Tuning Monte Carlo Event Generators to HERA Data

N. Brook[a], T. Carli[b], R. Mohr[b], M. Sutton[c], R.G. Waugh[a]

[a] Dept. of Physics & Astronomy, University of Glasgow, UK. G12 8 QQ
[b] Max Planck Institut für Physik, D-80805 München, Germany.
[c] University College London, UK.

Abstract: The Monte Carlo models ARIADNE, HERWIG and LEPTO are compared to deep-inelastic scattering data measured at the ep-collider HERA. An improved description of the data is achieved for all Monte Carlo models by tuning the available parameters.

1 Introduction

An adequate description of detailed event properties is a key ingredient for a reliable extraction of any physics observable from the data. Precision measurements of the structure function F_2 or the strong coupling constant α_s rely on a correct modelling of the event topology. Observables based on the hadronic final state need to be corrected for detector effects which is only possible if the Monte Carlo simulation describes the data. Therefore, improving the Monte Carlo model description of the large variety of hadronic final state data will play an important rôle in the future physics from HERA.

To allow an easy comparison of HERA data to model predictions, a test facility has been developed [1]. In this contribution, the influence of the model parameters to the hadronic properties of deep-inelastic events (DIS) is systematically investigated. The free parameters of the models are adjusted such that an optimal description of the HERA data is achieved. A list of the various parameters and switches, with a brief description, is given in table 1.

To quantify the description of a given data distribution by the model, a χ^2 is defined as:

$$\chi^2 = \frac{1}{n}\sum_{i=1}^{n} \frac{(\mathrm{MC}(i) - \mathrm{Data}(i))^2}{\sigma^2_{\mathrm{MC}(i)} + \sigma^2_{\mathrm{Data}(i)}} \tag{1}$$

where Data(i) and MC(i) are the values of the distribution in a given bin i for the data and Monte Carlo respectively. The sum runs over the n data bins.

From the data already published by the H1 and ZEUS experiments, a set of distributions is chosen that is sensitive to variations of the parameters in the Monte Carlo models. Care is taken that the various regions in the (x, Q^2) kinematic plane are covered. Sensitive parameters are identified for each Monte Carlo generator by considering either the variation of the χ^2 or the distribution $S = (\delta MC/MC)/(\delta P/P)$ weighted by the error on the experimental data after an extreme parameter variation (δP). Parameters already constrained by the recent LEP tuning [2] are not independently considered in our optimization procedure. Due to their sensitivity to the model parameters, the following distributions were chosen to perform the tuning on:

- the transverse energy (E_T) flow in function of the pseudo-rapidity ($\eta^* = -\ln\tan(\theta/2)$) and the mean E_T in the central rapidity region $-0.5 < \eta^* < 0.5$ in the hadronic center of mass frame (cms) [3] for selected (x, Q^2) bins;

	PARAMETER	default	tuned	DESCRIPTION
L	LST(34)	on	on	switch for new seaquark treatment
E	PARL(7)	0.5	0.1	probability of soft colour interaction
P	PARL(8)	0.01	0.04	cut-off in matrix element z_{qmin}
T	PARL(9)	1	25	cut-off in matrix element $\hat{s}_{min}$ in GeV^2
O	PYPAR(21)	0.2	0.2	Λ_{QCD} used in initial state parton shower in GeV
	PYPAR(22)	1	1.5	cut-off for initial state parton shower Q_0^2 in GeV^2
A	PARA(1)	0.22	0.22	Λ_{QCD} in GeV
R	PARA(3)	0.6	0.6	p_t cut-off for parton cascade in GeV
I	PARA(10)	1.0	1.5	power in soft suppression for remnant(α)
A	PARA(14)	1.0	1.0	factor for k_T to get extension of remnant (μ) in GeV
D	PARA(15)	1.0	0.5	power in soft suppression for struck quark (α)
N E	PARA(25)	2.0	2.0	$\frac{1}{b}$, governs probability of emissions outside the soft suppression cut-off
	PARA(27)	0.6	0.6	square root of primordial k_T^2 in GeV
H E R	CLMAX	3.35	5.5	maximum of the allowed cluster mass in GeV
W I G	PSPLT	1.0	0.65	exponent controlling the mass spectra of split clusters

Table 1: *Parameter values before and after tuning for the Monte Carlo models ARIADNE, HERWIG and LEPTO.*

- the scaled longitudinal momentum ($x_F = 2Pz/W$) of single particles in the cms at fixed values of the invariant mass of the hadronic system W [4];
- the logarithm of the inverse scaled momentum ($x_p = 2P/Q$) in the Breit frame in various (x, Q^2) bins [5];
- the transverse momentum of single particles in the positive hemisphere of the cms [6];
- the mean squared transverse momentum versus x_F ($\langle P_T^2 \rangle \times x_F$) in the positive hemisphere of the cms ('seagull' plot) [4, 6].

An overall combination of the χ^2 is formed by averaging over these distributions. The distributions entering this overall combination $\chi^2_{\rm comb}$ are given in table 2. The typical error on the χ^2 is about 0.04. The parameter combination giving the absolute minimum of $\chi^2_{\rm comb}$ is the tuned result. For all Monte Carlo models the parton density functions of GRV 94 is used [7]. The transverse energy-energy-correlation [4] turned out to be very sensitive to parameter variations. However, since the errors of this distribution are highly correlated, they were not included in $\chi^2_{\rm comb}$. Unfortunatley no jet data was included in this tuning procedure because, as yet, there is no published DIS jet data corrected to the hadron level.

2 ARIADNE

In ARIADNE [8] the QCD cascade is modelled by emitting gluons from a chain of independently radiating dipoles spanning colour connected partons [9], correcting the first emission to reproduce the first order matrix elements [10]. Hadronizing the partons into final state hadrons is achieved using the Lund string model as incorporated in JETSET [11]. Since the proton

TYPE of DATA		weight	ARIADNE		HERWIG		LEPTO	
		w_i	$\chi^2_{\rm default}$	$\chi^2_{\rm tuned}$	$\chi^2_{\rm default}$	χ^2_{tuned}	$\chi^2_{\rm default}$	$\chi^2_{\rm tuned}$
$\chi^2_{\rm comb}= \frac{1}{\sum w_i}\ \sum_{i=1}^{ndis} w_i\ \chi^2$			1.16	0.81	3.26	1.85	1.74	1.36
$\frac{1}{N}\frac{dN}{dx_F}$	at <W>= 77 GeV	0.5	0.58	0.43	0.51	0.14	0.41	0.22
	at <W>= 169 GeV	0.5	1.73	1.57	12.74	2.92	2.56	1.86
$\frac{1}{N}\frac{dN}{dp_T}$		1.0	0.24	0.13	0.70	1.28	0.86	0.75
$< p_T^2 >$ vs x_F	ZEUS	0.5	3.20	0.56	4.82	2.39	0.87	0.74
	H1	0.5	0.69	0.11	3.83	0.16	0.94	0.27
$\frac{1}{N}\frac{dE_T}{d\eta^*}$	low	0.25	1.00	2.47	2.19	1.83	7.25	6.60
	medium	0.25	0.96	0.34	2.81	2.33	2.01	1.35
	high	0.25	2.60	1.32	1.30	1.03	3.08	3.19
$\langle E_T \rangle$	all bins	0.25	0.86	1.85	2.32	1.24	5.33	3.37
$\frac{1}{N}\frac{dN}{d\ln 1/x_p}$	low	0.33	0.53	0.62	2.78	4.38	0.37	0.57
	medium	0.33	0.23	0.19	0.61	0.93	0.16	0.54
	high	0.33	2.13	2.04	2.89	3.49	2.18	1.55

Table 2: χ^2 *values obtained by comparing deep-inelastic HERA data to the prediction of the Monte Carlo models before and after tuning. The first line gives the* χ^2 *of the overall combination built on weighted data covering different aspects of the hadronic final state listed in the five following lines. Where available, kinematic bins equally distributed over the kinematic plane are selected. For* $dE_T/d\eta$ *the low bin corresponds to* $\langle x\rangle = 1.6\cdot 10^{-4}$ *and* $\langle Q^2\rangle = 6.8$ GeV^2*, the medium one to* $\langle x\rangle = 6.3\cdot 10^{-4}$ *and* $\langle Q^2\rangle = 14.2$ GeV^2*, the high one to* $\langle x\rangle = 2.1\cdot 10^{-3}$ *and* $\langle Q^2\rangle = 30.9$ GeV^2*.* $dN/d\ln 1/x_p$ *is calculated within the boundaries* $x = 1.2-2.5\cdot 10^{-3}$ *and* $Q^2 = 10-20$ GeV^2 *(low),* $x = 1.2-2.5\cdot 10^{-3}$ *and* $Q^2 = 20-40$ GeV^2 *(medium),* $x = 2.4-10\cdot 10^{-3}$ *and* $Q^2 = 80-160$ GeV^2 *(high).*

remnant at one endpoint of the parton chain is an extended object, the coherence condition allows only a fraction of this source to be involved in gluon radiation. Since the photon probing the proton only resolves the struck quark to a distance $\lambda \sim 1/Q$, the struck quark is also treated as an extended object. As a consequence gluon emissions in the proton and photon directions are suppressed. This phase space restriction is governed by $a = (\mu/k_T)^\alpha$ where a is the fraction of the colour antenna involved in the radiation, μ is a parameter related to a typical inverse size of a hadron and α governs the distribution of the energy along the dipole.

For tuning, the two α parameters for the proton remnant (PARA(10)) and the struck quark (PARA(15)) and the square root of the mean primordial k_T^2 in the proton (PARA(27)) are considered. All other parameters were left at their default values determined from a tuning of EMC and early DELPHI data [12]. The μ parameter, PARA(14), is highly correlated with PARA(10) and PARA(15) so was left at its default. Before the tuning, the $\chi^2_{\rm comb}$ was already acceptable, $\chi^2_{\rm comb}$=1.16, however it could be further reduced to $\chi^2_{\rm comb}$=0.81 by setting PARA(10)=1.5 and PARA(15)=0.5. The "quality" of the description for the individual distributions can be read from table 2. The improvement is mainly due to a much better description of $\langle P_T^2\rangle \times x_F$. (see Fig. 1/2b). The description of the E_T flows in the high x and Q^2 bin improved mainly in the current region (see Fig. 1/2d) however, at low x and Q^2 the tuned ARIADNE curve now lies slightly below the data (see Fig. 1/2c). The transverse energy-energy correlations are better described after the tuning although not included in $\chi^2_{\rm comb}$ (see Fig. 1/2e and f).

3 HERWIG

HERWIG [13] relies on a coherent parton branching algorithm with additional first order matrix element corrections [14] to populate the extremities of phase space which the partons from the conventional QCD cascade fail to occupy. The partons are transformed into hadrons using the cluster fragmentation model [15], whereby the primary hadrons are produced from an isotropic two body decay of colour-singlet clusters formed from partonic constituents.

All HERWIG parameters that are not constrained by the LEP data and are expected to be relevant to hadronic interactions were inspected for their sensitivity to HERA data. Only few had an impact on the description of the data. The distribution of intrinsic k_T of the incident hadronic constituents driven by PTRMS and the parameter BTCLM determining the mass distribution of the cluster containing the hadronic remnant are found only to modify the E_T flows at low η^* values close to or beyond the detector acceptance. The maximal allowed cluster mass CLMAX and the exponent controlling the mass spectra of split clusters PSPLT turned out to be most sensitive. The E_T flows grow with both increasing CLMAX and PSPLT, while the seagull plot decreases with increasing (decreasing) PSPLT (CLMAX). The x_F and the $\ln 1/x_p$ distributions fall with increasing CLMAX and decreasing PSPLT. However, the $\ln 1/x_p$ spectra in the intermediate (x, Q^2) range are systematically shifted to low values. A reduction in CLMAX, while nicely describing the E_T flows, at the same time reduces the level of the seagull plot such that varying CLMAX alone is unable to describe all data. PSPLT has the opposite effect on the seagull, so that *simultaneously* reducing PSPLT while increasing CLMAX reproduces the data best.

The optimal parameter settings are CLMAX=5.5 and PSPLT=0.65 leading to an improvement of $\chi^2_{\rm comb}$=3.26 to $\chi^2_{\rm comb}$=1.85. This is achieved by ameliorating the x_F spectra, the seagull plot and the E_T flows (see table 2). Such a high value for CLMAX would be potentially worrying, if not taken in conjunction with the reduction of PSPLT to compensate. Since for this HERWIG version and for HERA data it is also necessary to vary the parameter PSPLT, this result may still be consistent with the recent tuning of CLMAX below 4 achieved for LEP data, where the parameter PSPLT is not varied. Varying the parton density functions (GRV92 LO [16], MRS-H [17])) typically yields a variation smaller than 2% at the minimum of $\chi^2_{\rm comb}$. Although the transverse energy-energy correlations are not included in $\chi^2_{\rm comb}$, the new settings lead to a better agreement with the concave shape seen in the data (see Fig. 2e/f).

4 LEPTO

In LEPTO the hard parton processes are described by a leading order matrix element (ME). The soft and collinear divergences are regulated with a lower and upper cut in z_{qmin} where $z_q = p \cdot j_1 / p \cdot q$ where p (q) is the proton (photon) four-vector and j_1 the four-vector of one of the partons produced in the hard subprocess. In addition, the invariant mass squared of the two hard partons is required to exceed a minimal value, $\hat{s}_{min}$. Below the ME cut-offs, parton emissions are treated by parton showers based on the DGLAP evolution equations [20]. The amount of parton radiation depends on the virtuality chosen between a lower cut off (Q_0^2) and a maximum given by the scale of the hard process or the ME cut-off. LEPTO uses JETSET for the hadronization of the partons. In addition to this non-perturbative phase, LEPTO introduces two non-perturbative mechanisms. The first is in events where a sea-quark is involved in the hard subprocess, its partner is not simply rearranged to a meson or a baryon within the target remnant, but is used to stretch a string to one of the quarks remaining untouched in the proton. The additional E_T produced when hadronizing this string is needed to describe the data. The second is a soft (i.e. at a scale below Q_0^2) colour interaction which assumes that the colour configuration of the partonic system can be changed whilst traversing the colour field of the proton remnant. The probability that a parton pair exchanges their colour is controlled by a parameter, P_{sci}.

In the optimization procedure the parameters of JETSET and the ones governing the final state parton showers were only varied in sets defined by the four LEP collaborations. Sensitive parameters are filtered out by inspecting the relative change of the χ^2 after variation of the parameters within a reasonable range. The remaining parameters are shown in table 1. When increasing z_{qmin} and $\hat{s}_{min}$ the description of the seagull plot and the x_F spectra improves, but then the E_T flows are badly reproduced. The E_T flows do not allow a high value of the cut-off parameter for initial state parton showers $Q^2{}_{\rm ini}$ and prefer high values of P_{sci}. However, the x_F spectra are better described by low values of P_{sci}. High values of Λ_{QCD} in the initial state parton shower are able to better reproduce the E_T flows at low (x,Q^2), but are disfavored by the x_F spectra and the transverse energy-energy correlations.

When leaving z_{qmin} at the low default value of 0.01, $\hat{s}_{min}$ has to be raised to rather high values up to 100 GeV^2 to get a good $\chi^2_{\rm comb}$. For $0.04 < z_{qmin} < 0.08$ a $\chi^2_{\rm comb}$ minimum is found for a large range in $\hat{s}_{min}$ between 4 and 25 GeV^2. The best combination of z_{qmin} and $\hat{s}_{min}$ is used for a simultaneous variation of P_{sci} and $Q^2{}_{\rm ini}$. The last improvement is then achieved by varying Λ_{QCD} in the initial state parton shower and a stable $\chi^2_{\rm comb}$ minimum for values of 0.2-0.27 GeV is found. The parameter combination giving the lowest $\chi^2_{\rm comb}$=1.36 is shown in table 2. For this set the hadronization parameter used are those defined by L3. If one repeats the optimization procedure using the set from OPAL one gets compatible results, the set from ALEPH only achieved a $\chi^2_{\rm comb}$= 1.56.

The agreement of the E_T flows with the data is improved for all kinematic bins and the scaled longitudinal momenta are better described at high W (see table 2). Replacing the structure function to MRSH results in a worse $\chi^2_{\rm comb}$. In particular, it softens the transverse momentum distribution for single particles and the $\langle P_T^2\rangle\times\ x_F$ comes out too low for high x_F. With this parton density function the optimized $\chi^2_{\rm comb}$ value is only as good as the one obtained with GRV before tuning. The transverse energy-energy correlations are, even with the tuned parameters, poorly described for $x > 10^{-3}$. The shape of this distribution can be improved by using the parameter set from OPAL. A better description can be achieved by turning off the soft colour interactions or increasing the cut-off parameter of the final state parton shower. However, then the $\chi^2_{\rm comb}$ deteriorates mainly because the E_T flows fall below the data.

5 Conclusion

The Monte Carlo event generators ARIADNE, HERWIG and LEPTO have been compared with the hadronic final state of deep-inelastic scattering data from HERA. All the models' descriptions of the published data have improved with tuning, with ARIADNE giving the best fit to the data used in this study. Even though the description of the data has improved the situation is still far worse than that faced at LEP, where there is no initial state radiation or target remnant to model.

References

[1] J. Bromley et al., these proceedings.

[2] A. Blondel et al., 'QCD Event Generators', Proc. of LEP2 workshop, hep-ph/9601212.

[3] H1 Collab., S. Aid et al., Phys. Lett. B356 (1995) 118.

[4] H1 Collab., I. Abt et al., Z. Phys. C63, 377-389 (1994).

[5] ZEUS Collab., M.Derrick et al., Z. Phys. C67 (1995) 93-107.

[6] ZEUS Collab., M.Derrick et al., Z. Phys. C70 (1996) 1-15.

[7] M. Glück, E. Reya and A. Vogt, Z. Phys. C67 (1995) 433.

[8] L. Lönnblad, Comp. Phys. Comm. 71 (1992) 15, version used: 4.08.

[9] G. Gustafson, Ulf Petterson, Nucl. Phys. B306 (1988), G. Gustafson, Phys. Lett. B175 (1986) 453, B. Andersson et al., Z. Phys. C43 (1989) 625.

[10] L. Lönnblad, Z. Phys. C65 (1995) 285.

[11] T. Sjöstrand, Computer Phys. Comm. 39 (1986) 347, T. Sjöstrand and M. Bengtsson, Computer Phys. Comm. 43 (1987) 367,T. Sjöstrand, CERN-TH-6488-92 (1992), CERN-TH 7112/93 Dec. 93.

[12] W. de Boer and H. Fürstenau, Proc. of MC'91 w/s NIKHEF (1991) 616.

[13] G. Marchesini et al., Computer Phys. Commun. 67 (1992) 465, version used: 5.8d.

[14] M. H. Seymour, LU TP 94-17 hep-ph/9410414.

[15] B.R. Webber, Nucl. Phys. B238 (1984) 492.

[16] M. Glück, E. Reya and A. Vogt, Z. Phys. C53 (1992) 127.

[17] A. Martin, R. Roberts and J. Stirling, RAL-93-077.

[18] G. Ingelman, *in* Proc. HERA workshop, Eds. W. Buchmüller and G. Ingelman, Hamburg (1991) Vol. 3, 1366. G. Ingelman, A. Edin and J. Rathsman, DESY-preprint 96057, Hamburg, April 1996 and hep-ph/9605286, version used: 6.4.

[19] M. Bengtson, T. Sjöstrand, Z. Phys. C37 (1988) 465, M. Bengtson, G. Ingelman, T. Sjöstrand, *in* Proc. of the HERA Workshop 1987, Ed. R.D. Peccei, DESY Hamburg 1988, Vol. 1, 149.

[20] Yu. L. Dokshitzer, Sov. Phys. JETP 46 (1977) 641, V.N. Gribov and L.N. Lipatov, Sov. J. Nucl. Phys. 15 (1972) 438 and 675, G. Altarelli and G. Parisi, Nucl. Phys. 126 (1977) 297.

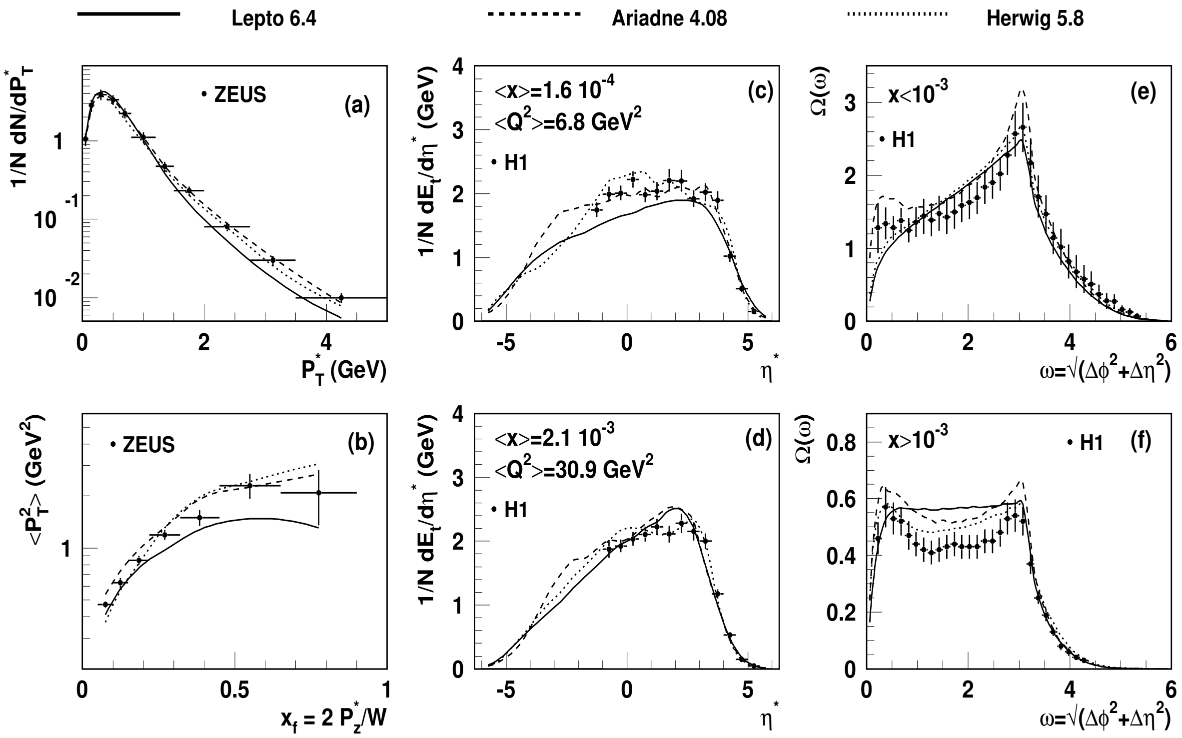

Figure 1: *The transverse momenta dN/dp_T (a) and the 'seagull' plot $\langle P_T^2 \rangle \times\ x_F$ (b) of single particles in the positive hemisphere of the hadronic center of mass. The transverse energy flow $dE_T/d\eta$ in a low (c) and high (d) x and Q^2 bin. The transverse energy-energy correlations for $x > 10^{-3}$ (e) and $x < 10^{-3}$ (f).*

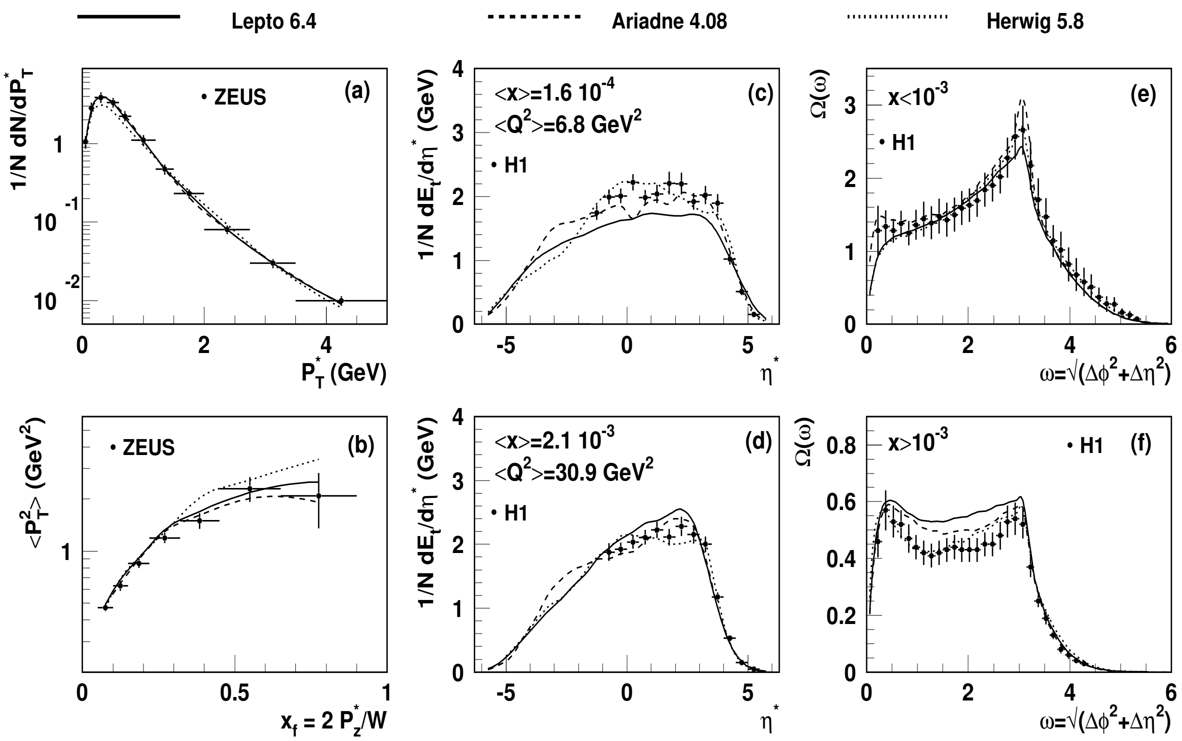

Figure 2: *As Fig.1, but Monte Carlo curves are produced with the tuned parameters.*