Combination of measurements and the BLUE method

Luca Lista[1,a]

[1] *INFN - Istituto Nazionale di Fisica Nucleare - Sezione di Napoli*

Abstract. The most accurate method to combine measurements from different experiments is to build a combined likelihood function and use it to perform the desired inference. This is not always possible for various reasons, hence approximate methods are often convenient. Among those, the best linear unbiased estimator (BLUE) is the most popular, allowing to take into account individual uncertainties and their correlations. The method is unbiased by construction if the true uncertainties and their correlations are known, but it may exhibit a bias if uncertainty estimates are used in place of the true ones, in particular if those estimated uncertainties depend on measured values. In those cases, an iterative application of the BLUE method may reduce the bias of the combined measurement.

1 Measurement combination

The most rigorous and accurate method to combine measurements from different experiments relies on the combination of the individual likelihood functions that have been used for each measurements. Imagine that an experiment A provides a measured data sample $\vec{x}_A$, whose likelihood function, characterized by a set of parameters $\vec{\theta}_A$, is

$$L(\vec{x}_A; \vec{\theta}_A),\qquad(1)$$

and an experiment B provides a sample $\vec{x}_B$, whose likelihood function, characterized by a set of parameters $\vec{\theta}_B$, is

$$L(\vec{x}_B; \vec{\theta}_B).\qquad(2)$$

The parameter sets $\vec{\theta}_A$ and $\vec{\theta}_B$ contain some experiment-specific parameters and some parameters common to both experiments. Among the latter, there are physical parameters of interest, nuisance parameters related to common source of systematic uncertainties, such as theory uncertainties, accelerator-specific component of luminosity uncertainty, etc. The global likelihood function, if A and B are independent experiments, is:

$$L(\vec{x}_A, \vec{x}_B; \vec{\theta}) = L(\vec{x}_A; \vec{\theta}_A) \cdot L(\vec{x}_B; \vec{\theta}_B),\qquad(3)$$

where $\vec{\theta}$ is the set of all common and non-common parameters. Any statistical method can be applied at this level to determine a combined measurement, upper limit and/or significance: a Bayesian or frequentist inference, profile likelihood, modified frequentist upper limit, etc.

[a] e-mail: luca.lista@na.infn.it

An example of such approach is provided by the combination of measurements of single top-quark production in the *s*-channel at the Tevatron performed by the CDF and D0 experiments [1–3].

CDF and D0 measurements of single-top production cross section were combined using as input the binned distributions of multivariate discriminator outputs for each individual measurement. Each bin in each data sample was used in a Bayesian analysis, assuming a Poisson distribution, to extract a central value of the cross-section estimate. A likelihood-ratio analysis using asymptotic formulae [4] was instead used to determine the combined significance of the observed *s*-channel signal.

2 Approximate approaches: the BLUE method

In many cases, unless experiments, or even analysis groups within the same experiment, agree in advance, individual likelihood functions may not be available, or are available under different software frameworks, etc. Approximate methods can be used to combine the individual results, which are usually provided in terms of a central value and an uncertainty:

$$\theta = \hat{\theta}_1 \pm \sigma_1 \,, \tag{4}$$

$$\theta = \hat{\theta}_2 \pm \sigma_2 \,. \tag{5}$$

Correlation between uncertainties, which is related to the non-diagonal elements of the covariance matrix **V**

$$\rho = \frac{V_{12}}{\sigma_1 \sigma_2} \,, \tag{6}$$

must be properly taken into account.

The most popular method to combine correlated measurements is the best linear unbiased estimate (BLUE). The method was formulated initially in the '30s [5] and proposed in high-energy physics in the '80s [6]. By definition, it is the unbiased linear estimator that provides the smallest possible variance assuming the true uncertainties and their correlation are known. The estimator is equivalent to a χ^2 minimization which, for Gaussian distributions, is also equivalent to a maximum-likelihood estimate.

Given two measurements, as in eq. (4) and (5), the BLUE estimate is given by the following linear combination of the individual measurements:

$$\hat{\theta} = \frac{\hat{\theta}_1(\sigma_2^2 - \rho\sigma_1\sigma_2) + \hat{\theta}_2(\sigma_1^2 - \rho\sigma_1\sigma_2)}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2} \tag{7}$$

with variance:

$$\sigma_{\hat{\theta}}^2 = \frac{\sigma_1^2\sigma_2^2(1 - \rho^2)}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2} \,. \tag{8}$$

Unlike the usual weighted average, for some values of the correlation coefficient ρ, the coefficients of the linear combination that appear in eq. (7) may be negative.

More in general, for *n* measurements $\hat{\theta}_1 \pm \sigma_1, \cdots, \hat{\theta}_n \pm \sigma_n$, with a covariance matrix **V**, the BLUE combination in eq. (7) can be generalized as:

$$\hat{\theta} = \sum_{i=1}^{n} w_i \hat{\theta}_i \,, \tag{9}$$

with variance:

$$\sigma_{\hat{\theta}}^2 = \vec{w}^{\mathrm{T}} \mathbf{V} \vec{w} \,, \tag{10}$$

The weights in eq. (9) can be computed as:

$$\vec{w} = \frac{\mathbf{V}^{-1}\vec{u}}{\vec{u}^{\mathrm{T}}\,\mathbf{V}\,\vec{u}} \tag{11}$$

where $\vec{u} = (1, \cdots, 1)$ is the vector with all elements equal to unity. The normalization condition for weights $\vec{w}$ holds:

$$\sum_{i=1}^{n} w_i = 1\,. \tag{12}$$

In case the weights w_i are positive, an interpretation of the BLUE combination in eq. (7) in term of weighted average is possible by introducing the "common error" [7], defined as:

$$\sigma_{\mathrm{c}} = \rho\sigma_1\sigma_2\,. \tag{13}$$

The two measurements in eq. (4) and (5) can be rewritten with uncertainties given by the sum in quadrature of fully uncorrelated contributions σ'_1 and σ'_2 and a 100% correlated contribution σ_{C}:

$$\theta = \hat{\theta}_1 \pm \sigma'_1 \pm \sigma_{\mathrm{C}}\,, \tag{14}$$

$$\theta = \hat{\theta}_2 \pm \sigma'_2 \pm \sigma_{\mathrm{C}}\,, \tag{15}$$

where the uncorrelated uncertainty contributions are defined by:

$$\sigma'^2_1 = \sigma^2_1 - \sigma^2_{\mathrm{C}}\,, \tag{16}$$

$$\sigma'^2_2 = \sigma^2_2 - \sigma^2_{\mathrm{C}}\,, \tag{17}$$

The BLUE combination in eq. (7) achieves an expression similar to a regular weighted average:

$$\hat{\theta} = \frac{\dfrac{\hat{\theta}_1}{\sigma'^2_1} + \dfrac{\hat{\theta}_2}{\sigma'^2_2}}{\dfrac{1}{\sigma'^2_1} + \dfrac{1}{\sigma'^2_2}}\,, \tag{18}$$

with weights that only take into account the uncorrelated uncertainty contributions, but in this case the uncertainty receives an additional contribution due to the correlated uncertainty, with respect to the uncertainty of the usual weighted average, and is given by:

$$\sigma^2_{\hat{\theta}} = \frac{1}{\dfrac{1}{\sigma'^2_1} + \dfrac{1}{\sigma'^2_2}} + \sigma^2_{\mathrm{C}}\,. \tag{19}$$

Cases with negative weights have a less intuitive interpretation than eq. (18), as will be more evident in section 4.

3 Quantifying the importance of individual measurements

In order to quantify the "importance" of each individual measurement used in a combination, the first approach adopted in literature was to quote the so-called "relative importance" (RI) of each individual measurement, proportional to the absolute value of the corresponding weight 'w_i, and defined as:

$$\mathrm{RI}_i = \frac{|w_i|}{\displaystyle\sum_{i=1}^{n} |w_i|}\,. \tag{20}$$

The definition is chosen in order to have a normalization condition: $\sum_i \mathrm{RI}_i = 1$. This approach was, for instance, used in combinations of top-quark mass measurements at Tevatron and at LHC [8, 9].

This choice is questionable, as observed in ref. [10]. In fact, imagine we have three measurement, say A, B_1 and B_2. The RI of measurement A changes whether the three measurements are combined all together or if B_1 and B_2 are first combined into B, and then A and the partial combination B are combined together.

Ref. [10] proposes alternatives to RI based on the Fisher information, which is defined as:

$$\mathcal{J}_{ij} = \left\langle \frac{\partial \ln L}{\partial \theta_i} \frac{\partial \ln L}{\partial \theta_j} \right\rangle , \tag{21}$$

the average being performed over all possible measurements, hence $\mathcal{J}$ does not depend on a specific measurement, but only on the form of the likelihood function and on the parameters choice. Fisher information sets a lower bound to the variance of an unbiased estimator [11, 12]:

$$\sigma_{\hat{\theta}_i}^2 \geq \mathcal{J}_{ii}^{-1} , \tag{22}$$

and for a single parameter, the Fisher information is given by:

$$\mathcal{J} = \mathcal{J}_{11} = \vec{u}^{\mathrm{T}} \mathbf{V} \vec{u} = \frac{1}{\sigma_{\hat{\theta}}^2} . \tag{23}$$

The alternative quantities to RI proposed in ref. [10] are the intrinsic information weight (IIW), defined as:

$$\mathrm{IIW}_i = \frac{1/\sigma_i^2}{1/\sigma_{\hat{\theta}}^2} = \frac{1/\sigma_i^2}{\mathcal{J}} \tag{24}$$

and the marginal information weight (MIW), defined as follows:

$$\mathrm{MIW}_i = \frac{\Delta \mathcal{J}_{\rangle}}{\mathcal{J}} = \frac{\mathcal{J} - \mathcal{J}_{\{1,\cdots,n\}-\{i\}}}{\mathcal{J}} , \tag{25}$$

i.e.: the relative difference of the Fisher information of the combination and the Fisher information of the combination excluding the ith measurement. Both IIW and MIW do not obey a normalization condition. For IIW the quantity $\mathrm{IIW}_{\mathrm{corr}}$ can be defined such that:

$$\sum_{i=1}^{n} \mathrm{IIW}_i + \mathrm{IIW}_{\mathrm{corr}} = 1 . \tag{26}$$

$\mathrm{IIW}_{\mathrm{corr}}$ represents the weight assigned to the correlation interplay, not assignable to individual measurements, and is given by:

$$\mathrm{IIW}_{\mathrm{corr}} = \frac{1/\sigma_{\hat{\theta}^2} - \sum_{i=1}^{n} 1/\sigma_i^2}{1/\sigma_{\hat{\theta}}^2} = \frac{\mathcal{J} - \sum_{i=1}^{n} 1/\sigma_i^2}{\mathcal{J}} . \tag{27}$$

Table 1 summarizes the properties of the different proposed quantities.

IIW and MIW are reported, instead of RI, in recent papers about combination of LHC and Tevatron measurements related to top-quark physics [13–15].

Table 1. Properties of the different quantities proposed for the BLUE combination

Weight type		≥ 0	$\sum_i = 1$	Consistent with partial combination
BLUE coefficient	w_i	✗	✓	✓
Relative importance	$\|w_i\|/ \sum_{i=1}^{n} \|w_i\|$	✓	✓	✗
Intrinsic Information Weight	IIW_i	✓	✗	✓
Marginal Information Weight	MIW_i	✓	✗	✓

4 Negative weights

Negative weights are always sign of a high-correlations regime. The maximum value of the ratio $\sigma_{\hat{\theta}^2}/\sigma_1^2$ (eq. (8)) as a function of ρ is obtained for $\rho = \sigma_1/\sigma_2$. For $\rho > \sigma_1/\sigma_2$, an increase in correlation implies a decrease of the uncertainty and a negative weight, and uncertainty becomes strongly dependent on ρ. For $\rho = \sigma_1/\sigma_2$, in particular, the weight w_2 becomes equal to zero, as well as $MIW_2 = 0$. But this does not imply that the measurement θ_2 is not used in the combination. Figure 1, from ref. [10], shows how the BLUE coefficient w_2 and the ratio of uncertainties $\sigma_{\hat{\theta}}^2/\sigma_1^2$ vary as a function of the correlation ρ for different fixed values of the ratio σ_2/σ_1 (note that the figure uses a different notation with respect to this text, as specified in the caption).
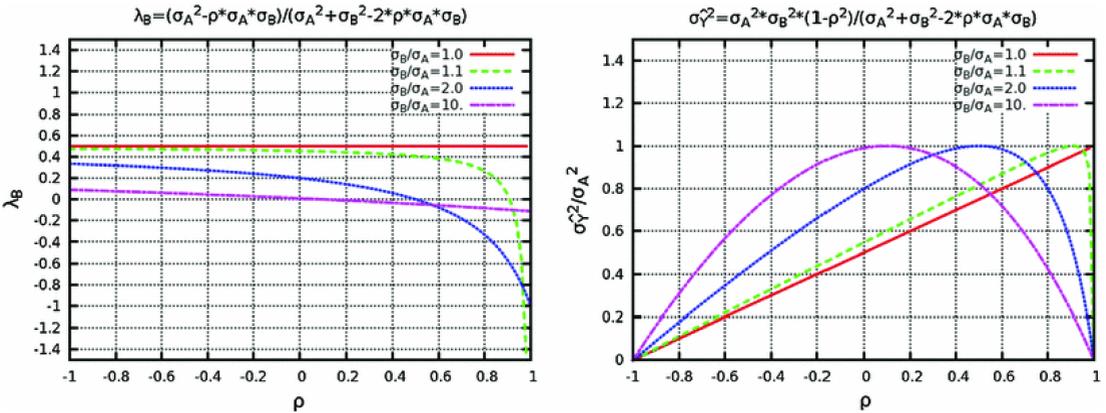

Figure 1. BLUE coefficient for the second measurement w_2 (left; λ_B in the original figure notation) and combined BLUE variance $\sigma_{\hat{\theta}}^2$ (right; $\sigma_{\hat{Y}}^2$ in the original figure notation) as a function of the correlation ρ between two measurements 1 and 2 for various fixed values of the ratio σ_2/σ_1 (σ_B/σ_A in the original figure notation). The figure is from ref. [10].

When the correlation coefficient ρ is not well know, assuming $\rho = 1$ is not always a conservative choice. The assumption $\rho = 1$ yields the largest possible uncertainty only if the uncorrelated contributions to the total uncertainty dominate. ρ should be accurately determined in case of negative weights in order to avoid the risk of underestimating uncertainties. Assume that the two measurements A and B have total uncertainties given by the sum in quadrature of uncorrelated contributions, σ_A(unc) and σ_B(unc) respectively, and correlated contributions, σ_A(cor) and σ_B(cor) respectively, whose correlation coefficient is ρ(cor). Figure 2 shows the most "conservative" value of the correlation coefficient ρ(cor), which is equal to 1 only for σ_B(cor)$/\sigma_A$(cor)$< (\sigma_A/\sigma_A$(cor)$)^2$.
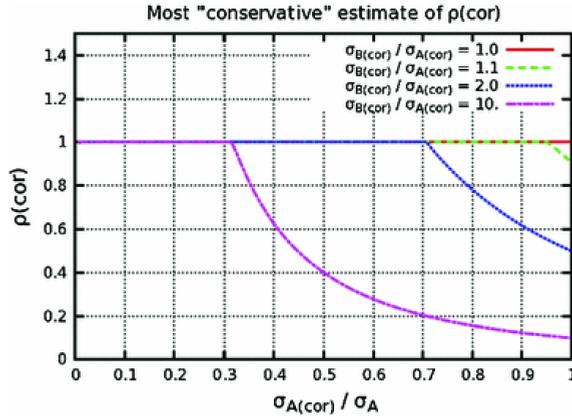
Figure 2. The most "conservative" value of an unknown correlation $\rho(\text{cor})$ between $\sigma_A(\text{cor})$ and $\sigma_A(\text{cor})$ as a function of $\sigma_A(\text{cor})/\sigma_A$, for several values of $\sigma_B(\text{cor})/\sigma_A(\text{cor}) \geq 1$. The figure is from ref. [10].

5 Bias with the BLUE method

The BLUE method provides an unbiased estimate only if uncertainties and their correlation are known exactly. This is not always a realistic scenario, since estimates of uncertainties and their correlation, and not their true values, are available in most of the cases. Moreover, the uncertainty estimates may depend on the assumed central value. One example of such a case in which the BLUE method provides a bias is the combination of two Poissonian estimates, each of which has an uncertainty estimate that depends on the central value through a square-root relation:

$$\hat{n}_1 = \hat{n}_1 \pm \sqrt{\hat{n}_1}, \tag{28}$$
$$\hat{n}_2 = \hat{n}_2 \pm \sqrt{\hat{n}_2}. \tag{29}$$

A maximum-likelihood estimate, using Poissonian distributions, would produce the following unbiased combined estimate:

$$\hat{n} = \frac{1}{2}\left(\hat{n}_1 + \hat{n}_2 \pm \sqrt{\hat{n}_1 + \hat{n}_2}\right). \tag{30}$$

The BLUE combination, instead, gives weights proportional to $1/\hat{n}_i$, which correspond to a harmonic average:

$$\hat{n} = \frac{2\hat{n}_1\hat{n}_2}{\hat{n}_1 + \hat{n}_2} \pm \sqrt{\frac{\hat{n}_1\hat{n}_2}{\hat{n}_1 + \hat{n}_2}}. \tag{31}$$

Equation (31), compared to eq. (30), exhibits a bias because, due to the dependence of uncertainties on the measured values, downward measurement fluctuations achieve larger weights pulling down the combination, while upward fluctuations produce a smaller opposite effect.

In many cases, relative uncertainty estimates are available, like for uncertainty contributions due to luminosity, efficiencies, etc. When performing a combination, the best estimate of the central value improves the individual measurements, hence one may argue whether the assumed uncertainties should change accordingly, re-evaluating them using the central-value estimate from the BLUE combination [16, 17]. The method can be applied iteratively until the combination converges to a stable

value. Namely:

$$\hat{\sigma}_i^{(0)} = \hat{\sigma}_i(\hat{\theta}_i)\,, \ \hat{\rho}_{ij}^{(0)} = \hat{\rho}(\hat{\theta}_i, \hat{\theta}_j) \xrightarrow{\text{BLUE}} \hat{\theta}^{(1)} \tag{32}$$

$$\hat{\sigma}_i^{(1)} = \hat{\sigma}_i(\hat{\theta}^{(1)})\,, \ \hat{\rho}_{ij}^{(1)} = \hat{\rho}(\hat{\theta}^{(1)}, \hat{\theta}^{(1)}) \xrightarrow{\text{BLUE}} \hat{\theta}^{(2)} \tag{33}$$

$$\hat{\sigma}_i^{(2)} = \hat{\sigma}_i(\hat{\theta}^{(2)})\,, \ \hat{\rho}_{ij}^{(2)} = \hat{\rho}(\hat{\theta}^{(2)}, \hat{\theta}^{(2)}) \xrightarrow{\text{BLUE}} \hat{\theta}^{(3)} \tag{34}$$

$$\cdots$$

In practice, convergence only needs few iterations in most of the cases.

Let us assume that a contribution to the total uncertainty is known as relative uncertainty. In this case, uncertainties can be written as the sum in quadrature of a contribution that does not depend on the central value and another contribution that is proportional to the central value[1]:

$$\theta = \hat{\theta}_1 \pm \sigma_1 \pm r_1 \hat{\theta}_1\,, \tag{35}$$

$$\theta = \hat{\theta}_2 \pm \sigma_2 \pm r_2 \hat{\theta}_2\,. \tag{36}$$

The covariance matrix estimate can be written as follows:

$$\hat{\mathbf{V}} = \begin{pmatrix} \sigma_1^2 + (r_1 \hat{\theta}_1)^2 & \rho_0 \sigma_1 \sigma_2 + \rho_r r_1 r_2 \hat{\theta}_1 \hat{\theta}_2 \\ \rho_0 \sigma_1 \sigma_2 + \rho_r r_1 r_2 \hat{\theta}_1 \hat{\theta}_2 & \sigma_2^2 + (r_2 \hat{\theta}_2)^2 \end{pmatrix} \tag{37}$$

where ρ_0 and ρ_r represent the correlation coefficients of the uncertainty contributions that do not depend on the central value and of the ones that are proportional to the central value, respectively.

A special case is when uncertainties are fully proportional to the central values, i.e.: $\sigma_1 = 0$ and $\sigma_2 = 0$. In that case, the iterative BLUE method converges in two iterations to the following central value:

$$\hat{\theta} = \frac{(r_2^2 - \rho_r r_1 r_2)\hat{\theta}_1 + (r_1^2 - \rho_r r_1 r_2)\hat{\theta}_2}{q_1^2 - 2\rho_r r_1 r_2 + r_2^2}\,, \tag{38}$$

which is similar to eq. (7) but relative uncertainties are used in place of the absolute ones.

A numerical Monte Carlo study [17] shows that in most of the cases the bias in the combination can be mitigated by applying the iterative procedure. Assuming for simplicity, and without loss of generality, a true value $\theta = 1$, uncertainties and their correlations are chosen randomly by spanning a wide ranges of values in the boundaries $\sigma_1, \sigma_2, r_1, r_2, < 1$ and $-1 < \rho_0, \rho_r < 1$. For each set of randomly-extracted uncertainties and correlation values, 500 000 random values of $\hat{\theta}_1$ and $\hat{\theta}_2$ are generated using the proper two-dimensional correlated Gaussian distribution. The BLUE method is then applied in its standard formulation and iteratively. Pulls distributions allow to study the bias and the correctness of uncertainty estimates. Bias is in general mitigated with the iterative BLUE method while the standard BLUE method tends to underestimate the central value, as shown in Fig. 3. The iterative BLUE application may provide overestimates in few of the cases with large uncertainties. More detailed comparisons of the two method are available in ref. [17]. In particular, the combination uncertainty may be overestimated (and in fewer cases underestimated) in both standard and iterative BLUE, but the iterative BLUE estimates tend to be more conservative, as visible in Fig. 4. A dedicated toy Monte Carlo may be useful in case of large individual uncertainties in order to achieve a proper combined uncertainty estimate.

---

[1]This was not the case for the combination of Poissonian measurements from eq. (28), (29), where the dependence on central value was not linear, but it is a convenient assumption in many realistic cases.
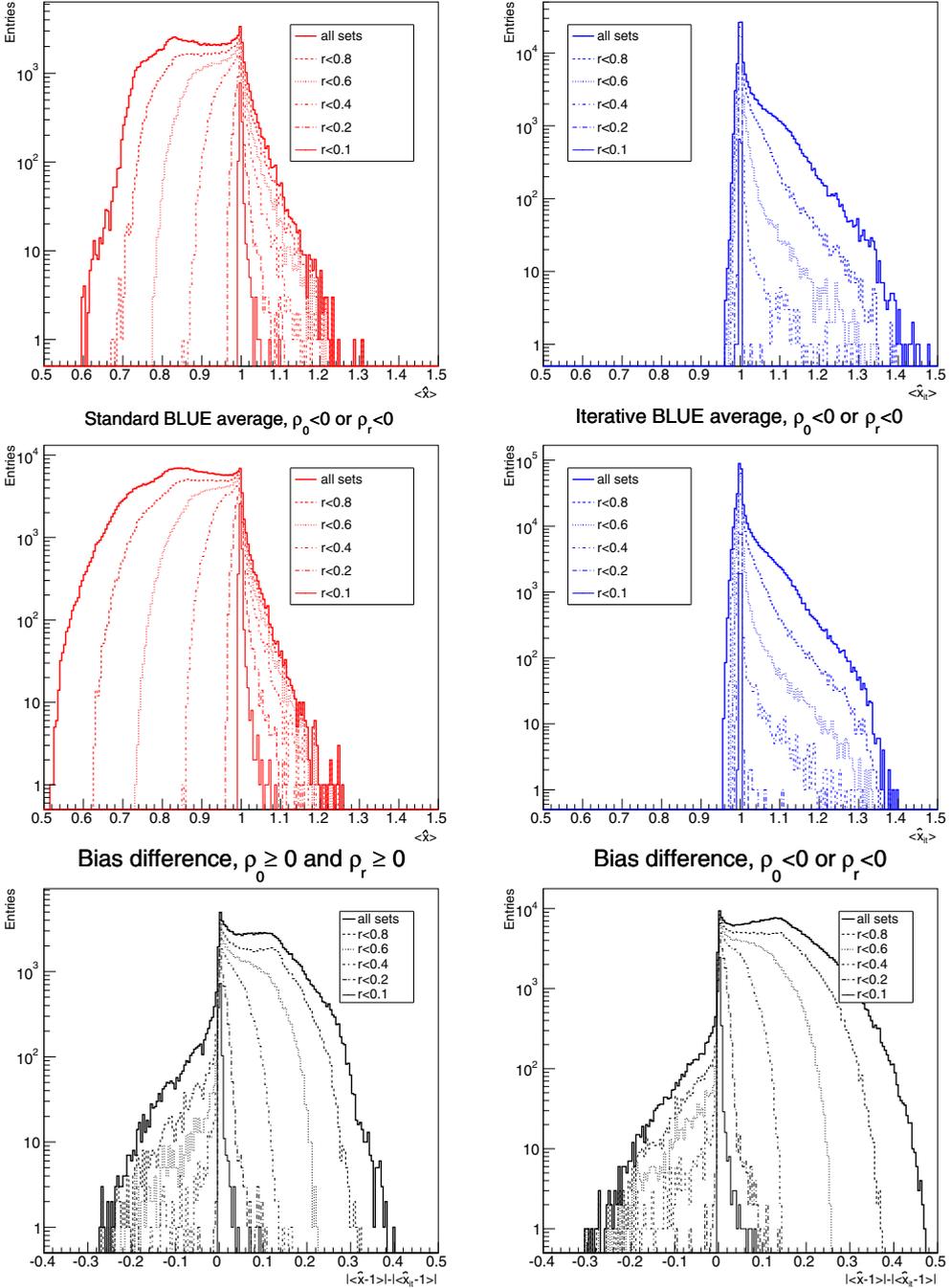
Figure 3. Distribution of the average value of the standard (top and middle, left) and iterative (top and middle, right) BLUE estimates for different limits on $r_{1,2}$ and for $\rho_0 \geq 0$ and $\rho_r \geq 0$ (top) and for $\rho_0 < 0$ and $\rho_r < 0$ (middle). Bottom plots show the difference of measured absolute value of the bias for the standard and iterative BLUE estimates for different limits on $r_{1,2}$ and for $\rho_0 \geq 0$ and $\rho_r \geq 0$ (left) and for $\rho_0 < 0$ and $\rho_r < 0$ (right). Positive values indicate a smaller bias in the iterative method compared to the standard method. Underestimate of the uncertainty (negative values) occur in fewer cases with the iterative BLUE method than with the standard BLUE method. The figure is from ref. [17].
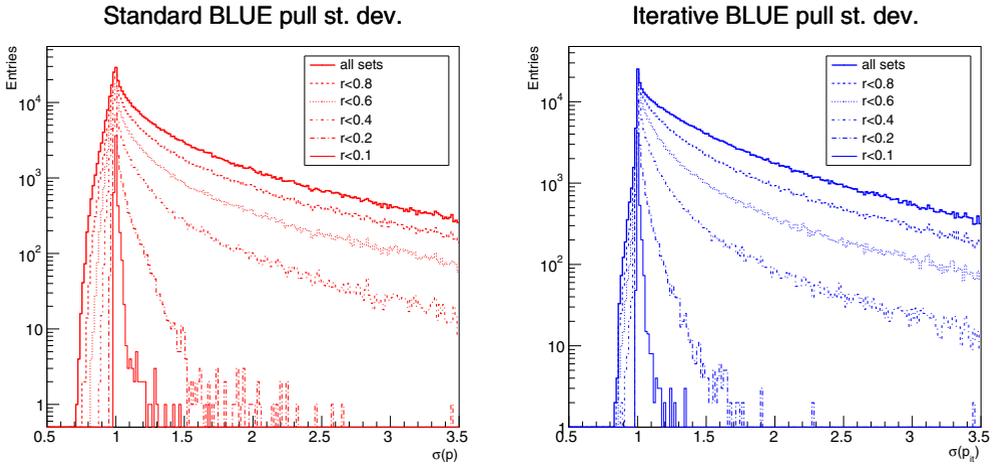
Figure 4. Distribution of the average value of the pull standard deviation for the standard (left) and iterative (right) BLUE estimates for different limits on $r_{1,2}$. The figure is from ref. [17].

6 Conclusions

Combination of measurements is a crucial task to improve the precision on the knowledge of important parameters. Combining individual likelihood functions before applying any statistical method is the most rigorous and precise approach, but often the full likelihood function of individual measurements is not available. When uncertainties and their correlations are available, the BLUE method is a simple and powerful tool to combine measurements. But BLUE may have counterintuitive behaviors such as negative weights or uncertainties that may decrease for increasing correlation. Correlation needs to be accurately evaluated in those cases, and an assumption of a 100% correlation is not necessary a conservative choice. In case relative uncertainty contributions are known, or more in general when uncertainty estimates depend on the central values, the BLUE method may exhibit a bias that can be mitigated, in most of the cases, with an iterative application of the method where uncertainties are rescaled at each iteration to the combined central value.

References

[1] CDF Collaboration, Phys. Rev. Lett **112**, 231804 (2014)
[2] CDF Collaboration, Phys. Rev. Lett **112**, 231805 (2014)
[3] D0 Collaboration, Phys. Lett B **726**, 656 (2013)
[4] G. Cowan *et al.*, Eur. Phys. J. C **71**, 1554 (2011)
[5] J. Aitken, Proc. Roy. Soc. Edinburgh **55**, 42 (1935)
[6] L. Lyons *et al.* Nucl. Instr. Meth. A **270**, 110 (1988)
[7] H. Greenlee, Fermilab Workshop on Confidence Limits (2000)
[8] CDF and D0 Collaborations, arXiv:1107.5255 (2011)
[9] ATLAS and CMS Collaborations, ATLAS-CONF-2012-095, CMS-PAS-TOP-12-001 (2012)
[10] A. Valassi, R. Chierici, Eur. Phys. J. C **74** 2717 (2014)
[11] H. Cramér, *Mathematical Methods of Statistics* (Princeton Univ. Press, Princeton, NJ, 1946)

[12] C. R. Rao, Bulletin of the Calcutta Mathematical Society, **37**, 81–89 (1945)

[13] ATLAS and CMS Collaborations, ATLAS-CONF-2013-102, CMS-PAS-TOP-13-005 (2013)

[14] ATLAS and CMS Collaborations, ATLAS-CONF-2014-012, CMS-PAS-TOP-14-006 (2014)

[15] ATLAS, CMS, CDF and D0 Collaborations, ATLAS-CONF-2014-008, CDF-NOTE-11071, CMS-PAS-TOP-13-014, D0-NOTE-6416, FERMILAB-TM-2582, arXiv:1403.4427 (2014)

[16] L. Lyons, A. J. Martin, D. H. Saxon, Phys. Rev. D **41** 982–985 (1990)

[17] L. Lista, Nucl. Instr. Meth. A **764** 82-93 (2014); corrig. ibid. A **773** 87-96 (2015)