Performance of Pythia6/Pythia8 Monte Carlo Generators in the Context of a Study of Double Parton Scattering

Nathaniel Odell

November 9, 2010

Abstract

A brief, qualitative comparison of Pythia6 and Pythia8 in the context of an analysis of double parton scattering is presented. While we see fairly good agreement between the two generators for standard kinematic variables, there are marked discrepancies when we examine a set of specialized discriminator variables.

1 Introduction

In the course of conducting a study of the rate of double parton scattering (DPS) at the LHC, it has been necessary to determine an appropriate choice of simulated data to use. The most commonly used generator among the CMS collaboration, Pythia6[*], is known to suffer from deficiencies when used to model data, in particular, its modelling of the underlying event/multiple parton interactions (MPI). In response to this, a newer version, Pythia8, incorporates a more sophisticated modelling of MPI[*]. We present a comparison of simulated data from the two generators with early LHC data to provide a qualitative illustration of this point.

2 DPS Analysis Background

A full description of the DPS study can be found elsewhere[*]. To summarize, we select on events containing at least 4 jets matching some criteria (explained below), and then characterize the event topology by utilizing so-called discriminator variables (DVs). In the case of DPS, we expect that we can pair jets in such a way that angular correlations between the two pairs are negligible, and that the pairs are each appoximately balanced, i.e., $(\vec{p}_{T,1} + \vec{p}_{T,2} \sim 0)$. Another way of thinking of this, is that 4-jet DPS events will resemble a superimposition of two dijet events. On the other hand, 4-jet events originating from a single parton scattering (SPS) should exhibit correlative features since the most significant contributor is assumed to be gluon bremsstrahlung radiation. One notable difficulty arises in the case that we use untagged jets, $jjjj$, because there is a combinatoric background. This is addressed in a crude way by choosing a pairing based on the value of the DVs, and a slightly more refined way by requiring that two of the jets be b-tagged, $bbjj$. This note will only address the former.

3 Discriminator Variables

Previous studies have shown that S-family DVs are effective for summarizing features of multijet events[*]. Our analysis currently relies on four variables of this type,

$$S = \frac{1}{\sqrt{2}} \sqrt{ \left(\frac{|\mathbf{p}_{T,1} + \mathbf{p}_{T,2}|}{\sqrt{p_{T,1} + p_{T,2}}} \right)^2 + \left(\frac{|\mathbf{p}_{T,3} + \mathbf{p}_{T,4}|}{\sqrt{p_{T,3} + p_{T,4}}} \right)^2 }, \tag{1}$$

$$\Delta S = \arccos \left(\frac{(\mathbf{p}_{T,1} + \mathbf{p}_{T,2})}{|\mathbf{p}_{T,1} + \mathbf{p}_{T,2}|} \cdot \frac{(\mathbf{p}_{T,3} + \mathbf{p}_{T,4})}{|\mathbf{p}_{T,3} + \mathbf{p}_{T,4}|} \right), \tag{2}$$

$$S_{p_T} = \frac{1}{\sqrt{2}} \sqrt{ \left(\frac{|\mathbf{p}_{T,1} + \mathbf{p}_{T,2}|}{p_{T,1} + p_{T,2}} \right)^2 + \left(\frac{|\mathbf{p}_{T,3} + \mathbf{p}_{T,4}|}{p_{T,3} + p_{T,4}} \right)^2 }, \tag{3}$$

$$S_\phi = \frac{1}{\sqrt{2}} \sqrt{\Delta\phi(j_1, j_2) + \Delta\phi(j_3, j_4)}. \tag{4}$$

Data	Run range
/MinimumBias/Comissioning10-SD15_JetMETTau-Jun14thSkim15_v1/RECO	132440 - 135807
/JetMETTau/Run2010A-Jun14thReReco15_v2/RECO	135808 - 137546
MC	
/QCD15_Pt-15_7TeV-pythia8/Spring10-START3X15_V26B-v2/GEN-SIM-RECO	
/QCD15_Pt-20_TuneD6T_7TeV-pythia6/Summer10-START36_V10-v1/GEN-SIM-RECO	

For the case of S and ΔS, we choose the pairing that minimizes S while for S_{p_T} and S_ϕ, we include the combinatoric background resulting from the 3 unique pairings of the 4 jets. The meaning of S, and consequently the very similar S_{p_T}, is obvious; they are affected by the balance and sum p_T of the jet pairings, and are expected to have smaller values for DPS events. ΔS is the angle between the vector sums of the pairs that minimize S, and should be approximately flat for DPS events. Finally, S_ϕ is the angle between the jets in each pair added in quadrature, and in the case of DPS, should peak at π.

4 Datasets and Selection Cuts

To avoid the effects of pile-up, which will likely produce event topologies that are indistinguishable from those produced in DPS events, we currently rely on the earlier, low-luminosity data collected by CMS. Datasets used for the comparison, as well as run ranges, are summarized in Table 1. The integrated luminosity for the run range utilized is $\sim$13 pb^{-1} (*get exact value from LumiCalc*).

To reduce contamination of our sample with background processes it is required that there be one and only one primary vertex (PV). To ensure that this vertex is well constructed and comes from collisions near the nominal interaction point, we require that the number of degrees of freedom of the vertex be at least 4 and $|z| \leq 15$ cm. One important point to note here is that there was a change in the merging criteria for PVs for data reprocessing in CMSSW versions 36X and 38X. We are currently using data reprocessed with 36X which has a larger PV merging window of $\sim$1 cm (*get exact numbers*).

Jets must be located in the region of the detector that is covered by the tracker ($|\eta| < 2.4$) and be composed of at least 2 charged constituents. (*Need to clarify further jetID cuts*). Each of the four jets must have $p_T > 20$ GeV. We select on trigger bits to remove the effects of beam gas and scraping events as well as the HLT trigger bit (Jet15U) that requires at least one jet in the event over 15 GeV. To address trigger inefficiencies we also require that the combined p_T of the two leading jets be greater than 70 GeV. Finally, to avoid complications due to energetic jets beyond the first four, we include a veto of 15 GeV on extra jets.

5 MC comparisons

Kinematic distributions are in figures 1-6. With our analysis cuts as presented, the generators exhibit fairly good agreement with the data except for at low-p_T values. In the low-p_T region, Pythia6 shows rather poor agreement for the leading two jets. This disagreement carries over to the related H_T distributions.

By raising the cut on the 2-jet H_T from 70 GeV to 75 GeV this disagreement is resolved.

The real difference arises when we inspect the DVs which can be seen in figures 7-10. The Pythia8 sample seems to model the data better for the cases of S, S_{p_T}, and S_ϕ though both samples show marked discrepancies. Both samples agree well for the case of ΔS with the Pythia6 sample fitting the data slightly better especially for values $\sim \pi$. The relative characteristics of these distributions are not affected by the change in the 2-jet H_T cut.

A very naive interpretation of the distributions, with the exception of ΔS, is that Pythia8 underestimates the amount of DPS and Pythia6 overestimates it. As an example, consider the S distribution. In the case of DPS, we expect jets to balance pairwise which will result in a value of S that is less. Pythia6 shows an enhancement in the number of events at lower values of the S spectra when compared to data, hence it overestimates DPS. Similar statements can be made for the other DV distributions. We are still in the process of reconciling the fact that the amount of DPS determined from ΔS is not compatible with the other distributions.

Figure 1: Comparisons of p_T distributions for leading two jets with 2-jet $H_T >$ 70 GeV

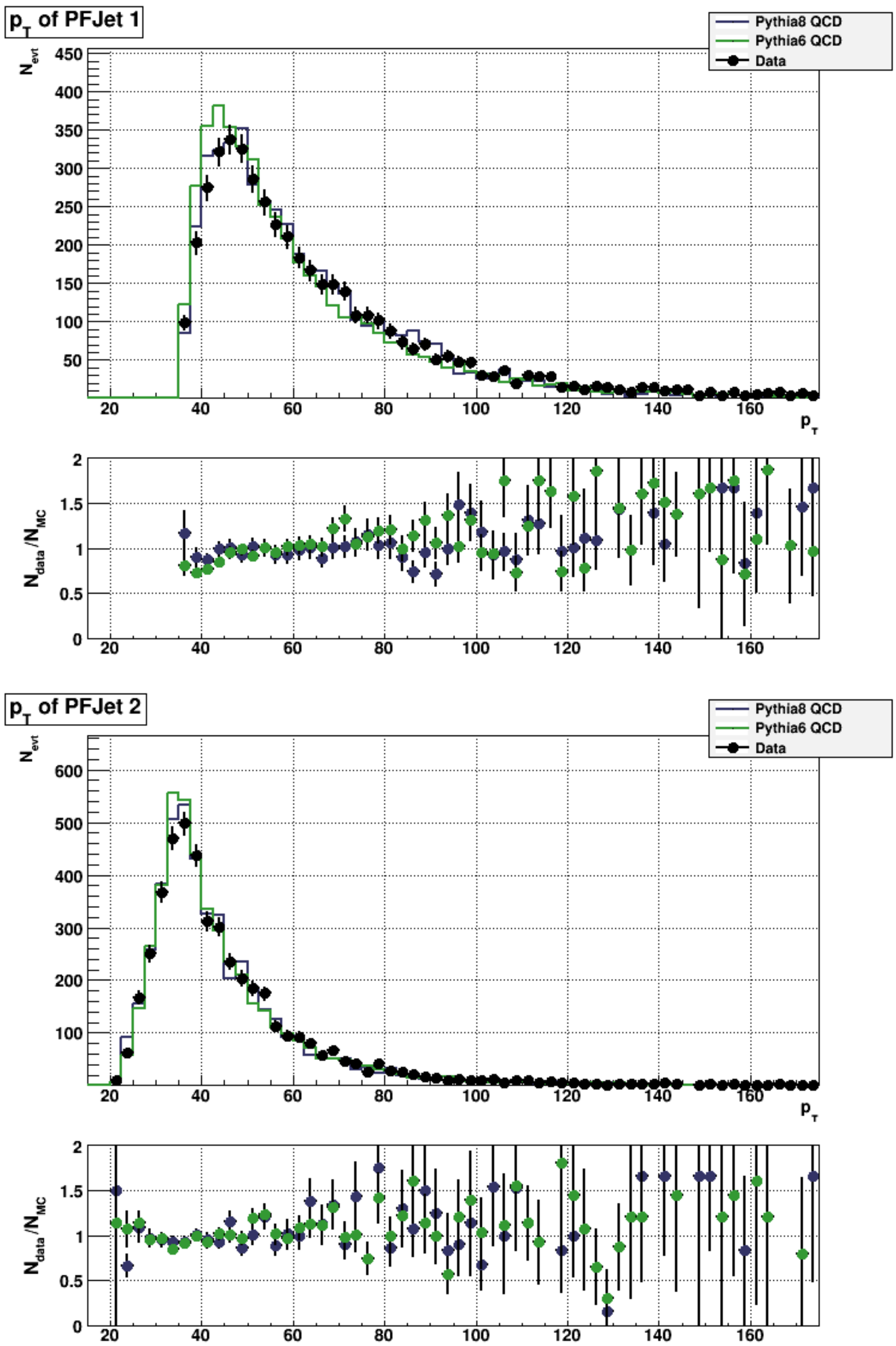


Figure 2: Comparisons of 4-jet H_T for leading two jets with 2-jet $H_T > 70$ GeV

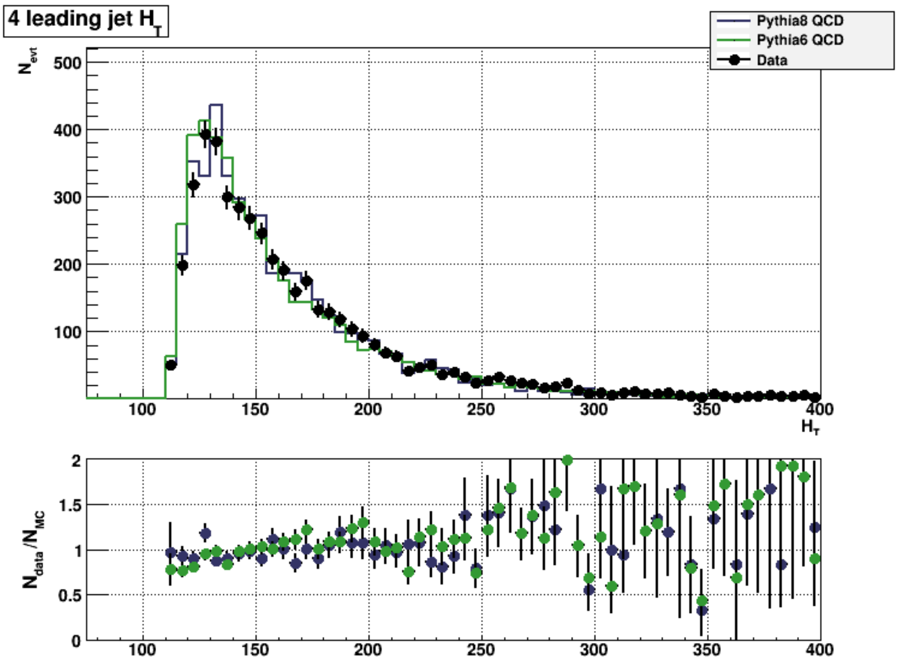


Figure 3: Comparisons of MH_T for leading two jets with 2-jet $H_T > 70$ GeV

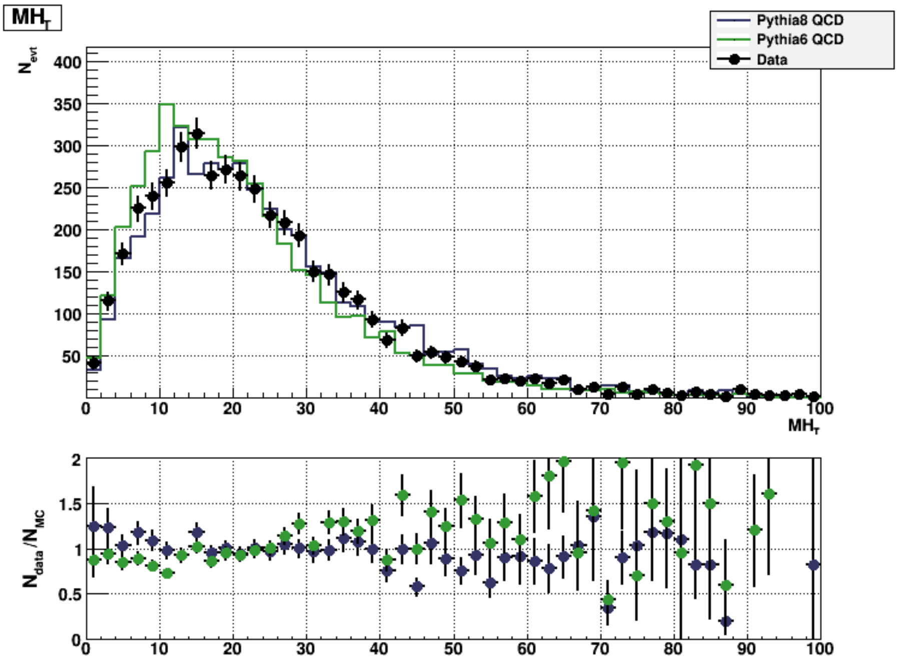
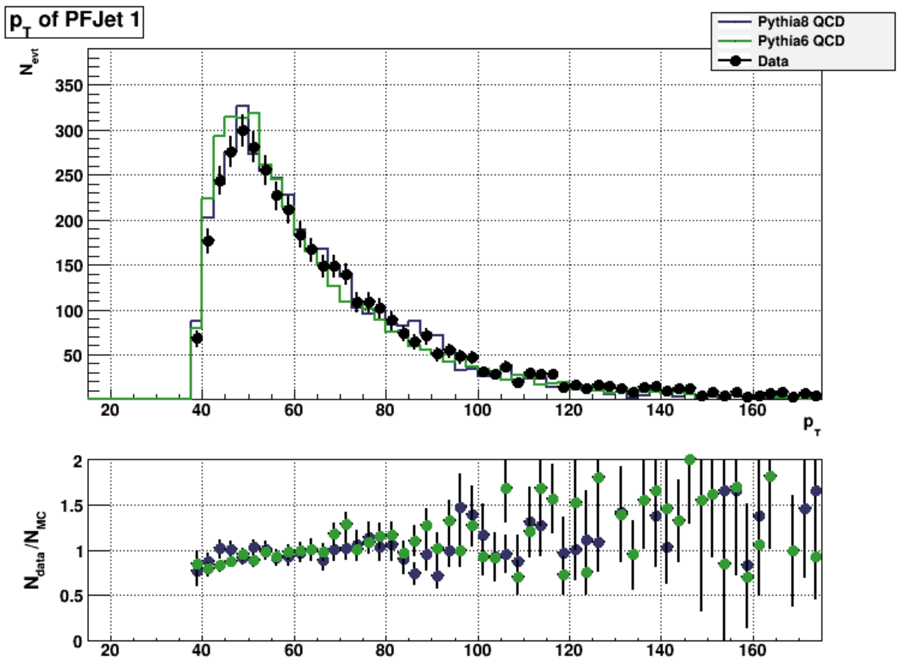


Figure 4: Comparisons of p_T distributions for leading two jets with 2-jet $H_T >$ 75 GeV



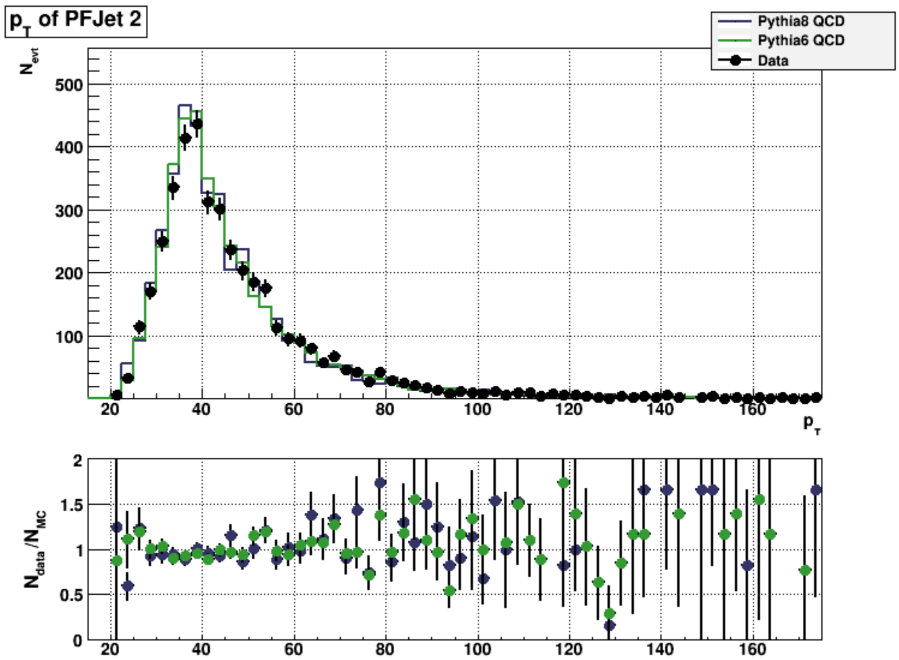


Figure 5: Comparisons of 4-jet H_T for leading two jets with 2-jet $H_T > 75$ GeV

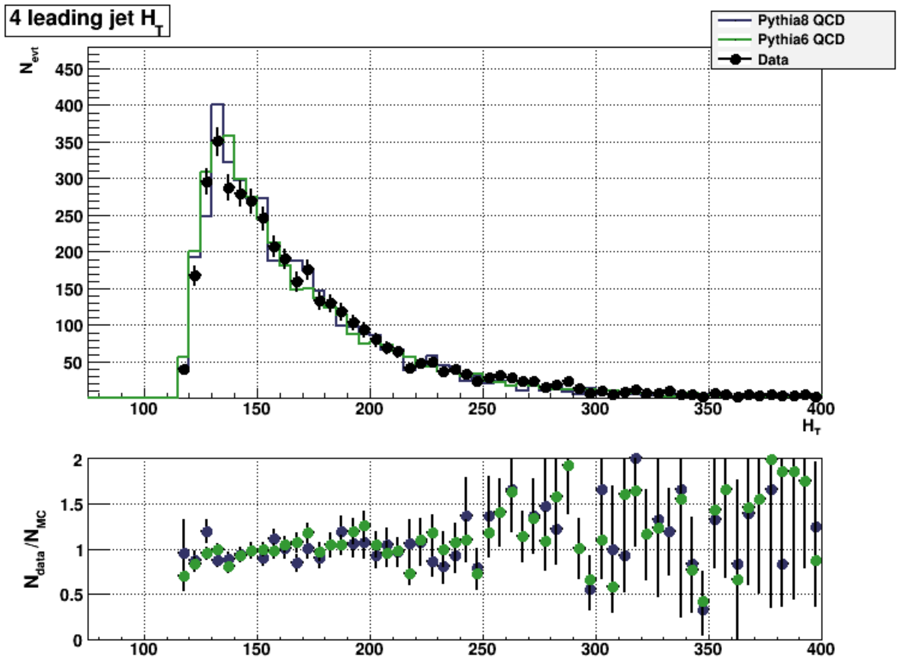


Figure 6: Comparisons of MH_T for leading two jets with 2-jet $H_T > 75$ GeV

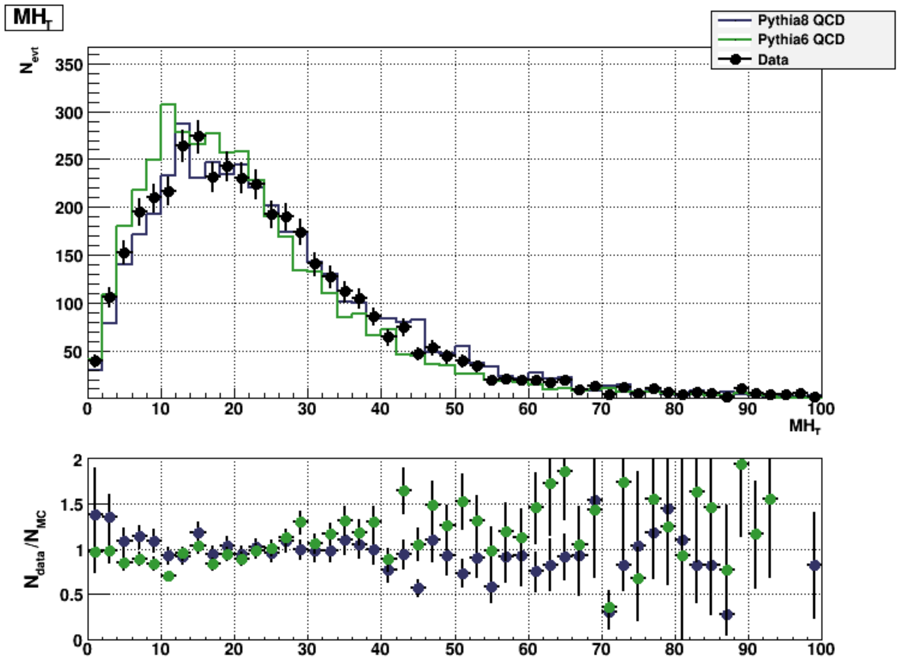


Figure 7: S for minimizing pairing and 2-jet $H_T > 70$ GeV

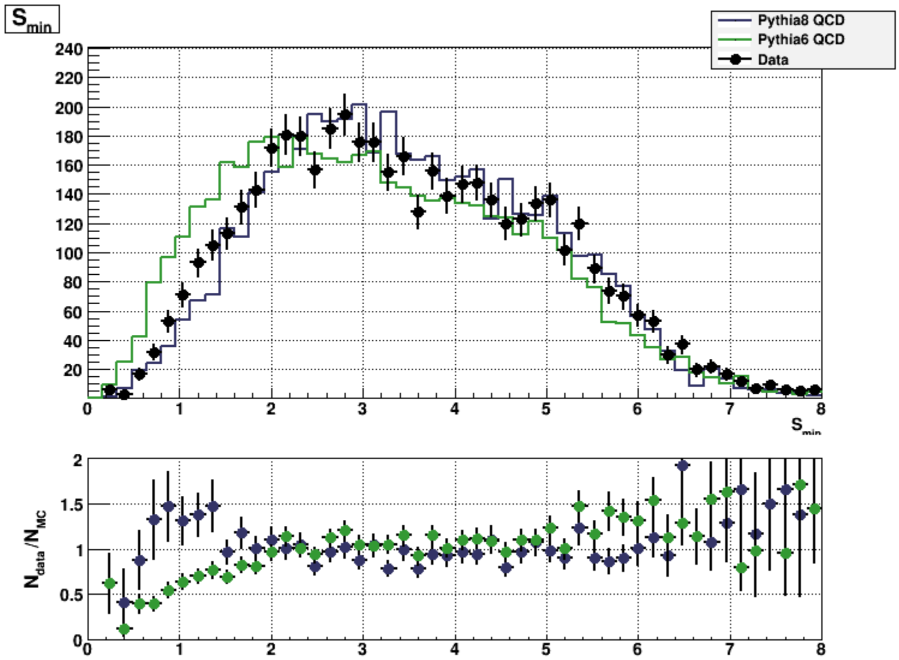


Figure 8: $S_{p_T,dem}$ DV with 2-jet $H_T > 70$ GeV

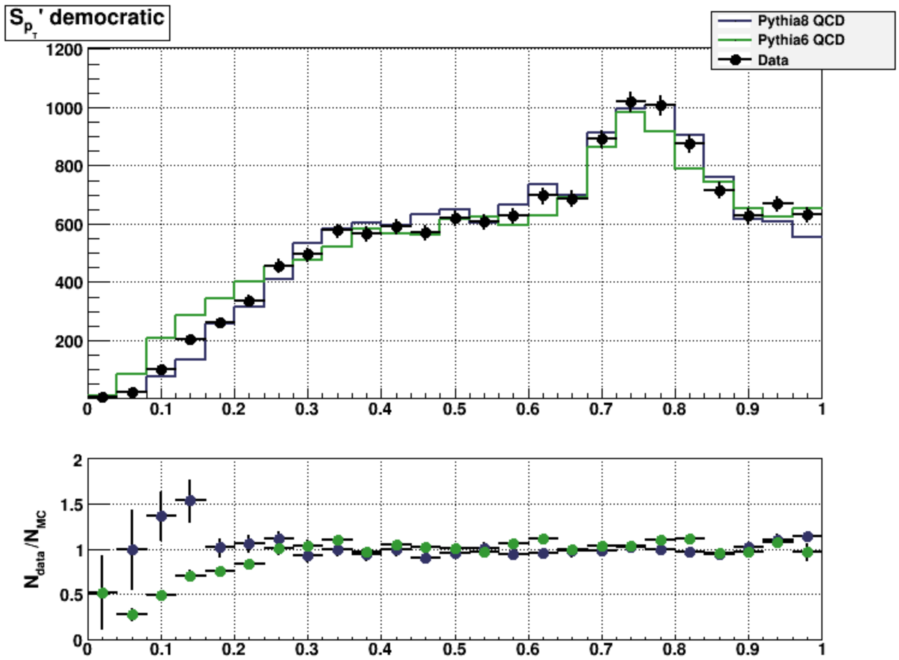


Figure 9: S_{phi} for minimizing pairing and 2-jet $H_T > 70$ GeV

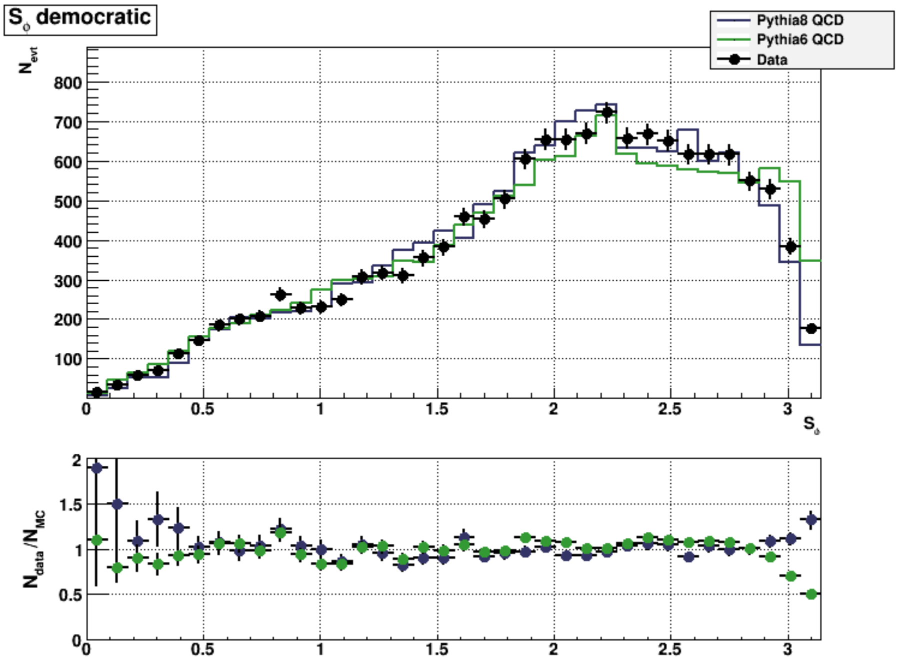


Figure 10: ΔS with 2-jet $H_T > 70$ GeV