Published by Institute of Physics Publishing for SISSA
Received: *December 23, 2004*
Accepted: *March 30, 2005*
Published: *April 20, 2005*

Unbiased determination of the proton structure function F_2^p with faithful uncertainty estimation

The NNPDF collaboration

Luigi Del Debbio,[a] Stefano Forte,[b] José I. Latorre,[c] Andrea Piccione[de] and Joan Rojo[c]

[a] *Theory Division, CERN*
CH-1211 Genève 23, Switzerland
[b] *Dipartimento di Fisica, Università di Milano and*
INFN, Sezione di Milano
Via Celoria 16, I-20133 Milano, Italy
[c] *Departament d'Estructura i Constituents de la Matèria, Universitat de Barcelona*
Diagonal 647, E-08028 Barcelona, Spain
[d] *INFN Sezione di Genova*
Via Dodecaneso 33, I-16146 Genova, Italy, and
[e] *Dipartimento di Fisica Teorica, Università di Torino*
Via P. Giuria 1, I-10125 Torino, Italy
Email: luigi.del.debbio@cern.ch, forte@mi.infn.it, latorre@ecm.ub.es, piccione@ge.infn.it, joanrojo@ecm.ub.es

Abstract: We construct a parametrization of the deep-inelastic structure function of the proton $F_2(x, Q^2)$ based on all available experimental information from charged lepton deep-inelastic scattering experiments. The parametrization effectively provides a bias-free determination of the probability measure in the space of structure functions, which retains information on experimental errors and correlations. The result is obtained in the form of a Monte Carlo sample of neural networks trained on an ensemble of replicas of the experimental data. We discuss in detail the techniques required for the construction of bias-free parameterizations of large amounts of structure function data, in view of future applications to the determination of parton distributions based on the same method.

Keywords: QCD, Deep Inelastic Scattering.

Contents

1. Introduction

The requirements of precision physics at hadron colliders have recently led to a rapid improvement in the techniques for the determination of parton distributions of the nucleon, which are mostly extracted from deep-inelastic structure functions [1]. Specifically, it is now mandatory to determine accurately the uncertainty on these quantities. The main problem to be faced here is that one is trying to determine an uncertainty on a function, i.e., a probability measure on a space of functions, and to extract it from a finite set of experimental data. This problem can be studied in a simpler context, namely, the determination from the pertinent data of a structure function and its associate error. This sidesteps the technical complication of extracting parton distributions from structure functions, but it does tackle the main issue, namely the determination of an error on a function. Furthermore, the determination of a structure function and associate error might be useful for a variety of applications, such as precision tests of QCD (determination of α_s [2], tests of sum rules) or the determination of polarized structure functions from asymmetry data [3]

A new approach to this problem was recently proposed in ref. [4], based on the use of neural networks as basic interpolating tools. The main idea of this approach is to train a set of neural networks on a set of Monte Carlo replicas of the experimental data which reproduces their probability distribution. Hence, whereas the Monte Carlo replicas reproduce faithfully the probability measure of the data for $F_2(x, Q^2)$ in the points of the (x, Q^2) plane where data are available, the neural networks provide an interpolation and extrapolation for all (x, Q^2) subject to the only requirement of smoothness. The set of neural networks thus provides the desired probability measure, at least in the measured (x, Q^2) region, provided the sampling of the (x, Q^2) plane is not too coarse.

Experiment	ref.	x range	Q^2 range	$N_{\rm dat}$	$\langle\sigma_{\rm stat}\rangle$	$\langle\sigma_{\rm sys}\rangle$	$\left\langle\frac{\sigma_{\rm sys}}{\sigma_{\rm stat}}\right\rangle$	$\langle\sigma_N\rangle$	$\langle\sigma_{tot}\rangle$	$\langle\rho\rangle$	$\langle{\rm cov}\rangle$
NMC	[6]	$2.0\ 10^{-3} - 6.0\ 10^{-1}$	$0.5 - 75$	288	3.7	2.3	0.76	2.0	5.0	0.17	3.8
BCDMS	[7]	$6.0\ 10^{-2} - 8.0\ 10^{-1}$	$7 - 260$	351	3.2	2.0	0.56	3.0	5.4	0.52	13.1
E665	[8]	$8.0\ 10^{-4} - 6.0\ 10^{-1}$	$0.2 - 75$	91	8.7	5.2	0.67	2.0	11.0	0.21	21.7
ZEUS94	[9]	$6.3\ 10^{-5} - 5.0\ 10^{-1}$	$3.5 - 5000$	188	7.9	3.5	1.04	2.0	10.2	0.12	6.4
ZEUSBPC95	[10]	$2.0\ 10^{-6} - 6.0\ 10^{-5}$	$0.11 - 0.65$	34	2.9	6.6	2.38	2.0	7.6	0.61	34.1
ZEUSSVX95	[11]	$1.2\ 10^{-5} - 1.9\ 10^{-3}$	$0.6 - 17$	44	3.8	5.7	1.53	1.0	7.1	0.10	4.1
ZEUS97	[12]	$6.0\ 10^{-5} - 6.5\ 10^{-1}$	$2.7 - 30000$	240	5.0	3.1	0.93	3.0	6.7	0.29	7.0
ZEUSBPT97	[13]	$6.0\ 10^{-7} - 1.3\ 10^{-3}$	$0.045 - 0.65$	70	2.6	3.6	1.40	1.8	4.9	0.41	8.8
H1SVX95	[14]	$6.0\ 10^{-6} - 1.3\ 10^{-3}$	$0.35 - 3.5$	44	6.7	4.6	0.74	3.0	8.9	0.36	28.1
H197	[15]	$3.2\ 10^{-3} - 6.5\ 10^{-1}$	$150 - 30000$	130	12.5	3.2	0.31	1.5	13.3	0.06	10.9
H1LX97	[16]	$3.0\ 10^{-5} - 2.0\ 10^{-1}$	$1.5 - 150$	133	2.6	2.2	0.87	1.7	3.9	0.30	3.9
H199	[17]	$2.0\ 10^{-3} - 6.5\ 10^{-1}$	$150 - 30000$	126	14.7	2.8	0.24	1.8	15.2	0.05	11.0
H100	[18]	$1.3\ 10^{-3} - 6.5\ 10^{-1}$	$100 - 30000$	147	9.4	3.2	0.42	1.8	10.4	0.09	8.6

Table 1: Experiments included in this analysis. All values of σ and cov are given as percentages.

In ref. [4] a parametrization of the proton, deuteron and nonsinglet F_2 structure functions based on the BCDMS and NMC fixed-target deep-inelastic scattering data was constructed in this way. Here, we extend the results of ref. [4] by constructing a parametrization of the proton F_2 structure function which includes all available data, in particular the HERA collider data. Besides the obvious motivation of having state-of-the art results for this quantity, the main aim of this work is to develop a set of techniques which are required for the application of the method of ref. [4] to cases where the handling of a large number of disparate data sets is required. This involves in particular the use of genetic algorithms for the training of neural networks.

In section 2 we summarize the features of the experimental data. In section 3 we review the fitting method of ref. [4], emphasizing the improvements which have been introduced here. In section 4 we discuss the details of the training of neural nets to the current data set. In section 5 our final results are presented and compared to those previously obtained in ref. [4].

2. Experimental data

We construct a parametrization of F_2 based on all available unpolarized charged lepton-proton deep-inelastic scattering data [5]. However, we do not include early SLAC data, for which the covariance matrix is not available, since they do not provide any extra kinematic coverage, and are anyway less precise than later data. This leaves a total of 13 experiments, listed in table 1, along with their main features. The coverage of the (x, Q^2) kinematic plane afforded by these data is shown in fig. 1.

Structure functions are defined by parametrizing the deep-inelastic neutral current scattering cross section as

$$\frac{d^2\sigma(x,Q^2)}{dxdQ^2} = \frac{2\pi\alpha^2}{xQ^4}\left(Y_+F_2(x,Q^2) \pm Y_-xF_3(x,Q^2) - y^2F_L(x,Q^2)\right) . \qquad (2.1)$$

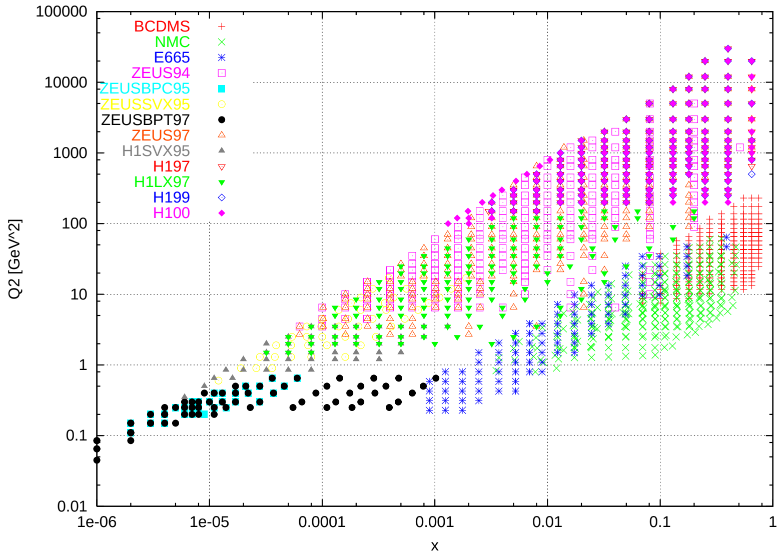

Figure 1: Kinematic range of the experimental data.

For the definition of kinematic variables see ref. [19]. We will construct a parametrization of the structure function $F_2(x, Q^2)$, which provides the bulk of the contribution to eq. (2.1). For all experiments the F_L contribution to the cross section has already been subtracted by the experimental collaborations, except for ZEUSBPC95, where we subtracted it using the values published by the same experiment. Note that the structure function F_2 receives contributions from both γ and Z exchange, though the Z contribution is only non-negligible for the high Q^2 datasets ZEUS94, H197, H199 and H100. We will construct a parametrization of the structure function F_2 defined in eq. (2.1), i.e. containing all contributions. When the experimental collaborations provide separately the contributions to F_2 due to γ or Z exchange we have recombined them in order to get the full F_2 eq. (2.1).

All the experiments included in our analysis provide full correlated systematics, as well as normalization errors. The covariance matrix can be computed from these as

$$\mathrm{cov}_{ij} = \left(\sum_{k=1}^{N_{\rm sys}} \sigma_{i,k}\sigma_{j,k} + F_i F_j \sigma_N^2 \right) + \delta_{ij}\sigma_{i,t}^2 \,, \tag{2.2}$$

where F_i, F_j are central experimental values, $\sigma_{i,k}$ are the $N_{\rm sys}$ correlated systematics, σ_N is the total normalization uncertainty, and the uncorrelated uncertainty $\sigma_{i,t}$ is the sum of the statistical uncertainty $\sigma_{i,s}$ and the N_u uncorrelated systematic uncertainties (when present):

$$\sigma_{i,t}^2 = \sigma_{i,s}^2 + \sum_{k=1}^{N_u} \sigma_{i,k}^2 \,. \tag{2.3}$$

The correlation matrix is then given by

$$\rho_{ij} = \frac{\mathrm{cov}_{ij}}{\sigma_{i,tot}\sigma_{j,tot}} \tag{2.4}$$

JHEP03(2005)080

where the total error $\sigma_{i,tot}$ for the i-th point is given by

$$\sigma_{i,tot} = \sqrt{\sigma_{i,t}^2 + \sigma_{i,c}^2 + F_i^2 \sigma_N^2}\,, \tag{2.5}$$

the total correlated uncertainty $\sigma_{i,c}$ is the sum of all correlated systematics

$$\sigma_{i,c}^2 = \sum_{k=1}^{N_{\rm sys}} \sigma_{i,k}^2\,. \tag{2.6}$$

For the ZEUS94, ZEUSSVX95 and ZEUSBPT97 experiments some uncertainties are asymmetric. As well known [20, 21], asymmetric errors cannot be combined in a simple multigaussian framework, and in particular they cannot be added to gaussian errors in quadrature. In the treatment of multigaussian errors, we will follow the approach of ref. [21], which, on top of several theoretical advantages, is closest to the ZEUS error analysis and thus adequate for a faithful reproduction of the ZEUS data. In this approach, a data point with central value x_0 and left and right asymmetric uncertainties σ_R and σ_L (not necessarily positive) is described by a symmetric gaussian distribution, centered at

$$\langle x \rangle = x_0 + \frac{\sigma_R - \sigma_L}{2} \tag{2.7}$$

and with width

$$\sigma_x^2 = \Delta^2 = \left(\frac{\sigma_R + \sigma_L}{2} \right)^2 . \tag{2.8}$$

The ensuing distribution can then be treated in the standard gaussian way.

3. Fitting technique

The construction of a parametrization of $F_2(x, Q^2)$ according to the method of ref. [4] consists of two steps: generation of a set of Monte Carlo replica of the original data, and training of a neural network to each replica. We summarize here the main features of these two steps, and the improvements that we introduced over the methods of ref. [4].

The Monte Carlo replicas of the original experiment are generated as a multigaussian distribution: each replica is given by a set of values

$$F_i^{({\rm art})(k)} = \left(1 + r_N^{(k)} \sigma_N\right) \left(F_i^{({\rm exp})} + \sum_{p=1}^{N_{\rm sys}} r_p^{(k)} \sigma_{i,p} + r_i^{(k)} \sigma_{i,t} \right) , \tag{3.1}$$

where $F_i^{({\rm exp})}$ is the i-th data point, we introduce an independent univariate gaussian random number $r^{(k)}$ for each independent error source, and the various errors are defined in eqs. (2.3)–(2.5).

The value of $N_{\rm rep}$ is determined in such a way that the Monte Carlo set of replicas models faithfully the probability distribution of the data in the original set. A comparison of expectation values, variance and correlation of the Monte Carlo set with the corresponding input experimental values as a function of the number of replicas is shown in fig. 2, where

JHEP03(2005)080

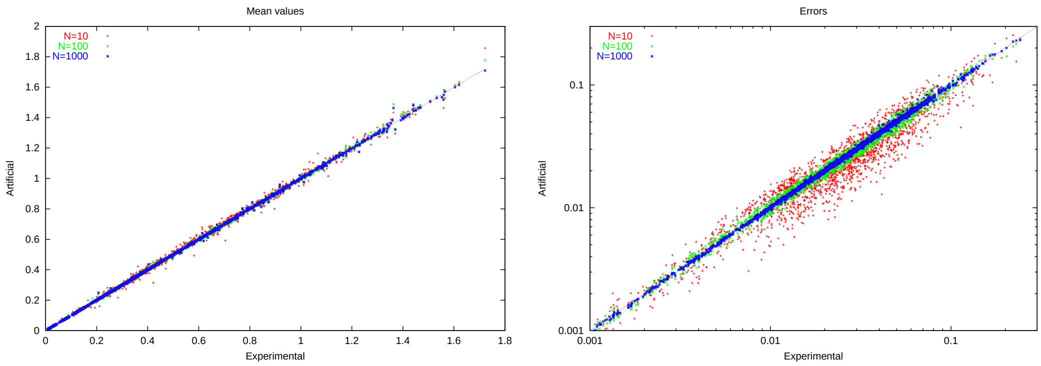

Figure 2: Scatter plot of experimental vs. Monte Carlo central values and errors.

we display scatter plots of the central values and errors for samples of 10, 100 and 1000 replicas. The corresponding plot for correlations is essentially the same as that shown in ref. [4]. A more quantitative comparison can be performed by defining suitable statistical estimators (see the appendix A). Results are presented in table 2. Note in particular the scatter correlations r for central values, errors and correlations, which indicate the size of the spread of data around a straight line. The table shows that a sample of 1000 replicas is sufficient to ensure average scatter correlations of 99% and accuracies of a few percent on structure functions, errors and correlations.

$N_{\rm rep}$ neural networks [22] are then trained on the Monte Carlo data, by training each neural network on all the $F_i^{(\rm art)}(k)$ data points in the k-th replica. The architecture of the networks is the same as in ref. [4]. The training is subdivided in three epochs, each based on the minimization of a different error function. In the first training epoch, the networks are trained to minimize the function

$$E^{(k)} = \frac{1}{N_{\rm dat}} \sum_{i=1}^{N_{\rm dat}} \left(F_i^{(\rm art)(k)} - F_i^{(\rm net)(k)} \right)^2 , \qquad (3.2)$$

i.e., the deviation from the central value per data point. In the second epoch the function to be minimized is the uncorrelated χ^2 per data point, namely, the χ^2 computed omitting correlated systematics:

$$\chi^2_{\rm diag}{}^{(k)} = \frac{1}{N_{\rm dat}} \sum_{i=1}^{N_{\rm dat}} \frac{\left(F_i^{(\rm art)(k)} - F_i^{(\rm net)(k)} \right)^2}{\sigma_{i,t}^{(\rm exp)2}} , \qquad (3.3)$$

where $\sigma_{i,t}^{(\rm exp)2}$ is defined in eq. (2.3). Finally, in the third epoch the full χ^2 per data point is minimized, namely

$$\chi^{2(k)} = \frac{1}{N_{\rm dat}} \sum_{i,j=1}^{N_{\rm dat}} \left(F_i^{(\rm art)(k)} - F_i^{(\rm net)(k)} \right) {\rm cov}_{ij}^{-1} \left(F_j^{(\rm art)(k)} - F_j^{(\rm net)(k)} \right) , \qquad (3.4)$$

where ${\rm cov}_{ij}$ is defined in eq. (2.2).

JHEP03(2005)080

$F_2^p(x,Q^2)$

$N_{\rm rep}$	10	100	1000
$\left\langle PE\left[\left\langle F^{(\rm art)}\right\rangle_{\rm rep}\right]\right\rangle$	1.88%	0.64%	0.20%
$r\left[F^{(\rm art)}\right]$	0.99919	0.99992	0.99999
$\left\langle V\left[\sigma^{(\rm art)}\right]\right\rangle_{\rm dat}$	6.7×10^{-4}	2.0×10^{-4}	6.9×10^{-5}
$\left\langle PE\left[\sigma^{(\rm art)}\right]\right\rangle_{\rm dat}$	37.21%	11.77%	3.43%
$\left\langle\sigma^{(\rm art)}\right\rangle_{\rm dat}$	0.0292	0.0317	0.0316
$r\left[\sigma^{(\rm art)}\right]$	0.945	0.995	0.999
$\left\langle V\left[\rho^{(\rm art)}\right]\right\rangle_{\rm dat}$	8.1×10^{-2}	7.8×10^{-3}	7.3×10^{-4}
$\left\langle\rho^{(\rm art)}\right\rangle_{\rm dat}$	0.3048	0.3115	0.2920
$r\left[\rho^{(\rm art)}\right]$	0.696	0.951	0.995
$\left\langle V\left[{\rm cov}^{(\rm art)}\right]\right\rangle_{\rm dat}$	5.2×10^{-7}	6.8×10^{-8}	6.9×10^{-9}
$\left\langle{\rm cov}^{(\rm art)}\right\rangle_{\rm dat}$	0.00013	0.00018	0.00015
$r\left[{\rm cov}^{(\rm art)}\right]$	0.687	0.941	0.994

Table 2: Comparison between experimental and Monte Carlo data. The experimental data have $\left\langle\sigma^{(\rm exp)}\right\rangle_{\rm dat}=0.0311$, $\left\langle\rho^{(\rm exp)}\right\rangle_{\rm dat}=0.2914$ and $\left\langle{\rm cov}^{(\rm exp)}\right\rangle_{\rm dat}=0.00015$. All statistical indicators are defined in the appendix A.

The rationale behind this procedure is that the true minimum which the fitting procedure must determine is that of the full χ^2 eq. (3.4). However, this is nonlocal and time consuming to compute. It is therefore advantageous to look for its rough features at first, then refine its search, and finally determine its actual location.

The minimum during the first two epochs is found using back-propagation (BP) (see ref. [4]). This method is not suitable for the minimization of the nonlocal function eq. (3.4). In ref. [4] BP was used throughout, and the third epoch was omitted. This is acceptable provided the total systematics is small in comparison to the statistical errors, and indeed it was verified that a good approximation to the minimum of eq. (3.4) could be obtained from the ensemble of neural networks. This is no longer the case for the present extended data set, as we shall see explicitly in section 5. Therefore, the full χ^2 eq. (3.4) is minimized in the third training epoch by means of genetic algorithms (GA), previously used and discussed by some of us for related purposes in ref. [23].

In comparison to previous work [4, 23], we have implemented several improvements, both in the BP and GA training epochs. In the BP epoch, we use on line training as in ref. [4], i.e. the parameters of the neural network are updated after each data point has been shown to it. This defines a training cycle. In ref. [4] it was shown that the length of training needed to achieve a given value of the χ^2 can differ significantly between experiments: it is larger for experiments which have smaller errors, contain more data, or cover kinematic regions where the structure function varies more rapidly. If one wants to end up with a similar value of the χ^2 for all experiments it is then necessary to adjust the relative length of training of different data sets. In ref. [4] this was achieved by finding by trial and error an optimal fixed weight for the two experiments included in the fit. This procedure is clearly not viable when the number of experiments is large. Therefore, we have implemented a

JHEP03(2005)080

dynamical weighted training, whereby the weight P_i of each experiment is chosen initially to be the same for all experiments, and then adjusted dynamically according to the relative contribution of each experiment χ_i^2 to the total χ^2 eq. (3.4):

$$P_i \to P_i \frac{\chi_i^2}{\sum_{j=1}^{N_{\rm exp}} \chi_i^2} \,. \tag{3.5}$$

The value of χ_i^2 is updated from the full data set every 1.25×10^6 training cycles; because there are ~ 1700 data points, this ensures that between updates each data point has been seen by the net about 700 times on average in the unweighted case. This method is not viable in the third (GA) training epoch, where the χ^2 can only be computed from the full set of data points (i.e., the training is necessarily batched, and not on-line). Therefore, one cannot choose to show a subset of data more often. One could in principle reweight the contribution of the single experiments to the χ^2, but this might distort the global minimum in an unpredictable way.

In the GA epoch, we have introduce two improvements in comparison to the methods described in ref. [23]. First, we have introduced multiple mutations, specifically three nested mutations for each training cycle.[1] The purpose of this is to avoid local minima, thereby increasing the speed of training. It is crucial that rates for these additional mutations are large, in order to allow for jumps from a local minimum to a deeper one. We find that one additional mutation with probability 20% and two additional mutations with probability 4% produce a significant improvement of the convergence rate. Second, we have introduced probabilistic selection. This entails that the sample of $N_{\rm sel}$ selected mutations is constructed by always selecting the mutation i_0 with the lowest χ^2, plus $N_{\rm sel} - 1$ mutations selected among the total $N_{\rm mut}$ mutations with probability

$$P_j = \exp\left(-\frac{\chi_j^2 - \chi_{i_0}^2}{\chi_{i_0}^2} \right) \,. \tag{3.6}$$

Namely, mutations with larger χ^2 are less likely to be selected but can still be selected with finite probability. This is helpful in allowing for mutations which only become beneficial after a combination of several individual mutations.

At the end of the GA training we are left with a sample of $N_{\rm rep}$ neural networks, from which e.g. the value of the structure function at (x, Q^2) can be computed as

$$F_2(x, Q^2) = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} F^{({\rm net})(k)}(x, Q^2) \,. \tag{3.7}$$

The goodness of fit of the final set is thus measured by the χ^2 per data point, which, given the large number of data points is essentially identical to the χ^2 per degree of freedom:

$$\chi^2 = \frac{1}{N_{\rm dat}} \sum_{i,j=1}^{N_{\rm dat}} \left(\left\langle F_i^{({\rm net})} \right\rangle_{\rm rep} - F_i^{({\rm exp})} \right) \left({\rm cov}^{-1} \right)_{ij} \left(\left\langle F_j^{({\rm net})} \right\rangle_{\rm rep} - F_j^{({\rm exp})} \right) \,, \tag{3.8}$$

where the average over replicas, denoted by $\langle\rangle_{\rm rep}$, is defined in the appendix A.

[1] Note that GA training cycles in ref. [23] are referred to as generations, as it is customary for genetic algorithms.

4. Training to the F_2^p data

In order to apply the general method discussed in section 3 to the F_2^p data presented in section 2 several specific issues must be addressed: the choice of training parameters and training length, the choice of the actual data set, and the choice of theoretical constraints. We now address these issues in turn.

The parameters and length for the first two training epochs have been determined by inspection of the fit of a neural network to the central experimental values. Clearly, this choice is less critical, in that it is only important in order for the training to be reasonably fast, but it doesn't impact on final result. We choose for the first BP epoch 5×10^7 training cycles with learning rate $\eta = 9 \times 10^{-2}$ and momentum term $\alpha = 0.9$, and for the second BP epoch 2.5×10^8 training cycles with learning rate $\eta = 9 \times 10^{-9}$ and momentum term $\alpha = 0.9$.

After these first two training epochs, the diagonal $\chi^2_{\rm diag}$ per data point eq. (3.3), which is being minimized, is of order two for the central data set. This is comparable to the length of training that was required to reach $\chi^2_{\rm diag} \approx 1.3$ for the smaller data set of ref. [4]. The χ^2 eq. (3.4), which is always bounded by it, $\chi^2 \leq \chi^2_{\rm diag}$ is accordingly smaller (see table 3). The training algorithm then switches to GA minimization of the χ^2 eq. (3.4). The determination of the length of this training epoch is critical, in that it controls the form of the final fit. This can only be done by looking at the features of the full Monte Carlo sample of neural networks.

	A		B	
Experiment	$\chi^2_{\rm diag}$	χ^2	$\chi^2_{\rm diag}$	χ^2
Total	2.01	1.49	8.29	1.24
NMC	1.96	1.54	6.05	1.46
BCDMS	1.63	1.54	13.27	1.09
E665	1.57	1.31	2.28	1.20
ZEUS94	3.05	2.23	6.15	1.94
ZEUSBPC95	4.30	1.27	11.06	1.09
ZEUSSVX95	3.18	1.79	10.97	1.92
ZEUS97	2.49	1.56	15.17	1.17
ZEUSBPT97	2.47	1.81	8.13	1.08
H1SVX95	3.24	0.72	8.55	0.90
H197	0.88	0.73	2.98	0.90
H1LX97	1.82	1.37	9.81	1.11
H199	1.13	1.04	2.17	1.00
H100	1.57	1.46	4.72	1.25

Table 3: The uncorrelated and total χ^2, eqs. (3.3) and eq. (3.4), respectively, for the fit to the central data points: (A) after the backpropagation training epoch and (B) after the final genetic algorithms training epoch.

Before addressing this issue, however, it turns out to be necessary to consider the possibility of introducing cuts in the data set. Indeed, consider the results obtained after a GA training of 4×10^4 cycles (with mutation rate 5×10^{-3}) to the central data set, displayed in table 3. This is a rather long training: indeed, in each GA training cycle all the data are shown to the nets. Hence in 4×10^4 GA cycles the data are shown to the nets 0.7×10^8 times, comparable to the number of times they are shown to the nets during BP training. It is apparent that whereas $\chi^2 \sim 1$ for most experiments, it remains abnormally high for NMC and especially ZEUS94 and ZEUSSVX95. Because of the weighted training which has been adopted, this is unlikely to be due to insufficient training of these data sets, and is more likely related to problems of these data sets.

Whereas ZEUSSVX95 only contains a small number of data points, NMC and ZEUS94 account each for more than 10% of the total number of data points, and thus they can bias final results considerably. The case of NMC was discussed in detail in ref. [4]. This data set is the only one to cover the medium-x, medium-small Q^2 region (compare figure 1) and thus it cannot be excluded from the fit. As discussed in ref. [4], the relatively large value of the χ^2 for this experiment is a consequence of internal inconsistencies within the data set. A value of $\chi^2 \approx 1.5$ indicates that the neural nets do not reproduce the subset of data which are incompatible with the bulk, as it should be, whereas a value $\chi^2 \approx 1$ could only be obtained by overlearning, i.e. essentially by fitting irrelevant fluctuations (see ref. [4]).

Let us now consider the case of ZEUS94. The kinematic region of this experiment is entirely covered by ZEUS97, H197, H199, H100. We can therefore study the impact of excluding this experiment from the global fit, without information loss. The results obtained in such case are displayed in table 4: when the experiment is not fitted the χ^2 of all experiments with which it overlaps improves and so does the global χ^2, whereas the χ^2 for ZEUS94 itself only deteriorates by a comparable amount, despite the fact that the experiment is now not fitted at all. We conclude that the experiment should be excluded from the fit, since its inclusion result in a deterioration of the fit quality, whereas its exclusion does not entail information loss. Difficulties in the inclusion of this experiment in global fits were already pointed out in refs. [24, 25], where it was suggested that they may be due to underestimated or nongaussian uncertainties. It is likely that ZEUSSVX95 has similar problems. However, its inclusion in the fit is no reason of concern, even if its high χ^2 were due to incompatibility of this experiment with the others or underestimate of its experimental uncertainties, because of the small number of data points. It is therefore retained in the data set. Our final data set thus includes all experiments in table 1, except ZEUS94. We are thus left with 1698 data points.

Experiment	χ^2
Total	1.14
NMC	1.42
BCDMS	1.11
E665	1.06
ZEUS94	2.17
ZEUSBPC95	1.02
ZEUSSVX95	1.97
ZEUS97	1.17
ZEUSBPT97	0.97
H1SVX95	0.84
H197	0.83
H1LX97	1.09
H199	0.94
H100	1.17

Table 4: The same fit as the last column of table 3 if the ZEUS94 data are excluded from the fit.

For the sake of future applications, it is interesting to ask how the procedure of selecting experiments in the data set can be automatized. This can be done in an iterative way as follows: first, a neural net (or sample of neural nets) is trained on only one experiment; then, the total χ^2 for the full data set is computed using this neural net (or sample of nets); the procedure is then repeated for all experiments, and the experiment which leads to the smallest total χ^2 is selected. In the second iteration, the net (or sample of nets) is trained on the experiment selected in the first iteration plus any of the other experiments, thereby leading to the selection of a second experiment to be added to that selected previously, and so on. The process can be terminated before all experiments are selected, for instance if it is seen that the addition of a new experiment doesn't lead to a significant improvement in χ^2 for given length of training.

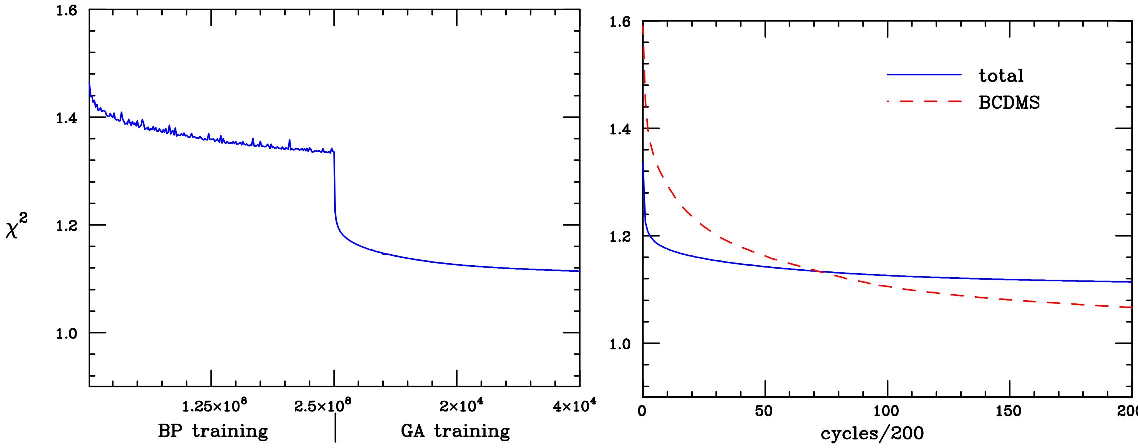

Figure 3: Dependence of the χ^2 eq. (3.8) on the length of training: (left) total training (right) detail of the GA training.

We now proceed to discuss the length of training for our final data set. The χ^2 eq. (3.8) is shown in figure 3 as a function of the number of GA training cycles. The χ^2 decreases very rapidly during the first few hundreds of training cycles. After about 5000 training cycles, the χ^2 as a function of the training length essentially flattens for all experiments but BCDMS. The further decrease of the total χ^2 is then due essentially to the decrease of the contribution from BCDMS. A training length of 4×10^4 GA cycles is necessary in order for the χ^2 of BCDMS to flatten out at $\chi^2 \sim 1$. As discussed in ref. [4], the BCDMS data can only be learnt with a longer training because they have high precision while being located in the intermediate x (valence) region, where the parton distribution displays significant variation.

The χ^2 for the fit of a neural net to the central data with this training is given in table 3. It shows that all experiments are well reproduced with the exceptions discussed above. It is interesting to observe that while the χ^2 eq. (3.4) decreases significantly during the GA training, the uncorrelated χ^2_{diag} eq. (3.3) increases. This means that correlations are sizable for the HERA experiments, so that the approach of ref. [4], based on the minimization of χ^2_{diag} eq. (3.3), is not adequate in this case. GA minimization appears to be very efficient in reducing the χ^2 relatively fast.

We finally turn to the issue of theoretical constraints. The only theoretical assumption on the shape of $F_2(x, Q^2)$ is that it satisfies the kinematic constraint $F_2(1, Q^2) = 0$ for all Q^2. As this constraint is local, its implementation is straightforward: it can be enforced by including in the data set a number of artificial data points which satisfy the constraint with a suitably tuned error. In the present fit we have checked that the best choice is to add a number of artificial points at $x = 1$, equal to 2% of the experimental trained points (33 points with ZEUS94 excluded from the fits), and with error equal to one tenth of the mean statistical error of the trained points. These points are equally spaced in $\ln Q^2$, within the range covered by the trained points.

JHEP03(2005)080

5. Results

The χ^2 for the training of a single neural net to the central data points is shown in table 4. The results for the final set of 1000 neural networks are displayed in table 5, while in table 6 we give the details of results for each experiment. Besides the χ^2 we also list the values of various quantities, defined in the appendix A, which can be used to assess the goodness of fit.

The quality of the final fit is somewhat better than that of the fit to the central data points shown in table 4. In particular, with the exception of NMC (which is likely to have internal inconsistencies [4]) and ZEUSSVX95 (which is likely to have the same problems as those of ZEUS94 discussed in section 4) the χ^2 per degree of freedom is of order 1 for all experiments. It is interesting to note that the χ^2 for the neural network average is rather better than the average $\langle\chi^2\rangle$ eq. (A.11). The (scatter) correlation between experimental data and the neural network prediction equals one to about 1% accuracy, with the exception of NMC, ZEUSSVX95 (which have the aforementioned problems) and E665. The E665 kinematic region overlaps almost entirely (apart from very small $Q^2 < 1\,\mathrm{GeV}^2$) with that of NMC and BCDMS, while having lower accuracy (this is why the experiment was not included in the fits of ref. [4]). The data points corresponding to this experiment are therefore essentially predicted by the fit to other experiments, thus explaining the somewhat smaller scatter correlation.

$F_2^p(x,Q^2)$	
N_{rep}	1000
χ^2	1.12
$\langle\chi^2\rangle$	2.13
$r\left[F^{(\mathrm{net})}\right]$	0.993
$\mathcal{R}$	0.54
$\left\langle V\left[\sigma^{(\mathrm{net})}\right]\right\rangle_{\mathrm{dat}}$	$1.3\ 10^{-3}$
$\left\langle PE\left[\sigma^{(\mathrm{net})}\right]\right\rangle_{\mathrm{dat}}$	93%
$\left\langle\sigma^{(\mathrm{exp})}\right\rangle_{\mathrm{dat}}$	0.0273
$\left\langle\sigma^{(\mathrm{net})}\right\rangle_{\mathrm{dat}}$	0.0060
$r\left[\sigma^{(\mathrm{net})}\right]$	0.80
$\left\langle V\left[\rho^{(\mathrm{net})}\right]\right\rangle_{\mathrm{dat}}$	0.13
$\left\langle\rho^{(\mathrm{exp})}\right\rangle_{\mathrm{dat}}$	0.31
$\left\langle\rho^{(\mathrm{net})}\right\rangle_{\mathrm{dat}}$	0.54
$r\left[\rho^{(\mathrm{net})}\right]$	0.49
$\left\langle V\left[\mathrm{cov}^{(\mathrm{art})}\right]\right\rangle_{\mathrm{dat}}$	$3.4\ 10^{-7}$
$\left\langle\mathrm{cov}^{(\mathrm{exp})}\right\rangle_{\mathrm{dat}}$	$1.3\ 10^{-4}$
$\left\langle\mathrm{cov}^{(\mathrm{net})}\right\rangle_{\mathrm{dat}}$	$1.5\ 10^{-5}$
$r\left[\mathrm{cov}^{(\mathrm{net})}\right]$	0.57

Table 5: Estimators of the final $F_2(x,Q^2)$ results.

The average neural network variance is in general substantially smaller than the average experimental error, typically by a factor $3-4$. This is the reason why $\langle\chi^2\rangle > \chi^2$: the neural nets fluctuate less about central experimental values than the Monte Carlo replicas. In the presence of substantial error reduction, the (scatter) correlation between network covariance and experimental error is generally not very high, and can take low values when a small number of data points from one experiment is enough to determine the outcome of the fit, such as in the case of the NMC experiment, even more so for E665. [4]

As discussed extensively in ref. [4] it is important to make sure that this is due to the fact that information from individual data points is combined through an underlying law by the neural networks, and not due to parametrization bias. To this purpose, the $\mathcal{R}$-estimator has been introduced in ref. [4], where it was shown that in the presence of substantial error reduction $\mathcal{R} \gtrsim 1$ if there is parametrization bias, whereas $\mathcal{R} \approx 0.5$ in the absence of parametrization bias.[2] It is apparent from tables 5–6 that indeed $\mathcal{R} \approx 0.5$ for all

[2] Note that in ref. [4] the $\mathcal{R}$-ratio was defined in terms of the diagonal χ^2_{diag} eq. (3.3), because neural

Experiment	NMC	BCDMS	E665
χ^2	1.39	1.08	1.09
$\langle\chi^2\rangle$	2.38	2.11	2.11
$r\left[F^{(\rm net)}\right]$	0.95	0.99	0.91
$\mathcal{R}$	0.59	0.54	0.53
$\left\langle V\left[\sigma^{(\rm net)}\right]\right\rangle_{\rm dat}$	0.003	$6.4\ 10^{-5}$	0.0014
$\left\langle PE\left[\sigma^{(\rm net)}\right]\right\rangle_{\rm dat}$	0.92	0.92	0.94
$\left\langle\sigma^{(\rm exp)}\right\rangle_{\rm dat}$	0.017	0.007	0.032
$\left\langle\sigma^{(\rm net)}\right\rangle_{\rm dat}$	0.004	0.002	0.005
$r\left[\sigma^{(\rm net)}\right]$	0.074	0.98	-0.06
$\left\langle V\left[\rho^{(\rm net)}\right]\right\rangle_{\rm dat}$	0.32	0.035	0.18
$\left\langle\rho^{(\rm exp)}\right\rangle_{\rm dat}$	0.17	0.52	0.21
$\left\langle\rho^{(\rm net)}\right\rangle_{\rm dat}$	0.68	0.59	0.48
$r\left[\rho^{(\rm net)}\right]$	0.19	0.80	0.12
$\left\langle V\left[{\rm cov}^{(\rm art)}\right]\right\rangle_{\rm dat}$	$3.8\ 10^{-9}$	$2.1\ 10^{-9}$	$2.5\ 10^{-9}$
$\left\langle{\rm cov}^{(\rm exp)}\right\rangle_{\rm dat}$	$4.5\ 10^{-5}$	$3.4\ 10^{-5}$	$1.7\ 10^{-4}$
$\left\langle{\rm cov}^{(\rm net)}\right\rangle_{\rm dat}$	$1.2\ 10^{-5}$	$3.0\ 10^{-6}$	$1.0\ 10^{-5}$
$r\left[{\rm cov}^{(\rm net)}\right]$	0.11	0.97	0.17

Experiment	ZEUSBPC95	ZEUSSVX95	ZEUS97	ZEUSBPT97	H1SVX95	H197	H1LX97	H199	H100
χ^2	0.97	1.90	1.17	0.89	0.79	0.83	1.02	0.92	1.09
$\langle\chi^2\rangle$	1.96	2.88	2.19	2.91	1.82	1.84	2.02	1.93	2.11
$r\left[F^{(\rm net)}\right]$	0.98	0.96	0.99	0.99	0.97	0.99	0.99	0.98	0.99
$\mathcal{R}$	0.51	0.66	0.55	0.55	0.44	0.46	0.53	0.48	0.54
$\left\langle V\left[\sigma^{(\rm net)}\right]\right\rangle_{\rm dat}$	$4.7\ 10^{-4}$	0.0037	0.0018	$1.3\ 10^{-4}$	0.0043	0.0031	0.0007	0.003	0.0013
$\left\langle PE\left[\sigma^{(\rm net)}\right]\right\rangle_{\rm dat}$	0.92	0.96	0.94	0.72	0.97	0.97	0.90	0.97	0.93
$\left\langle\sigma^{(\rm exp)}\right\rangle_{\rm dat}$	0.021	0.060	0.036	0.012	0.064	0.040	0.027	0.051	0.032
$\left\langle\sigma^{(\rm net)}\right\rangle_{\rm dat}$	0.006	0.011	0.085	0.006	0.010	0.007	0.008	0.008	0.008
$r\left[\sigma^{(\rm net)}\right]$	0.85	0.63	0.87	0.72	0.77	0.87	0.41	0.80	0.90
$\left\langle V\left[\rho^{(\rm net)}\right]\right\rangle_{\rm dat}$	0.07	0.24	0.06	0.10	0.118	0.14	0.1	0.16	0.14
$\left\langle\rho^{(\rm exp)}\right\rangle_{\rm dat}$	0.61	0.24	0.28	0.40	0.36	0.06	0.29	0.05	0.09
$\left\langle\rho^{(\rm net)}\right\rangle_{\rm dat}$	0.80	0.62	0.42	0.67	0.61	0.31	0.55	0.33	0.31
$r\left[\rho^{(\rm net)}\right]$	0.56	0.39	0.65	0.61	0.41	0.46	0.64	0.36	0.56
$\left\langle V\left[{\rm cov}^{(\rm art)}\right]\right\rangle_{\rm dat}$	$8.9\ 10^{-8}$	$2.2\ 10^{-6}$	$6.0\ 10^{-7}$	$2.8\ 10^{-9}$	$3.0\ 10^{-6}$	$3.8\ 10^{-7}$	$5.3\ 10^{-8}$	$3.0\ 10^{-7}$	$1.8\ 10^{-7}$
$\left\langle{\rm cov}^{(\rm exp)}\right\rangle_{\rm dat}$	$2.8\ 10^{-4}$	$8.5\ 10^{-4}$	$3.4\ 10^{-4}$	$5.8\ 10^{-5}$	0.0014	$1.0\ 10^{-4}$	$2.1\ 10^{-4}$	$1.4\ 10^{-4}$	$9.7\ 10^{-5}$
$\left\langle{\rm cov}^{(\rm net)}\right\rangle_{\rm dat}$	$2.8\ 10^{-5}$	$8.2\ 10^{-5}$	$3.2\ 10^{-5}$	$2.3\ 10^{-5}$	$7.1\ 10^{-5}$	$1.4\ 10^{-5}$	$3.9\ 10^{-5}$	$1.6\ 10^{-5}$	$2.0\ 10^{-5}$
$r\left[{\rm cov}^{(\rm net)}\right]$	0.71	0.51	0.78	0.67	0.48	0.61	0.55	0.51	0.60

Table 6: Final $F_2(x,Q^2)$ results for the individual experiments: fixed target (top) and HERA (bottom).

experiments. Note that, contrary to what was found in ref. [4], there is now error reduction also for the BCDMS experiment, by an amount comparable to that of other experiments. We will come back to this issue when comparing results to those of ref. [4].

Further evidence that the error reduction is not due to parametrization bias can be obtained by studying the dependence of $\left\langle\sigma^{(\rm net)}\right\rangle_{\rm dat}$ on the length of training. This dependence is shown in fig. 4 for the BCDMS experiment. It is apparent that the error reduction is correlated with the goodness of fit displayed in fig. 3, and it occurs during

networks were trained by minimizing that quantity. It is easy to see that the results of ref. [4] for $\mathcal{R}$ remain true when the full χ^2 eq. (3.4) is minimized, provided $\mathcal{R}$ is redefined accordingly.

JHEP03(2005)080

the GA training, thereby suggesting that error reduction occurs when the neural networks start reproducing an underlying law. If error reduction were due to parametrization bias it would be essentially independent of the length of training.

The point-to-point correlation ρ of the neural nets is somewhat larger than that of the data, as one might expect as a consequence of an underlying law which is being learnt by the neural nets. In fact, for the NMC experiment the increase in correlation essentially compensates the reduction in error, in such a way that the average covariance of the nets and the data are essentially the same. This again shows that in the case of the NMC experiment a small number of points is sufficient to predict the remaining ones. For all other experiments, however, the covariance of the nets is substantially smaller than that of the data. As a consequence the (scatter) correlation of covariance remains relatively high for all experiments, except NMC, and especially E665 whose points are essentially predicted by the fit to other experiments.

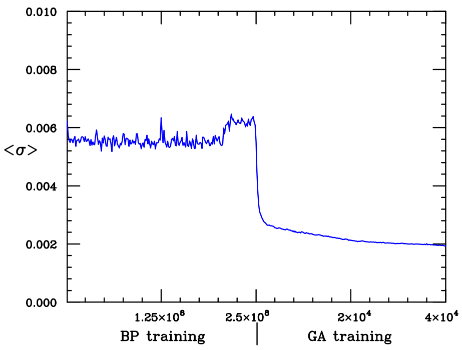

Figure 4: Dependence of $\langle \sigma^{(\mathrm{net})} \rangle_{\mathrm{dat}}$ on the length of training for the BCDMS experiment.

The structure function and associated one-σ error band is compared to the data as a function of x for a pair of typical values of Q^2 in fig. 5. In fig. 6 the behaviour of the structure function as a function of x at fixed Q^2 and as a function of Q^2 at fixed x is also shown. It is apparent that in the data region the error on the neural nets is rather smaller than that on the data used to train them. The error however grows rapidly as soon as the nets are extrapolated outside the region of the data. At large x, however, the extrapolation is kept under control by the kinematic constraint $F_2(1, Q^2) = 0$.

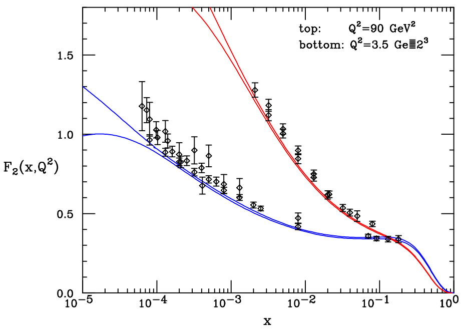

Figure 5: Final results for $F_2(x, Q^2)$ compared to data. For the neural net result, the one-σ error band is shown.

Let us finally compare the determination of $F_2(x, Q^2)$ presented here with that of ref. [4], which was based on pre-HERA data. In fig. 7 one-σ error bands for the two parametrizations are compared, whereas in fig. 8 we display the relative pull of the two parametrizations, defined as

$$P(x, Q^2) \equiv \frac{F_2^{new}(x, Q^2) - F_2^{old}(x, Q^2)}{\sqrt{\sigma_{old}^2(x, Q^2) + \sigma_{new}^2(x, Q^2)}}, \tag{5.1}$$

where $\sigma(x, Q^2)$ is the error on the structure function determined as the variance of the neural network sample. In view of the fact that the old fit only included BCDMS and NMC data, it is interesting to consider four regions: (a) the BCDMS region (large x,

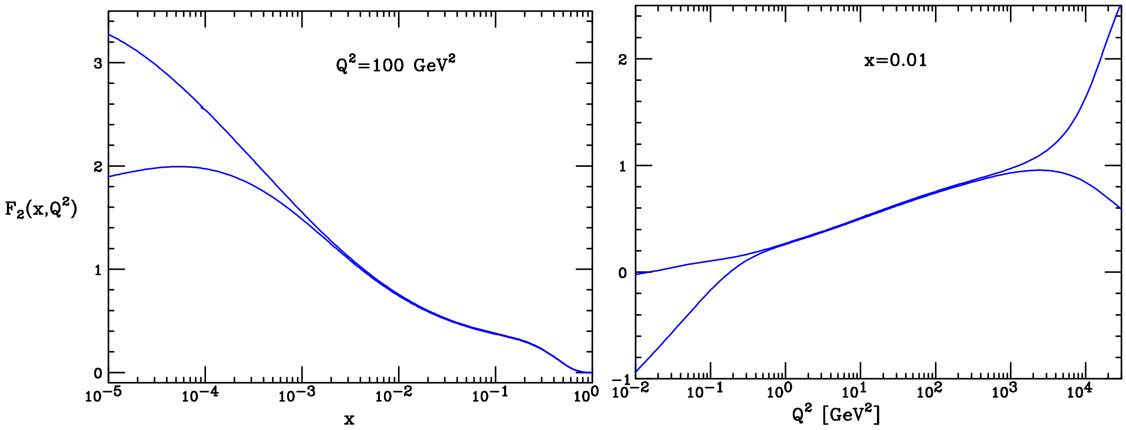

Figure 6: One-σ error band for the structure function $F_2(x, Q^2)$ computed from neural nets. Note the different scale on the y axis in the two plots.

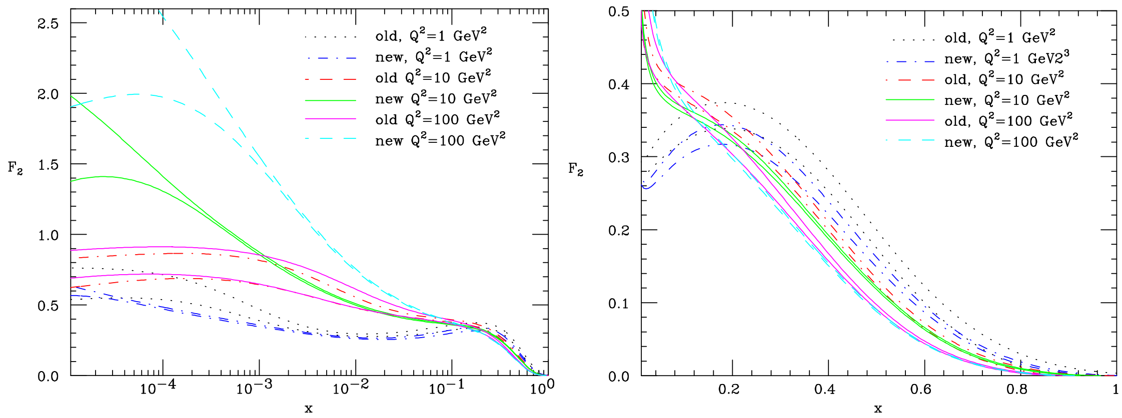

Figure 7: Comparison of the parametrization of $F_2(x, Q^2)$ of ref. [4] (old) with that of the present paper (new). The pairs of curves correspond to a one-σ error band.

intermediate Q^2, e.g. $x = 0.3$, $Q^2 = 20\,\mathrm{GeV}^2$); (b) the NMC region (intermediate x, not too large Q^2, e.g. $x = 0.1$, $Q^2 = 2\,\mathrm{GeV}^2$); (c) the HERA region (small x and large Q^2, e.g. 0.01 and $Q^2 > 10\,\mathrm{GeV}^2$); (d) the region where neither the old nor the new fit had data (very large or very small Q^2). In region (a) the new fit is rather more precise than the old one, for reasons to be discussed shortly, while central values agree, with $P \lesssim 1$). In region (b) the new fit is significantly more precise than the old one, and central values only agree to within a couple of standard deviations ($P \sim 2$), due to the relatively poor quality of the NMC data ($\chi^2 \sim 1.5$, see the discussion in section 4 above). In region (c) the new fit is rather accurate while the old fit had large errors, but $P \gg 1$ nevertheless, because the HERA rise of F_2 is outside the error bands extrapolated from NMC. This shows that even though errors on extrapolated data grow rapidly they become unreliable when extrapolating far from the data. Finally in region (d) all errors are very large and P is consequently small, except at small x and large Q^2, where the new fits extrapolate the rise in the HERA data, which is missing altogether in the old fits.

Let us finally come to the issue of the BCDMS error, which, as already mentioned, is considerably reduced in the current fit in comparison to the data and the previous fit. This may appear surprizing, in that the new fit does not contain any new data in the BCDMS region. However, as is apparent from fig. 4, this error reduction takes place in the GA training stage, when the full χ^2 eq. (3.4) is minimized. Furthermore, we have verified that if the uncorrelated $\chi^2_{\rm diag}$ eq. (3.3) is minimized during the GA training no error reduction is observed any longer for BCDMS: in such case we get $\langle \sigma^{(\rm net)} \rangle_{\rm dat} = 0.0055$ instead of the value given in table 6. Hence, no error reduction was observed for BCDMS in ref. [4] because there neural networks were trained by minimizing $\chi^2_{\rm diag}$. In fact, as discussed in section 4, this experiment turns out to require the longest time to learn, especially after inclusion of the HERA data. Error reduction only obtains after this lenghty minimization process.

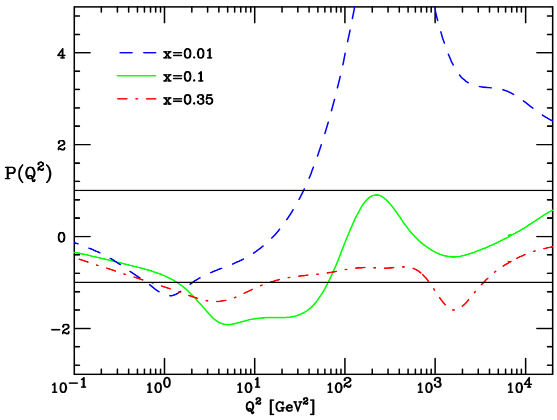

Figure 8: The relative pull eq. (5.1) of the new and old F_2 parametrizations. The one-σ band is also shown.

6. Conclusion

We have presented a determination of the probability density in the space of structure functions for the structure function $F_2(x, Q^2)$ for the proton, based on all available data from the NMC, BCDMS, E665, ZEUS and H1 collaborations. Our results take the form of a Monte Carlo sample of 1000 neural networks, each of which provides a determination of the structure function for all (x, Q^2). The structure function and its statistical moments (errors, correlations and so on) can be determined by averaging over this sample. Results are made available as a FORTRAN routine which gives $F_2(x, Q^2)$, determined by a set of parameters, and 1000 sets of parameters corresponding to the Monte Carlo sample of structure functions. They can be downloaded from the web site *http://sophia.ecm.ub.es/f2neural/*.

This work updates and upgrades that of ref. [4], where similar results were obtained from the BCDMS and NMC data only. The main improvements in the present work are related to the need of handling a large number of experimental data, affected by large correlated systematics. Apart from many smaller technical aspects, the main improvement introduced here is the use of genetic algorithms to train neural networks on top of backpropagation. This has allowed for a more accurate handling of correlated systematics.

Whereas the results of this paper may be of direct practical use for any application where an accurate determination of $F_2^p(x, Q^2)$ and its associate error are necessary, its main motivation is the development of a set of techniques which will be required for the construction of a full set of parton distributions with faithful uncertainty estimation based on the same method. This will be presented in a forthcoming publication.

Acknowledgments

We thank M. Arneodo, E. Rizvi, E. Tassi and F. Zomer for informations on the HERA data, and G. d'Agostini, L. Garrido and G. Ridolfi for discussions. This work has been supported by the project GC2001SGR-00065 and by the Spanish grant AP2002-2415.

A. Statistical estimators

We define various statistical estimators which have been used in the paper. The superscripts (*dat*), (art) and (net) refer respectively to the original data, to the $N_{\rm rep}$ Monte Carlo replicas of the data, and to the $N_{\rm rep}$ neural networks. The subscripts *rep* and *dat* refer respectively to whether averages are taken by summing over all replicas or over all data.

Replica averages

- Average over the number of replicas for each experimental point i

$$\left\langle F_i^{(\rm art)}\right\rangle_{\rm rep} = \frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}} F_i^{(\rm art)(k)}\,. \tag{A.1}$$

- Associated variance

$$\sigma_i^{(\rm art)} = \sqrt{\left\langle \left(F_i^{(\rm art)}\right)^2\right\rangle_{\rm rep} - \left\langle F_i^{(\rm art)}\right\rangle_{\rm rep}^2}\,. \tag{A.2}$$

- Associated covariance

$$\rho_{ij}^{(\rm art)} = \frac{\left\langle F_i^{(\rm art)} F_j^{(\rm art)}\right\rangle_{\rm rep} - \left\langle F_i^{(\rm art)}\right\rangle_{\rm rep}\left\langle F_j^{(\rm art)}\right\rangle_{\rm rep}}{\sigma_i^{(\rm art)}\sigma_j^{(\rm art)}}\,. \tag{A.3}$$

$$\mathrm{cov}_{ij}^{(\rm art)} = \rho_{ij}^{(\rm art)}\sigma_i^{(\rm art)}\sigma_j^{(\rm art)}\,. \tag{A.4}$$

- Mean variance and percentage error on central values over the $N_{\rm dat}$ data points.

$$\left\langle V\left[\left\langle F^{(\rm art)}\right\rangle_{\rm rep}\right]\right\rangle_{\rm dat} = \frac{1}{N_{\rm dat}}\sum_{i=1}^{N_{\rm dat}}\left(\left\langle F_i^{(\rm art)}\right\rangle_{\rm rep} - F_i^{(\rm exp)}\right)^2 \tag{A.5}$$

$$\left\langle PE\left[\left\langle F^{(\rm art)}\right\rangle_{\rm rep}\right]\right\rangle_{\rm dat} = \frac{1}{N_{\rm dat}}\sum_{i=1}^{N_{\rm dat}}\left[\frac{\left\langle F_i^{(\rm art)}\right\rangle_{\rm rep} - F_i^{(\rm exp)}}{F_i^{(\rm exp)}}\right]\,. \tag{A.6}$$

  We define analogously $\left\langle V\left[\langle\sigma^{(\rm art)}\rangle_{\rm rep}\right]\right\rangle_{\rm dat}$, $\left\langle V\left[\langle\rho^{(\rm art)}\rangle_{\rm rep}\right]\right\rangle_{\rm dat}$, $\left\langle V\left[\langle\mathrm{cov}^{(\rm art)}\rangle_{\rm rep}\right]\right\rangle_{\rm dat}$ and $\left\langle PE\left[\langle\sigma^{(\rm art)}\rangle_{\rm rep}\right]\right\rangle_{\rm dat}$.

- Scatter correlation:

$$r\left[F^{(\mathrm{art})}\right]=\frac{\left\langle F^{(\exp)}\left\langle F^{art}\right\rangle_{\mathrm{rep}}\right\rangle_{\mathrm{dat}}-\left\langle F^{(\exp)}\right\rangle_{\mathrm{dat}}\left\langle\left\langle F^{art}\right\rangle_{\mathrm{rep}}\right\rangle_{\mathrm{dat}}}{\sigma_s^{(\exp)}\sigma_s^{(\mathrm{art})}} \quad (A.7)$$

where the scatter variances are defined as

$$\sigma_s^{(\exp)} = \sqrt{\left\langle\left(F^{\exp}\right)^2\right\rangle_{\mathrm{dat}}-\left(\left\langle F^{\exp}\right\rangle_{\mathrm{dat}}\right)^2} \quad (A.8)$$

$$\sigma_s^{(\mathrm{art})} = \sqrt{\left\langle\left(\left\langle F^{art}\right\rangle_{\mathrm{rep}}\right)^2\right\rangle_{\mathrm{dat}}-\left(\left\langle\left\langle F^{art}\right\rangle_{\mathrm{rep}}\right\rangle_{\mathrm{dat}}\right)^2}\,. \quad (A.9)$$

We define analogously $r\left[\sigma^{(\mathrm{art})}\right]$, $r\left[\rho^{(\mathrm{art})}\right]$ and $r\left[\mathrm{cov}^{(\mathrm{art})}\right]$. Note that the scatter correlation and scatter variance are not related to the variance and correlation eqs. (A.2)–(A.4).

- Average variance:

$$\left\langle\sigma^{(\mathrm{art})}\right\rangle_{\mathrm{dat}}=\frac{1}{N_{\mathrm{dat}}}\sum_{i=1}^{N_{\mathrm{dat}}}\sigma_i^{(\mathrm{art})}\,. \quad (A.10)$$

We define analogously $\left\langle\rho^{(\mathrm{art})}\right\rangle_{\mathrm{dat}}$ and $\left\langle\mathrm{cov}^{(\mathrm{art})}\right\rangle_{\mathrm{dat}}$, as well as the corresponding experimental quantities.

Neural network averages

- Average χ^2 over nets

$$\left\langle\chi^2\right\rangle=\frac{1}{N_{\mathrm{rep}}}\sum_{k=1}^{N_{\mathrm{rep}}}\chi^{2(k)}\,, \quad (A.11)$$

where $\chi^{2(k)}$ is given by eq. (3.4).

- Mean variance and percentage error on central values over the N_{dat} data points.

$$\left\langle V\left[\left\langle F^{(\mathrm{net})}\right\rangle_{\mathrm{rep}}\right]\right\rangle_{\mathrm{dat}}=\frac{1}{N_{\mathrm{dat}}}\sum_{i=1}^{N_{\mathrm{dat}}}\left(\left\langle F_i^{(\mathrm{net})}\right\rangle_{\mathrm{rep}}-F_i^{(\exp)}\right)^2 \quad (A.12)$$

$$\left\langle PE\left[\left\langle F^{(\mathrm{net})}\right\rangle_{\mathrm{rep}}\right]\right\rangle_{\mathrm{dat}}=\frac{1}{N_{\mathrm{dat}}}\sum_{i=1}^{N_{\mathrm{dat}}}\left[\frac{\left\langle F_i^{(\mathrm{net})}\right\rangle_{\mathrm{rep}}-F_i^{(\exp)}}{F_i^{(\exp)}}\right]\,. \quad (A.13)$$

We define analogously percentage errors on the correlation and covariance.

- Scatter correlation

$$r\left[F^{(\mathrm{net})}\right]=\frac{\left\langle F^{(\exp)}\left\langle F^{(\mathrm{net})}\right\rangle_{\mathrm{rep}}\right\rangle_{\mathrm{dat}}-\left\langle F^{(\exp)}\right\rangle_{\mathrm{dat}}\left\langle\left\langle F^{(\mathrm{net})}\right\rangle_{\mathrm{rep}}\right\rangle_{\mathrm{dat}}}{\sigma_s^{(\exp)}\sigma_s^{(\mathrm{net})}}\,. \quad (A.14)$$

We define analogously $\left\langle\rho^{(\mathrm{net})}\right\rangle_{\mathrm{dat}}$ and $\left\langle\mathrm{cov}^{(\mathrm{net})}\right\rangle_{\mathrm{dat}}$.

- $\mathcal{R}$-ratio

$$\mathcal{R} = \frac{\langle\tilde{\chi}^2\rangle}{\langle\chi^2\rangle}\,, \tag{A.15}$$

where $\langle\chi^2\rangle$ is given by eq. (A.11) and

$$\left\langle\widetilde{\chi^2}\right\rangle = \frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}}\widetilde{\chi^2}^{(k)} \tag{A.16}$$

$$\widetilde{\chi^2}^{(k)} = \frac{1}{N_{\rm dat}}\sum_{i,j=1}^{N_{\rm dat}}\left(F_i^{({\rm net})(k)} - F_i^{({\rm exp})(k)}\right){\rm cov}^{-1}_{ij}\left(F_j^{({\rm net})(k)} - F_j^{({\rm exp})(k)}\right)\,. \tag{A.17}$$

References

[1] W.-K. Tung, *Status of global QCD analysis and the parton structure of the nucleon*, hep-ph/0409145.

[2] S. Forte, J.I. Latorre, L. Magnea and A. Piccione, *Determination of α_s from scaling violations of truncated moments of structure functions*, *Nucl. Phys.* **B 643** (2002) 477 [hep-ph/0205286].

[3] Spin Muon collaboration, B. Adeva et al., *A next-to-leading order QCD analysis of the spin structure function G_1*, *Phys. Rev.* **D 58** (1998) 112002.

[4] S. Forte, L. Garrido, J.I. Latorre and A. Piccione, *Neural network parametrization of deep-inelastic structure functions*, *JHEP* **05** (2002) 062 [hep-ph/0204232].

[5] See e.g. the HEPDATA database, http://durpdg.dur.ac.uk/HEPDATA/.

[6] New Muon collaboration, M. Arneodo et al., *Measurement of the proton and deuteron structure functions,F_2^p and F_2^d) and of the ratio σ_l/σ_t*, *Nucl. Phys.* **B 483** (1997) 3 [hep-ph/9610231].

[7] BCDMS collaboration, A.C. Benvenuti et al., *A high statistics measurement of the proton structure functions $F_2(x,Q^2)$ and r from deep inelastic muon scattering at high Q^2*, *Phys. Lett.* **B 223** (1989) 485.

[8] E665 collaboration, M.R. Adams et al., *Proton and deuteron structure functions in muon scattering at 470 GeV*, *Phys. Rev.* **D 54** (1996) 3006.

[9] ZEUS collaboration, M. Derrick et al., *Measurement of the F_2 structure function in deep inelastic e_+p scattering using 1994 data from the Zeus detector at HERA*, *Z. Physik* **C 72** (1996) 399 [hep-ex/9607002].

[10] ZEUS collaboration, J. Breitweg et al., *Measurement of the proton structure function F_2 and $\sigma_{tot}^{\gamma^*p}$ at low Q^2 and very low x at HERA*, *Phys. Lett.* **B 407** (1997) 432 [hep-ex/9707025].

[11] ZEUS collaboration, J. Breitweg et al., *Zeus results on the measurement and phenomenology of F_2 at low x and low Q^2*, *Eur. Phys. J.* **C 7** (1999) 609 [hep-ex/9809005].

[12] ZEUS collaboration, S. Chekanov et al., *Measurement of the neutral current cross section and F_2 structure function for deep inelastic e_+p scattering at HERA*, *Eur. Phys. J.* **C 21** (2001) 443 [hep-ex/0105090].

[13] ZEUS collaboration, J. Breitweg et al., *Measurement of the proton structure function F_2 at very low Q^2 at HERA*, *Phys. Lett.* **B 487** (2000) 53 [hep-ex/0005018].

JHEP03(2005)080

[14] H1 collaboration, C. Adloff et al., *A measurement of the proton structure function $F_2(x, Q^2)$ at low x and low Q^2 at HERA*, *Nucl. Phys.* **B 497** (1997) 3 [hep-ex/9703012].

[15] H1 collaboration, C. Adloff et al., *Measurement of neutral and charged current cross-sections in positron proton collisions at large momentum transfer*, *Eur. Phys. J.* **C 13** (2000) 609 [hep-ex/9908059].

[16] H1 collaboration, C. Adloff et al., *Deep-inelastic inclusive ep scattering at low x and a determination of α_s*, *Eur. Phys. J.* **C 21** (2001) 33 [hep-ex/0012053].

[17] H1 collaboration, C. Adloff et al., *Measurement of neutral and charged current cross sections in electron proton collisions at high Q^2*, *Eur. Phys. J.* **C 19** (2001) 269 [hep-ex/0012052].

[18] H1 collaboration, C. Adloff et al., *Measurement and QCD analysis of neutral and charged current cross sections at HERA*, *Eur. Phys. J.* **C 30** (2003) 1 [hep-ex/0304003].

[19] Particle Data Group collaboration, S. Eidelman et al., *Review of particle physics*, *Phys. Lett.* **B 592** (2004) 1.

[20] R. Barlow, *Asymmetric errors*, *eConf* **C030908** (2003) WEMT002 [physics/0401042]; *Asymmetric systematic errors*, physics/0306138.

[21] G. D'Agostini, *Bayesian reasoning in data analysis: a critical introduction*, World Scientific, Singapore, 2003.

[22] C. Peterson and T. Rognvaldsson, *An introduction to artificial neural networks*, lectures at the *1991 CERN School of Computing*, LU-TP-91-23.

B. Müller, J. Reinhardt and M.T. Strickland, *Neural Networks: an introduction*, Springer, Berlin, 1995);

G. Stimpfl-Abele and L. Garrido, *Fast track finding with neural nets*, *Comput. Phys. Commun.* **64** (1991) 46.

[23] J. Rojo and J.I. Latorre, *Neural network parametrization of spectral functions from hadronic tau decays and determination of QCD vacuum condensates*, *JHEP* **0401** (2004) 055 [hep-ph/0401047].

[24] W.T. Giele, S.A. Keller and D.A. Kosower, *Parton distribution function uncertainties*, hep-ph/0104052.

[25] S.I. Alekhin, *Global fit to the charged leptons dis data: α_s, parton distributions and high twists*, *Phys. Rev.* **D 63** (2001) 094022 [hep-ph/0011002].

JHEP03(2005)080