# Lab5

Sierra Wilde and Michael Higgins

November 2019

## 1 Overview of the ATLAS Experiment at the LHC

The experiment at ATLAS can detect Higgs Bosons from proton-proton collisions at the Large Hardron Collider (LHC). The Higgs itself does not last long enough to hit one of ATLAS's many calorimeters, but its daughter particles can. ATLAS has spatial resolving capabilities, meaning it can trace the jets back to their starting point. Jets here refer to the collimated stream of particles created from the decay of a certain particle. The jets that ATLAS is interested in measuring are "Higgs-jets," which occur after the Higgs boson decays into a bottom-anti-bottom pair.

The detector is made up of a superconducting solenoid and electromagnetic and hadronic calorimeters. Outside of that is a muon spectrometer with three superconducting toroids. The calorimeters are spread out over a different range of angles around the axis of the proton's trajectory. All of the calorimeters together cover a pseudorapidity of $|\eta| < 4.9$, which is defined by

$$\eta = -ln[tan(\frac{\theta}{2})],$$

where $\theta$ is the is the angle above the x-axis, which points toward the center of the LHC ring. The muon spectrometer measures the deflected muons over $|\eta| < 2.7$.

## 2 Simulated Data

Simulations that mimic real data are made of both the signal and the background, in order to optimize discrimination. The signal here is the Higgs boson decaying into a bottom quark and an anti-bottom quark. Both the Higgs data and the background, QCD data are simulated with 14 parameters, the main two discussed in the ATLAS paper were $p_t$ and $\eta$.

# 3 Event Reconstruction

There are multiple ways that ATLAS reconstructs the produced data. The first is jet labelling. There are many different types of jets, some of which are described below:

- **Truth jets:** Truth jets carry the true information about the jets in simulations. They don't include information on low-interacting particles such as muons and neutrinos, because they rarely leave energy deposits in the calorimeters.

- **Calorimeter jets:** Calorimeter jets are made by reclustering parts of the original jet and removing anything with a transverse momentum, $p_T$, that is less than 5% of the parent's jet $p_T$.

In simulations, the truth jets are reconstructed using the same methods as those used for real data. Calorimeter tracks and other types of tracks are made from the simulations and are then analyzed.

Muons and photons from the events are reconstructed differently. The muons are reconstructed from measurements from the spectrometer, and have $p_t > 5GeV$ and $|\eta| < 2.5$ in order to be triggered. Photons are measured from energy deposit clusters in the electromagnetic calorimeter, and are labeled differently depending on whether they can be traced back to a track. The photons must have $E_T > 175GeV$ and $|\eta| < 1.37$ or $1.52 < |\eta| < 2.37$ to be sent to analysis.

Triggering a Higgs jet candidate relies largely on $p_T$. There is an angular separation between a Higgs decay that depends on the transverse momentum:

$$\Delta R \approx \frac{2m_H}{p_T}$$

$m_H$ here is the mass of the Higgs boson. Higgs with $p_T > 250GeV$ and $|\eta| < 2.0$ are considered for analysis, and bottom quarks from the Higgs decay with $p_T > 5GeV$ and $|\eta| < 2.5$ are triggered. In order to obtain a signal set, only the Higgs with the highest $p_T$ are used in order to reduce the systematic effects from initial-state radiation, which is energy created from annihilating particles.

# 4 Exploring the Data

Since the ATLAS paper discusses transverse momentum as the main way of tagging jets, I first want to plot histograms of the transverse momenta of both the Higgs data and the background, QCD data. Then I can see if there are any major differences in distributions, or places where the distributions look very similar.

The Higgs data has a very sharp, sharks-tooth distribution, whereas the QCD data has a distribution that does not have this sharp peak. In fact, the shape of the QCD distribution looks completely different, as seen in Figure 1. There is a greater probability of finding a Higgs with a $p_T \approx 1100$, but this

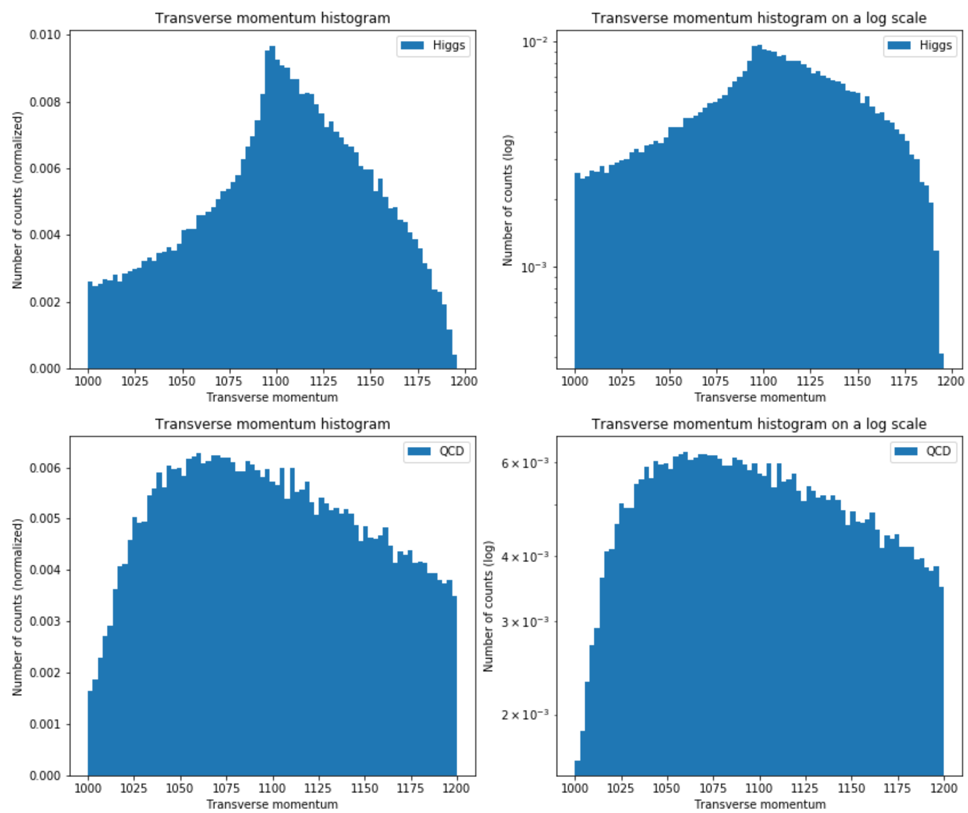


Figure 1: Comparison of the distributions of the normalized histograms of transverse momentum of the Higgs boson data (top) and the transverse momentum of the QCD data (bottom). The logarithmic scaled plots of each are on the right.

probability is only 0.004 greater. Both datasets cover the same range, and thus it might be hard to distinguish a single data point as a Higgs boson.

Since the paper also uses pseudorapidity as a metric, I decided to plot histograms of the pseudorapidity for each dataset against each other in order to spot any differences in this data.

The distributions of each dataset look very similar. However, the Higgs data more closely follows a Gaussian distribution than the QCD data. There is the characteristic bell-shaped curve in the linearly-scaled plot, and a parabolic

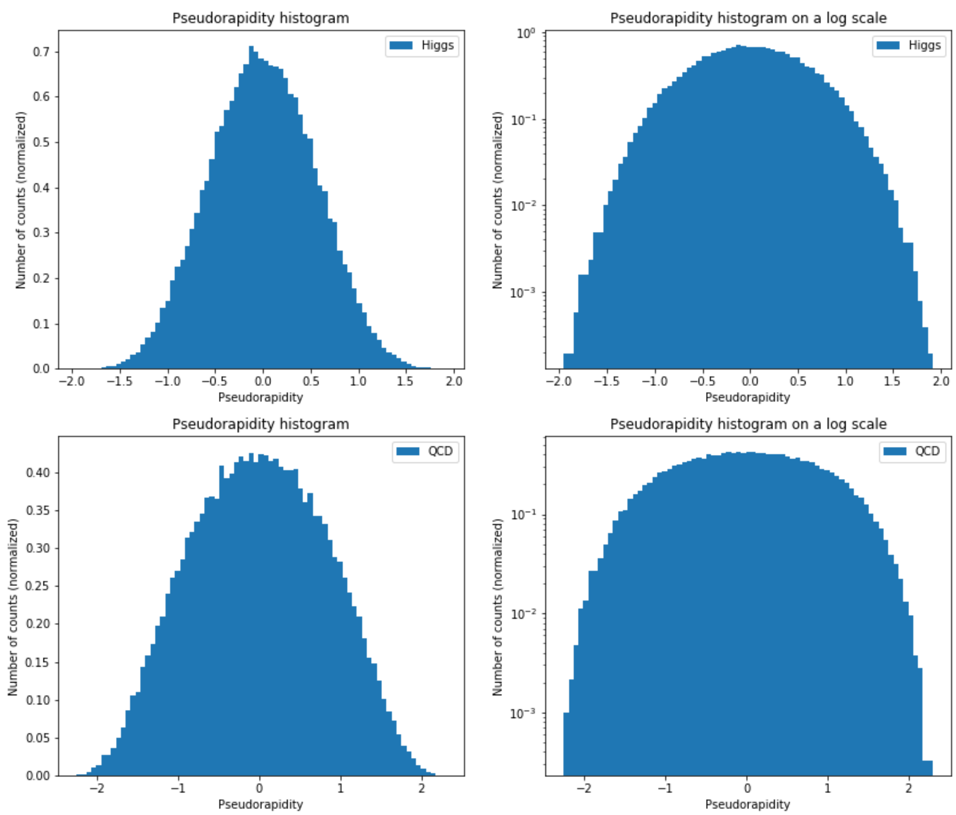


Figure 2: Comparison of the distributions of the normalized histograms of pseudorapidity of the Higgs boson data (top) and the pseudorapidity of the QCD data (bottom). The logarithmic scaled plots of each are on the right.

shape in the log plot. The log plot of the QCD data is too flat in the center to be considered a parabola. The pseudorapidity has the same problem as the transverse momentum in that both span almost same range of values, though the tails of the Higgs data drop off much more quickly than the tail of the QCD data. This means that it is much less likely to measure a Higgs with $|\eta| \approx 2$ than it is a background particle. In order try to find more distinguishing features, I must look at the other parameters in the data to see if there are larger differences.

Before looking at the rest of the data, I wanted to see if there was a relationship between $p_T$ and $\eta$. If there is a difference in this relationship between the two datasets, this could help with Higgs signal discrimination.

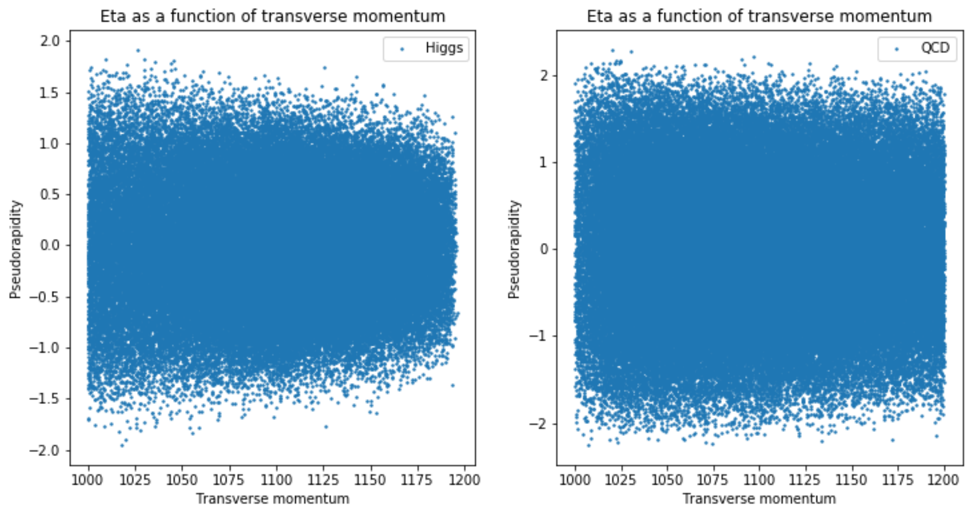


Figure 3: Pseudorapidity as a function of the transverse momentum in the Higgs data (left) and the QCD data (right).

The distribution of $\eta$ as a function of $p_T$ is approximately uniform across the entire range of data for the QCD data. There is a visible difference in the Higgs data's pseudorapidity distribution, which tapers as transverse momentum increases. With the use of both of these parameters, the difference between the distributions of the Higgs and QCD data increases with transverse momentum.

I am plotting histograms of the rest of the parameters included in the datasets to compare the signal distribution from the background. $\phi$ is defined to be the azimuthal angle, or the angle around the z-axis, which is the direction of the beamline. The mass is the rest mass of the particle in question. $e_2$ and $e_3$ are the 2- and 3-point ECF ratios, respectively, which are energy correlation functions. $D_2$ is the ratio between $e_3$ and $e_2^3$, also called the 3- to 2-point ECF ratio. The equations associated with these are as follows:

$$
\begin{aligned}
\Delta R &= \sqrt{\Delta\eta^2 + \Delta\phi^2} \\
e_2 &= \sum_{i<j\in J} p_{T,i} p_{T,j} \Delta R_{ij} \frac{1}{p_{T,J}^2} \\
e_3 &= \sum_{i<j<k\in J} p_{T,i} p_{T,j} p_{T,k} \Delta R_{ij} \Delta R_{ik} \Delta Rjk \frac{1}{p_{T,J}^3} \\
D_2 &= \frac{e_3}{e_2^3}
\end{aligned}
$$

Each $t_n$ refers to n-subjettiness and $k_t\Delta R$ is the $\Delta R$ of two subjets within a single jet.

Only the distributions of $\phi$ look the same for both the Higgs and QCD datasets. Both seem to be approximately uniform over the entire range. This makes sense, because the jets would be equally likely to travel in any $\phi$ direction, because of the cylindrical symmetry of the proton beams. The shapes of the rest of the distributions vary. Many of the QCD distributions ($e_2$, $e_3$, $D_2$, and angularity) look to be approximately exponential, given the steep decrease on the regular scale plots, and the approximately linear decrease on the log plot. The mass, $e_2$, $e_3$, and $D_2$ distributions all have a greater spread in the QCD data. This makes sense, because these all have to do with energy. Since the QCD data has multiple particles, there is going to be a greater variety of measure masses, and therefore a wider array of energies as well. The remaining Higgs parameters all have about the same spread as their QCD counterparts, but the shapes are different, which is made clear by looking at both the regular and the log plots. Just by looking at the histograms themselves, most of the distributions are not centered at the same place either. This indicates that the mean values of the Higgs and the QCD data differ, unlike the transverse velocity and pseudorapidity, which were both centered at around the same value for both datasets.

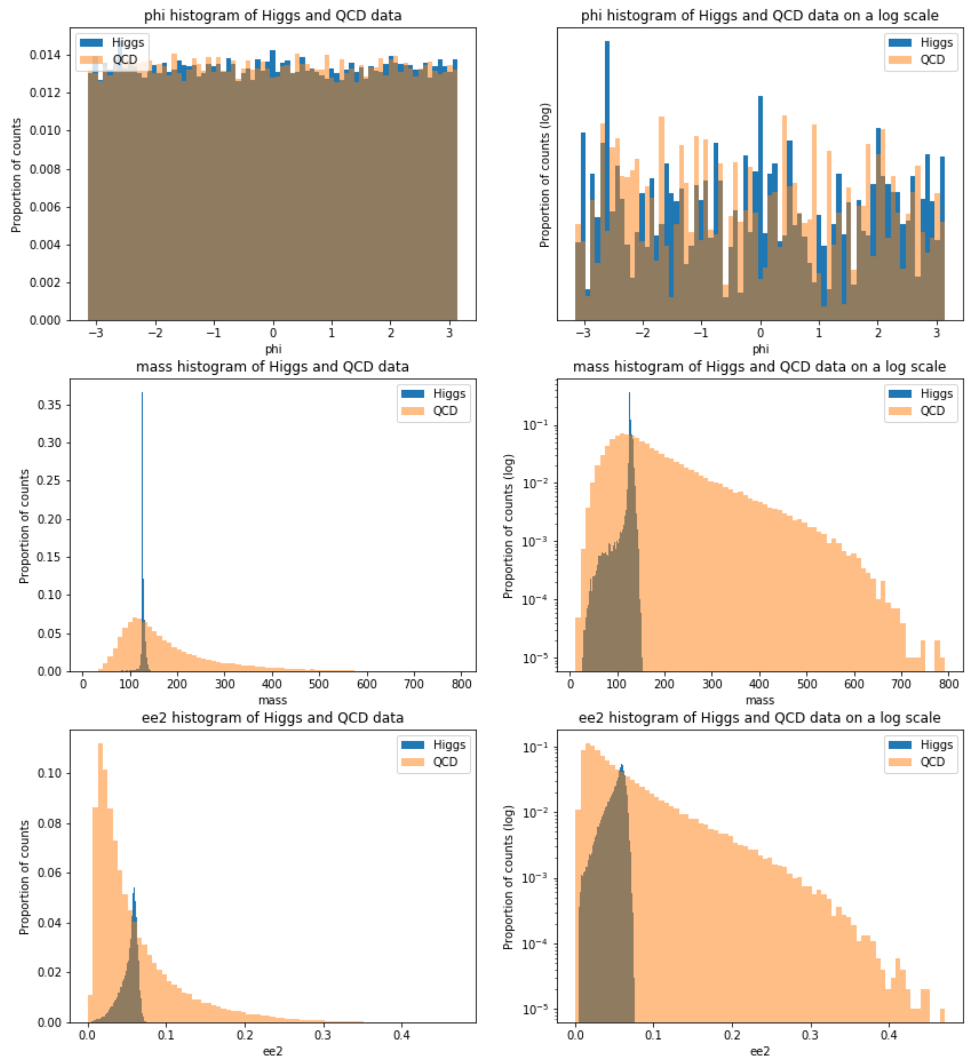


Figure 4: Histograms of the parameters $\phi$, mass, and $e_2$, included in the Higgs and QCD data. All are normalized so that it turns each in to a probability distribution.

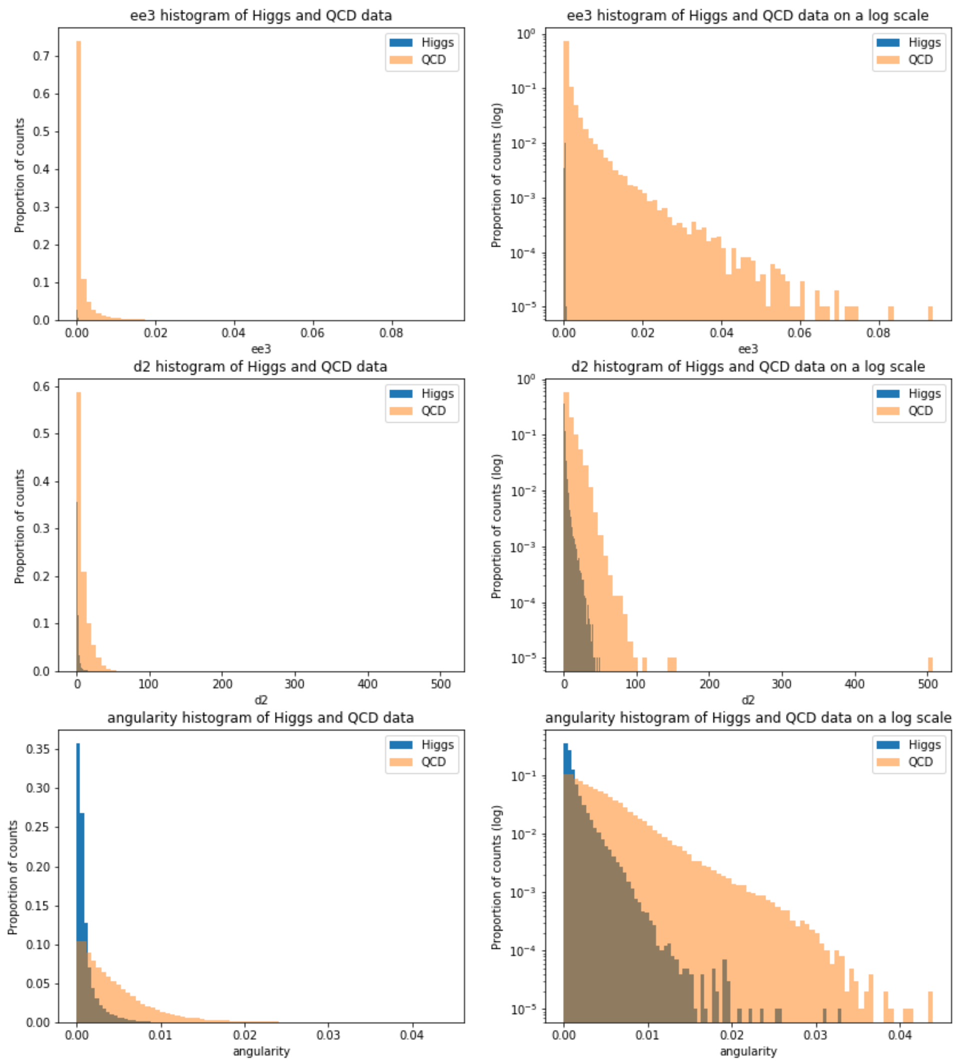


Figure 5: Histograms of the parameters $e_3$, $D_2$, and angularity, included in the Higgs and QCD data. All are normalized so that it turns each in to a probability distribution.

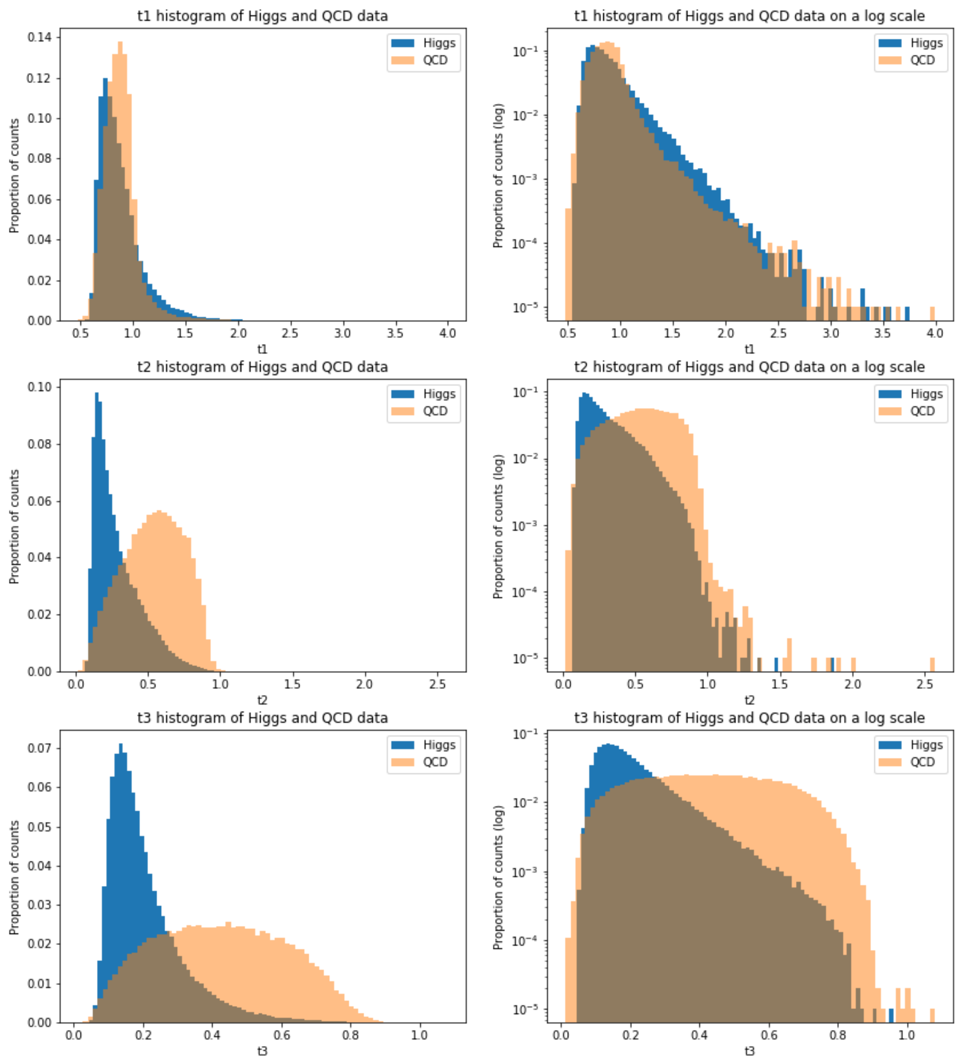


Figure 6: Histograms of the parameters $t_1$, $t_2$, $t_3$, included in the Higgs and QCD data. All are normalized so that it turns each in to a probability distribution.

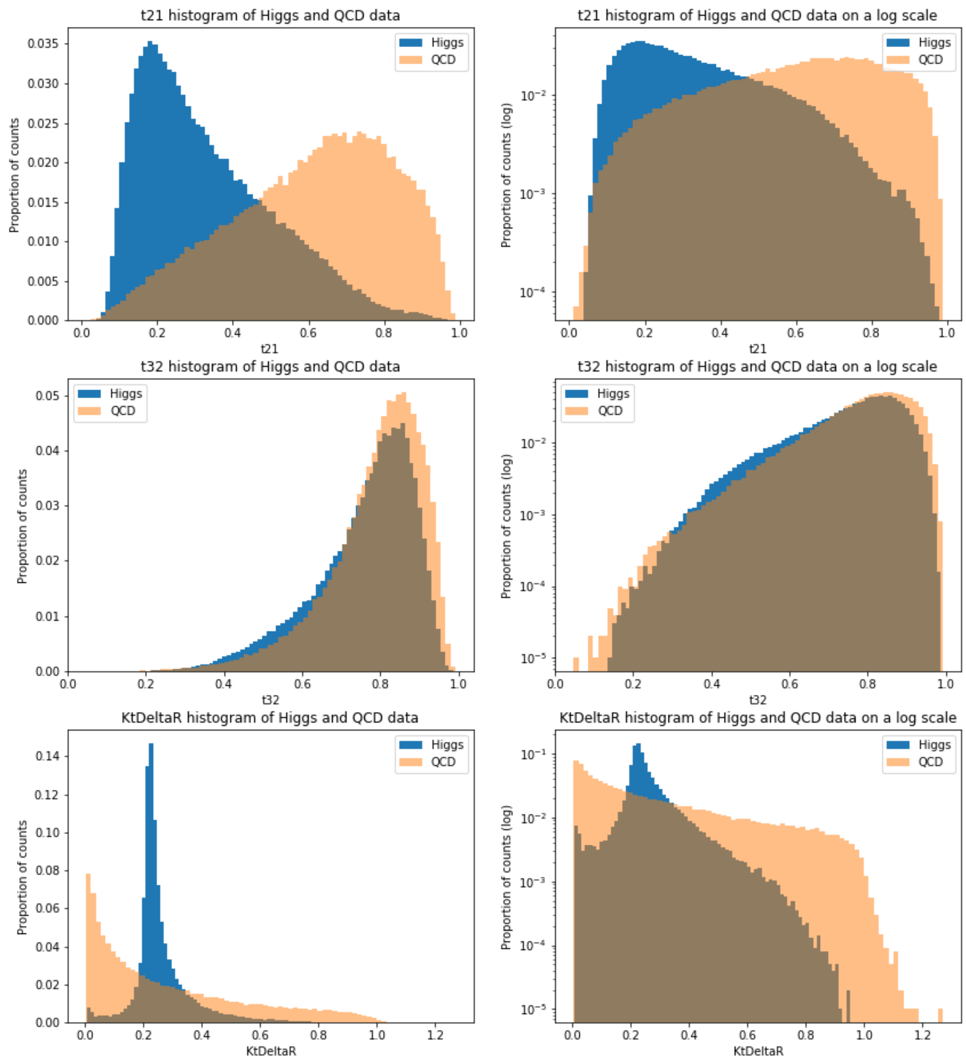


Figure 7: Histograms of the parameters $t_{21}$, $t_{32}$, $K_t \Delta R$, included in the Higgs and QCD data. All are normalized so that it turns each in to a probability distribution.