SI - 97 - 12

Unfolding the Photon Structure Function†

Klaus Affholderbach and Glen Cowan

Universität Siegen, Fachbereich Physik, D-57068 Siegen, Germany

Abstract. Because of finite acceptance and resolution, the measured distribution of the hadronic mass W in single-tag $\gamma\gamma$ collisions (or equivalently, the distribution of $x = Q^2/(Q^2 + W^2)$) is distorted with respect to what one would obtain with a perfect detector. The dominant uncertainties in current measurements of the x distribution are related to the procedure that must be applied to correct for these distortions (unfolding). Specifically, large uncertainties stem from the model dependence of the detector response matrix, which relates the true and observed distributions. In this paper we discuss the unfolding problem and investigate some ideas for reducing the model dependence of unfolded x distributions in $\gamma\gamma$ collisions.

The large integrated luminosity expected from the LEP II programme will provide new opportunities to investigate $\gamma\gamma$ collisions. Single-tag events, where one electron remains in the beam pipe and the other is measured with an angle $\theta_{\rm tag}$ and energy $E_{\rm tag}$, can be viewed as the deep inelastic scattering of an electron and an (almost) real target photon. Of particular interest is the joint distribution of the negative four-momentum squared of the probing photon, $Q^2 = 4E_{\rm tag}E_{\rm beam}\sin^2(\theta_{\rm tag}/2)$, and the invariant mass of the hadronic system, W. Equivalently, one usually measures the distribution Q^2 and the variable $x = Q^2/(Q^2+W^2)$; this can be directly related to the photon structure function $F_2^\gamma(x, Q^2)$. (The formalism of single-tag $\gamma\gamma$ collisions is described in e.g. [1]). In this paper we concentrate on the question of measuring the distribution of x for single-tag $\gamma\gamma$ collisions within a given narrow range of Q^2.

The dominant uncertainties in current measurements of the x distribution are related to corrections that must be introduced to account for finite acceptance and resolution of the detector. These effects stem mainly from hadrons at low angles with respect to the beam line that escape detection. This results in a lower measured hadronic mass compared to its true value, and hence in a corresponding distortion of the variable x. This is in contrast to the situation with Q^2, which is entirely determined by the tag electron's angle and energy. It can therefore be measured to within several percent, and

† Contribution to the Proceedings of the Phenomenology Workshop on LEP 2 Physics, Oxford, April 14 – 18, 1997, submitted to Journal of Physics G.

corrections for resolution effects do not pose a major problem. Here we will concentrate on unfolding the distribution of x.

A detailed description of unfolding problems can be found in [2]. We can represent a sample of measured values of x by means of a histogram with M bins. The expectation values of the numbers of events $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)$ that would be obtained with a perfect detector are thus the parameters which we want to estimate. What we obtain from the experiment is a histogram $\mathbf{n} = (n_1, \ldots, n_N)$, which is distorted with respect to $\boldsymbol{\mu}$ both because of statistical fluctuations as well as from the effects of acceptance and resolution. The latter can cause an event with a true value of x in bin i to be observed in some different bin j. (Note that in general the M bins for the true histogram $\boldsymbol{\mu}$ need not be the same as the N bins of the observed histogram $\mathbf{n}$. In the following, however, we will use the same binning for both.)

The number of entries n_i observed in a given bin i can be treated as a Poisson variable with expectation value $\nu_i = E[n_i]$. The vectors $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are related by

$$\boldsymbol{\nu} = R\boldsymbol{\mu} + \boldsymbol{\beta} \,, \tag{1}$$

where the matrix element R_{ij} represents the probability for an event to be observed in bin i given that its true x value was in bin j, and the vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_N)$ gives the expected background. Here for simplicity we will neglect the background, and thus equation (1) becomes $\boldsymbol{\nu} = R\boldsymbol{\mu}$. Note that a true value in bin j need not lead to any measured value at all, i.e. the efficiencies

$$\varepsilon_j = \sum_{i=1}^{N} R_{ij} \tag{2}$$

are in general less than unity.

If we had the vector $\boldsymbol{\nu}$, then we could simply invert R to obtain $\boldsymbol{\mu} = R^{-1}\boldsymbol{\nu}$. What we have instead, however, are the data values $\mathbf{n}$, which are subject to random fluctuations. The estimators $\hat{\boldsymbol{\mu}} = R^{-1}\mathbf{n}$ are unbiased, but have extremely large variances. (In the following, estimators will be denoted by hats.) These are in fact the estimators that one obtains by maximizing the log-likelihood function based on Poisson-distributed data,

$$\log L(\boldsymbol{\mu}) = \sum_{i=1}^{N} \log \left(\frac{\nu_i^{n_i}}{n_i!} \, e^{-\nu_i} \right) \,, \tag{3}$$

which becomes

$$\log L(\boldsymbol{\mu}) = \sum_{i=1}^{N} (n_i \log \nu_i - \nu_i) \tag{4}$$

after dropping terms not depending on the parameters. (Note that this is regarded as a function of $\boldsymbol{\mu}$, since one has $\boldsymbol{\nu} = R\boldsymbol{\mu}$.)

The idea behind unfolding is to construct estimators $\hat{\boldsymbol{\mu}}$ with much smaller variances than the maximum likelihood estimators, at the cost of introducing a small bias. In regularized unfolding, this is done by choosing the smoothest solution (according to some criterion) out of those for which the log-likelihood is within some $\Delta \log L$ of its maximum value. This is equivalent to maximizing a linear combination of $\log L(\boldsymbol{\mu})$ and a regularization function $S(\boldsymbol{\mu})$,

$$\varphi(\boldsymbol{\mu}, \lambda) = \alpha \log L(\boldsymbol{\mu}) \,+\, S(\boldsymbol{\mu}) \,+\, \lambda \left[n_{\mathrm{tot}} - \sum_{i=1}^{N} \nu_i \right] , \tag{5}$$

with respect to the parameters $\boldsymbol{\mu}$ and the Lagrange multiplier λ. The function $S(\boldsymbol{\mu})$ must be defined to reflect the smoothness of the solution. The regularization parameter α determines the trade-off between likelihood and smoothness, and can be chosen to correspond to a given value of $\Delta \log L = \log L_{\max} - \log L(\boldsymbol{\mu})$. The final term in φ restricts the solution to satisfy $\sum_i \nu_i = \sum_{i,j} R_{ij} \mu_j = n_{\mathrm{tot}}$, where n_{tot} is the total number of events observed.

A commonly used regularization function is based on the mean squared second derivative of the unfolded distribution (Tikhonov regularization). Computer programs based on this technique have been used in previous structure function measurements [3, 4]. The regularization function can be implemented by approximating the second derivative of the distribution using finite differences. This leads to a function of the form

$$S(\boldsymbol{\mu}) = - \sum_{i,j=1}^{M} G_{ij} \, \mu_i \, \mu_j \,, \tag{6}$$

where coefficients G_{ij} must be determined so as to take into account differences in the bin widths; cf. [2]. Another possible regularization function is

$$S(\boldsymbol{\mu}) = - \sum_{j=1}^{M} \mu_j \log \frac{\mu_j}{\mu_{\mathrm{tot}}} = -\mu_{\mathrm{tot}} \sum_{j=1}^{M} p_j \log p_j = \mu_{\mathrm{tot}} H \,, \tag{7}$$

where $\mu_{\mathrm{tot}} = \sum_{i=1}^{M} \mu_i$, $p_j = \mu_j / \mu_{\mathrm{tot}}$, and H is the Shannon entropy. The entropy-based regularization function makes no reference to the relative locations of any of the bins, which has the advantage that the bins at the edges of the distribution are treated on the same footing as those in the middle. In addition, use of (7) (referred to in the following as MaxEnt) can be directly applied to multidimensional distributions.

One can show that the estimators $\hat{\boldsymbol{\mu}}$ obtained from any regularized unfolding technique are biased. An additional systematic error is related to the model dependence of the response matrix R. This matrix must in general be determined by a Monte Carlo calculation where events are generated and processed by a detector simulation program.

By construction, the conditional probability for an event to be observed at x' given that it was generated at x is independent of the distribution of x. The corresponding statement for finite bins in x is approximately true as long as the bins are not too large.

The true value of x is not, however, the only variable that has an influence on the probability to measure a given value x'. For example, if the hadrons are mostly at low angles with respect to the beam line, then the resolution for x will be poor, since on average more particles will be lost. Since different models have in general different distributions for all of the variables that characterize the final state hadrons, each will lead to somewhat different response matrices R, and hence to different results for the unfolded x distribution.

A method to reduce this model dependence is to measure for each event not only x, but in addition some other variable y (to be defined) that also characterizes the final state. The response matrix gives the probability for an event with true values of x and y in bins i and j to be observed in bins k and l. This matrix is now by construction independent of the model's joint distribution of x and y. In principle this idea could be extended to an arbitrary number of additional variables, and eventually one would eliminate all model dependence from the response matrix. Since the individual cells of the multidimensional space must be populated with enough Monte Carlo events to determine the response matrix, however, the total number of bins, and hence the number of variables, is necessarily limited. Here we will only consider the case of two measured variables.

In order to reduce the model dependence of R, the second variable y should be related to the detector's ability to measure x. Here, we consider $y = |\cos\theta_{\text{had}}|$, where θ_{had} is the polar angle of the total hadronic system in the laboratory frame. In fact, large discrepancies between observed and predicted angular distributions of hadrons have been observed in single-tag events [5]. Thus some reduction in the uncertainty from model dependence can be expected by obtaining this information directly from the data. Other possibilities are the orientation of the hadronic system with respect to the $\gamma\gamma$ axis, or a quantity related to hard gluon radiation, such as the thrust or jet structure of the hadrons measured in the $\gamma\gamma$ rest frame. After unfolding the distribution of x and $|\cos\theta_{\text{had}}|$, we will integrate over $|\cos\theta_{\text{had}}|$ to obtain the distribution of x alone.

In order to investigate these ideas quantitatively, single-tag $\gamma\gamma$ events were generated with the HERWIG Monte Carlo generator [6]. This allows the user to choose from a variety of parametrizations for the photon structure function from the package PDFLIB [7].

The response matrix was determined by means of a highly simplified detector simulation program, roughly corresponding to a typical LEP detector. An electromagnetic calorimeter was assumed to be sensitive to photons and electrons with energies $E > 200$ MeV, to be hermetic down to 30 mrad from the beam line, and was taken to have an energy resolution of $\sigma_E = 0.18\sqrt{E}$. A hadron calorimeter in the angular range $|\cos\theta| < 0.99$ was assumed to be fully efficient for energies above 1.0 GeV, and to have an energy resolution of $\sigma_E = 0.80\sqrt{E}$. Charged particles with $|\cos\theta| < 0.94$ and transverse momentum $p_\perp > 150$ MeV are measured with a resolution of $\sigma_{p_\perp}/p_\perp = 0.003 \oplus 0.0008 \cdot p_\perp$. It was assumed that electrons with energies of at least 10 GeV and $\theta > 30$ mrad could be identified.

First, large samples of single-tag events (approximately 12 fb^{-1}) were generated at a centre-of-mass energy of $E_{\rm cm} = 184$ GeV with HERWIG using the GRV LO [8], SaS1D [9] and LAC1 [10] structure functions. True single-tag events were defined to be those with a tag electron in the range $60 < \theta < 100$ mrad, which for the GRV structure functions gives a mean Q^2 of 40.0 GeV^2 and a cross section of 11.6 pb. These events were processed by the detector simulation routine in order to determine a response matrix R based on each of the three sets of structure functions.

Scatter plots of the true and observed values of x are shown in Fig. 1 for two different ranges of $|\cos\theta_{\rm had}|$. In Fig. 1(a), the (true) hadronic system has an angle of at least 60° with the beam line. This results in a relatively good measurement the hadronic mass W, and hence in a good resolution for x. In Fig. 1(b), the hadronic system has $|\cos\theta_{\rm had}| > 0.9$, so that a larger fraction of the hadrons escape detection. This leads to a visibly weaker correlation between the measured and true values of x.

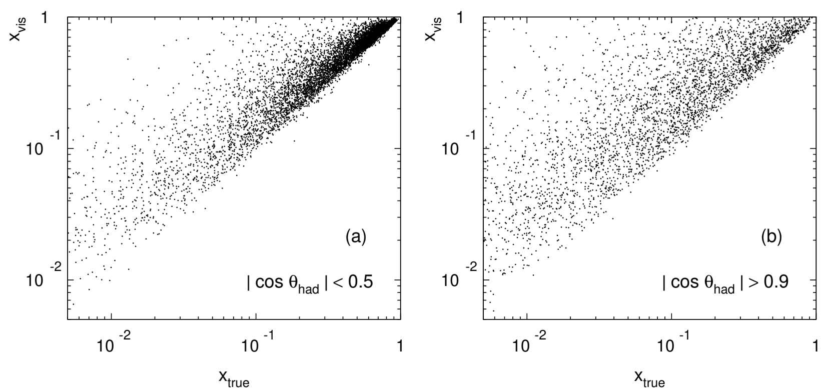

Figure 1. Scatter plots of the visible and true values of x from the HERWIG Monte Carlo (with GRV structure function) and a simplified detector simulation, for two different ranges of the angle of the hadronic system with respect to the beam line.

The effect of the direction of the hadronic system on the x resolution can also be seen in Fig. 2, showing (a) the mean and (b) the standard deviation σ_x of $x_{\rm vis}$ as a function of $x_{\rm true}$ for the same ranges of $\cos\theta_{\rm had}$ as in Fig. 1. For x less than around 0.3, the typical x resolution is between 0.12 and 0.18 for $|\cos\theta_{\rm had}| > 0.9$, but is around 0.08 – 0.13 for $|\cos\theta_{\rm had}| < 0.5$. Thus if two models differ in their distributions of $\cos\theta_{\rm had}$, they will result in different response matrices for x.

In order to minimize the errors in the unfolded result, the bin size should not be too much smaller than the resolution. The bins should be sufficiently small, however, so that the resolution is approximately constant over a bin. At low x, these two considerations come into conflict. The bin boundaries for x were chosen to be

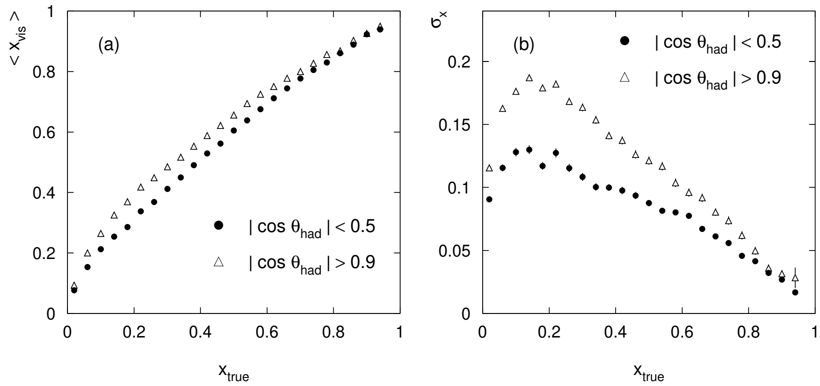

Figure 2. (a) The mean and (b) the standard deviation of the measured x distribution as a function of the true value of x, for events in two ranges of $\cos\theta_{\rm had}$.

$[0.0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1.0]$ (8 bins). The resolution for $|\cos\theta_{\rm had}|$ is around $0.20 - 0.25$ for $|\cos\theta_{\rm had}| < 0.5$, and improves to around 0.18 as $|\cos\theta_{\rm had}|$ increases. The bins for $|\cos\theta_{\rm had}|$ were chosen to be $[0.0, 0.3, 0.5, 0.7, 0.8, 0.9, 1.0]$ (6 bins).

To test the unfolding procedure, a smaller independent sample of events was generated using the GRV structure functions and processed by the detector simulation routine; the resulting events were treated as real data. The test sample corresponds to 207 pb^{-1}, and consists of 2444 true single-tag events (according to the definition above). Of these, 1697 passed relatively loose event selection criteria, having a tag electron measured in the range $60 < \theta < 100$ mrad, no electron on the opposite side above 30 mrad, and in addition, at least three charged particle tracks.

Figure 3(a) shows the distribution dN/dx based on large samples of data generated with the GRV, SaS1D and LAC1 structure functions, normalized to the expected number of entries for an integrated luminosity of 207 pb^{-1}. Figure 3(b) shows the GRV predicted distribution of $x_{\rm true}$, $dN_i/dx = \mu_i/\Delta x_i$, where Δx_i is the width of bin i. Also shown are the expectation values of what one would observe including the effects of detector acceptance and resolution, corresponding to the histogram $\boldsymbol{\nu}$. The histogram of $x_{\rm vis}$ from the simulated data, i.e. the vector $\mathbf{n}$ divided by the bin widths, is shown as points with error bars. These are subject to both detector effects and statistical fluctuations.

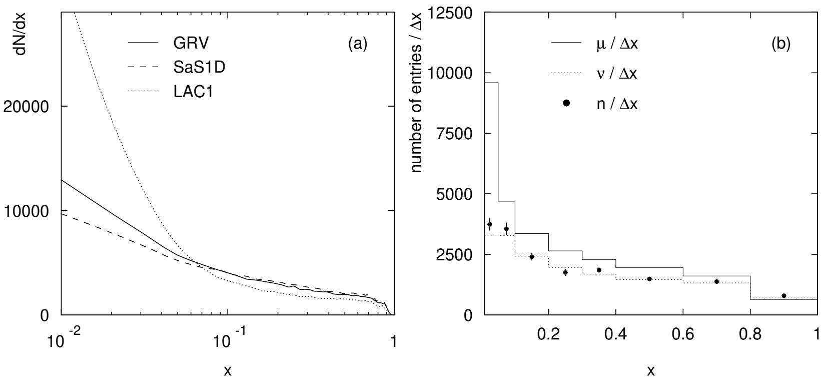

Figure 3. (a) Distributions dN/dx from different structure functions. (b) The expected dN/dx from HERWIG with GRV (solid histogram), the expectation values after detector smearing (dotted histogram) and the observed number of entries from the test data sample (points).

First, the difference between Tikhonov and MaxEnt regularization was investigated using the test data sample and response matrix both based on GRV. Figure 4 shows the result of a one-dimensional unfolding of the x distribution, with statistical error bars, using (a) Tikhonov and (b) entropy-based regularization functions. The regularization parameter α was determined in both cases such that the biases of the estimators, $b_i = E[\hat{\mu}_i] - \mu_i$, are consistent with zero within their own statistical errors. This is done by constructing estimators $\hat{b}_i$ for the biases; cf. [2]. Here this criterion leads to larger errors for the MaxEnt case. Because of this fact, however, MaxEnt gives a better level of agreement between the estimated and true distributions within the statistical errors. The bias in the Tikhonov result can be reduced by increasing the regularization parameter, in which case it becomes similar to that shown here from MaxEnt. The prescription for setting α is by no means unique, and a corresponding systematic uncertainty should be assigned to the final result. A reasonable measure of this uncertainty is given by the estimated bias.

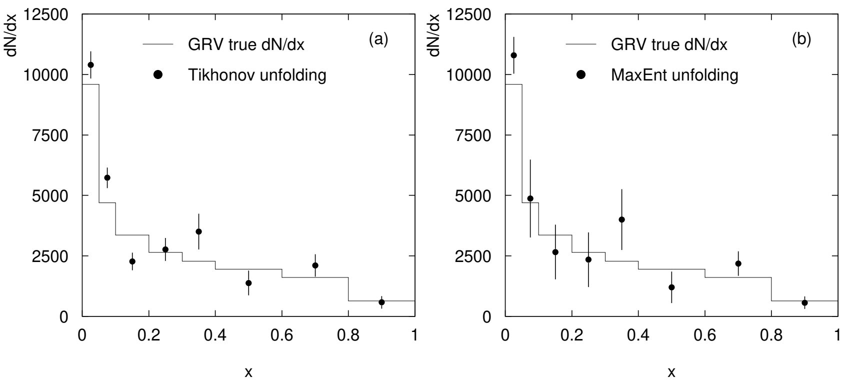

Figure 4. Unfolded distributions dN/dx using (a) Tikhonov and (b) entropy-based regularization functions.

Next, the question of model dependence of the response matrix was investigated using MaxEnt regularization. Figure 5(a) shows results obtained using response matrices from HERWIG with GRV, SaS1D and LAC1 structure functions, where the test data sample was in all three cases based on GRV. All of the results are in reasonable agreement with the true values within their statistical errors. From the ratio of unfolded to true distributions in Fig. 5(b), one can see that the unfolded result indeed depends on which response matrix is used.

Figure 6 shows the result of a two-dimensional unfolding of the variables x and $|\cos\theta_{\mathrm{had}}|$, again using the entropy-based regularization function, and the same criterion as before for determining the regularization parameter. Again, all of the results are in reasonable agreement with the true values within the statistical errors. Here, however, the spread between the points for different models is reduced with respect to what was obtained from the one-dimensional unfolding. The differences between results from GRV and SaS1D matrices differ by 1 – 5%, except for a 9% change for the second bin; these shifts are much smaller than the relative statistical errors of the unfolded $\hat{\mu}_i$.

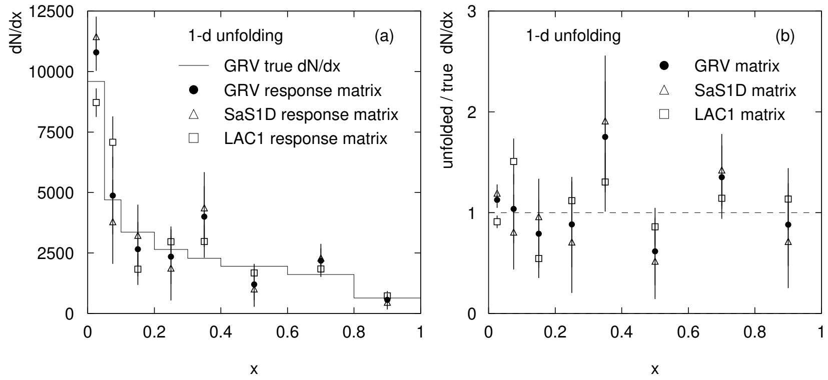

Figure 5. Results from a one-dimensional unfolding of the variable x. (a) The true and unfolded distributions and (b) their ratios.

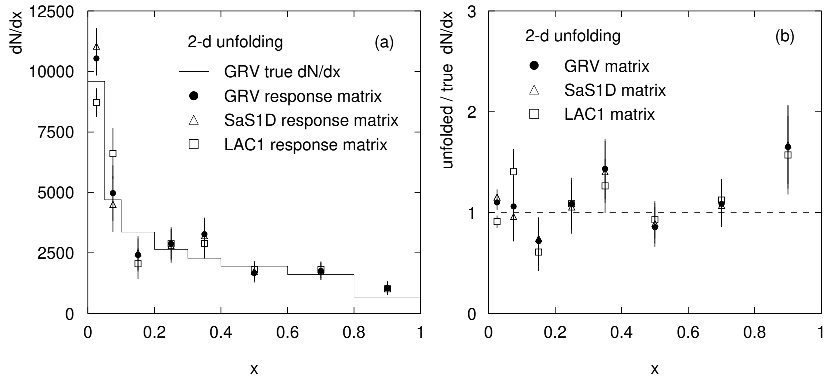

Figure 6. Results of a two-dimensional unfolding of the variables x and $\cos\theta_{\mathrm{had}}$. (a) The true and unfolded distributions and (b) their ratios.

It is not entirely clear why the entropy-based regularization gives smaller statistical errors in the two-dimensional case than when unfolding in one dimension. Although the variable $\cos\theta_{\mathrm{had}}$ should reduce the model dependence of the response matrix, x and $\cos\theta_{\mathrm{had}}$ are only weakly correlated, and hence the angle of the hadronic system does not directly provide information about x. It appears here that the prescription for determining the regularization parameter led to a solution with smaller errors in the two-dimensional case. It is not obvious whether this is particular to the example done here, or rather is true more generally.

Although some improvement with two-dimensional unfolding was anticipated, the exact reasons for the differences between the one and two-dimensional cases here are not yet fully understood. A large reduction in model dependence would be expected if the models had very different distributions of $\cos\theta_{\mathrm{had}}$. This distribution does not change much, however, when the HERWIG model is used with different structure functions, and hence the three models considered here are not very different in this regard. Further investigation with other models in which the $\cos\theta_{\mathrm{had}}$ distributions show greater differences will indicate whether the two-dimensional procedure in fact gives a significant improvement in the result.

Acknowledgements

Many thanks are due to Gerald Myatt and the other organizers of the Oxford Workshop on LEP 2 Physics for a pleasant and informative meeting, and to Mike Seymour and Susan Cartwright for stimulating discussions and for their organizing of the $\gamma\gamma$ working group.

References

[1] P. Aurenche and G.A. Schuler (conveners), *$\gamma\gamma$ Physics*, in G. Altarelli, T. Sjöstrand and F. Zwirner (eds.), *Physics at LEP2*, CERN 96-01, vol. 1, (1996) 291.

[2] Glen D. Cowan, *Statistical Data Analysis with Applications in Particle Physics*, Clarendon Press, Oxford, forthcoming.

[3] V. Blobel, *Unfolding methods in high energy physics experiments*, Proceedings of the 1984 CERN School of Computing, CERN 85-09 (1985).

[4] Andreas Höcker and Vakhtang Kartvelishvili, Nucl. Instr. and Meth. **A372** (1996) 469.

[5] R. Akers et al., OPAL Collaboration, Z. Phys. **C61** (1994) 199.

[6] G. Marchesini, B.R. Webber, G. Abbiendi, I.G. Knowles, M.H. Seymour, and L. Stanco, Comp. Phys. Comm. **67** (1992) 465.

[7] H. Plothow-Besch, Comp. Phys. Comm. **75** (1993) 396.

[8] M. Glück, E. Reya and A. Vogt, Phys. Rev. **D46** (1992) 1973.

[9] G.A. Schuler and T. Sjöstrand, Z. Phys. **C68** (1995) 607.

[10] H. Abramowicz, K. Charchula and A. Levy, Phys. Lett. **269B** (1991) 458.