeltaPhi

En *Histogram Analyser* (en adelante HA,) disponible en [http://opendata.atlas.cern/release/2020/documentation/visualization/the\\_display\\_histograms\\_13TeV.html](http://opendata.atlas.cern/release/2020/documentation/visualization/the_display_histograms_13TeV.html) se muestran los histogramas a partir del dataset *outreach.csv* previamente convertido a formato ROOT.

A continuación, vamos a reproducir algunos de los resultados de esta aplicación, explorando los parámetros físicos que permitan filtrar los datos relacionados con la señal del Higgs  $H$ .

Para analizar estos datos, se usan cortes que permiten eliminar procesos que corresponden al fondo de la data obtenida, e intentan dejar el proceso deseado.

Los procesos físicos que han sido simulados y se encuentran en los datos son:

$$H \rightarrow WW, WW, t\bar{t}, Z \quad (1)$$

En donde la señal principal es el proceso del bosón de Higgs y los otros tres procesos corresponden al fondo. Veamos brevemente cada proceso:

### 3.1. Decaimiento $H \rightarrow WW$

El bosón de Higgs es una pieza fundamental en el modelo estándar de la física de partículas, ya que la interacción de este con las partículas elementales resulta en la aparición de la masa. Como este bosón interactúa con todas las partículas con masa, hay muchas formas en las que se puede producir. Este bosón es muy inestable, tiende a decaer en solo  $1,56 \cdot 10^{-22} s$ . La probabilidad que uno de los procesos posibles de decaimiento se de, está dado por el *branching ratio* que es la fracción respecto al número total de desintegraciones. Para  $H^0$  el *branching ratio* para cada proceso, depende de la masa y se muestra en la figura 2.

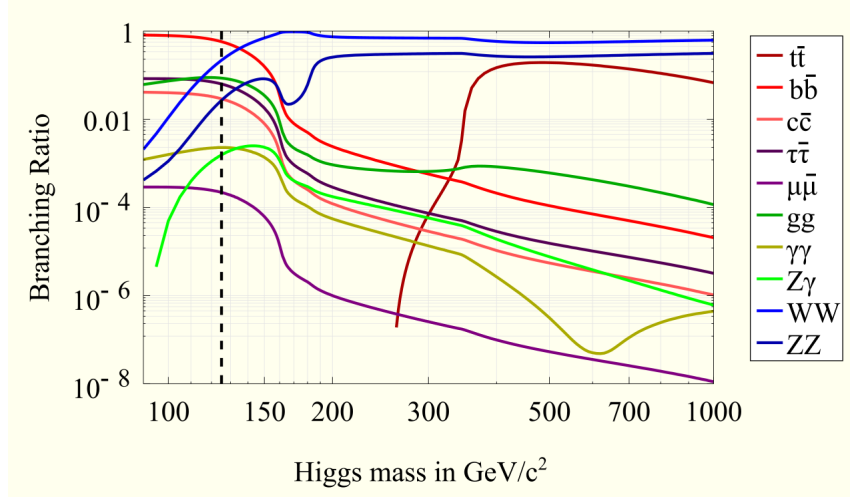


Figura 2: La predicción del Modelo Estándar para los diferentes modos de desintegración del Higgs dependiendo del valor de su masa. Tomado de [4].

Asimismo los bosones  $W$  decaen principalmente en un par quark antiquark ( $q\bar{q}$ ), un leptón-antineutrino ( $l^-\bar{\nu}$ ) o un antileptón-neutrino ( $l^+\nu$ ) [5] donde el leptón puede ser un muón o un electrón. En esencia, el proceso que vamos a buscar es:

$$H \rightarrow W^+W^- \rightarrow l^+l^-\nu\bar{\nu} \quad (2)$$

En este proceso, una huella clave es la detección de un leptón aislado de alto momento transversal o perpendicular a la dirección del haz de protones durante la colisión  $p_T$ . Sin embargo, deben establecerse criterios físicos que permitan filtrar otros procesos más frecuentes que también producen leptones con alto  $p_T$ . A continuación una breve descripción de estos procesos.

### 3.2. Producción de bosones $WW$

El bosón  $W$  junto con el  $Z$  son los responsables de la fuerza nuclear débil, su producción es a través de los siguientes canales:

$$q\bar{q} \rightarrow W^+W^- \quad (3)$$

$$\gamma\gamma \rightarrow W^+W^- \quad (4)$$

$$gg(\text{gluones}) \rightarrow W^+W^- \quad (5)$$

### 3.3. Producción de par quark antiquark $t\bar{t}$

Los quarks top decaen principalmente en un bosón  $W$  y un quark bottom. El  $W$ , a su vez, decae en un par de leptones o en un par de quarks como se describió anteriormente.

### 3.4. Producción de bosones $Z$

El bosón  $Z$  decae como:

$$q\bar{q} \quad (70 \%) \quad (6)$$

$$\nu\bar{\nu} \quad (20 \%) \quad (7)$$

$$l\bar{l} \quad (10 \%) \quad (8)$$

En resumen todos los procesos dentro de los datos simulados tienen partículas de tipo  $ee, \mu\mu, e\mu, \nu$  y  $b_{jet}$ . Por lo tanto se debe considerar propiedades físicas de estas interacciones que permitan filtrar los eventos relacionados con el Higgs [1]. Estos son:

- Para eventos de Higgs, la masa reconstruida de los dos leptones alcanza un máximo de 25-45 GeV.
- La mayoría de los eventos de Higgs contienen 0 o 1 chorro.
- La mayoría de eventos no tienen jets b.
- El momento transversal total de los leptones alcanza un máximo en torno a los 60 GeV.
- El ángulo de apertura entre el momento transverso faltante y los leptones tiende a ser grande.
- El ángulo de apertura de los leptones de las desintegraciones de Higgs tiende a ser pequeño.

Vamos a aplicar estos cortes para reconstruir el momento transversal del leptón. Los histogramas obtenidos en *Atlas Open data* a partir del dataset *outreach.root* se muestran en la figura 4, donde la gráfica roja corresponde a la data total y la azul a la data relacionada principalmente con eventos de Higgs.

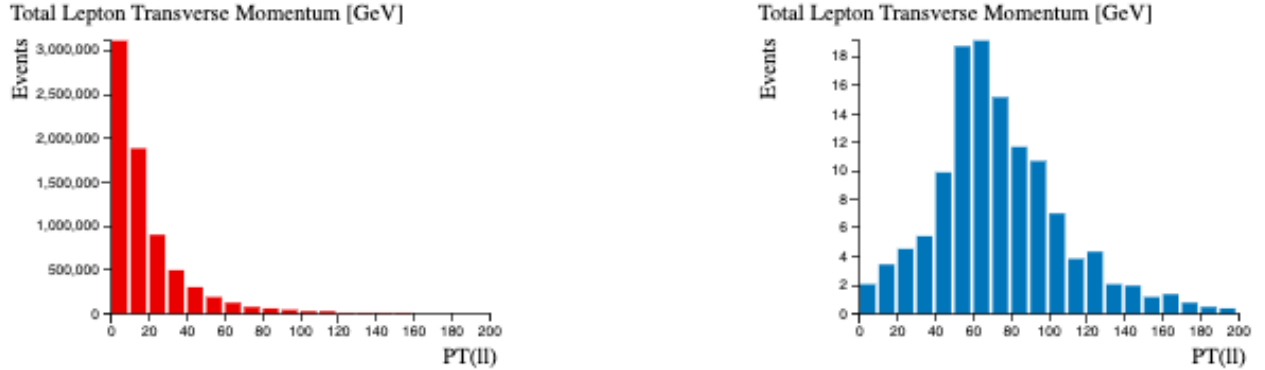


Figura 3: Suma vectorial de los momentos transversales de los leptones cargados observados. Histogramas obtenidos en Atlas Open data a partir del dataset *outreach.root*, la gráfica roja corresponde a la data total y la azul a la data relacionada con eventos de Higgs. (Tomado de [1])

## 4. Reconstrucción de histogramas

*Observación:* Los histogramas tomados del HA se mostrarán en inglés (tal como puedo acceder a ellos), los histogramas propios, estarán en español como el resto de este informe.

Un detalle bastante importante entre los datos usados para este trabajo y los albergados en el HA es que la cantidad de eventos total en el archivo *outreach.root* es un orden de magnitud menor, tendremos en cuenta este factor en el análisis de algunos histogramas, para comenzar comparemos la masa de dileptón reconstruida:

Se observa el pico del primer histograma cerca a 90 y el reconstruido entre 83 y 90 GeV, también es muy prominente un segundo pico entre 0 y 20 GeV.

### 4.1. Filtrado de eventos de Higgs

Vamos a filtrar los datos para obtener histogramas del Higgs, usando parámetros físicos de corte. Para acceder a los datos de un tree creado, usamos un bucle for. Antes de ello, se crea el canva donde se dibujará el histograma y se define el objeto histograma de ROOT:

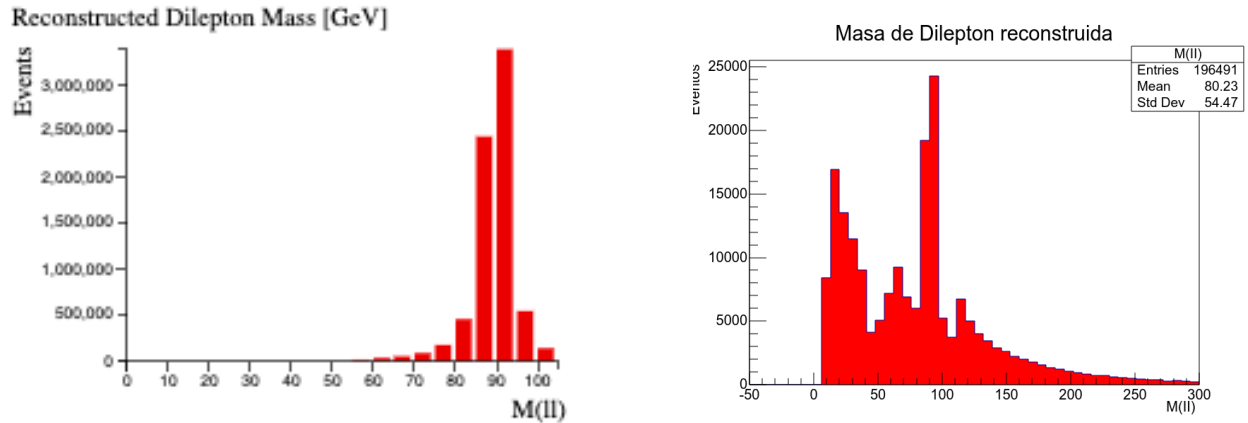


Figura 4: Los histogramas muestran la masa de dileptón reconstruida. A la izquierda, la obtenida a partir de los datos de HA [1], y a la derecha con los datos del archivo outreach.root

```

canvasJT = ROOT.TCanvas("CanvasJT","NJets",800,600)
histJT = ROOT.TH1F("NJets","Histograma NJets Total; NJets ; Eventos ",10,0,10)

for event in tree:
    histJT.Fill(tree.NJets)
print("Done!")

histJT.SetFillColor(4)
histJT.Draw()
canvasJT.Draw()

```

Para realizar el filtrado uso para todos los casos condicionales, que especifiquen los parámetros listados anteriormente. La comparación entre los datos totales y filtrados se puede observar en la figura 5

```

canvasJ = ROOT.TCanvas("CanvasJ","NJets",800,600) #Creo el canvas
histJ = ROOT.TH1F("NJets","Histograma NJets; NJets ; Eventos ",10,0,10)//Creo el histograma

for event in tree:
    if tree.NJets <= 1: //Condicional, para los eventos menores e iguales a 1 jet
        histJ.Fill(tree.NJets) //Almaceno los datos en el histograma
print("Done!")

histJ.SetFillColor(4) //Ajuste de color
histJ.Draw() // Dibujo el histograma
canvasJ.Draw() //Dibujo la canva

```

Para obtener el histograma del momento transverso del leptón para el bosón de Higgs, se deben filtrar todas las otras variables, de ser posible. Sin embargo, en este proceso de filtrado, observé que los datos de  $LepDelta\Phi$  y  $METLLDelta\Phi$  no sólo son menos numerosos sino que están cortados para valores poco característicos de un evento de Higgs como se muestra en la figura 6

Sin embargo para efectos de este reporte, vamos a prescindir de estos dos criterios, y procederemos a filtrar las demás variables como se muestra:

```

#Histograma momento transverso del lepton filtrado
canvasH = ROOT.TCanvas("CanvasH","PT",800,600)
histH = ROOT.TH1F("PT","Histograma momento transverso del lepton; PT ; Eventos ",50,0,300)

for event in tree:
    if tree.Channel == 2:
        if tree.NJets <= 1:

```

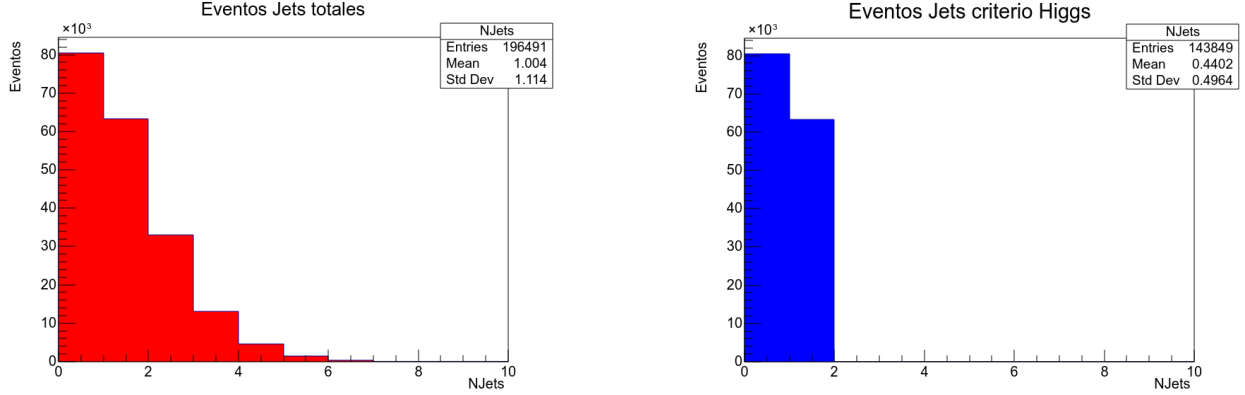


Figura 5: Número de eventos de Jets. A la izquierda, la obtenida a partir de todos los eventos, y a la derecha con el criterio de máximo 1 jet en un evento Higgs.

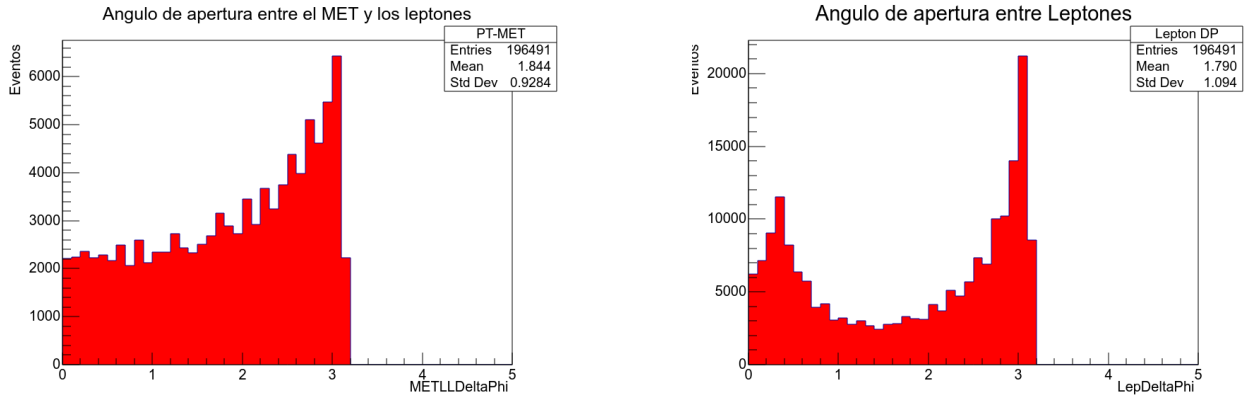


Figura 6: La gráfica de la izquierda corresponde a Ángulo de apertura entre el momento transverso perdido MET y los leptones  $\rightarrow$  METLLDeltaPhi. La gráfica de la derecha es el ángulo de apertura entre leptones. Se observa una anomalía en la distribución de los histogramas.

```

if tree.BTags ==0:
    if tree.MET > 20:
        if 20<tree.Mll and tree.Mll<50:
            histH.Fill(tree.SumLepPt)

print("Done!")

histH.SetFillColor(4)
histH.Draw()
canvasH.Draw()

```

Para facilitar el entendimiento de los resultados, veamos el histograma PT (figura 7) para el total de eventos versus el histograma filtrado:

## 5. Comentarios finales

Se observan varias cosas interesantes, la figuras izquierdas de 4 y 7 ¡difieren enormemente! esto puede tener dos razones no excluyentes la una de la otra: tenemos un orden de magnitud menos de eventos, que estos datos hayan sido preseleccionados para mostrar menos eventos de Z, por ende más eventos de decaimientos de bosones W y quarks tau, que podría explicar los dos picos. Sin embargo, a pesar de la diferencia en los histogramas totales, vemos que las gráficas de la derecha, que corresponde a los candidatos a higgs, son muy similares, incluso sus promedios están cercanos a 60, sin embargo aún tenemos muchos posibles eventos de fondo.

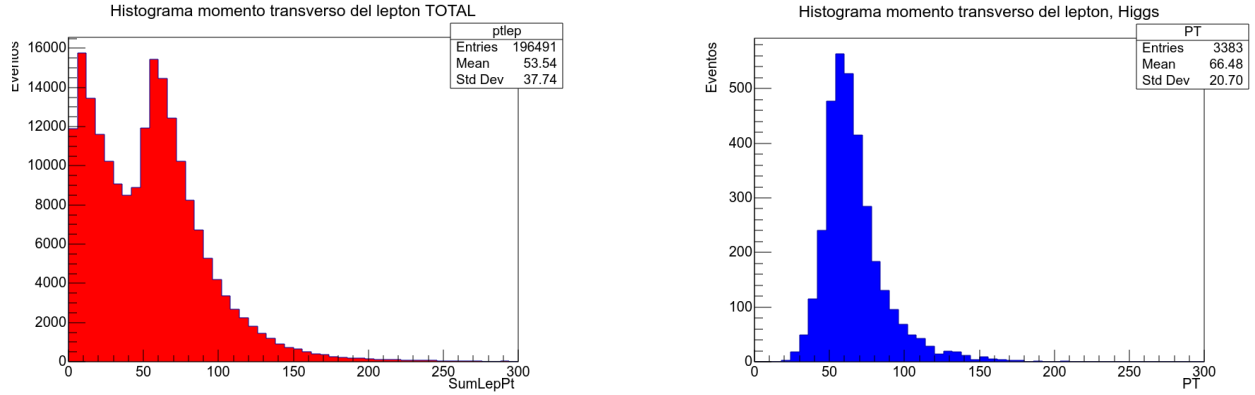


Figura 7: La gráfica de la izquierda corresponde al momento transverso para todos los eventos. La gráfica de la derecha corresponde al momento transverso para los eventos candidatos a Higgs.

## Referencias

- [1] CERN. Atlas open data 13 tev documentation. <http://opendata.atlas.cern/release/2020/documentation/visualization/index.html>, 2021.
- [2] Arturo Sanchez. Atlas-opendata-vm-root6.18-ubuntu-18-server-2020-v2. <https://zenodo.org/record/3629875>, 2020.
- [3] CERN. Root, data analysis framework. [https://root.cern/manual/storing\\_root\\_objects/](https://root.cern/manual/storing_root_objects/), 2021.
- [4] Timothy Rias. The standard model prediction for the branching ratios. [https://en.wikipedia.org/wiki/Higgs\\_boson#/media/File:HiggsBR.svg](https://en.wikipedia.org/wiki/Higgs_boson#/media/File:HiggsBR.svg), 2012.
- [5] et al. Patrignani, C. *Review of Particle Physics*, volume C40. 2016.