Early LHC data preparations for beyond-the-standard-model searches at CMS

Didar Dobur[*][†]

University of Florida

E-mail: ddidar@cern.ch

Searches for supersymmetry (SUSY) and other phenomena beyond the standard model involve a broad range of signatures with jets, leptons, photons, and missing transverse momentum (MET). These searches require careful control over backgrounds from standard model processes. We present several methods for data-driven background determinations that have been tested on early LHC data collected by the CMS experiment at $\sqrt{s} = 7$ TeV. These data allow us to study QCD backgrounds, to evaluate methods to suppress the effects of jet-energy mis-measurement, to validate data-driven methods for predicting the MET distribution, and to measure background contributions from processes producing fake leptons.

---

[*]Speaker.

[†]On behalf of the CMS Collaboration

1. Introduction

The data of pp collisions at $\sqrt{s} = 7$ TeV collected by the CMS detector [1] at the LHC in 2010 enable searches for topologies beyond the Standard Model (SM) at an energy regime which was never reached before [2]. CMS has conducted searches in a verity of topologies which can be grouped in, fully hadronic searches where an access of events with di(multi)-jet plus large MET is searched for, di-photon plus MET searches and leptonic searches where the signature can include single, di-lepton with same or opposite charge and multi-leptons in association with jets and large MET. For all these search topologies the sensitivity to new physics depends strongly on how well the SM backgrounds are understood. Therefore CMS puts a great emphasis on developing data-driven methods for controlling and measuring the backgrounds directly from data. In this document, we present early results on the performance of several such methods which were exercised using the data corresponding to integrated luminosities between 11 and 65 nb^{-1} [3]. Note that with this limited luminosity it was not yet possible to probe the kinematic phase space relevant to SUSY, nevertheless these studies represent an important test of these methods.

2. Results

The understanding of MET tails from background, both due to real MET in physics processes (e.g., $t\bar{t}$, W/Z+jets) and artificial MET due to mis-measurements of jets are of a great importance, in particular for fully hadronic search topologies. One of the promising observables to suppress artificial MET in QCD events is defined as [4], $\alpha_T \equiv p_{T2}/M_T$ where p_{T2} is the transverse momentum of the jet with the lower p_T and M_T is the transverse mass of the di-jet system, defined as $M_T = \sqrt{2p_{T1}p_{T2}(1-\cos\Delta\phi)}$. For a perfectly measured di-jet event this observable should have a value $\alpha_T = 0.5$. If the p_T of one (and only one) of the jets is mis-measured, then by construction $\alpha_T < 0.5$. If, however, a third jet is completely lost, then $\alpha_T > 0.5$ for the apparent di-jet event. On the other hand, in SUSY-like events which contain real MET α_T distribution is expected to extend to higher values. In Figure 1 α_T distribution in di-jet events is shown for $H_T > 120$ GeV, where H_T is the scalar sum of the jet's p_T. Data are in a good agreement with Monte Carlo (MC) and exhibits the expected falling behavior of α_T around 0.5. Note that selecting events with a cut on α_T, e.g., 0.55 will reduce events with artificial MET significantly. Another important feature of this variable is the fact that its rejection power increases as H_T increases where jet energy resolutions get better. This is illustrated in Figure 1 (right), where the fraction of events with $\alpha_T > 0.55$ as a function of H_T is shown. The falling behavior of this fraction is tested under various conditions, by smearing jet energies, emulating extreme jet losses and in γ+jet events where γ is treated as jet. α_T observable was modified to be also used in multi-jet plus MET topologies where α_T is calculated with pseudojets that are formed from multi-jets.

A complimentary topological observable that is useful for rejecting events with fake MET from jet energy mis-measurements is $\Delta\phi^*$, which is defined as the minimum angle, between jet k and the MHT calculated with the remaining jets: $\Delta\phi^* \equiv \min_{\text{jets } k}\left(|\Delta\phi(\vec{p}_k, -\sum_{\text{jets } i\neq k}\vec{p}_i)|\right)$ where $\vec{p}$ is the momentum of jets. MHT that is analogous to MET, and is defined as MHT $\equiv |\sum_{\text{jets } j} -\vec{p}_{Tj}|$. $\Delta\phi^*$ variable effectively tests whether the MHT in the event is due mis-measurement of at least one jet. In Figure 2 (left) the distribution of $\Delta\phi^*$ in data is shown. In generic QCD events, the MHT
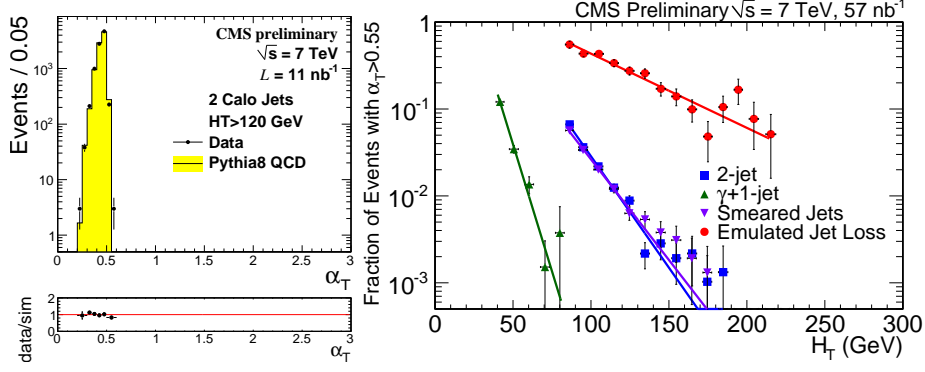
Figure 1: Left: α_T distribution in di-jet events. Data/MC ratio is shown at the bottom panel. Right: Fraction of events passing an $\alpha_T > 0.55$ cut as a function of H_T.
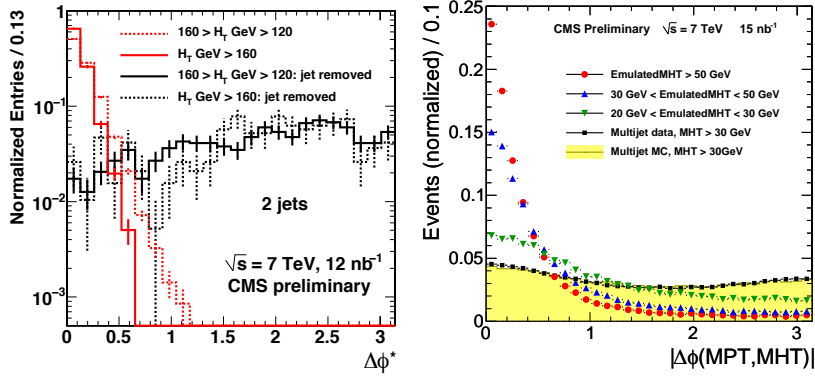

Figure 2: Left: Distributions of $\Delta\phi^*$ for di-jet under various conditions. Right: $|\Delta\phi(MPT, MHT)|$ distribution in multi-jet events and in events with real MHT emulated by removal of one jet.

direction generally coincides with the direction of one of the jets, which has been mis-measured and the $\Delta\phi^*$ therefore peaks at zero and becomes narrower at higher H_T. To emulate the behavior in a SUSY sample with real MET, we randomly remove one of the jets and the resulting $\Delta\phi^*$ distributions are more uniform.

Another way to control artificial MET is to compare measurements from two independent detectors, i.e., calorimeter and tracker. The distribution of the angle between MHT and MPT, $\Delta\phi(MPT, MHT)$ where $MPT \equiv |-\sum_i \vec{p}_T(\text{track}_i)|$ is shown in Figure 2 (right). In the case of fake missing energy due to mis-measurement or arising from instrumental noise a small correlation between the angles of MPT and MHT is expected and data confirms this behavior. In contrast, $|\Delta\phi(MPT, MHT)|$ peaks at zero in events with real MHT that is emulated here by removing jets randomly. As MHT increases, this distribution narrows.

Despite the strong suppression of QCD backgrounds using observables described above it is still necessary to measure the residual MET tails from data. A data-driven method to predict the artificial MET in γ+jets events using QCD multi-jet events was exercised in data. The basic assumption of this method is that artificial MET is a function of jet multiplicity, N_{jet} and HT. Therefore MET templates are measured in bins of, N_{jet} and HT in a jet triggered control sample and for each γ+jets event depending on its N_{jet} and HT the corresponding template from control sample is added. The final prediction is the sum of N MET templates where N is the number of

selected γ+jets. In Figure 3 (left) predicted and observed MET distribution is presented where a good agreement between prediction and observation can be seen. In Figure 3 (right) the results of another method that is designed to predicts artificial MET background to di- γ plus MET search topology is presented. In this method, a control sample that is selected by inverting the isolation requirement on the photons, hence dominated by QCD, is used to measure MET background. The observed number of events in the signal sample with MET< 10 GeV is used to fix the normalization of the measured MET distribution and the predicted number of background events for MET> 20 GeV is obtained as 4.2 ± 1.5 while 4 events are observed.

The dominant background for same-sign di-lepton plus jets and MET signatures is expected to be due to $t\bar{t}$ production where one lepton comes from W and one from b-quark decays. The isolation property of the latter lepton can provide a good handle to measure $t\bar{t}$ background from data. A tag-and-probe method using $b\bar{b}$ events selected with one b-tagged jet and a muon in the opposite hemisphere assuming that it originates from the other b-quark is studied. In Figure 4 the measured Relative isolation (defined as the sum of the transverse energy within a cone of $R = 0.3$ around the muon direction, divided by the muon's p_T) distribution is shown for two N_{jet} ranges. The measured relative isolation distributions can be used to predict the $t\bar{t}$ contribution to same-sign di-muon signature. Although the MC predictions for QCD background for this signature seem to be small it forms the most uncertain one due to poorly known QCD production cross-section and therefore should be measured from data. A data-driven method that is based on the factorization principle of certain cut efficiencies was tested. If the factorization principle holds the individual efficiencies can be measured in a QCD dominated control sample and can be applied to a preselected number of events to obtain the final prediction for QCD background. In Figure 5 (left) the validation of the the factorization of the isolation cut efficiency on two muons is shown. The efficiency for both muons to pass the isolation cut is found to be consistent with the prediction obtained from the square of the single muon isolation efficiency.

For the leptonic search topologies the isolation requirement on leptons strongly suppresses QCD backgrounds, however the residual contribution needs to be measured from data. Electrons coming from QCD processes are due to heavy-flavor decays, jets being misidentified as electrons or due to photon conversions. We measure the relative isolation templates in jet triggered control samples that are further enriched with electrons of each of these sources. Using these templates together with a prompt lepton template taken from MC to account for W contribution (this template can also be measured from data using tag-and-probe methods once a sufficient data is collected) in
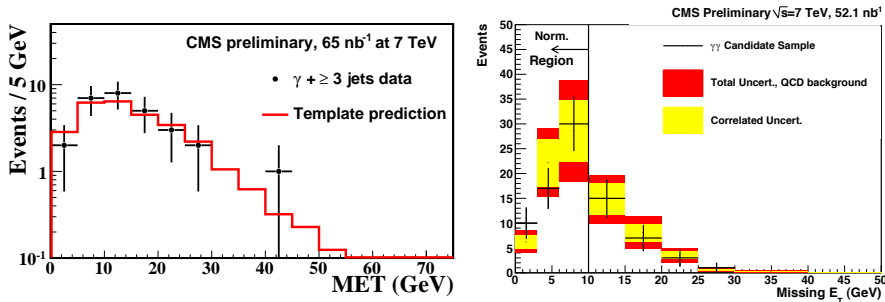

Figure 3: Left: MET predictions, based on templates, compared to the observed MET in $\gamma+ \geq 3$ jet events. Right: Predicted MET distribution for di-γ plus MET topology.
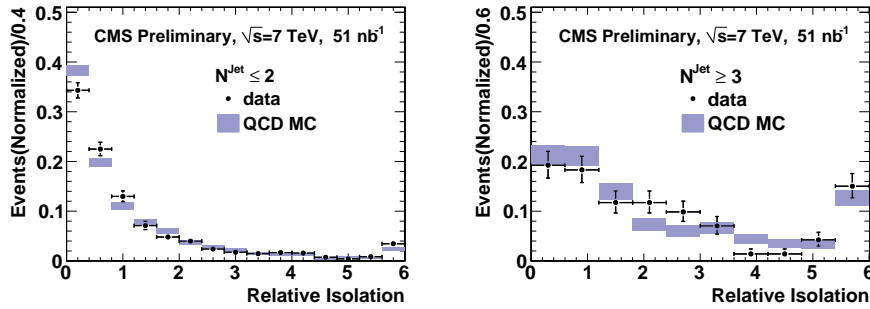
Figure 4: Measured muon isolation distributions using the tag-and-probe method in data (black markers) and in a muon enriched QCD MC (blue shaded curve) for two different N_{Jet} ranges. The last bins include overflow entries.
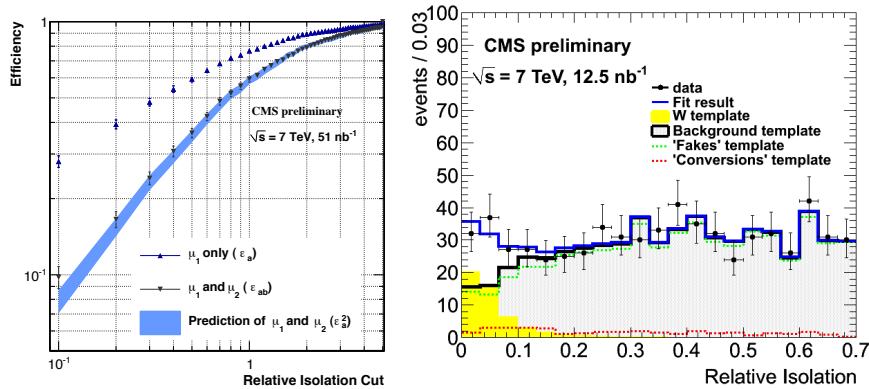

Figure 5: Left: Efficiency as a function of the muon isolation cut in loosely selected same-sign di-muon data. Right: Fits to the distribution of relative isolation of electrons.

a linear combination a fit to data in the signal region is performed. The right plot in Figure 5 shows the results of this method. The fit provides a good description of the data.

3. Summary

Searches for signatures of SUSY require a good understanding of the SM backgrounds. CMS puts a great emphasis on data-driven techniques to measure these backgrounds from data. Despite the limited amount of data, a number of such methods were examined. Only a selected set of these results are summarized here. A more comprehensive summary of the results can be found elswhere [3].

References

[1] CMS Collaboration,**JINST 3:S08004,2008**, *"The CMS experiment at the CERN LHC"*

[2] CMS Collaboration,**CMS Note-2010-008**, *"The CMS physics reach for searches at 7 TeV "*

[3] CMS Collaboration,**CMS Physics Analysis Summary: SUS-10-001**, *"Performance of Methods for Data-Driven Background Estimation in SUSY Searches"*

[4] L.Randall and D. Tucker-Smith,**Phys.RevLett. 101221803 (2008)**, *"Dijet searches for Supersymmetry at the LHC "*