

JetWeb and HepData

Global tuning of MC event generators

Andy Buckley

Institute for Particle Physics Phenomenology
Durham University, UK

CHEP06, Mumbai 2006-02-13

Outline

Introduction

Event generators and tuning

JetWeb — a system for global MC tuning

HepData — the HEP reaction database

CEDAR: Collaborative e-Science Data Analysis Resource

JetWeb and HepData are components of **CEDAR**:

- ▶ **JetWeb**: global tuning of Monte Carlo generator parameters
- ▶ **HepData**: archival of published experimental data
- ▶ HepForge: development environment for HEP software
- ▶ HepCode: centralised repository of pheno code/programs
- ▶ HepML: set of XML data formats for data sets and MC config



www.cedar.ac.uk

Event generators: motivations

Physics motivations of event generators:

- ▶ Provide bg estimates for new physics searches
- ▶ Simulating new physics processes
- ▶ Detector calibration
- ▶ Designing new experimental facilities

Event generators: necessary inputs

(no pain, no gain)

But some inputs are needed:

- ▶ Matrix elements for hard process
- ▶ **Perturbative parton showers (matched to MEs)**
- ▶ **Phenomenological models for non-perturbative physics**

These last two contain parameters which must be tuned to experimental data.

Event generators: tuneable parameters

- ▶ Parton density functions (PDFs)
- ▶ Matrix elements (kinematics, phase space, coupling/scale)
- ▶ Parton showers (and matching to MEs)
- ▶ Threshold effects in heavy flavour production
- ▶ Fragmentation functions
- ▶ Hadronisation (string params., multiplicities, strangeness production. . .)

Large number of params: nonsensical tunings dominate parameter space

Data for event generator tuning

Useful constraints from \sim all data:

- ▶ **Tevatron:** jets, photons, EW, charm, beauty
- ▶ **HERA:** DIS inclusive, photo-production, jets, c, b
- ▶ **Fixed target DIS**
- ▶ **LEP:** annihilation, γe^- , $\gamma\gamma$
- ▶ **SPS:** min bias, inclusive hadrons

Caveats

There are always caveats, especially in high dimensional problems!

- ▶ *Global* tuning to data required — selective tunings are dangerous
- ▶ Expect *sets* of reasonable configurations, rather than one “ultimate” tuning
- ▶ Any tuning depends on your definition of “best”. Both in terms of what phenomenological features a given expt. considers important and in terms of the fitting measure used.

Introducing JetWeb

JetWeb is a centralised system for global MC tunings

- ▶ Written in Java, run on Apache+Tomcat framework
- ▶ Database of global data fit qualities for MC models (generator & params)
- ▶ Web interface to view fit qualities and plots
- ▶ Web interface to request generation of a particular model if not (sufficient) in db
- ▶ Comparisons via a set of routines, each corresponding to a published paper



JetWeb in action (1)

Welcome screen

NB. This is the *frozen* copy of JetWeb. . .

JetWebWelcome - Galeon

File Edit View Tab Settings Go Bookmarks Tools Help

Back / Stop 100 http://jetweb.hep.ucl.ac.uk/welcome.html

Google Dictionary [Merkuraw](#) [fpm](#)

Downloads Software Hardware Developers Help Search

simulations

- [HERWIG](#)
- [PYTHIA](#)

experiments

- [HERA\(H1,ZEUS\)](#)
- [LEP\(OPAL\)](#)
- [Tevatron\(CDF, D0\)](#)

best fits, all data

- [HERWIG](#)
- [PYTHIA](#)

summaries, all fits

- [HERWIG](#)
- [PYTHIA](#)

JetWeb

Automated Data Comparisons for High Energy Physics

Search the DataBase [Maintenance](#)

Searches Prepared Earlier...

- [HERWIG fragmentation parameters \(CLMAX,PSPLT\)](#)
- [Multiparton interactions/underlying event](#)
- [Intrinsic KT photon/proton](#)
- [PYTHIA parton showers PARF67](#)
- [Parton Distribution Functions in Photon](#)

If you do use any results from here, please reference the URL

JetWeb in action (2)

Searching

NB. This is the *frozen* copy of JetWeb. . .

The screenshot shows a browser window titled "Search the JetWeb DataBase - Galeon". The address bar shows the URL "http://jetweb.hep.uct.ac.uk/JetWeb/JWSearch". The page content is on a yellow background and includes a search bar, a "Welcome" message, and a section for "Common parameters".

Search the JetWeb DataBase

Welcome

Get results Sort results by: Only show me results with data from:

Common parameters

Generator herwig <input type="checkbox"/> pythia <input type="checkbox"/>	Version v6.400 <input type="checkbox"/> v6.206 <input type="checkbox"/> v6.100 <input type="checkbox"/>	Minimum transverse momentum of hard scatters (GeV) <input type="text"/>	Underlying event model(Integer 0-5) <input type="text"/> More info	Photon PDF GRVLO <input type="checkbox"/> Ss1D <input type="checkbox"/> Ss2D <input type="checkbox"/> WHIT2 <input type="checkbox"/>	Proton PDF GRVLO <input type="checkbox"/> CTEQ5L <input type="checkbox"/> CTEQ4L <input type="checkbox"/>	Intrinsic transverse momentum in photon (GeV) <input type="text"/>	Intrinsic transverse momentum in proton (GeV) <input type="text"/>
---	--	---	--	---	--	--	--

Java hztool fitter, J. Butterworth, S. Butterworth

JetWeb in action (3)

Fit list

NB. This is the *frozen* copy of JetWeb. . .

Results sorted by Fit (All ET) - Galeon

File Edit View Tab Settings Go Bookmarks Tools Help

Back / Stop 100 http://jetweb.hep.ucl.ac.uk/JetWeb/JWSearch

Google Dictionary

Downloads Software Hardware Developers Help Search

Results sorted by Fit (All ET)

Last updated 05-Oct-2002 at 12:24:15

HERWIG v6.100 run 30/09/2002 PDFs: Proton **GRVLO** Proton **CTEQSL** PTMIN 3 0GeV UE **JIMMY** Proton kt.0.0 Proton kt.0.0 Scale **1.55** Model ID **97** : [Plots etc](#)

Combined: Chi2/Dof: High ET: **1.47** Low ET: **2.41** Jet Shape: **16.63** Charm: **8.13** All ET: **2.1**

HERA Lumi 5.0(+) pb⁻¹ Chi2/Dof: High ET: 1.47 Low ET: 2.12 Jet Shape: 7.9 Charm: 8.13 All ET: 1.84

LEP Lumi 400.0(+) pb⁻¹ Chi2/Dof: High ET: ? Low ET: 3.52 Jet Shape: 3.73 Charm: ? All ET: 3.52

Tevatron Lumi 0.00003(+) pb⁻¹ Chi2/Dof: High ET: ? Low ET: ? Jet Shape: 27.87 Charm: ? All ET: 2.57

HERWIG v6.100 run 30/09/2002 PDFs: Proton **WHIT2** Proton **CTEQSL** PTMIN 3 0GeV UE **JIMMY** Proton kt.0.0 Proton kt.0.0 Scale **1.65** Model ID **241** : [Plots etc](#)

Combined: Chi2/Dof: High ET: **1.9** Low ET: **2.46** Jet Shape: **14.33** Charm: **3.05** All ET: **2.23**

HERA Lumi -0.0(+) pb⁻¹ Chi2/Dof: High ET: 1.9 Low ET: 2.22 Jet Shape: 1.15 Charm: 3.05 All ET: 2.08

LEP Lumi 300.0 pb⁻¹ Chi2/Dof: High ET: ? Low ET: 3.38 Jet Shape: 9.16 Charm: ? All ET: 3.38

Tevatron Lumi 0.00003(+) pb⁻¹ Chi2/Dof: High ET: ? Low ET: ? Jet Shape: 27.87 Charm: ? All ET: 1.97

HERWIG v6.100 run 30/09/2002 PDFs: Proton **SaS2D** Proton **CTEQSL** PTMIN 3 0GeV UE **JIMMY** Proton kt.0.0 Proton kt.0.0 Scale **1.55** Model ID **76** : [Plots etc](#)

Combined: Chi2/Dof: High ET: **1.92** Low ET: **2.64** Jet Shape: **19.29** Charm: **13.54** All ET: **2.39**

HERA Lumi 5.0(+) pb⁻¹ Chi2/Dof: High ET: 1.92 Low ET: 2.39 Jet Shape: 12.02 Charm: 13.54 All ET: 2.19

LEP Lumi 200.0 pb⁻¹ Chi2/Dof: High ET: ? Low ET: 3.57 Jet Shape: 11.84 Charm: ? All ET: 3.57

Tevatron Lumi 0.00003(+) pb⁻¹ Chi2/Dof: High ET: ? Low ET: ? Jet Shape: 27.87 Charm: ? All ET: 2.57

Done.

JetWeb in action (4)

Fit details

NB. This is the *frozen* copy of JetWeb...

JetWeb Fit No:269
HERWIG v6.100 run

Date of last fit 08/11/2002
[Examine the fitted papers](#)
[HERA fit](#)
[LEP fit](#)
[Tevatron fit](#)

Request higher statistics for
 Request similar data
 Search for similar data

Compared this for all fitted experiments: χ^2/Dof at an overall scale factor of 1.55

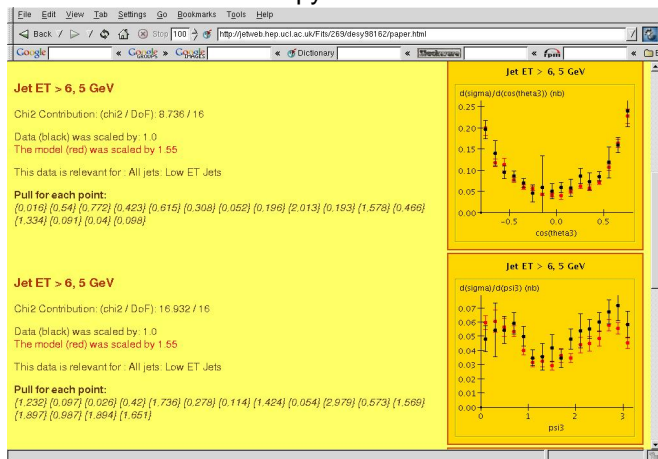
- High ET: **1.5952063**
- Low ET: **2.0411307**
- Jet Shape: **4.0698969**
- Charm: **8.0909271**
- All ET: **2.1205643**

Parton distribution functions: Photon **GRVLO** Proton **CTEQ5L**
 PTMIN (Minimum transverse momentum for hard scatters) **3GeV**
 Underlying Event Model **JIMMY**
 Intrinsic KT in the photon is 0.0
 Intrinsic KT in the proton is 0.0
 Parton shower cutoff is 2.5
 Photon radius: 1.0
 Proton radius: 3.0
 PHad: 300
 Fragmentation parameters CLMAX,PSPLT(1),(2): 3.35,1.1
 PRSOF: 0
 QCCLAM: 0.18

JetWeb in action (5)

Paper view (1)

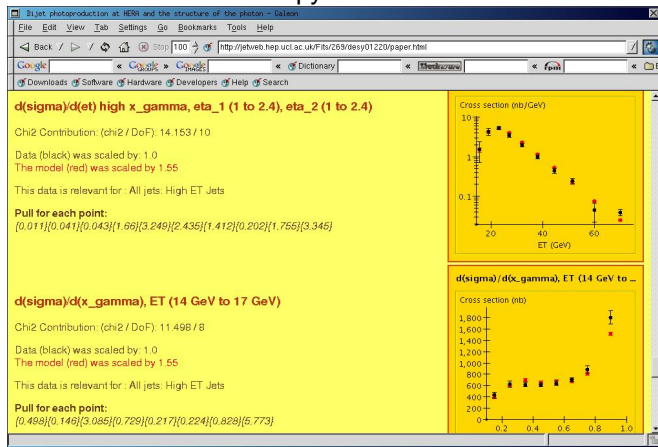
NB. This is the *frozen* copy of JetWeb...



JetWeb in action (6)

Paper view (2)

NB. This is the *frozen* copy of JetWeb. . .



HzTool

JetWeb is only the front-end, distribution and comparison system. Reproduction of experimental analyses done by **HzTool**

- ▶ HzTool is a library of Fortran routines for reproducing observables
- ▶ Typically one HzTool routine per paper
- ▶ Typically written by paper authors
- ▶ Generator independent (at least in principle!)
- ▶ Due to HzTool's background, most routines are from HERA (at present)

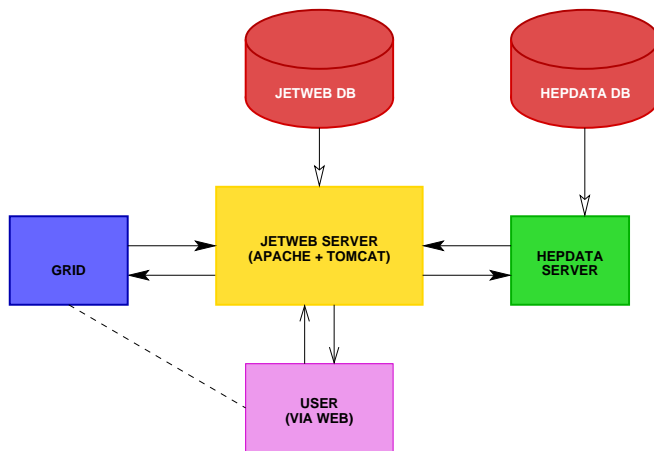
JetWeb now accesses HzTool via a 'HzSteer' package into which most (all?) generator dependencies have been moved.

JetWeb enhancements

CEDAR is re-writing / enhancing JetWeb in several ways:

- ▶ Removal of experimental data from JetWeb database: now connecting directly to new HepData system. Single source for scattering data
- ▶ Framework re-factor to make it easier to add new generators/modules (the 'build one to throw away — you will anyway' rule in action!)
- ▶ Plans to use Grid authentication and job distribution
- ▶ New histogramming modules using the AIDA interfaces
- ▶ (Scanning parameter space, re-weighted fits. . .)

(New) JetWeb behind the scenes



Rivet

- ▶ Rivet is a C++ re-implementation of HzTool
- ▶ Based on concept of ‘projections’: maps from final state event record to observables
- ▶ Both a collection of tools (jet algorithms etc.) and a library of analyses
- ▶ Like HzTool, an external steering package will be used to pass params to generators
- ▶ **Ben Waugh will talk about HzTool and Rivet in detail on Thursday**

Developers: AB, Jon Butterworth, Leif Lonnblad, Ben Waugh, James Monk

Introducing (legacy) HepData

- ▶ Searchable ~30 years' worth of scattering data, PDFs etc.
- ▶ Historically stored in hierarchical BDMS database: very inflexible and little modern support
- ▶ Data available as text files, PAW kumacs or GIF images
- ▶ Legacy db stores pretty much everything as a string
- ▶ Adding records requires writing a Fortran routine
- ▶ Mirrors to SPIRES, PDG info

HepData enhancements

- ▶ Migration to relational SQL database with Java servlet front-end, sharing object model components with JetWeb
- ▶ Database sanitising e.g. axis-level properties, uniform units, measurement classes. . .
- ▶ Using “HepML” and AIDA interfaces: XSLT transformations to many data formats possible
- ▶ Will use Grid authentication for expts. to submit data directly (modulo sanity checking)

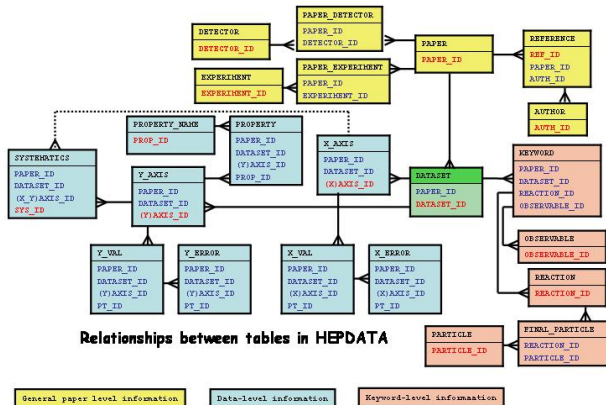


PostgreSQL



New HepData db relational structure

Note this is not final — in fact users shouldn't even know about it!



A brief aside on “HepML”

- ▶ XML representation for generator configs and data sets
- ▶ Defined by a set of XML schema docs in the `http://www.cedar.ac.uk/xml/ namespace`
- ▶ Will also contain Python API and Java XSLT transformer classes using HD object model
- ▶ Intended for use by JetWeb etc., also by external projects (e.g. Professor)
- ▶ Current sub-schemas for HepData data I/O and for generator configurations
- ▶ We'd like to work with generator authors (and MCDB) to make HepML a common interfacing/config format

HepML fragment (1)

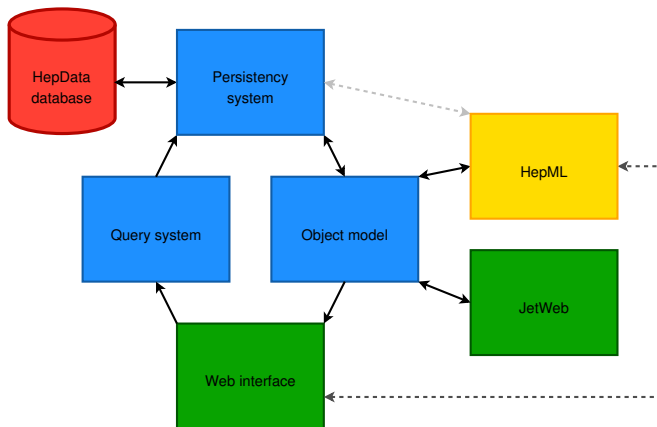
```
<?xml version="1.0" encoding="UTF-8"?>
<hepml xmlns="http://www.cedar.ac.uk/xml/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.cedar.ac.uk/xml/schemas hepdata.xsd"
  <data timestamp="2006-02-12 00.14.50">
    <paper irn="3326047" paperId="3543">
      <dataset datasetId="1">
        <xaxis description="SQRT(S) IN GEV" xaxisId="1">
          <points>
            <point pointId="1" value="0.7726" />
            <point pointId="2" value="0.7776" />
            <point pointId="3" value="0.7801" />
            <point pointId="4" value="0.7826" />
            <point pointId="5" value="0.7851" />
            <point pointId="6" value="0.7876" />
            <point pointId="7" value="0.8026" />
          </points>
        </xaxis>
      </dataset>
    </paper>
  </data>
</hepml>
```

...

HepML fragment (2)

```
...
<yaxis description="SIG(Q=HAD) IN MUB" yaxisId="1">
  <points>
    <point pointId="1" value="0.269" />
    <point pointId="2" value="0.74" />
    ...
    <point pointId="7" value="0.065" />
  </points>
  <axiserror type="PCT_SYS" plus="11.0" minus="11.0"
    description="FROM NUCLEAR ABSORPTION CORRECTION" />
  ...
  <pointerrors type="NULL">
    <pointerror plus="0.096" minus="0.096" pointId="1" />
    <pointerror plus="0.192" minus="0.192" pointId="2" />
    ...
    <pointerror plus="0.036" minus="0.036" pointId="7" />
  </pointerrors>
</yaxis>
```

HepData behind the scenes



First HepData demos (1)

Query interface (breaking the rules — spot the SQL!)

HepData - CEDAR - Mozilla Firefox

http://hepdata.cedar.ac.uk/serve/hepdata-test/Query

CEDAR

• Home

HepData query testing

Reaction:

Observable:

Params:

- reac = null
- obs = null

```
SELECT DISTINCT kr.paper_id paperId, kr.ds_id dsId, o.description obsDesc, r.description reacDesc FROM (keyword kr
JOIN keyword ko USING (paper_id,ds_id)) JOIN observable o ON o.observable_id=ko.observable_id JOIN reaction r ON
r.reaction_id=kr.reaction_id JOIN final_particle fp ON fp.reaction_id=r.reaction_id JOIN particle p ON
p.particle_id=fp.particle_id WHERE (ko.observable_id IS NOT NULL and kr.reaction_id IS NOT NULL) LIMIT 10;
```

- 1-1 : SIG / GAMMA P --> ETA P Dataset ID:1
- 1-2 : DSIG/DOMEGA / GAMMA P --> ETA P Dataset ID:2
- 1-3 : DSIG/DOMEGA / GAMMA P --> ETA P Dataset ID:3
- 1-4 : DSIG/DOMEGA / GAMMA P --> ETA P Dataset ID:4
- 1-5 : DSIG/DOMEGA / GAMMA P --> ETA P Dataset ID:5

Done

First HepData demos (2)

HepData \rightarrow HepML $\xrightarrow{\text{XSL}}$ HTML

The screenshot shows a Mozilla Firefox browser window with the address bar at `http://hepdata.cedar.ac.uk/server/hepdata-test/XSL`. The page title is "HepData XSL test". The main content area displays "Reformatted HepML for HepData:" followed by "Paper: 3326047" and "Dataset:". Below this is a table with three columns: "SQRT(S) IN GEV", "SIG(Q=HAD) IN MUB", and "SIG(Q=HAD)/SIG(Q=MU)". Each column has a header row with error bars: $\pm 11\% \pm 5\% \pm 6.6\% \pm 7\%$. The table contains seven rows of data.

SQRT(S) IN GEV	SIG(Q=HAD) IN MUB $\pm 11\% \pm 5\% \pm 6.6\% \pm 7\%$	SIG(Q=HAD)/SIG(Q=MU) $\pm 11\% \pm 5\% \pm 6.6\% \pm 7\%$
0.7726	0.269 \pm 0.096	1.85 \pm 0.66
0.7776	0.74 \pm 0.192	5.15 \pm 1.34
0.7801	1.13 \pm 0.225	7.92 \pm 1.58
0.7826	1.63 \pm 0.166	11.5 \pm 1.17
0.7851	1.07 \pm 0.226	7.59 \pm 1.61
0.7876	0.625 \pm 0.149	4.46 \pm 1.06
0.8026	0.065 \pm 0.036	0.48 \pm 0.27

Summary

- ▶ CEDAR is primarily a generator tuning system, combining JetWeb and HepData
- ▶ Hopefully greater than the sum of its parts!
- ▶ Aim is that any expt MC configuration should be “CEDAR-blessed” to be considered trustworthy
- ▶ Modular design and many spin-offs (see also HepForge talk on Thursday)
- ▶ CEDAR also hopes to encourage code re-use and uptake of common interfaces in HEP software
- ▶ Timescale: must be ready for LHC data. See <http://dev.cedar.ac.uk> and HepForge sites for roadmaps