Summer project memo: sensitivity study of Bump Hunter for semi-visible jets searching

Zhiyuan Huang, Marcella Bona, Giuliano Gustavino

1 Introduction

Many astrophysical experimental observations, such as gravitational lensing[1], have indicated the existence of dark matter. However, dark matter is still unknown in many perspectives. Also, the standard model fails to predict the existence of dark matter. Therefore, dark matter could be a breakthrough in physics beyond the standard model. There are many kinds of theoretical models that describe dark matter. The prevalent model being studied is Weakly Interacting Massive Particles (WIMPs)[2]. However, in this project, the model called "dark sector" will be studied. Based on "dark sector" theory[3, 4, 5, 6, 7, 8, 9], this hidden sector can communicate with the visible, i.e., Standard Model, sector through portal interactions—the renormalizable examples are the Higgs, photon, and lepton portals. Because the dynamics in the hidden sector can be arbitrarily complicated, these models tend to yield LHC signatures characterized by high-multiplicity final states, displaced vertices, and novel collider objects such as lepton, photon, or emerging jets[10]. This project will focus on another exotic possibility, semi-visible jets. It is assumed that a mediator $Z^{'}$ which can interact both with standard model particles and dark matter could be generated at LHC and decay to dark matter. And part of the dark matter can decay back to the standard model particles via the portal coupling. It is assumed that the $Z^{'}$ is generated statically. And it decays into a pair of dark quarks. The dark quarks will shower into two back-to-back jets. Part of the dark partons in the jets can decay back to standard model particles and the rest will stay in the dark sector which makes the jets consist of both visible particles and invisible particles, namely semi-visible jets. Therefore, there will be missing energy in these two jets. The fractions of dark matter in these two jets might vary, which means the missing energy can be aligned along the direction of one of the jets. The Feyman diagram is shown in Fig1. [10]. To simplify, the main parameters for the search are invisible fraction r_{inv} and the mass of $Z^{'}$. The selection of the final states of jets which might enhance the significance is studied. The topic of the summer project is to study the sensitivity using pyBumpHunter and use it to validate the selections.

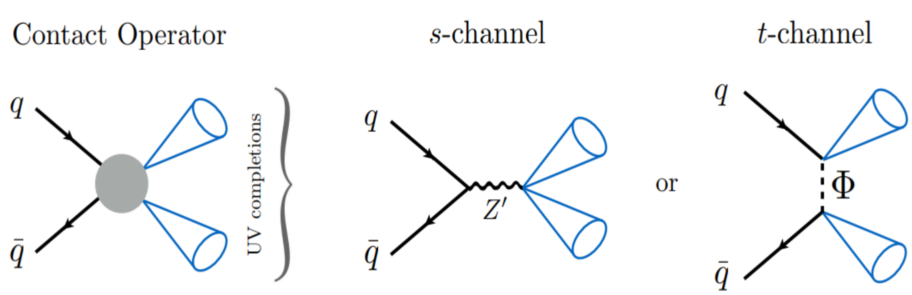


Figure 1: Feyman Diagram[10]

2 Setting up pyBumpHunter and generating histogram using TH1 fit

First, I have set up the pyBumpHunter package and written code for fitting histogram and generating histogram samples based on the fit result. The method is TH1F.fitto() which is a binned function fitting. The file used was **bkg_rinv_2d_cms.root** which is QCD background applied with CMS selection. The first two setting up commands in **install_pyBumpHunter.sh** for setting up pyBumpHunter should be changed to:

- lsetup "root 6.26.08-x86_64-centos7-gcc11-opt" #Activate python3.9.12 environment and the corresponding ROOT. The main purpose is making sure ROOT and Python are the same version. After that, you can install python3.9.12 virtual environment.
- cd pyBumpHunter
- python3 -m venv pyBH_env #Install virtual python3 environment which is a copy of python3.9.12

And no need to execute the syntax "source install_pyBumpHunter.py" again after "source install_FrequentistFramework.sh" because the syntax is already included in install_FrequentistFramework.sh. For more details, see /eos/user/z/zhhuang/svj/ install_pyBumpHunter.sh.

The file used as input is /afs/cern.ch/user/b/biliu/public/signals_svj/bkg_rinv_2d_cms.root. The fitting range is 1200 - 5000GeV. And the distribution is shown in Fig 2.

The background model for fitting is shown below:

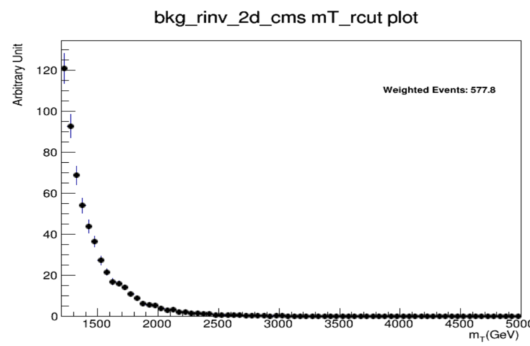
$$f(x) = p_0(1-x)^{p_1} x^{p_2+p_3\ln(x)+p_4\ln^2(x)} \tag{1}$$

Where $x = m_{jj}/\sqrt{s}$ and p_i are free parameters. However, the first parameter p_0 which is a normalization factor is not useful in the TH1 fit method. So, p_0 is fixed to 1 here. And the fit result is shown in Table 1.

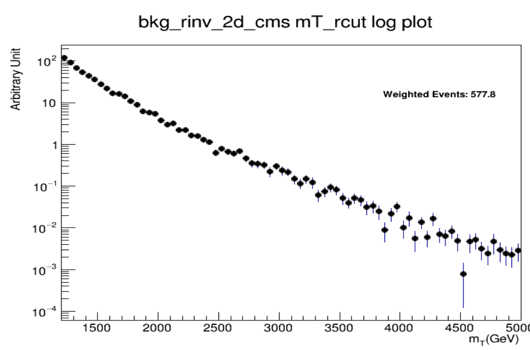
PARAMETER NAME	VALUE	ERROR	STEP SIZE
p_1	1.84108e+01	9.06685e-01	8.32798e-05
p_2	-1.98163e+00	9.39560e-01	7.33365e-06
p_3	-1.52277e-01	7.75174e-01	7.10034e-07
p_4	-2.11037e-01	1.68647e-01	1.67169e-06

Table 1: Fit result of TH1 fit for QCD background distribution with CMS selection

And the plots of the fit result are shown in Fig 3.

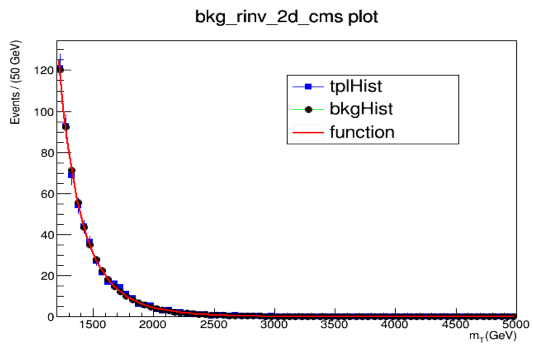


(a) Plot of QCD background distribution with CMS selection

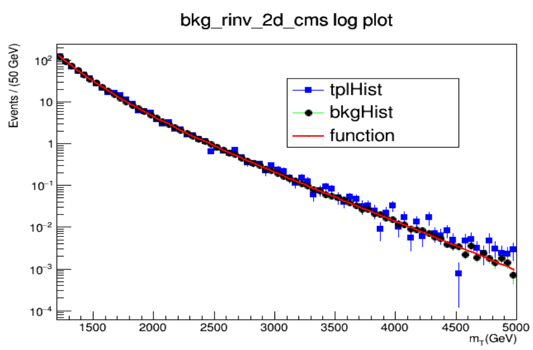


(b) Log plot of distribution of QCD background distribution with CMS selection

Figure 2: Data distribution of QCD background distribution with CMS selection



(a) Plot of fit result for QCD background distribution with CMS selection

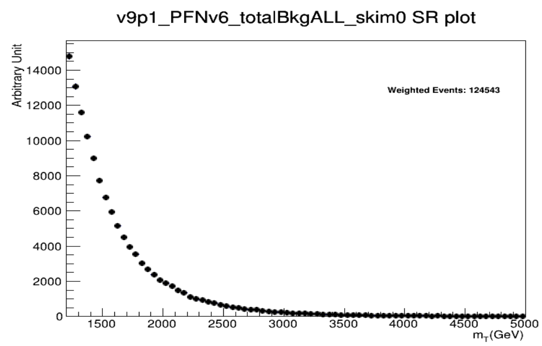


(b) Log plot of fit result QCD background distribution with CMS selection

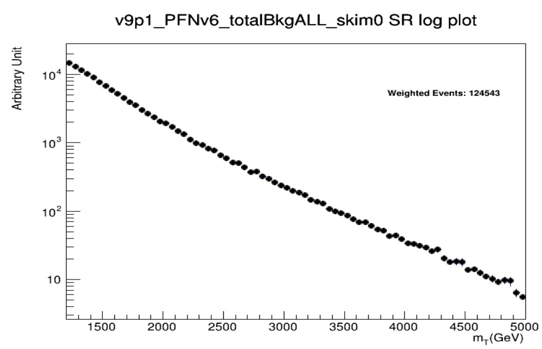
Figure 3: Fit result of QCD background distribution with CMS selection

3 Bump Hunter Study

The types of Monte Carlo background taken into consideration are QCD, diboson, $t\bar{t}$, $Z\nu\nu$, and Wjets. The selection studied is PFN score > 0.6 in signal region which is defined by the width of the subleading jet should be larger than 0.05. The background distribution is shown in Fig 4. A background toy is generated by TH1 fitted function. The parameters of the function are shown in Table 2. The fit result is shown in Fig5.



(a) Plot of m_T distribution of all kinds of background described above in signal region



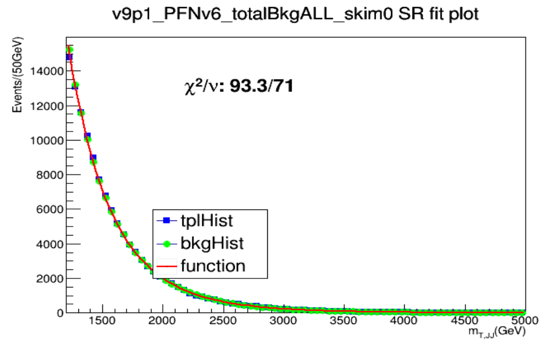
(b) Log plot m_T distribution of all kinds of background described above in signal region

Figure 4: m_{T} distribution in signal region of all background

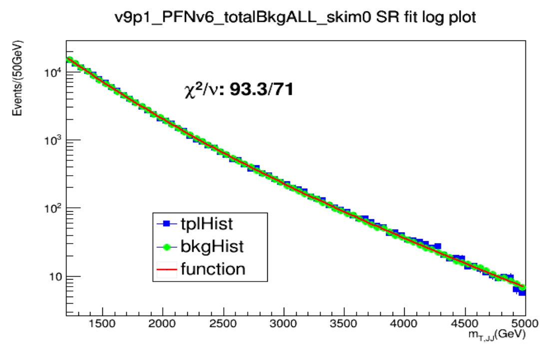
PARAMETER NAME	**VALUE**	**ERROR**	**STEP SIZE**
p_1	5.20211e+00	2.86742e-01	1.53896e-05
p_2	-4.36238e+00	3.22146e-01	1.40557e-06
p_3	5.29696e-01	2.68167e-01	6.80584e-07
p_4	2.36368e-01	5.84238e-02	3.20104e-07

Table 2: Fit result of TH1 fit for all kinds of background described above in signal region

The toy is tested by Bump Hunter to check if the toy can simulate the background Monte Carlo well. And the result is shown in Table 3 and Fig 6. The significance is 1. and there



(a) Plot of distribution of all background



(b) Plot of binned fit of all background in signal region

Figure 5: Log plot of binned fit of all background in signal region

is a bump around 1475 GeV, which indicates that the fit does not model the background perfectly.

Global p-value	**Significance**	**BUMP MIN**	**BUMP MAX**	width	number of signal events
0.0247	1.96512	1350 GeV	1600 GeV	250 GeV	728

Table 3: Result of background check by Bump Scan

The signal injection is used to evaluate the validation of selection. The algorithm of signal injection is used to calculate how many signals needed to be injected to reach the expected significance. The expected limit of significance can be set to low enough to be used to calculate the significance when the signal strength is 1 (The signal strength is defined by the ratio of the expected number of signals which reaches the expected limit of significance over the number of input signals). In the signal injection machinery, the background toy generated by TH1 fit and shown in Fig 5(a) is used as the background histogram, and the histogram of signal Monte Carlo is used as the signal histogram. The signal injection will inject the signals into the background automatically. The number of pseudo-experiments is set to 10,000. The scan width range is set to 100 GeV to 1500 GeV. The scan step is set

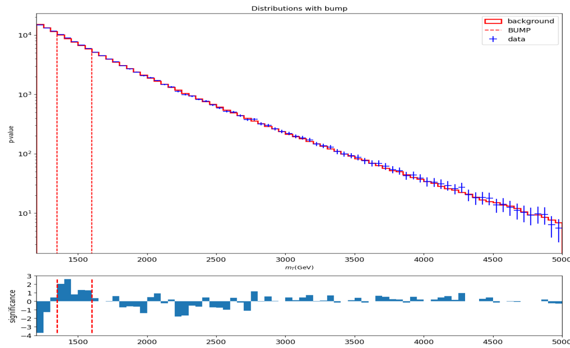


Figure 6: Check of background simulation in Bump Scan

to 50 GeV. The machinery will generate 10,000 pseudo-background-only histograms and compare each of them with the toy histogram to calculate the minimum p-value and get a pseudo-background-only p-value array. And then generate 10,000 pseudo-background-signal histograms and compare each of them with the toy histogram to calculate the minimum p-value and get a pseudo-background-signal p-value array. Then compare the pseudo-background-only p-value array with the median in the pseudo-background-signal p-value array, get the number of pseudo-background-signal p-values which are larger than the median in the pseudo-background-signal p-value array as S. So, the global p-value is the ratio of S over the number of pseudo-background-only experiments. The significance is calculated by np.norm.ppf(1 - global p-value). The example of signal distribution where $m_{Z'}$ is 2.5 TeV and r_{inv} is 0.4 is shown in Fig 7.
Fig 8 shows the significance in all possible signal scenarios.

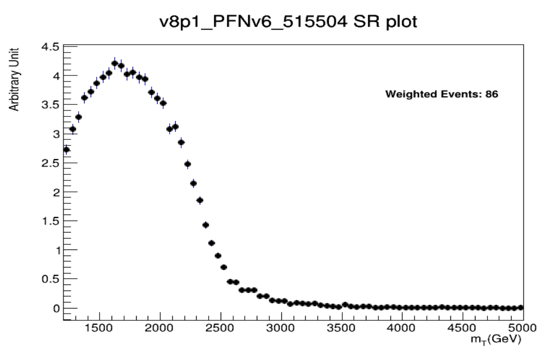


Figure 7: Signal distribution in signal region in $m_{Z'} = 2.5$TeV and $r_{\text{inv}} = 0.4$ scenario

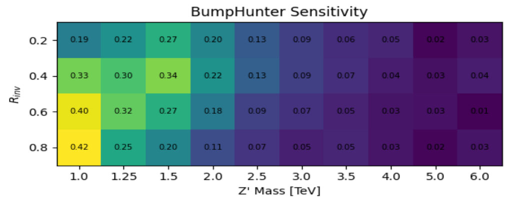


Figure 8: Sensitivity studies in signal region in all signal scenarios

4 Conclusion

I have set up the Bump Hunter, used the TH1 fit function to model the background and generate a background toy, then used the background toy and signal Monte Carlo to study the sensitivity calculated by signal injection in different physics scenarios. The next step we are going to do is to compare the CMS selection and PFN selection.

References

[1] Douglas Clowe, Maruš a Bradač, Anthony H. Gonzalez, Maxim Markevitch, Scott W. Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *The Astrophysical Journal*, 648(2):L109–L113, aug 2006.

[2] Gerard Jungman, Marc Kamionkowski, and Kim Griest. Supersymmetric dark matter. *Physics Reports*, 267(5):195–373, 1996.

[3] Haim Goldberg and Lawrence J. Hall. A new candidate for dark matter. *Physics Letters B*, 174(2):151–155, 1986.

[4] Matthew J. Strassler and Kathryn M. Zurek. Echoes of a hidden valley at hadron colliders. *Physics Letters B*, 651(5-6):374–379, aug 2007.

[5] Maxim Pospelov, Adam Ritz, and Mikhail Voloshin. Secluded WIMP dark matter. *Physics Letters B*, 662(1):53–61, apr 2008.

[6] Jonathan L. Feng and Jason Kumar. Dark-matter particles without weak-scale masses or weak interactions. *Phys. Rev. Lett.*, 101:231301, Dec 2008.

[7] Nima Arkani-Hamed, Douglas P. Finkbeiner, Tracy R. Slatyer, and Neal Weiner. A theory of dark matter. *Phys. Rev. D*, 79:015014, Jan 2009.

[8] Yonit Hochberg, Eric Kuflik, Hitoshi Murayama, Tomer Volansky, and Jay G. Wacker. Model for thermal relic dark matter of strongly interacting massive particles. *Physical Review Letters*, 115(2), jul 2015.

[9] Manuel A. Buen-Abad, Gustavo Marques-Tavares, and Martin Schmaltz. Non-abelian dark matter and dark radiation. *Physical Review D*, 92(2), jul 2015.

[10] Timothy Cohen, Mariangela Lisanti, Hou Keong Lou, and Siddharth Mishra-Sharma. LHC searches for dark sector showers. *Journal of High Energy Physics*, 2017(11), nov 2017.