Search for New Physics Using Quaero: A General Interface to - D0 Event Data


To cite this version:


HAL Id: in2p3-00011036
http://hal.in2p3.fr/in2p3-00011036

Submitted on 29 Nov 2001

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
1 Universidad de Buenos Aires, Buenos Aires, Argentina
2 LAFEX, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, Brazil
3 Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
4 Institute of High Energy Physics, Beijing, People's Republic of China
5 Universidad de los Andes, Bogotá, Colombia
6 Charles University, Center for Particle Physics, Prague, Czech Republic
7 Institute of Physics, Academy of Sciences, Center for Particle Physics, Prague, Czech Republic
8 Universidad San Francisco de Quito, Quito, Ecuador
9 Instituto des Sciences Nucléaires, IN2P3-CNRS, Université de Grenoble 1, Grenoble, France
10 CPPM, IN2P3-CNRS, Université de la Méditerranée, Marseille, France
11 Laboratoire de l'Accélérateur Linéaire, IN2P3-CNRS, Orsay, France
12 LPNHE, Universités Paris VI and VII, IN2P3-CNRS, Paris, France
13 DAPNIA/Service de Physique des Particules, CEA, Saclay, France
14 Universität Mainz, Institut für Physik, Mainz, Germany
15 Panjab University, Chandigarh, India
16 Delhi University, Delhi, India
17 Tata Institute of Fundamental Research, Mumbai, India
18 Seoul National University, Seoul, Korea
19 CINVESTAV, Mexico City, Mexico
20 FOM-Institute NIKHEF and University of Amsterdam/NIKHEF, Amsterdam, The Netherlands
21 University of Nijmegen/NIKHEF, Nijmegen, The Netherlands
22 Institute of Nuclear Physics, Kraków, Poland
23 Joint Institute for Nuclear Research, Dubna, Russia
24 Institute for Theoretical and Experimental Physics, Moscow, Russia
25 Moscow State University, Moscow, Russia
26 Institute for High Energy Physics, Protvino, Russia
27 Lancaster University, Lancaster, United Kingdom
28 Imperial College, London, United Kingdom
29 University of Arizona, Tucson, Arizona 85720
30 Lawrence Berkeley National Laboratory and University of California, Berkeley, California 94720
31 University of California, Davis, California 95616
32 California State University, Fresno, California 93740
33 University of California, Irvine, California 92697
34 University of California, Riverside, California 92521
35 Florida State University, Tallahassee, Florida 32306
36 University of Hawaii, Honolulu, Hawaii 96822
37 Fermi National Accelerator Laboratory, Batavia, Illinois 60510
38 University of Illinois at Chicago, Chicago, Illinois 60607
39 Northern Illinois University, DeKalb, Illinois 60115
40 Northwestern University, Evanston, Illinois 60208
41 Indiana University, Bloomington, Indiana 47405
42 University of Notre Dame, Notre Dame, Indiana 46556
43 Iowa State University, Ames, Iowa 50011
44 University of Kansas, Lawrence, Kansas 66045
45 Kansas State University, Manhattan, Kansas 66506
46 Louisiana Tech University, Ruston, Louisiana 71272
47 University of Maryland, College Park, Maryland 20742
48 Boston University, Boston, Massachusetts 02215
49 Northeastern University, Boston, Massachusetts 02115
50 University of Michigan, Ann Arbor, Michigan 48109
51 Michigan State University, East Lansing, Michigan 48824
52 University of Nebraska, Lincoln, Nebraska 68588
53 Columbia University, New York, New York 10027
54 University of Rochester, Rochester, New York 14627
55 State University of New York at Stony Brook, New York 11794
56 Brookhaven National Laboratory, Upton, New York 11973
57 Langston University, Langston, Oklahoma 73050
58 University of Oklahoma, Norman, Oklahoma 73019
59 Brown University, Providence, Rhode Island 02912
60 University of Texas, Arlington, Texas 76019
Abstract

We describe QUAERO, a method that i) enables the automatic optimization of searches for physics beyond the standard model, and ii) provides a mechanism for making high energy collider data generally available. We apply QUAERO to searches for standard model $WW$, $ZZ$, and $tt$ production, and to searches for these objects produced through a new heavy resonance. Through this interface, we make three data sets collected by the DØ experiment at $\sqrt{s} = 1.8$ TeV publicly available.
It is generally recognized that the standard model, a successful description of the fundamental particles and their interactions, must be incomplete. Models that extend the standard model often predict rich phenomenology at the scale of a few hundred GeV, an energy regime accessible to the Fermilab Tevatron. Due in part to the complexity of the apparatus required to test models at such large energies, experimental responses to these ideas have not kept pace. Any technique that reduces the time required to test a particular candidate theory would allow more such theories to be tested, reducing the possibility that the data contain overlooked evidence for new physics.

Once data are collected and the backgrounds have been understood, the testing of any specific model in principle follows a well-defined procedure. In practice, this process has been far from automatic. Even when the basic selection criteria and background estimates are taken from a previous analysis, the reinterpretation of the data in the context of a new model often requires a substantial length of time.

Ideally, the data should be “published” in such a way that others in the community can easily use those data to test a variety of models. The publishing of experimental distributions in journals allows this to occur at some level, but an effective publishing of a multidimensional data set has, to our knowledge, not yet been accomplished by a large particle physics experiment. The problem appears to be that such data are context-specific, requiring detailed knowledge of the complexities of the apparatus. This knowledge must somehow be incorporated either into the data or into whatever tool the non-expert would use to analyze those data.

Many data samples and backgrounds have been defined in the context of SLEUTH, a quasi-model-independent search strategy for new high $p_T$ physics that has been applied to a number of exclusive final states in the data collected by the DØ detector during 1992–1996 in Run I of the Fermilab Tevatron. In this Letter we describe a tool (QUAERO) that automatically optimizes an analysis for a particular signature, using these samples and standard model backgrounds. SLEUTH and QUEREO are complementary approaches to searches for new phenomena, enabling analyses that are both general (SLEUTH) and focused (QUAERO). We demonstrate the use of QUAERO in eleven separate searches: standard model $WW$ and $ZZ$ production; standard model $t\bar{t}$ production with leptonic and semileptonic decays; resonant $WW$, $ZZ$, $WZ$, and $t\bar{t}$ production; associated Higgs boson production; and pair production of first generation scalar leptoquarks. The data described here are accessible through QUAERO on the World Wide Web, for general use by the particle physics community.

The signals predicted by most theories of physics beyond the standard model involve an increased number of predicted events in some region of an appropriate variable space. In this case the optimization of the analysis can be understood as the selection of the region in this variable space that minimizes $\sigma^{95\%}$, the expected 95% confidence level (CL) upper limit on the cross section of the signal in question, assuming the data contain no signal. The optimization algorithm consists of a few simple steps:

(i) Kernel density estimation is used to estimate the probability distributions $p(x|s)$ and $p(x|b)$ for the signal and background samples in a low-dimensional variable space $\mathcal{V}$, where $x \in \mathcal{V}$. The signal sample is contained in a Monte Carlo file provided as input to QUAERO. The background sample is constructed from all known standard model and instrumental sources.

(ii) A discriminant function $D(x)$ is defined by

$$D(x) = \frac{p(x|s)}{p(x|s) + p(x|b)}.$$  \hspace{1cm} (1)

The semi-positive-definiteness of $p(x|s)$ and $p(x|b)$ restricts $D(x)$ to the interval $[0, 1]$ for all $x$.

(iii) The sensitivity $S$ of a particular threshold $D_{cut}$ on the discriminant function is defined as the reciprocal of $\sigma^{95\%}$. $D_{cut}$ is chosen to maximize $S$.

(iv) The region of variable space having $D(x) > D_{cut}$ is used to determine the actual 95% CL cross section upper limit $\sigma^{95\%}$.

When provided with a signal model and a choice of variables $\mathcal{V}$, QUAERO uses this algorithm and DØ Run I data to compute an upper limit on the cross section of the signal. Instructions for use are available from the QUAREO web site.

Table II shows the data available within QUAREO, and Table III summarizes the backgrounds. These data and their backgrounds are described in more detail in Ref. [3]. The final states are inclusive, with many events containing one or more additional jets. Kolmogorov-Smirnov tests have been used to demonstrate agreement between data and the expected backgrounds in many distributions. The fraction of events with true final state objects satisfying the cuts shown that satisfy these cuts after reconstruction is given as an “identification” efficiency ($\epsilon_{ID}$). Because electrons are more accurately measured and more efficiently identified than muons in the DØ detector, the corresponding muon channels $\mu E_T 2j$ and $\mu \mu 2j$ have been excluded from these data.

To check standard model results, we remove $WW$ and $ZZ$ production from the background estimate and search (i) for standard model $WW$ production in the space defined by the transverse momentum of the electron ($p_T^e$) and missing transverse energy ($E_T$) in the final state $e\mu E_T$, and (ii) for standard model $ZZ$ production in the space defined by the invariant mass of the two electrons ($m_{ee}$) and two jets ($m_{jj}$) in the final state $ee 2j$. Removing $t\bar{t}$ production from the background estimate, we search for this process (iii) in the final state $e E_T 4j$ using the two variables laboratory aplanarity ($A$) and $\sum p_T^j$. The signals predicted by most theories of physics beyond the standard model involve an increased number of predicted events in some region of an appropriate variable space. In this case the optimization of the analysis can be understood as the selection of the region in this variable space that minimizes $\sigma^{95\%}$, the expected 95% confidence level (CL) upper limit on the cross section of the signal in question, assuming the data contain no signal. The optimization algorithm consists of a few simple steps:

(i) Kernel density estimation is used to estimate the probability distributions $p(x|s)$ and $p(x|b)$ for the signal and background samples in a low-dimensional variable space $\mathcal{V}$, where $x \in \mathcal{V}$. The signal sample is contained in a Monte Carlo file provided as input to QUAERO. The background sample is constructed from all known standard model and instrumental sources.

(ii) A discriminant function $D(x)$ is defined by

$$D(x) = \frac{p(x|s)}{p(x|s) + p(x|b)}.$$  \hspace{1cm} (1)

The semi-positive-definiteness of $p(x|s)$ and $p(x|b)$ restricts $D(x)$ to the interval $[0, 1]$ for all $x$.

(iii) The sensitivity $S$ of a particular threshold $D_{cut}$ on the discriminant function is defined as the reciprocal of $\sigma^{95\%}$. $D_{cut}$ is chosen to maximize $S$.

(iv) The region of variable space having $D(x) > D_{cut}$ is used to determine the actual 95% CL cross section upper limit $\sigma^{95\%}$.

When provided with a signal model and a choice of variables $\mathcal{V}$, QUAERO uses this algorithm and DØ Run I data to compute an upper limit on the cross section of the signal. Instructions for use are available from the QUAREO web site.

Table II shows the data available within QUAREO, and Table III summarizes the backgrounds. These data and their backgrounds are described in more detail in Ref. [3]. The final states are inclusive, with many events containing one or more additional jets. Kolmogorov-Smirnov tests have been used to demonstrate agreement between data and the expected backgrounds in many distributions. The fraction of events with true final state objects satisfying the cuts shown that satisfy these cuts after reconstruction is given as an “identification” efficiency ($\epsilon_{ID}$). Because electrons are more accurately measured and more efficiently identified than muons in the DØ detector, the corresponding muon channels $\mu E_T 2j$ and $\mu \mu 2j$ have been excluded from these data.

To check standard model results, we remove $WW$ and $ZZ$ production from the background estimate and search (i) for standard model $WW$ production in the space defined by the transverse momentum of the electron ($p_T^e$) and missing transverse energy ($E_T$) in the final state $e\mu E_T$, and (ii) for standard model $ZZ$ production in the space defined by the invariant mass of the two electrons ($m_{ee}$) and two jets ($m_{jj}$) in the final state $ee 2j$. Removing $t\bar{t}$ production from the background estimate, we search for this process (iii) in the final state $e E_T 4j$ using the two variables laboratory aplanarity ($A$) and $\sum p_T^j$. The signals predicted by most theories of physics beyond the standard model involve an increased number of predicted events in some region of an appropriate variable space. In this case the optimization of the analysis can be understood as the selection of the region in this variable space that minimizes $\sigma^{95\%}$, the expected 95% confidence level (CL) upper limit on the cross section of the signal in question, assuming the data contain no signal. The optimization algorithm consists of a few simple steps:

(i) Kernel density estimation is used to estimate the probability distributions $p(x|s)$ and $p(x|b)$ for the signal and background samples in a low-dimensional variable space $\mathcal{V}$, where $x \in \mathcal{V}$. The signal sample is contained in a Monte Carlo file provided as input to QUAERO. The background sample is constructed from all known standard model and instrumental sources.

(ii) A discriminant function $D(x)$ is defined by

$$D(x) = \frac{p(x|s)}{p(x|s) + p(x|b)}.$$  \hspace{1cm} (1)

The semi-positive-definiteness of $p(x|s)$ and $p(x|b)$ restricts $D(x)$ to the interval $[0, 1]$ for all $x$.

(iii) The sensitivity $S$ of a particular threshold $D_{cut}$ on the discriminant function is defined as the reciprocal of $\sigma^{95\%}$. $D_{cut}$ is chosen to maximize $S$.

(iv) The region of variable space having $D(x) > D_{cut}$ is used to determine the actual 95% CL cross section upper limit $\sigma^{95\%}$.
and (iv) in the final state $e\mu E_T 2j$, using the two variables $p_T^e$ and $\sum p_T^j$, assuming a top quark mass of 175 GeV.

Including all standard model processes in the background estimate, we look for evidence of new heavy resonances. We search (v) for resonant $WW$ production in the final state $e\mu E_T 2j$, using the single variable $m_{e\mu jj}$ after constraining $m_{e\mu}$ and $m_{jj}$ to $M_W$, and (vi) for resonant $ZZ$ production in the final state $e\mu E_T 2j$, using the variable $m_{eejj}$ after constraining $m_{jj}$ to $M_Z$. In both cases we remove events that cannot be so constrained. To obtain a specific signal prediction, we assume that the resonance behaves like a standard model Higgs boson in its couplings to the $W$ and $Z$ bosons. Constraining $m_{e\mu}$ to $M_W$ and $m_{jj}$ to $M_Z$, we use the quality of the fit and $m_{e\mu jj}$ to search (vii) for a massive $W'$ boson in the extended gauge model of Ref. [13]. Using $m_{e\mu jj}$ after constraining $m_{e\mu}$ to $M_W$, we search (viii) for a massive narrow $Z'$ resonance with $Z$-like couplings decaying to $t\bar{t} \to W^+bW^-\bar{b} \to e\mu 4j$.

Non-resonant new phenomena are also considered. The variables $m_{jj}$ and either $m_T^e$ or $m_{ee}$ are used to search for a light Higgs boson produced (ix) in association with a $W$ boson, and (x) in association with a $Z$ boson. Finally, we search (xi) for first generation scalar leptoquarks with mass 225 GeV in the final state $ee 2j$ using $m_{ee}$ and $S_T$, the summed scalar transverse momentum of all electrons and jets in the event. The numerical results of these searches are listed in Table III. Figures 1 and 2 present plots of the signal density, background density, and selected region in the variables considered.

<table>
<thead>
<tr>
<th>Final State</th>
<th>Selection criteria</th>
<th>$\epsilon_{FD}$</th>
<th>$\int L , dt$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e\mu$</td>
<td>$p_T^{e\mu} &gt; 15$ GeV, $</td>
<td>\eta^{e\mu}</td>
<td>&lt; 1.7$</td>
</tr>
<tr>
<td>$e\mu E_T 2j$</td>
<td>$p_T^{e\mu} &gt; 20$ GeV, $E_T &gt; 30$ GeV, $p_T^{e\mu} &gt; 40$ GeV</td>
<td>0.61</td>
<td>$115 \pm 6$ pb$^{-1}$</td>
</tr>
<tr>
<td>$ee 2j$</td>
<td>$p_T^{ee}, \sum p_T^j &gt; 20$ GeV</td>
<td>0.70</td>
<td>$123 \pm 7$ pb$^{-1}$</td>
</tr>
</tbody>
</table>

**TABLE I.** A summary of the data available within QUAERO, including the selection cuts applied and the efficiency of identification requirements. The final states are inclusive, with many events containing one or more additional jets. Reconstructed jets satisfy $p_T^j > 15$ GeV and $|\eta^j| < 2.5$, and reconstructed electrons satisfy $p_T^e > 15$ GeV and $(|\eta^e| < 1.1 \text{ or } 1.5 < |\eta^e| < 2.5)$, where $\eta$ is the pseudorapidity measured from the center of the detector.

<table>
<thead>
<tr>
<th>Final State</th>
<th>Standard model backgrounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e\mu$</td>
<td>data data $W$ $Z$ $VV$ $t\bar{t}$</td>
</tr>
<tr>
<td>$e\mu E_T 2j$</td>
<td>data $VECBO$ $PYTHIA$ $HERWIG$</td>
</tr>
<tr>
<td>$ee 2j$</td>
<td>data $PYTHIA$ $PYTHIA$</td>
</tr>
</tbody>
</table>

**TABLE II.** Standard model backgrounds (often produced with accompanying jets) to the final states considered. $V$ denotes $WW$, $WZ$, and $ZZ$; “data” indicates backgrounds from jets misidentified as electrons estimated using data. Monte Carlo programs ($ISAET$, $PYTHIA$, $HERWIG$, and $VECBO$) are used to estimate several sources of background.

We note slight indications of excess in the searches for $t\bar{t} \to eE_T 4j$ and $t\bar{t} \to e\mu E_T 2j$ (corresponding to cross section $\times$ branching fraction of $\sigma \times B = 0.39^{+0.21}_{-0.19}$ pb and $0.14^{+0.09}_{-0.05}$ pb) that are consistent with our measured $t\bar{t}$ production cross section of $5.5 \pm 1.8$ pb [14] and known $W$ boson branching fractions. Observing no compelling excess in any of these processes, limits on $\sigma \times B$ are determined at the 95% CL. As expected, we find these data insensitive to standard model $ZZ$ production (with predicted $\sigma \times B \approx 0.05$ pb), and to associated Higgs boson production (with predicted $\sigma \times B \lesssim 0.01$ pb). As a check of the method, QUAERO almost exactly duplicates a previous search for $LQ_{225}LQ_{225} \to ee 2j$ [13].

QUAERO is a method both for automatically optimizing searches for new physics and for allowing DO to make a subset of its data available for general use. In this Letter we have outlined the algorithm used in QUAERO, and we have described the final states currently available for analysis using this method. QUAERO’s performance on several examples, including both standard model and resonant $WW$, $ZZ$, and $t\bar{t}$ production, has been demonstrated. The limits obtained are comparable to those from previous searches at hadron colliders, and the search for $W' \to WZ$ is the first of its kind. This tool should in-
FIG. 1. The background density (a), signal density (b), and selected region (shaded) (c) determined by quaero for the standard model processes discussed in the text. From top to bottom the signals are: \(WW \rightarrow e\mu E_T\), \(ZZ \rightarrow ee2j\), \(t\bar{t} \rightarrow e\mu E_T 4j\), and \(t\bar{t} \rightarrow e\mu E_T 2j\). The dots in the plots in the rightmost column represent events observed in the data.

crease the facility with which new models may be tested in the future.

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the Department of Energy and National Science Foundation (USA), Commissariat à L’Energie Atomique and CNRS/Institut National de Physique Nucléaire et de Physique des Particules (France), Ministry for Science and Technology and Ministry for Atomic Energy (Russia), CAPES and CNPq (Brazil), Departments of Atomic Energy and Science and Education (India), Colciencias (Colombia), CONACyT (Mexico), Ministry of Education and KOSEF (Korea), CONICET and UBACyT (Argentina), The Foundation for Fundamental Research on Matter (The Netherlands), PPARC (United Kingdom), Ministry of Education (Czech Republic), and the A.P. Sloan Foundation.

Visitor from University of Zurich, Zurich, Switzerland.

[8] \(\sigma^{95\%}\) is a Bayesian limit, computed assuming a flat prior.
[9] F. Paige and S. Protopopescu, BNL Report No. 38304,
1986 (unpublished); we used v7.22.


