SLEUTH: A QUASI-MODEL-INDEPENDENT SEARCH STRATEGY FOR NEW PHYSICS

BRUCE KNUTESON
Enrico Fermi Institute, University of Chicago, 5640 S. Ellis Ave., Chicago, IL 60637

for the DØ collaboration

How can we search for new physics when we only vaguely know what it should look like? How can we perform an unbiased yet data-driven search? If we see apparently anomalous events in our data, how can we quantify their "interestingness" *a posteriori*? We present an analysis strategy (SLEUTH) that simultaneously addresses each of these questions, and we demonstrate its application to over thirty exclusive final states in data collected by DØ in Run I of the Fermilab Tevatron.

1 Motivation

It is generally recognized that the standard model, an extremely successful description of the fundamental particles and their interactions, must be incomplete. Although there is likely to be new physics beyond the current picture, the possibilities are sufficiently broad that the first hint could appear in any of many different guises. This suggests the importance of performing searches that are as model-independent as possible.

Most recent searches for new physics have followed a well-defined set of steps: first selecting a model to be tested against the standard model, then finding a measurable prediction of this model that differs as much as possible from the prediction of the standard model, and finally comparing the predictions to data. This is clearly the procedure to follow for a small number of compelling candidate theories. Unfortunately, the resources required to implement this procedure grow almost linearly with the number of theories. Although broadly speaking there are currently only three models with internally consistent methods of electroweak symmetry breaking — supersymmetry, strong dynamics, and theories incorporating large extra dimensions — the number of specific models (and corresponding experimental signatures) is in the hundreds. Of these many specific models, at most one is a correct description of nature.

In these proceedings we describe an explicit prescription for searching for the physics responsible for stabilizing electroweak symmetry breaking, in a manner that relies only upon what we are sure we know about electroweak symmetry breaking: that its natural scale is on the order of the Higgs mass. When we wish to emphasize the generality of the approach, we say that it is quasi-model-independent, where the "quasi" refers to the fact that the correct model of electroweak symmetry breaking should become manifest at the scale of several hundred GeV.

New sources of physics will in general lead to an excess over the expected background in some final state. A general signature for new physics is therefore a region of variable space in which the probability for the background to fluctuate up to or above the number of observed events is small. Because the mass scale of electroweak symmetry breaking is larger than the mass scale of most standard model backgrounds, we expect this excess to populate regions of high transverse momentum (p_T). The method we will describe involves a systematic search for such excesses. Although motivated by the problem of electroweak symmetry breaking, this method is generally sensitive to any new high p_T physics.

2 SLEUTH

SLEUTH,[1,2,3,4] a quasi-model-independent prescription for searching for high p_T physics beyond the standard model, has three components: the definitions of physical objects and exclusive final states; the choice of variables relevant for each final state; and an algorithm that systematically hunts for an excess in the space of those variables, and quantifies the likelihood of any excess found. We consider each in turn.

2.1 *Final states*

The data are partitioned into exclusive final states using standard criteria that identify isolated and energetic electrons (e), muons (μ), and photons (γ), as well as jets (j), missing transverse energy ($\not\!\!E_T$), and the presence of W and Z bosons. We expect the first sign of new physics to appear in one of these final states, but which final state it will be is anyone's guess. We analyze each of these final states independently.

2.2 *Variables*

For each exclusive final state, we consider a small set of variables summarized in Table 1.[1,3]

Table 1: A quasi-model-independently motivated list of interesting variables for any final state. The set of variables to consider for any exclusive channel is the union of the variables in the second column for each row that pertains to that final state.

If the final state includes	then consider the variable
$\not\!\!E_T$	$\not\!\!E_T$
one or more charged leptons	$\sum p_T^{\ell}$
one or more electroweak bosons	$\sum p_T^{\gamma/W/Z}$
one or more jets	$\sum p_T^{j}$

2.3 *Algorithm*

The SLEUTH algorithm requires as input a data sample, a set of events modeling each background process i, and the number of background events $\hat{b}_i \pm \delta\hat{b}_i$ from each background process expected in the data sample. From these we determine the region $\mathcal{R}$ of greatest excess and quantify the degree $\mathcal{P}$ to which that excess is interesting. The algorithm itself, applied to each individual final state, consists of seven steps:

1. We construct a mapping from the d-dimensional variable space defined by Table 1 into the d-dimensional unit box (i.e., $[0,1]^d$) that flattens the total background distribution. We use this to map the data into the unit box.

2. We define a "region" R about a set of N data points to be the volume within the unit box closer to one of the data points in the set than to any of the other data points in the sample. The arrangement of data points themselves thus determines the regions. A region containing N data points is called an N-region.

3. Each region contains an expected number of background events $\hat{b}_R$, numerically equal to the volume of the region $\times$ the total number of background events expected, and an associated systematic error $\delta\hat{b}_R$, which varies within the unit box according to the systematic errors assigned to each contribution to the background estimate. We can therefore compute the probability p_N^R that the background in the region fluctuates up to or beyond the observed number of events. This probability is the first measure of the degree of interest of a particular region.

4. The rigorous definition of regions reduces the number of candidate regions from infinity to $\approx 2^{N_{\rm data}}$. Imposing explicit criteria on the regions that the algorithm is allowed to consider further reduces the number of candidate regions. We apply geometric criteria that favor high values in at least one dimension of the unit box, and we limit the number of events in a region to fifty. The number of remaining candidate regions is still sufficiently large that an exhaustive search is impractical, and a heuristic is employed to search for regions of excess. In the course of this search, the N-region $\mathcal{R}_N$ for which p_N^R is minimum is determined for each N, and $p_N = \min_R (p_N^R)$ is noted.

5. In any reasonably-sized data set, there will always be regions in which the probability for $\hat{b}_R$ to fluctuate up to or above the observed number of events is small. We determine the fraction P_N of *hypothetical similar experiments* (hse's) in which p_N found for the hse is smaller than p_N observed in the data by generating random events drawn from the background distribution and computing p_N by following steps 1–4.

6. We define P and $N_{\rm min}$ by $P = P_{N_{\rm min}} = \min_N (P_N)$, and identify $\mathcal{R} = \mathcal{R}_{N_{\rm min}}$ as the most interesting region in this final state.

7. We use a second ensemble of hse's to determine the fraction $\mathcal{P}$ of hse's in which P found in the hse is smaller than P observed in the data. The most important output of the algorithm is this single number $\mathcal{P}$, which may loosely be said to be the "fraction of hypothetical similar experiments in which you would see an excess as interesting as what you actually saw in the data." $\mathcal{P}$ takes on values between zero and unity, with values close to zero indicating a possible hint of new physics. The computation of $\mathcal{P}$ rigorously takes into account the many regions that have been considered within this final state.

The smallest $\mathcal{P}$ found in the many different final states considered ($\mathcal{P}_{\rm min}$) determines $\tilde{\mathcal{P}}$, the "fraction of *hypothetical similar experimental runs* (hser's) that would have produced an excess as interesting as actually observed in the data," where an hser consists of one hse for each final state considered. $\tilde{\mathcal{P}}$ is calculated by simulating an ensemble of hypothetical similar experimental runs, and noting the fraction of these hser's in which the smallest $\mathcal{P}$ found is smaller than the smallest $\mathcal{P}$ observed in the data. Because $\tilde{\mathcal{P}}$ depends only on the single final state that defines $\mathcal{P}_{\rm min}$, correlations among final states may be neglected in this calculation. Like $\mathcal{P}$, $\tilde{\mathcal{P}}$ takes on values between zero and unity, and the potential presence of new high p_T physics would be indicated by finding $\tilde{\mathcal{P}}$ to be small. The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ is that in computing $\tilde{\mathcal{P}}$ we account for the many final states that have been considered.

3 Results

SLEUTH's performance on representative signatures has been studied.[1,2,3,4] When ignorance of $t\bar{t}$ is feigned in the $e\mu X$ final states, we find $\mathcal{P}_{e\mu\not{E}_T 2j} = 1.9\sigma$ in DØ data, correctly suggesting the presence of $t\bar{t}$. Feigning ignorance of $t\bar{t}$ in the W+jets-like final states, we find $\mathcal{P}_{\min} > 3\sigma$ in 30% of an ensemble of mock experimental runs on the final states $W\,3j$, $W\,4j$, $W\,5j$, and $W\,6j$. Dedicated searches for the top quark in these channels[5] yield an excess of 2.75σ in $e\mu\not{E}_T\,2j$, 2.6σ in $W\,4j(nj)$ with no b-tag, and 3.6σ in $W\,3j(nj)$ with a b-tag. We see that SLEUTH performs surprisingly well despite being denied b-tagging information and any information of the properties of the signal it is supposed to be looking for. These and similar studies performed in other final states suggest that SLEUTH is sensitive to a variety of new physics signatures.

We have applied SLEUTH to over thirty exclusive final states at DØ, determining $\mathcal{P}$ for each. Upon taking into account the many final states (both populated and unpopulated) that are considered, we find $\tilde{\mathcal{P}}$=0.89, implying that 89% of an ensemble of hypothetical similar experimental runs would have produced a final state with a candidate signal more interesting than the most interesting observed in these data.

4 Conclusions

We have applied SLEUTH to search for new high p_T physics in data spanning over thirty exclusive final states collected by the DØ experiment during Run I of the Fermilab Tevatron. A quasi-model-independent, systematic search of these data has produced no evidence of physics beyond the standard model. The strategy proposed here may prove useful in future searches for new phenomena.

Acknowledgments

The author's attendance at Moriond was supported in part by Boaz Klima and funds from the National Science Foundation.

References

1. DØ Collaboration, B. Abbott *et al.*, Phys. Rev. D **62**, 92004 (2000).
2. DØ Collaboration, B. Abbott *et al.*, to be published in Phys. Rev. D, hep-ex/0011067 (2000).
3. DØ Collaboration, B. Abbott *et al.*, Phys. Rev. Lett., **86**, 3712 (2001).
4. B. Knuteson, Ph.D. thesis, University of California at Berkeley, 2000 (unpublished).
5. DØ Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **79**, 1203 (1997).