1 Significance

Missing transverse momentum may arise from many sources, some of which, like instrumental defects and measurement resolution, represent unwanted contributions that we would like to suppress. In this section we discuss a significance variable, $\mathcal{S}$, which assesses on an event by event basis the likelihood that the observed $\not\!\!E_T$ is a fluctuation from zero arising from finite measurement resolution. $\mathcal{S}$ provides information beyond the raw value of $\not\!\!E_T$ that the user may deploy to improve S/B in some physics cases.

1.1 Definition

The determination of missing transverse momentum involves computation of the vector sum of selected $\vec{\not\!\!E}_T$ vectors, and its significance requires evaluation of the total uncertainty associated with that sum. Using the notation $\vec{\varepsilon} = (\varepsilon_x, \varepsilon_y)$ to indicate a generic $\vec{E}_T$ vector, we will characterize the measurement uncertainty associated with $\vec{\varepsilon}$ by a function $\mathcal{L}(\vec{\varepsilon})$, which defines likelihood contours in the $(\varepsilon_x, \varepsilon_y)$ plane. To find the likelihood function associated with the sum over many $\vec{\varepsilon}_i$, we consider first the case of two vectors, $\vec{\varepsilon}_1$ and $\vec{\varepsilon}_2$; the resultant $\vec{\varepsilon} = \vec{\varepsilon}_1 + \vec{\varepsilon}_2$ has a likelihood distribution given by:

$$\mathcal{L}(\vec{\varepsilon}) = \int \mathcal{L}_1(\vec{\varepsilon}_1)\mathcal{L}_2(\vec{\varepsilon}_2)\delta(\vec{\varepsilon} - \vec{\varepsilon}_1 - \vec{\varepsilon}_2)\, d\vec{\varepsilon}_1\, d\vec{\varepsilon}_2. \tag{1}$$

For an arbitrary number of input vectors the final likelihood is obtained by applying Eq. 1 recursively.

This formulation in terms of likelihoods is completely general and will accomodate any likelihood distributions $\mathcal{L}(\vec{\varepsilon}_i)$, but in practice we can often assume gaussian errors for measured quantities – in which case the integrals of Eq. (1) may be done analytically. Given a 2×2 covariance matrix $\mathbf{V}_i$ describing the measurement uncertainties in $\vec{\varepsilon}_i$, a gaussian likelihood distribution is given by[1]

$$\mathcal{L}_i(\vec{\varepsilon}_i) \sim \exp(-\frac{1}{2}\vec{\varepsilon}_i^{\,T}\mathbf{V}_i^{-1}\vec{\varepsilon}_i)$$

and the integration of Eq. 1 yields

$$\mathcal{L}(\vec{\varepsilon}) \sim \exp(-\frac{1}{2}(\vec{\varepsilon} - \vec{\varepsilon}_1 - \vec{\varepsilon}_2)^T\, \mathbf{V}^{-1}\, (\vec{\varepsilon} - \vec{\varepsilon}_1 - \vec{\varepsilon}_2))$$

with $\mathbf{V} = \mathbf{V}_1 + \mathbf{V}_2$. This is just ordinary propagation of errors. For many contributing

[1]The superscript T indicates the transpose of a matrix or column vector; subscript T indicates "transverse", as in "transverse energy".

vectors the expression becomes

$$\mathcal{L}(\vec{\varepsilon}) \sim \exp(-\frac{1}{2}(\vec{\varepsilon} - \sum_i \vec{\varepsilon}_i)^T (\sum_i \mathbf{V}_i)^{-1} (\vec{\varepsilon} - \sum_i \vec{\varepsilon}_i). \tag{2}$$

In the case of $\not\!\!E_T$ significance, the sums are taken over detector elements (such as calotowers) or higher level objects (such as jets). Each $\vec{\varepsilon}_i$ is the $-\vec{\not\!\!E}_T$ of the i^{th} element or object, and $\mathbf{V}_i$ its corresponding covariance matrix. Each covariance matrix in turn is most easily specified in a natural coordinate system having one axis aligned with the object's $\vec{E}_T$ vector, $\vec{E}_T \equiv (E_T \cos\phi,\ E_T \sin\phi)$:

$$\mathbf{U}_i = \begin{pmatrix} \sigma^2_{E_T i} & 0 \\ 0 & E^2_{T i}\,\sigma^2_{\phi_i} \end{pmatrix}. \tag{3}$$

Once specified in this way, the matrix must be rotated into the standard CMS xy reference frame with a rotation matrix $R(\phi_i)$,

$$\mathbf{V}_i = R(\phi_i)\mathbf{U}_i R^{-1}(\phi_i) \tag{4}$$

before summing with others.

We then define the significance as the log-likelihood ratio,

$$\mathcal{S} \equiv -2\ln\left(\frac{\mathcal{L}(\vec{\varepsilon} = \vec{\not\!\!E}_T^{\text{observed}})}{\mathcal{L}(\vec{\varepsilon} = 0)}\right), \tag{5}$$

which compares the likelihood of the observed $\vec{\not\!\!E}_T$ with the likelihood of the null hypothesis $\vec{\not\!\!E}_T = 0$. Combining Eqs. 2, 4, and 5 we obtain

$$\mathcal{S} = (\sum_i \vec{\varepsilon}_i)^T \left(\sum_i R(\phi_i)\mathbf{U}_i R^{-1}(\phi_i)\right)^{-1} (\sum_i \vec{\varepsilon}_i). \tag{6}$$

Note that

$$\vec{\not\!\!E}_T = -\sum_i \vec{\varepsilon}_i.$$

Equation 6 is the basis upon which we build the significance algorithm. In general $\mathcal{S}$ will be small when the $\not\!\!E_T$ can be attributed to measurement resolution, and large otherwise. The value of $\mathcal{S}$ depends on resolutions which are presently not well known, but it may still be used effectively as a discriminant to distinguish hypothesis pairs ("signal" and "background") in appropriate cases. The discrimination power is something the user should evaluate for his or her own analysis.

Two comments are in order:

- It is useful to realize that in the gaussian case $\mathcal{S}$ is nothing more than a χ^2 value. In fact if we choose to work in a coordinate system with the x axis aligned with the $\vec{\not\!\!E}_T$ axis, instead of the CMS horizontal axis, then Eq. 6 simplifies to a one-dimensional statement, $\mathcal{S} = E_T^2/\sigma^2_{E_T}$. This clarifies the essential meaning of $\mathcal{S}$, but it tends also to obscure an important fact – namely that through its denominator, $\mathcal{S}$ embodies the full topological information in the event. Essential features such as the angles between the measured MET vector and the various jets and unclustered calotowers in the event are embedded in the definition of $\sigma^2_{E_T}$. This fact is more evident in the representation of Eq. 6.

- The specialization to gaussian likelihood functions adopted here is somewhat less restrictive than it may appear. Eq. (1) is linear in the likelihoods and consequently any likelihood that can be expressed as linear combination of gaussians is easily handled. This is the situation commonly seen, as for example in cases where resolutions are parametrized by combinations of "narrow" and "wide" gaussians. Asymmetric distributions can also sometimes be handled in this way, using displaced gaussians. In any case Eq. (1) remains valid for arbitrary distributions, even if numerical integration may be required in difficult cases.

1.2 Implementation

To apply Eq. (6) one must specify the domain of the sums; different flavors of significance can be constructed by choosing different sets of objects to sum over. For succinctness, we restrict ourselves here to discussing MHT Significance, $\mathcal{S}_H$, in which the sums are taken over jets, electrons, and muons. Another obvious possibility is MET Significance, $\mathcal{S}_E$, in which one sums directly over calotowers. $\mathcal{S}_E$ carries additional information embedded in the distribution of unclustered energy in the event, and in the long run may prove to be the more useful formulation, but $\mathcal{S}_H$ has the advantage of relying only on well-studied objects whose resolutions are known. In Monte Carlo studies (of CSA07 vintage) $\mathcal{S}_E$ is somewhat more effective than $\mathcal{S}_H$. Nevertheless, unless specifically noted otherwise, all further discussion is about $\mathcal{S}_H$.

In the current implementation the algorithm sums jets, electrons, and muons. We require the jets to pass loose criteria, $p_T > 20\,\mathrm{GeV}$, $\eta < 5$, and $EMF < 0.9$, where EMF is the electromagnetic fraction in the jet, the ratio of ecal to hcal energy deposition. Electrons must satisfy $p_T > 10\,\mathrm{GeV}$, $\eta < 3$, and muons, $p_T > 10\mathrm{GeV}$, $\eta < 2.5$. These criteria are settable[2] and may be subject to future optimization, as well as individual user control. The algorithm uses `SelectedLayer1Jets, SelectedLayer1Electrons,` and `SelectedLayer1Muons` as its input so the user may impose additional and/or more restrictive criteria through the PAT Layer1 selection.

In the implementation in CMSSW 1.6.12 reported here, we parametrize the jet resolutions

[2] `/CMSSW/PhysicsTools/PatAlgos/data/producersLayer1/mhtProducer.cfi`

with an expansion in $1/\sqrt{E_T}$ of the reconstructed jet,

$$\left(\frac{\sigma_{E_T}}{E_T}\right)^2 = \left(\frac{a}{E_T}\right)^2 + \left(\frac{b}{\sqrt{E_T}}\right)^2 + c^2 \tag{7}$$

with $a = 5.6$, $b = 1.25$, and $c = 0.033$ as given in the Physics TDR (Ch. 11.4). Similarly the jet angular resolution σ_ϕ is taken to have the same form, with $a = 4.75$, $b = 0.426$, and $c = 0.023$. These values are set in the `mhtProducer.cfi` file and enter the formalism here through the covariance matrix $\mathbf{U}_i$ of Eq. (3). For each jet, the parameterization of Eq. 7 is evaluated at the measured value of the jet E_T and is then taken to be constant, *i.e.*, it is not a variable in the integration of Eq. 1. The entries of the total covariance matrix $\mathbf{V} \equiv \sum_i R(\phi_i)\mathbf{U}_i R^{-1}(\phi_i)$ are usually dominated by the jet resolutions, which go roughly as $\sigma_{E_T} \sim 1.25\sqrt{E_T}$ per jet, while the contributions from electrons and muons, characterized by $\sigma_{p_T} \sim 0.02 p_T$, are significantly smaller. We therefore treat the lepton resolutions as negligible (*i.e.*, zero) for now. This is approximation cannot be justified at large p_T and/or η, and will need to be improved upon in the future. Note that this *only* applies to resolutions used in computing the covariance matrix, and *not* to the lepton $\vec{\varepsilon}$ vectors entering the $\vec{\not{E}}_T$ sum!

1.3 Performance

To illustrate the use of the significance variable, we explore its performance in two quite different physics cases: $W \to e\nu$ and SUSY (LM1, all hadronic). In the first case the typical $\not{E}_T$ is around 40 GeV; while in the latter case it is generally above 200 GeV.

For the $W \to e\nu$ analysis we follow the strategy similar to that laid out in CMS Note 2007/026, and for LM1 we follow CR2007/053. $W \to e\nu$ must first pass the single isolated electron HLT trigger, and we then require GSF electrons of $p_T > 20\,\mathrm{GeV}$ with mild isolation and ID requirements. In addition, for the purpose of exploring MHT Significance, we require the event to contain at least one jet because $\mathcal{S}_H$ isn't defined otherwise. For LM1 there must be three or more jets ($p_{T1,2,3} > 180,\ 110,\ 30\,\mathrm{GeV}$). The analysis cuts are deliberately kept simple to illustrate generic characteristics. Signal samples are CSA07, background samples are CSA07 Gumbo and Chowder:

```
/Wenu/CMSSW_1_6_7-ReRecoIdeal-1198082363/RECO
/CSA07Electron/CMSSW_1_6_7-CSA07-Tier0-A1-Gumbo/RECO
/LM1_sftsdkpyt/CMSSW_1_6_7-CSA07-1200560744/RECO
/CSA07JetMET/CMSSW_1_6_7-CSA07-Tier0-A1-Chowder/RECO
```

In each of the two physics cases we examine the distributions for $\mathcal{S}_H$ and MHT, and construct plots of signal efficiency versus background efficiency for two different analysis strategies: cutting only on MHT, or cutting only on $\mathcal{S}_H$ (above and beyond the cuts noted above). The results are shown in the four panels of Fig. 1. For $W \to e\nu$ the scatterplot of $\mathcal{S}_H$ versus MHT in panel (a) indicates that $W \to e\nu$ signal events exhibit large $\mathcal{S}_H$ at the expected

$MHT \sim 40\,\mathrm{GeV}$, and correspondingly the signal efficiency versus background efficiency shown in panel (b) confirms that the $\mathcal{S}_H$ cut is the (slightly) more favorable choice. The non-smoothness in the efficiency curves in panel (b) is due to the use of weighted events in the background sample. For LM1, the scatter plot in panel (c) reveals that $\mathcal{S}_H$ will have little impact on the analysis as MHT is by itself an effective variable and significance has little to add. The efficiency plot in panel (d) shows that in fact the $\mathcal{S}_H$ cut is less useful than the direct cut on MHT for this mode.

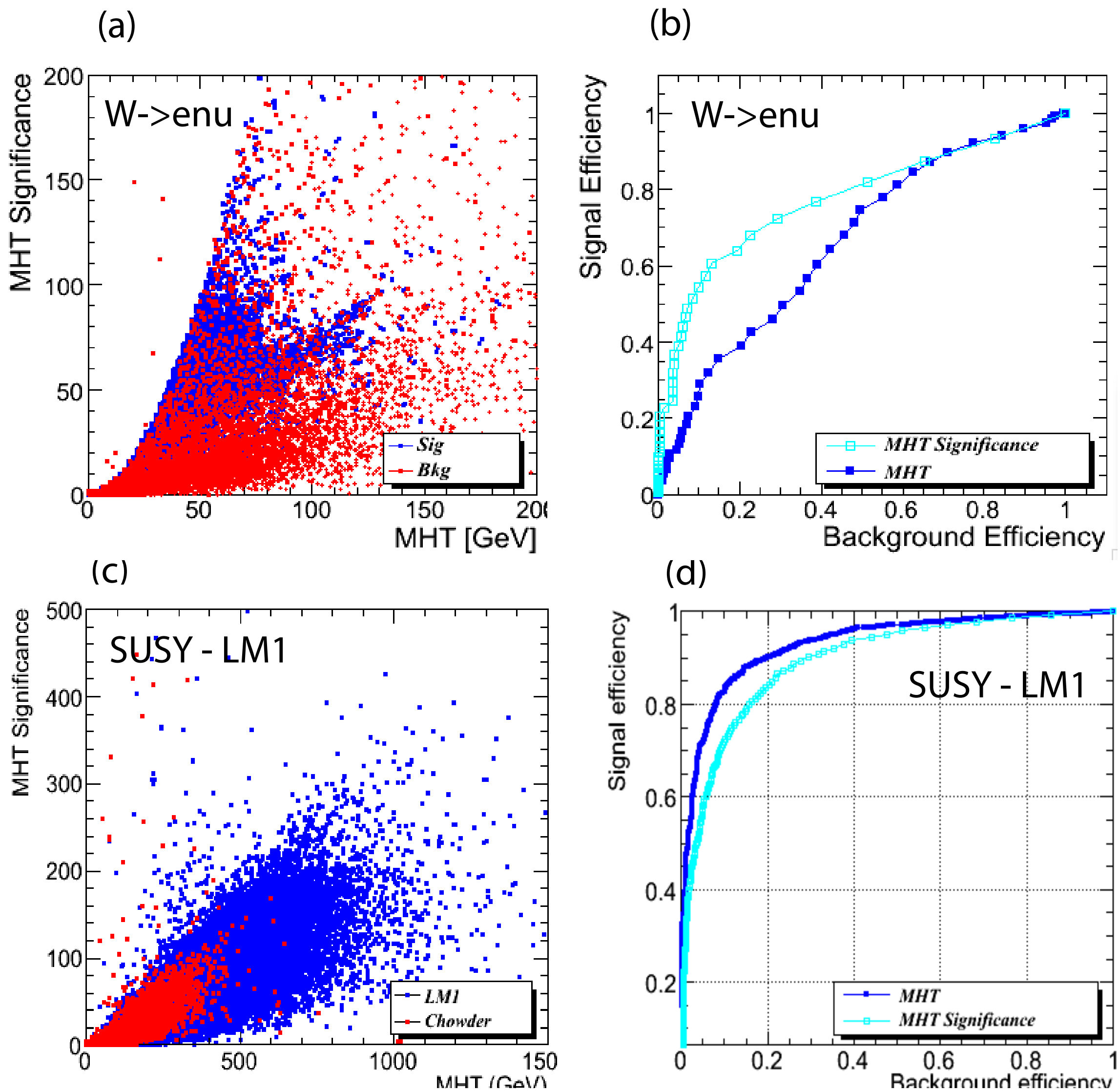

Figure 1: (a) $\mathcal{S}_H$ versus MHT for $W \to e\nu$; (b) signal efficiency versus background efficiency for $W \to e\nu$; (c) $\mathcal{S}_H$ versus MHT for LM1; (d) signal efficiency versus background efficiency for LM1.