Eur. Phys. J. C 10, 395–403 (1999)
Digital Object Identifier (DOI) 10.1007/s100529900114

THE EUROPEAN PHYSICAL JOURNAL C

Extraction of parton distributions and α_s from DIS data within a Bayesian treatment of systematic errors

S.I. Alekhin

Institute for High Energy Physics, 142284 Protvino, Russia (e-mail: alekhin@mx.ihep.su)

Received: 5 May 1997 / Revised version: 12 October 1998 / Published online: 8 September 1999

Abstract. We have performed a NLO QCD global fit of BCDMS, NMC, H1 and ZEUS data with full account of point-to-point correlations using the Bayesian approach to the treatment of systematic errors. Parton distributions in the proton with their experimental uncertainties, including both statistical and systematic, are obtained. The gluon distribution in a wide region of x is found to be softer than the gluon distribution extracted in standard global analyses which include prompt photon data. We obtain a robust estimate of $\alpha_s(M_Z) = 0.1146 \pm 0.0036$ ($\geq 75\%$ C.L.) based on Chebyshev's inequality, which is compatible with an earlier determination of α_s from the DIS data, but is less dependent on high-twist effects.

1 Introduction

Recently it has been argued in [1] that the parton distribution functions (PDFs) obtained from global data analyses (e.g. [2–4]) have principal shortcoming because the errors of the parameters are not known. Indeed, a procedure often used to evaluate the spread of the predictions given by these PDFs is to compare the results of the calculations with various PDF parametrizations. It is evident that if different authors use the same theoretical model and similar data sets, this procedure cannot account for the real uncertainties originating from the statistical and systematic fluctuations of the data samples which are used in the extraction of the PDFs. These uncertainties should be evaluated by propagating these uncertainties into the errors of the PDF parameters. In modern experiments the dominating sources of uncertainty are usually the systematic errors. A complete treatment of the systematics is not possible if only the combined errors from various sources are presented in the publication. Fortunately, for the recent deep inelastic scattering (DIS) data from HERA, and for the data from previous experiments at the SPS, the full error matrices are available. Deep inelastic scattering of charged leptons remains the most direct source of information on PDFs and a careful analysis of these data including propagation of the systematics is valuable in exploring nucleon structure. The handling of statistical fluctuations is well understood on the basis of probability theory. In contrast, the handling of the systematic errors has been the subject of various approaches.

In one of the approaches, based on the classical treatment of probability, one considers the systematic shifts as additional unknown systematic error parameters arising from a lack of complete understanding of the experimental apparatus. Within this approach one usually tries to determine these parameters using some statistical estimator, e.g. χ^2, which is minimized as a function of these free parameters. The extracted values are then considered as a reasonable approximation to the true values of the systematic-error parameters, and the data are corrected for these systematic shifts. The systematic errors of the theoretical model parameters are evaluated by inverting the full error matrix, which includes both the physical and systematic-error parameter derivatives. In most cases, the only kinds of systematic errors which can be determined in this classical approach are the errors connected to the overall relative normalization of the data. Other systematic-error parameters are often strongly correlated with each other and also with the physical parameters. Consequently, as a result of these correlations, the fits to these systematic-error parameters can result in unreasonable central values with large errors.

The approach used in this paper is based on the Bayesian treatment of systematic uncertainties. In this approach these are considered as random variables with postulated or evaluated probability-distribution functions. The systematic errors are estimated within the general statistical procedures along with the statistical errors. In the analysis of modern DIS data, which generally have a number of significant systematic errors, this approach offers a unique possibility to account for point-to-point correlations in the data. We use this Bayesian approach to perform a complete propagation of the systematic uncertainties in the DIS data into the uncertainties of the resulting PDF parametrizations.

2 Theoretical and experimental input

2.1 Data samples used in the global fit

Cross-section data for the deep inelastic scattering of muons and electrons on hydrogen and deuterium targets

Table 1. The number of the data points (NDP) and χ^2/NDP for the analysed data sets. The number of the independent systematic errors (NSE) and the diagonalized average bias ($b^{\rm D}$) are also given

Experiment	BCDMS	NMC	H1	ZEUS	total
NDP	558	190	147	166	1061
χ^2/NDP	0.97	1.43	0.91	2.00	1.20
NSE	10	13	5	20	48
$b^{\rm D}$	-0.04	-0.05	0.23	0.20	0.02

[5–8] are used. The following cuts are used to reduce the contribution of higher-twist effects:

$$W \geq 4\,\mathrm{GeV}, \qquad Q^2 \geq 9\,\mathrm{GeV}^2,$$

where W and Q^2 are the usual DIS variables. The number of data points for each experiment after the cuts are given in Table 1. For the data from the ZEUS collaboration the asymmetric systematic errors have been averaged. For the BCDMS data we assume a complete correlation of systematic errors for the proton and deuterium cross sections. As is evident from the table, the total number of independent systematic errors is rather large, and it would be difficult to make a fit in the classical approach.

2.2 Probability model of the data

If experimental data with K sources of multiplicative systematic errors are explicitly described by a particular theoretical model, the data can be presented in the Bayesian approach as follows:

$$y_i = (f_i + \mu_i\sigma_i)\cdot\left(1+\sum_{k=1}^{K}\lambda_k s_i^k\right),$$

where $f_i = f_i(\theta^0)$ is the value predicted by the theoretical model with parameter θ^0, μ_i and λ_k are independent random variables, σ_i are the statistical errors and s_i^k are the systematic errors from the k-th source. Here, $i = 1\cdots N$, where N is the total number of points in the data set. If the data originate from a data sample with a large number of events per bin, then the μ values are normally distributed. As regards the λ parameters, the only assumption we make is that they have zero average value and dispersions which are equal to 1. Within this ansatz the individual measurements are correlated and the correlation matrix C_{ij} is given by

$$C_{ij} = \sum_{k=1}^{K} f_i s_i^k f_j s_j^k + \delta_{ij}\sigma_i^2,$$

where δ_{ij} is the Kronecker symbol. To obtain the estimator of the parameter θ^0 we minimize the quadratic form

$$\chi^2(\theta) = \sum_{i,j=1}^{N}[f_i(\theta)-y_i]E_{ij}[f_j(\theta)-y_j], \qquad (1)$$

where E_{ij} is the inverted correlation matrix. It should be noted that throughout this paper the normalization errors within this formalism are treated in the same way as the other systematic errors. All systematic errors are regarded as multiplicative, which is usually the case for counting experiments. The minimization is performed using the MINUIT package [9] augmented with the subroutines of [10] that improve the numerical stability of the calculations.

If the λ_k are normally distributed and $s_i^k \ll 1$, the values $\{y_i\}$ obey the multidimensional Gaussian distribution with correlations, and our estimator $\hat{\theta}$ has a minimal possible dispersion. One may believe that the systematic errors calculated from the propagation of the uncertainties in the parameters of the apparatus or in the Monte Carlo corrections are distributed in a Gaussian way. It was shown in [11] that even if this is not the case, the dispersion of our extracted parameters is smaller than the dispersion of the parameters extracted from the simple χ^2 classical approach without account of correlations. The statistical properties of the parameter estimators based on the covariance matrix of the measurements are discussed in [12]. The main conclusion of that paper is that these estimators can be biased even in the limit of large statistics. A bias arises if measured values (i.e. fluctuating values, rather than expected values) are used in the estimate of the covariance matrix. In this paper, since the covariance matrix is constructed using the predicted average values of the measurements, our estimators are asymptotically unbiased.

2.3 QCD input

The physical model for fitting the data samples is based on the parton model with a pQCD evolution of the gluon and light-quark distributions. The contributions of c-quark and b-quark are calculated using the LO formula[1] of [13], and setting $m_c = 1.5\,\mathrm{GeV}$ and $m_b = 4.5\,\mathrm{GeV}$. The renormalization/factorization scale is set to $(Q^2+4m_{c,b}^2)^{1/2}$. The initial PDFs are defined at the value of $Q_0^2 = 9\,\mathrm{GeV}^2$ and are evolved using NLO DGLAP equations [15] within the modified minimal subtraction ($\overline{\rm MS}$) factorization scheme [16]. The QCD evolution program has been tested as suggested in [17], and checked to a numerical precision of $O(0.1\%)$ in the kinematic region of the data included in this analysis. We start from the rather general and widely used expressions for the PDFs

$$xq_i(x,Q_0) = A_i x^{a_i}(1-x)^{b_i}(1+\gamma_1^i\sqrt{x}+\gamma_2^i x), \qquad (2)$$

and then reduce the number of free parameters, while keeping the quality of the fit to the data. The resulting expressions for the PDFs at Q_0 are

$$xu_V(x,Q_0) = \frac{2}{N_u^V}x^{a_u}(1-x)^{b_u}(1+\gamma_2^u x),$$

[1] We have checked that accounting for the NLO contribution to the heavy-quark production calculated in accordance with [14] changes the fitted parameters by less that one standard deviation

$$xu_S(x,Q_0) = \frac{A_S}{N_S}\eta_u x^{a_{su}}(1-x)^{b_{su}},$$

$$xd_V(x,Q_0) = \frac{1}{N_d^V}x^{a_d}(1-x)^{b_d},$$

$$xd_S(x,Q_0) = \frac{A_S}{N_S}x^{a_{sd}}(1-x)^{b_{sd}},$$

$$xs_S(x,Q_0) = \frac{A_S}{N_S}\eta_s x^{a_{ss}}(1-x)^{b_{ss}}.$$

$$xG(x,Q_0) = A_G x^{a_G}(1-x)^{b_G}.$$

We do not consider N_u^V, N_d^V and A_G as free parameters. These are calculated from the other parameters using number and momentum conservation for the partons. The value of N_S is defined by the relation

$$2\int_0^1 x\big[u_s(x,Q_0)+d_s(x,Q_0)+s_s(x,Q_0)\big]\mathrm{d}x = A_S.$$

After a few trial fits it is found that η_u for the sea is well compatible with unity and it is then fixed at this value. We fix $\eta_s = 0.5$ for the strange sea, which is compatible with CCFR findings [18] and fix $a_{su} = a_{sd} = a_{ss}$, $b_{ss} = (b_{su} + b_{sd})/2$, because the electron- and muon-scattering data used in this analysis do not allow for a separate determination of these parameters.

We obtain the strong coupling constant $\alpha_\mathrm{s}(Q)$ from the fitted parameter $\alpha_\mathrm{s}(M_Z)$ by numerically solving the NLO renormalization equation

$$\frac{1}{\alpha_\mathrm{s}(Q)} - \frac{1}{\alpha_\mathrm{s}(M_Z)} = \frac{\beta_0}{2\pi}\ln\left(\frac{Q}{M_Z}\right) + \beta\ln\left[\frac{\beta+1/\alpha_\mathrm{s}(Q)}{\beta+1/\alpha_\mathrm{s}(M_Z)}\right],$$

where

$$\beta_0 = 11 - \frac{2}{3}n_\mathrm{f}, \quad \beta = \frac{2\pi\beta_0}{51-\frac{19}{3}n_\mathrm{f}}.$$

This approach avoids the uncertainties originating from approximations based on the expansion in inverse powers of $\ln(Q)$. The uncertainties are ~ 0.001 if α_s is evolved from $O(\mathrm{GeV})$ to M_Z (cf. [19]), i.e. are comparable with the standard deviation of $\alpha_\mathrm{s}(M_Z)$. The number of the active fermions n_f is changed from 4 to 5 at $Q = m_b$, while keeping the continuity of $\alpha_\mathrm{s}(Q)$.

2.4 Corrections to the basic formulae and data

2.4.1 Target-mass correction (TMC)

In addition to the pure pQCD evolution we apply a target-mass correction [20] using the relation

$$F_2^{\mathrm{TMC}}(x,Q) = \frac{x^2}{\tau^{3/2}}\frac{F_2(\xi,Q)}{\xi^2} + 6\frac{M^2}{Q^2}\frac{x^3}{\tau^2}\int_\xi^1 \mathrm{d}z\frac{F_2(z,Q)}{z^2},$$

where

$$\xi = \frac{2x}{1+\sqrt{\tau}}, \quad \tau = 1+\frac{4M^2x^2}{Q^2}$$

and M is the nucleon mass. The contribution to this correction of the order of M^4/Q^4 as given in [20] is negligible in the kinematic range of the data used in this analysis. The target-mass correction is most significant for the BCDMS data, where it ranges from -1% to $+7\%$. The relative average value (RAV) of this correction, defined as the mean of the ratio of the absolute value of the correction to the statistical error, is 0.16 for the BCDMS data set. Note that our way of introducing this correction differs from the simple approximate ansatz $F_2^{\mathrm{TMC}}(x,Q) = F_2(\xi,Q)$ used in [21] (since we perform the complete integrals shown above). In our case the TMC exhibits cross-over from negative to positive values at $x \approx 0.5$ instead of $x \approx 0.4$ as in [21]. Our correction also differs in magnitude (with a maximal difference of about 4%). For the NMC data this correction is significantly smaller (range $[-1\%, 0\%]$, the RAV is 0.05). For the ZEUS and H1 data, the TMC correction is absolutely negligible.

2.4.2 Reduction to a common $R = \sigma_L/\sigma_T$

The usual practice in global fits of the DIS data is to use only the information on F_2 as experimental input. In this analysis, since the fit is performed to the measured cross sections, a model for the value of $R = \sigma_L/\sigma_T$ is needed. We calculate R as the sum of the NLO contribution from the light quarks and gluons, corrected for the target-mass effects[2]:

$$R(x,Q) = \frac{\frac{4M^2x^2}{Q^2}F_2^{\mathrm{TMC}}(x,Q) + F_L^{\mathrm{TMC}}(x,Q)}{F_2^{\mathrm{TMC}}(x,Q) - F_L^{\mathrm{TMC}}(x,Q)},$$

where

$$F_L^{\mathrm{TMC}}(x,Q) = F_L(x,Q) + \frac{x^2}{\tau^{3/2}}\frac{F_2(\xi,Q)}{\xi^2}(1-\tau)$$

$$+\frac{M^2}{Q^2}\frac{x^3}{\tau^2}(6-2\tau)\int_\xi^1 \mathrm{d}z\frac{F_2(z,Q)}{z^2},$$

$$F_L(x,Q) = \frac{\alpha_\mathrm{s}(Q)}{2\pi}\frac{8}{3}x^2\int_x^1\frac{\mathrm{d}z}{z^3}\big[F_2(z,Q)+(z-x)G(z,Q)\big].$$

The value of R is calculated iteratively in the fit, for every new set of PDF parameters (its final form is given in Fig. 1). Essentially, our procedure reduces the data on F_2 given in the original experimental papers to a common value of R. This reduction is most important at the smallest x region in each experiment. For the BCDMS data the value of the reduction correction is in the range of $[-3.5\%, 0\%]$, the RAV of the the R correction is 0.10. The BCDMS collaboration has calculated R from pQCD predictions, but used larger gluon distributions than in our final set. For the NMC data the RAV of this correction is 0.10 (range $[-1.5\%, 2\%]$). For the ZEUS data, which exhibit the highest sensitivity to the choice of R at some

[2] The contribution from c and b quarks to the value of R has a negligible effect on the results of the fit

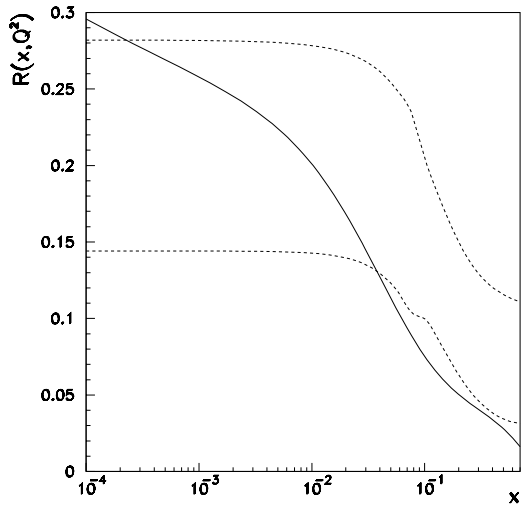

Fig. 1. $R = \sigma_L/\sigma_T$ calculated using our extracted PDF parameters (solid line) and the band of R^{1990}_{SLAC} of [22] (dashed lines) at $Q^2 = 9\,\mathrm{GeV}^2$

Table 2. The fitted PDF parameters with total experimental errors (both statistic and systematic)

Valence	a_u	0.745 ± 0.024
	b_u	3.823 ± 0.069
	γ_2^u	0.56 ± 0.28
	a_d	0.876 ± 0.066
	b_d	5.32 ± 0.22
Glue	a_G	-0.267 ± 0.043
	b_G	8.2 ± 1.5
Sea	A_S	0.159 ± 0.036
	a_{sd}	-0.1885 ± 0.0072
	b_{sd}	7.5 ± 1.3
	η_u	1.0 ± 0.12
	b_{su}	10.61 ± 0.96
	η_s	0.5 ± 1.0
	$\alpha_s(M_Z)$	0.1146 ± 0.0018

points due to the large span in the lepton scattering variable y, this correction ranges from -3% to 0% with the RAV of 0.04. The H1 data are affected to the same extent on average.

In summary, we note that although the R correction is not very large on average, it is significant for the data points on the edge of the experimental acceptance. Since at small x the value of R is heavily dependent on the gluon distribution, our approach imposes additional constraints on its value. The uncertainty in the value of the cross sections, which originates from the change of the radiative corrections for a different value of R, is usually accounted for in the corresponding systematic errors specified by the experimental groups, and the residual effects are believed to be small.

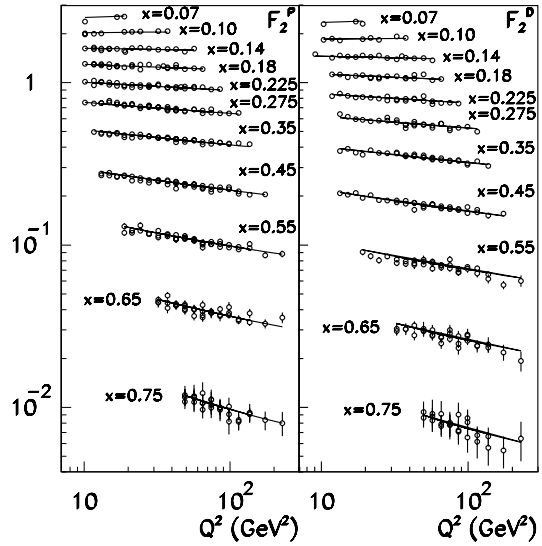

Fig. 2. The description of the BCDMS data with our PDFs. The data and curves are scaled by a factor of 1.2^{11-i}, where i runs from 1 for the highest x bin to 11 for the lowest one

2.4.3 Fermi motion correction for deuterium

The deuterium data are corrected for the Fermi motion using the procedure of [23] with the Paris wave function for deuterium [24]. This correction is also calculated iteratively to obtain a fully consistent set of PDFs. For the calculation of the relevant integrals we use a program [25], which exhibits a better numerical stability than the standard procedures based on the simple Gaussian algorithm. The value of $R = \sigma_L/\sigma_T$ for the deuteron was adopted to be unchanged under this correction; we have checked that this assumption is of minor importance for the final results. The Fermi motion correction is maximum at large x, ranging from -2% to $+15\%$ for the BCDMS data and from -2% to -1% for the NMC data, and its RAV is about 0.6 for both experiments.

3 Results

The fitted central values with total (statistical and systematic) experimental errors of the adjustable PDF parameters extracted from the minimization of (1) are given in Table 2. The full correlation error matrix is given in Table 3. To decrease the model dependence in our predictions, we calculate the covariance matrix of the fitted parameters for η_u and η_s as free parameters, although the values of these parameters used in the calculations have been kept intact, i.e. equal to 1 and 0.5, respectively.

The resulting χ^2 values are given in Table 1. On average the model describes the data fairly well. For a detailed analysis of the statistical consistency of the results we calculate the diagonalized residuals b_i^{D} using the relations

$$b_i^{\mathrm{D}} = \sum_{j=1}^{N} D_{ij}(f_j - y_j),$$

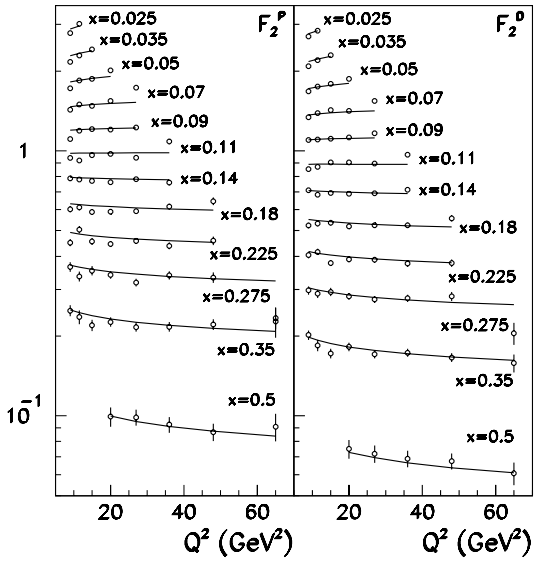

Fig. 3. The same as in Fig. 2 for the NMC data (i runs from 1 to 12). For presentation purposes only we show combined energy bins

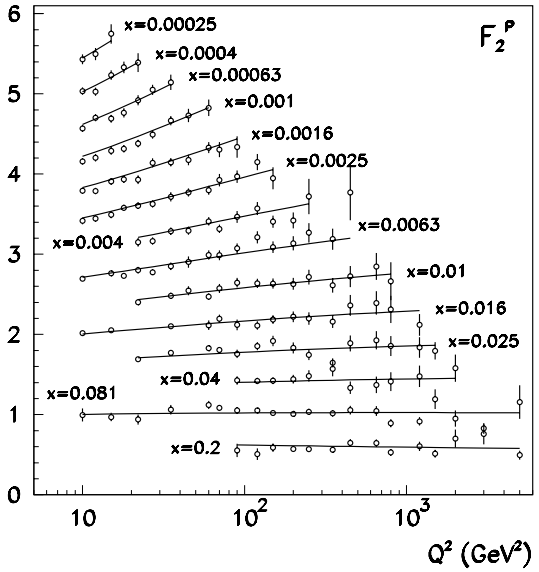

Fig. 5. The description of the ZEUS data with our PDFs. The data and curves are shifted by $4.5 - 0.3i$, where i runs from 1 for the highest x bin to 14 for the lowest one

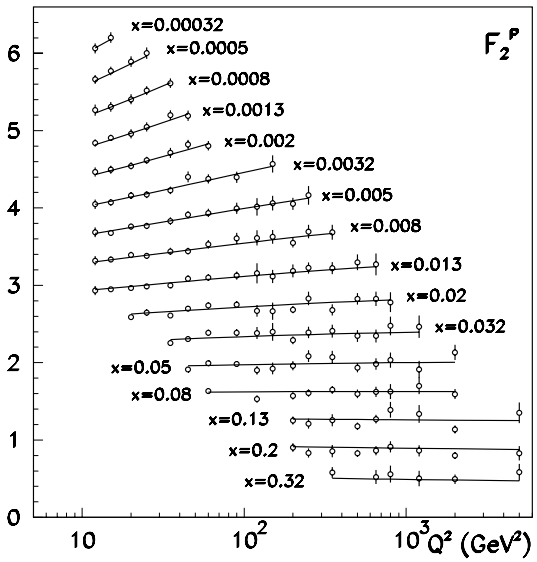

Fig. 4. The description of the H1 data with our PDFs. The data and curves are shifted by $5.1 - 0.3i$, where i runs from 1 for the highest x bin to 16 for the lowest one

where D_{ij} is the square root of the positively defined error matrix E_{ij}. It can be shown that under the assumptions made in Sect. 2.2 these residuals are not correlated, and have averages of zero with dispersions equal to 1. The distributions of these residuals for each experiment are given in Fig. 6. The average values are given in Table 1. The average of b^{D} over all experiments is well below its estimated standard deviation ($\approx 1/(\mathrm{NDP})^{1/2} \sim 0.03$), which supports the expectation that our estimator is asymptotically unbiased.

It can be seen that while the BCDMS and H1 residuals are in good agreement with the Gaussian distribution, the NMC and especially the ZEUS residuals show visible deviations from the curves. These discrepancies are a direct consequence of the poorer description of the NMC and ZEUS data by our PDF fit. Note that the averages of the residuals for ZEUS and NMC do not significantly contribute to χ^2. It is difficult to ascribe these deviations to the shortcoming of the theoretical model, because in this case one is to suppose that the ZEUS and NMC data have some additional fluctuations compared with the BCDMS and H1 data. (The two sets of data sets have a similar statistical significance and are in a similar kinematic region.) One possibility is that the errors are underestimated (but maybe remain Gaussian). The procedure used in similar cases by PDG [19] is to scale the errors so that χ^2/NDP equals unity. This recipe improves the situation with the NMC data, but is unsuccessful for ZEUS (see dashed curves on Fig. 6). Driven by this observations one can suppose that systematic errors of the ZEUS data (and maybe of the NMC data), follow non-Gaussian distributions (but with zero average values). In this case the extracted PDF parameters would not follow the Gaussian distribution, and the standard rule for the evaluated confidence level becomes invalid. We cannot be sure of the non-Gaussian nature of the experimental fluctuations. However, given the indications above, we prefer to use in the following a more robust estimate of the confidence level, based on the Chebyshev's inequality[3] [26]. It may be worthwhile to scale the errors of the ZEUS and NMC data. However, since these experiments have a number of independent sources of the systematics which contribute to the results of the fit in a nontrivial combination, this procedure is not straightforward.

[3] This approach can be recommended for most phenomenological investigations if the Gaussian distribution of the experimental values is not proven, e.g. in the case discussed in [27]

Table 3. The correlation matrix of the fitted PDFs parameters

	a_u	b_u	γ_2^u	a_d	b_d	A_S	a_{sd}	b_{sd}	η_u	b_{su}	η_s	a_G	b_G	$\alpha_s(M_Z)$
a_u	1.000	-0.731	-0.904	-0.365	-0.430	0.000	0.508	-0.352	0.389	0.576	-0.034	-0.260	-0.184	0.113
b_u	-0.731	1.000	0.917	0.201	0.258	0.012	-0.303	0.168	-0.251	-0.564	-0.020	0.198	0.250	-0.239
γ_2^u	-0.904	0.917	1.000	0.360	0.410	0.058	-0.425	0.298	-0.392	-0.695	0.040	0.262	0.260	-0.083
a_d	-0.365	0.201	0.360	1.000	0.965	-0.143	-0.057	0.933	-0.798	-0.614	-0.114	0.069	0.093	-0.064
b_d	-0.430	0.258	0.410	0.965	1.000	-0.148	-0.072	0.870	-0.702	-0.554	-0.105	0.059	0.080	-0.117
A_S	0.000	0.012	0.058	-0.143	-0.148	1.000	0.286	-0.335	0.091	0.273	0.988	0.391	0.667	0.503
a_{sd}	0.508	-0.303	-0.425	-0.057	-0.072	0.286	1.000	-0.119	0.270	0.533	0.291	-0.132	0.224	-0.007
b_{sd}	-0.352	0.168	0.298	0.933	0.870	-0.335	-0.119	1.000	-0.865	-0.617	-0.298	0.036	-0.003	-0.197
η_u	0.389	-0.251	-0.392	-0.798	-0.702	0.091	0.270	-0.865	1.000	0.725	0.101	-0.243	-0.209	0.019
b_{su}	0.576	-0.564	-0.695	-0.614	-0.554	0.273	0.533	-0.617	0.725	1.000	0.329	-0.180	-0.066	0.060
η_s	-0.034	-0.020	0.040	-0.114	-0.105	0.988	0.291	-0.298	0.101	0.329	1.000	0.372	0.640	0.483
a_G	-0.260	0.198	0.262	0.069	0.059	0.391	-0.132	0.036	-0.243	-0.180	0.372	1.000	0.784	0.109
b_G	-0.184	0.250	0.260	0.093	0.080	0.667	0.224	-0.003	-0.209	-0.066	0.640	0.784	1.000	0.006
$\alpha_s(M_Z)$	0.113	-0.239	-0.083	-0.064	-0.117	0.503	-0.007	-0.197	0.019	0.060	0.483	0.109	0.006	1.000

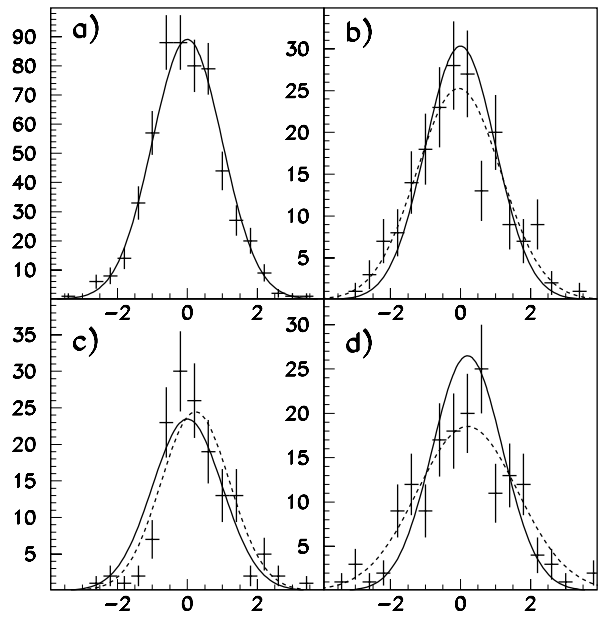

Fig. 6a–d. The distribution of the diagonalized residuals for the **a)** BCDMS, **b)** NMC, **c)** H1 and **d)** ZEUS experiments. The solid curves correspond to the Gaussian distribution with zero average and unit dispersion. The dashed curves correspond to the Gaussian in which the average and the dispersion are calculated from the distribution of the residuals. All curves are normalized to the number of entries in each histogram

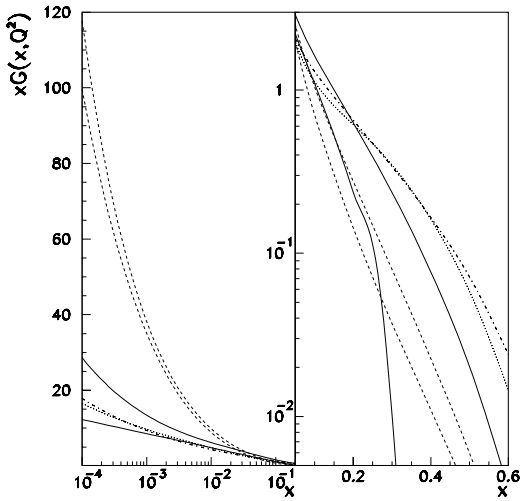

Fig. 7. The gluon distribution extracted in our analysis. The solid lines correspond to $Q^2 = 9\,\mathrm{GeV}^2$, the dashed ones to $Q^2 = 10,000\,\mathrm{GeV}^2$. The dotted line and dashed lines show the predictions of the MRS(R1) and CTEQ4M PDFs (at $Q^2 = 9\,\mathrm{GeV}^2$), respectively

The principal difference of our analysis from the other global fits is that we do not renormalize the data samples. Note that in other analyses the BCDMS data are usually shifted down. Therefore, our fits to F_2 are slightly higher at large x. All data for F_2 (corrected to the common value of R) are compared to our fit in Figs. 2–5. Here the error bars correspond to the sum of statistics and systematics (added in quadrature). Our fits are compared to a few selected set of standard PDFs in Figs. 7–10. The strange sea is not shown because only a weak upper limit can be extracted from this set of data. As previously mentioned, the distributions of our PDF parameters are defined mainly by the distributions of the systematic uncertainties. Therefore, there is a possibility that these may be non-Gaussian. In this case it is better to use Chebyshev's inequality to extract more robust estimates for the error bands. The bands given in the figures are two standard deviations contours, which correspond to a robust confidence level larger than 75%.

Prompt photon data were not used in our analysis. Such data are often used to constrain the gluon distribution at moderate x. In our analysis, the gluon distri-

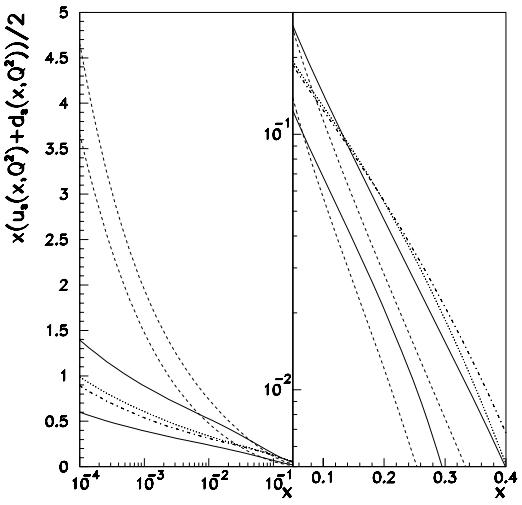

Fig. 8. The same as in Fig. 7 for the nonstrange sea

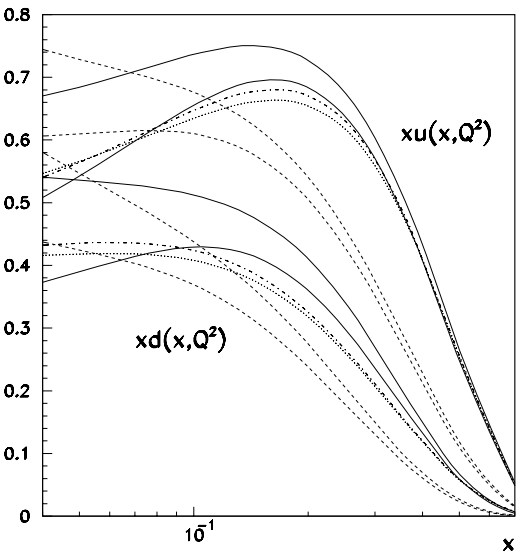

Fig. 9. The same as in Fig. 7 for the up and down quarks

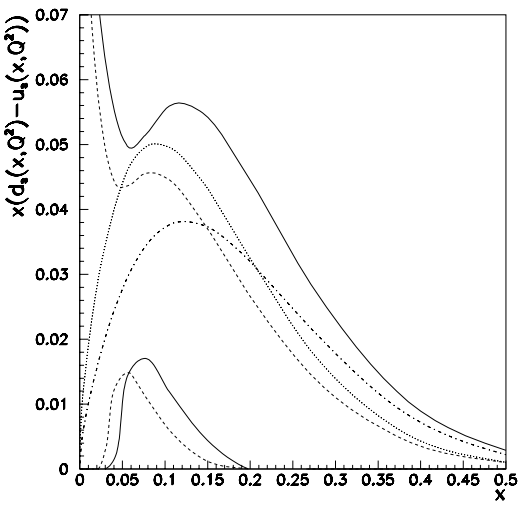

Fig. 10. The same as in Fig. 7 for the non-strange sea asymmetry

bution is determined rather well in the kinematic region $x = [0.0001, 0.5]$, with better precision than in an earlier analysis [28]. This has been achieved by incorporating data on F_2 at small x. The small x data define the gluon distribution in that region, and provide the momentum constraint to determine the distribution at larger x as well. The quark distributions are determined more precisely. However, it should be pointed out that the extracted PDFs and errors are model dependent. For example, if the condition $a_{su} = a_{sd} = a_{ss}$ is relaxed, it significantly increases the errors of the sea distribution at small x. An analogous effect arises if additional polynomial terms are added to the initial PDF parametrizations. This kind of model dependence is inevitable because it is not possible to determine a continuous functional form of the distributions with only a limited set of measurements (without additional constraints). The dependence on the functional form of the initial PDFs is more pronounced for the quark distributions than for the gluons or for $\alpha_s(M_Z)$. The main reason of this difference is that only the sum of the sea and the valence quarks is constrained by the F_2 measurements . Therefore, these data are more limited in the ability to discriminate between different PDFs. In contrast, the gluon distribution and $\alpha_s(M_Z)$ are defined by the derivatives of F_2, which are less sensitive to the variation in the contributions from the various quarks. Consequently, the correlation coefficients between PDF parameters corresponding to the quark distributions are generally larger than for the gluons (as shown in Table 3). At small x and large Q a shrinking in the error bands of the extracted gluon distribution is observed. This reflects the well-known property of the DGLAP kernel that the dominance for it of the singular terms leads to a focusing of any input gluon distribution to the universal form of [29].

For comparison the MRS(R1) [2] and CTEQ4M [4] parametrizations are also shown in the figures. This comparison is of limited importance because the other PDFs do not have any error bands. However, it can be assumed that the error bands for the MRS and CTEQ PDFs are smaller than for our analysis, because more experimental data have been included in the analyses by these two other groups. We observe a statistically significant difference between our gluon distribution with that from the MRS and the CTEQ fits at large x. This difference may be ascribed to the inclusion of prompt photon data in these other analyses. It has recently been recognized in [30] that the interpretation of the prompt photon data is suspect, because of large effects from multiple initial-state gluon radiations. An alternative analysis of the prompt photon data with the improved theoretical treatment of these effects [31] yields a much lower gluon distribution at moderate x, which is more compatible with ours. The discrepancies in the d quark and, to a lesser extent, the u quark distributions at moderate x may partially be explained by the influence of the target mass and the Fermi motion corrections. Taking into account the nucleon-binding effects in deuterium (this was not done in our analysis) can even more increase the d quark distribution at large x (see [32]). Meanwhile we can estimate that this increase would

be within the quoted errors, because the data used in our analysis do not span the region of very large x or have large errors there. It should be noted that the larger values of the quark distributions in this region of x may help to explain the excess of high transverse momentum jets in recent Fermilab collider data (above the NLO QCD predictions). This excess is observed in the $E_T = 200$–$400\,\mathrm{GeV}$ region, where the contribution from quark–quark scattering (viz. [2,4]) is important. It may be that some of the discrepancies also originate from possible numerical inaccuracies in the QCD evolution codes reported recently in [17] and in a difference between the extracted values of α_s. The difference in the sea distribution at $x \sim 0.3$ is not statistically significant and may disappear after the inclusion of more data in the analysis.

The robust estimate of the value of $\alpha_\mathrm{s}(M_Z)$ from this analysis of

$$\alpha_\mathrm{s}(M_Z) = 0.1146 \pm 0.0036 \ (\geq 75\%\ \mathrm{C.L.}),$$

is higher, but compatible with that extracted from the original SLAC-BCDMS analysis [33]. It is less sensitive to the higher-twist contribution because of the stringent cut on Q^2 and W. The estimate of $\alpha_\mathrm{s}(M_Z)$ has a weak model dependence, i.e. it is not changed much if the PDF functional form is changed from (2) to our final form. Our value of $\alpha_\mathrm{s}(M_Z)$ is lower than the world average. The difference is not very significant, especially if one uses a robust estimate. At the same time it may indicate additional systematic errors on the BCDMS data, since these data basically define the value of α_s.

4 Conclusion

The Bayesian treatment of systematic errors is a clear and efficient method for the analysis of data with numerous sources of systematic errors, and for DIS scattering data in particular. The approach allows for the straightforward and correct account of point-to-point correlations as opposed to the widely used "simplification", which consists of combining statistical and systematic errors in quadratures. The concerns that the estimator using the covariance matrix suffers from possible bias is not relevant if the covariance matrix is constructed using the predicted averages (instead of the measured quantities).

This is the first time that quark and gluon distributions have been extracted from a global fit with full account of all statistical and systematic experimental errors. These extracted PDFs are useful in further phenomenological studies. Having a procedure that yields estimates of PDF error bands enables us to perform comparisons of various global fits, and of PDFs extracted from different processes. In addition, the calculation of theoretical cross sections for various processes (which are based on PDFs) are more meaningful if one is able to correctly account for the uncertainties from the PDF parametrizations.

Acknowledgements. I am indebted to R.M. Barnett and P.S. Gee for providing the computing facilities, and to R. Eichler and J. Feltesse for the valuable comments on the experimental data. I thank Dr. L.W. Whitlow for providing the code for the calculation of R^{1990}_{SLAC}, A.S. Nikolaev for the code to calculate the Paris wave function of the deuteron, and S. Riemersma for the code to calculate the coefficient functions for heavy-quark production. Special thanks are to A. Bodek for careful reading of the manuscript and valuable comments. The work was partially supported by RFFI grant 96-02-18897.

References

1. D.E. Soper, J.C. Collins, Report No. CTEQ NOTE 94/01, hep-ph/9411214, (1994); D.E. Soper, talk given at 30th Moriond, Meribel les Allues, France, hep-ph/9506218, (1995)
2. A.D. Martin, R.G. Roberts, W.J. Stirling, Phys. Lett. B **387**, 419 (1996)
3. M. Glück, E. Reya, A. Vogt, Z. Phys. C **67**, 433 (1995)
4. CTEQ collaboration, H.L. Lai et al., Phys. Rev. D **55**, 1280 (1997)
5. BCDMS collaboration, A.C. Benvenuti et al., Phys. Lett. B **223**, 485 (1989); BCDMS collaboration, A.C. Benvenuti et al., Phys. Lett. B **237**, 592 (1990)
6. NM collaboration, M. Arneodo et al., Nucl. Phys. B **483**, 3 (1997)
7. ZEUS collaboration, M. Derrick et al., Z. Phys. C **72**, 399 (1996)
8. H1 collaboration, S. Aid et al., Nucl. Phys. B **470**, 3 (1996)
9. F. James, M. Roos, CERN Program Library Long Writeup D 506, Version 92.1, (1992)
10. S.I. Alekhin, Report No. IHEP 94-70, CERN-SCAN-9501137, (1994)
11. S.I. Alekhin, Report No. IHEP 95-65, CERN-SCAN-9508274, (1995); Report No. IHEP 95-48, CERN-SCAN-9511190, (1995)
12. G. D'Agostini, Nucl. Instrum. Methods A **346**, 306 (1994)
13. M. Glück, E. Reya, M. Stratmann, Nucl. Phys. B **422**, 37 (1994)
14. E. Laenen, S. Riemersma, J. Smith, W.L. van Neerven, Nucl. Phys. B **392**, 162 (1993)
15. V.N. Gribov, L.N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972); V.N. Gribov, L.N. Lipatov, Sov. J. Nucl. Phys. **15**, 675 (1972); G. Altarelli, G. Parisi, Nucl. Phys. B **126**, 298 (1977); Yu.L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977)
16. W. Furmanski, R. Petronzio, Z. Phys. C **11**, 293 (1982); Phys. Lett. B **97**, 437 (1980); G. Curci, W. Furmanski, R. Petronzio, Nucl. Phys. B **175**, 27 (1980)
17. J. Blümlein et al., Report No. DESY 96-199, hep-ph/9609400 (1996)
18. CCFR collaboration, A.O. Bazarko et al., Z. Phys. C **65**, 189 (1995)
19. Particle Data Group, R.M. Barnett et al., Phys. Rev. D **54**, 77 (1996)
20. H. Georgi, H.D. Politzer, Phys. Rev. D **14**, 1829 (1976)
21. A.D. Martin, R.G. Roberts, W.J. Stirling, Phys. Rev. D **51**, 4756 (1995)
22. L.W. Whitlow et al., Phys. Lett. B **250**, 193 (1990)
23. G.B. West, Ann. Phys. (NY) **74**, 464 (1972)
24. M. Lacombe et al., Phys. Rev. 1980, C **21**, 861 (1980); M. Lacombe et al., Phys. Lett. B **101**, 139 (1981)

25. S.N. Sokolov, Report No. IFVE 88-110 (1988)
26. W.T. Eadie, D. Drijard, F.E. James, M. Roos, B. Sadoulet, Statistical Methods in Experimental Physics (North-Holland, Amsterdam, 1971)
27. A.P. Bukhvostov, hep-ph/9705387, (1997)
28. NM collaboration, M. Arneodo et al., Phys. Lett. B **309**, 222 (1993)
29. L.V. Gribov, E.M. Levin, M.G. Ryskin, Phys. Rep. C **100**, 1 (1981)
30. CTEQ collaboration, J. Huston et al., Phys. Rev. Lett. **77**, 444 (1996)
31. W. Vogelsang, A. Vogt, Nucl. Phys. B **453**, 334 (1995)
32. U.K. Yang, A. Bodek, Report No. UR-1543, ER-40685-929, hep-ph/9809480, (1998)
33. M. Virchaux, A. Milsztajn, Phys. Lett. B **274**, 221 (1992)