On the combination of correlated estimates of physics observables

Richard Nisius

Abstract The combination of a number of correlated estimates of various observables is a common task in particle physics. This is frequently performed using the Best Linear Unbiased Estimate (BLUE) method. Most features of such a combination can already be seen by analysing the special case of a pair of estimates from two correlated estimators of the same observable, which is discussed in some detail here. Two important parameters of this combination are the weight of the less precise estimate and the ratio of uncertainties of the combined result and the more precise estimate. Derivatives of these quantities are derived with respect to the correlation and the ratio of uncertainties of the two estimates, which themselves are typically only known with some uncertainty.

Using an example, a critical assessment is performed of suggested methods to deal with this fact. A proposal is made to decide on the usefulness of a combination and to perform it. The proposal is based on possible improvements with respect to the most precise estimate by including additional estimates. This procedure can be applied to the general case of several observables in view of the uncertainties on the correlation and relative uncertainties of the estimates.

Keywords BLUE · combination · conditional probability · Peelle’s Pertinent Puzzle

1 Introduction

The combination of a number of correlated estimates of a single observable is discussed in Ref. [1]. Here, the term estimate denotes a particular outcome (measurement) based on an estimator of the observable, which follows a probability density distribution (pdf). The particular estimate obtained may be a likely or unlikely outcome given that distribution. Repeating the measurements numerous times under identical conditions, the estimates will follow the underlying multi-dimensional pdf of the estimators\(^1\). The analysis [1] makes use of a \(\chi^2\) minimisation to obtain the combined value expressed in the mathematically equivalent BLUE language.

Provided the estimators are unbiased, when applying this formalism the Best Linear Unbiased Estimate of the observable is obtained with the following meaning: Best: the combined result for the observable obtained this way has the smallest variance; Linear: the result is a linear combination of the individual estimates; Unbiased Estimate: when the procedure is repeated for a large number of cases consistent with the underlying multi-dimensional pdf, the mean of all combined results equals the true value of the observable. For a real situation, for which the estimates are obtained by experiments that cannot be repeated numerous times, when performing a combination one has to rely on this fact, although the combined value obtained from the particular estimates may be far away from the true value. This fact however should not be mistaken for a bias inherent to the method.

The equations to solve the problem for the general case of \(m\) estimates and \(n\) observables with \(m \geq n\) are given in Ref. [2]. They have been implemented in a software package [3] that is embedded into the ROOT analysis framework [4], but are not repeated here. However, the special case of two correlated estimates of the same observable is discussed in some detail. This is because already from this case the main features of the combination can easily be understood.

This paper is organised as follows: the case of two estimators and the consequences of the conditional probability is explained in Section 2. The description of the relations for two estimates is given in Section 3, where also the derivatives are derived. This is followed by a discussion of the properties of the estimates to be combined in Section 4. The impact of assigning relative uncertainties is reviewed in Section 5. The concept of re-

\(^1\)In this paper, the discussion is restricted to Gaussian estimator pdfs.
duced correlations is outlined in Section 6, and other methods constructed to maximise the variance of the combined result are discussed in Section 7. Based on an example, the consequences of using these methods are discussed in Section 8. A detailed proposal on how to decide on a combination and how to perform investigations of its stability is given in Section 9. Finally, conclusions are drawn in Section 10.

2 Correlated estimators and conditional probabilities

Let $X_1$ and $X_2$ with variances $\sigma_1^2$ and $\sigma_2^2$ be two unbiased, but correlated Gaussian estimators of a true value $x_T$. They obey the two-dimensional pdf $P(X_1, X_2)$, with identical mean values $\langle X_1 \rangle = \langle X_2 \rangle = x_T$ for the two estimators if calculated based on the entire pdf.

For a vanishing true value $x_T = 0$ and with a correlation of the two estimators of $\rho$ the pdf reads:

$$P(X_1, X_2) = \frac{1}{\sqrt{2\pi\sigma_1}\sqrt{2\pi\sigma_2}} \frac{1}{\sqrt{1-\rho^2}} \exp \left\{ \frac{1}{2(\rho^2-1)} \left( \frac{X_1^2}{\sigma_1^2} + \frac{X_2^2}{\sigma_2^2} - 2\rho X_1 X_2 \right) \right\}.$$  

(1)

The outcome of a pair of data analysis using these estimators will be two estimates denoted with $x_1$ and $x_2$ that will occur according to this pdf. The estimates will have variances of $\sigma_1^2$ and $\sigma_2^2$ assigned, and their correlation is $\rho$. Without loss of generality it is assumed that $X_1$ is as least as precise an estimator of $x_T$ than $X_2$ is, such that $z \equiv \sigma_2/\sigma_1 \geq 1$.

In combinations of estimates of physics observables the typical situation is that one estimate, here $x_1$, is available, and the question arises what the improvement will be if also the information from another estimate, here $x_2$, is used, rather than determining $x_T$ and its uncertainty solely based on $x_1$. Therefore, it is important to understand what is the likely outcome of $x_2$ given the existence of $x_1$. This is most directly seen by analysing the conditional pdf for $X_2$ given $X_1 = x_1$ which reads:

$$P(x_1, X_2) = \frac{1}{\sqrt{2\pi\sigma_1}} \frac{1}{\sqrt{2\pi\sigma_2\sqrt{1-\rho^2}}} \exp \left\{ -\frac{1}{2} \left( \frac{x_1}{\sigma_1} \right)^2 \right\} \cdot \exp \left\{ -\frac{1}{2} \left( \frac{X_2 - \rho x_1}{\sigma_2\sqrt{1-\rho^2}} \right)^2 \right\}.$$  

(2)

A few facts are worth noticing, see also Refs. [5,6], and a related discussion in Ref. [7]. Firstly, this conditional pdf for $X_2$ at a given fixed value of $x_1$ is no longer centred at $\langle X_2 \rangle = x_T = 0$ but at $\langle X_2 \rangle = \rho x_1$. Although $X_2$ in itself is an unbiased estimator, given the existence of the estimate $x_1$ and the correlation of the estimators, it is no longer distributed around the true value, except for the situation in which the value of the more precise estimate coincides with the true value, i.e. $x_1 = x_T$.

This is a mere consequence of the correlation. As intuitively expected, in the case of positively correlated estimates, if one estimate is larger (smaller) than $x_T$ the other also more likely will be larger (smaller). For negatively correlated estimates the situation is reversed.

For $\rho > 0$, and depending on whether $\rho z$ is larger (smaller) than unity, the mean $\langle X_2 \rangle$ is even further away from (closer to) the true value $x_T$ than $x_1$ is. Given that the distribution in $X_2$ is still symmetric around its mean, for $\rho > 1/z$ in more than half of the cases in which $x_T < x_1$ also $x_T < x_1 < X_2$ is fulfilled. Secondly, the variance of $X_2$ no longer amounts to the initial value of $\sigma_2^2$ but it is reduced to $(1-\rho^2)\sigma_2^2$ which vanishes for $\rho = \pm 1$, again a consequence of the correlation. Finally, for $\rho = 0$ the original values of the mean and width of the pdf for $X_2$ are recovered.

Simulating the two-dimensional pdf $P(X_1, X_2)$ using five million pairs of estimates, the consequences of the conditional probability for the example of individually unbiased estimators obeying $\langle X_1 \rangle = \langle X_2 \rangle = x_T = 0$ are discussed. For uncertainties of $\sigma_1 = 0.85$ and $\sigma_2 = 1.15$, i.e. for $z = 1.35$, the results are shown in Figure 1 for three different values of the correlation, $\rho = 0, 0.9, -0.9$. For the uncorrelated case, Figure 1(a), the half axes of the ellipses coincide with the coordinate axes. For any value of $x_1$, e.g. along the vertical red line shown, the conditional pdf is centred around $X_2 = x_T = x_T$. A hypothetical outcome, namely the pair of estimates $x_1$ and $x_2$, is shown by the red dot. Since for the chosen value of $x_T$ this point lies in the upper right (i.e. first) quadrant, both estimates are larger than $x_T$. Since the point is above the diagonal line, $x_2$ has been chosen to be larger than $x_1$, such that the order is $x_T < x_1 < x_2$. This means the true value is outside the interval given by the two estimates. Analysing the entire two-dimensional pdf one finds that, even for the uncorrelated case, for which the pdf is equally shared by the four quadrants in the $X_1$-$X_2$ plane, in half of all possible outcomes (namely in quadrants one and three), the true value does not fall within the interval spanned by the estimates, despite the fact that both estimators are unbiased and not correlated.

Although discussed here for the two-dimensional case, it similarly applies to a pair of estimates from the same Gaussian estimator.
The information from next.

The black line corresponds to $X_1 = X_2$, the red line to $X_1 = x_1$, and finally the dot to a particular pair of estimates chosen to be $x_1 = 0.30$ and $x_2 = 0.95$. A value of out = 50% means that in half of the cases $x_T$ does not lie within the interval spanned by the pair of estimates. Shown are (a) $\rho = 0$, (b) $\rho = 0.9$, and (c) $\rho = -0.9$. In b (c) the half axes shown in blue are changed and rotated (counter) clockwise from the positive $X_2$ axis.

3 The special case of two correlated estimates

Again, $x_1$ and $x_2$ with variances $\sigma_1^2$ and $\sigma_2^2$ obeying $z = \sigma_2/\sigma_1 \geq 1$ are two Gaussian estimates from two unbiased estimators of the true value $x_T$ of the observable, and $\rho$ denotes their total correlation with $-1 \leq \rho \leq 1$.

In this situation the BLUE of $x_T$ is:

$$x = (1 - \beta) x_1 + \beta x_2,$$

where $\beta$ is the weight of the less precise estimate, and, by construction, the sum of weights is unity. The variable $x$ is the combined result and $\sigma_z^2$ denotes its variance, i.e. the uncertainty assigned to the combined value $\sigma_z$.

In the following the derivation of the formulas for $\beta$ and $\sigma_z/\sigma_1$ within the BLUE formalism is repeated, see also [1]. The covariance matrix for the general solution of the linear combinations in the BLUE formalism is given by Eq. 5 of Ref. [2]. For the studied case of two estimates of one observable it reduces to:

$$\sigma_z^2 = \left(1 - \beta\right)^T \left(\begin{array}{cc} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{array}\right) \cdot \left(1 - \beta\right),$$

(4)

dividing by $\sigma_1^2$ and inserting $z$ yields:

$$\frac{\sigma_z^2}{\sigma_1^2} = \left(1 - \beta \beta\right) \cdot \left(\frac{1}{\rho z} z^2\right) \cdot \left(1 - \beta\right),$$

(5)

multiplication results in:

$$\frac{\sigma_z^2}{\sigma_1^2} = \left(1 - \beta\right)^2 + 2 \rho z \beta \left(1 - \beta\right) + \beta^2 z^2$$

$$= 1 - 2 \beta (1 - \rho z) + \beta^2 (1 - 2 \rho z + z^2),$$

(6)
taking the derivative with respect to $\beta$ equal to zero (i.e. the $\chi^2$ minimisation) gives:

$$\frac{\partial}{\partial \beta} \left( \frac{\sigma^2}{\sigma_1^2} \right) = -2(1 - \rho z) + 2\beta(1 - 2\rho z + z^2) = 0.$$  \hfill (7)

Finally, after solving for $\beta$ one obtains:

$$\beta = \frac{1 - \rho z}{1 - 2\rho z + z^2} = \frac{1 - \rho z}{(1 - \rho z)^2 + z^2(1 - \rho^2)}$$ \hfill (8)

which is valid for $-1 \leq \rho \leq 1$ and $z \geq 1$, but for $\rho = z = 1$.

The last term in Eq. 8 shows that the denominator of $\beta$ is always positive such that the sign of $\beta$ is determined by the sign of the numerator. The resulting $\beta$ as a function of $\rho$, and for various $z$ values is shown in Figure 2(a). Identifying Eq. 3 and Eq. 8 yields:

$$\frac{1}{2} \geq \beta = \frac{x - x_1}{x_2 - x_1} = \frac{1 - \rho z}{1 - 2\rho z + z^2} \geq \frac{1}{1 - z},$$ \hfill (9)

where the left limit has been derived at $\rho \neq 1$ $z = 1$, and the right limit at $\rho = 1$.

A few features are important to understand the results of the combination. As expected, the value of $\beta$ has to be smaller or equal than 0.5, because otherwise $x_2$ would be the more precise estimate. Since the denominator in Eq. 8 is positive for all allowed values of $\rho$ and $z$, the function for $\beta$ turns negative for $\rho > 1/z$ as shown in Figure 2(a). This is exactly the point at which for a given $x_1$ the conditional probability for $X_2$ to be even further away from $x_T$ than $x_1$ is, exceeds 50%, see Section 2.
The first equal sign in Eq. 9 means that the value of $\beta$ can be interpreted as the difference of the combined value from the more precise estimate in units of the difference of the two estimates. If $\beta$ is positive, the signs of the numerator and denominator are identical and $x$ lies within the interval spanned by $x_1$ and $x_2$. Given $\beta \leq 0.5$ it never lies further away from the more precise estimate than half the difference of the two. Again, this is expected since the more precise estimate should dominate the combination. In contrast, if $\beta$ is negative, the signs of the numerator and denominator are different. This means the value of $x$ lies on the opposite side of $x_1$ than $x_2$ does, or in other words, the combined value lies outside the interval spanned by the two estimates. Given the discussion about the conditional pdf in Section 2, a very desirable feature.

Inserting the result for $\beta$ into Eq. 6 yields:

$$\frac{\sigma_x^2}{\sigma_1^2} = 1 - 2 \frac{(1 - \rho z)^2}{1 - 2\rho z + z^2} + \frac{(1 - \rho z)^2}{1 - 2\rho z + z^2}$$

$$= \frac{(1 - 2\rho z + z^2) - (1 - \rho z)^2}{1 - 2\rho z + z^2},$$

(10)

which after evaluating the numerator and taking the square root gives:

$$\frac{\sigma_x}{\sigma_1} = \sqrt{\frac{z^2(1 - \rho^2)}{1 - 2\rho z + z^2}}.$$  

(11)

The resulting $\sigma_x/\sigma_1$, as a function of $\rho$, and for various $z$ values is shown in Figure 2(b). This variable quantifies the uncertainty of the combined value in units of the
uncertainty of the more precise estimate, i.e. $1 - \sigma_x/\sigma_1$ is the relative improvement achieved by also using $x_2$, i.e. including the information contained in the less precise estimator. Consequently, $\sigma_x/\sigma_1$ can be used to decide whether it is worth combining.

Since in the numerator of Eq. 10 the first term is identical to the denominator (which is always positive, see Eq. 8), and the second term is positive for all values of $\rho$ and $z$, the value of $\sigma_x/\sigma_1$ is always smaller or equal to unity, as shown in Figure 2(b). Again this is expected, since including the information from the estimate $x_2$ should improve the knowledge on $x$, which means its precision $\sigma_x$. Not surprisingly, the value of $\sigma_x/\sigma_1$ is exactly one for $\rho = 1/z$, i.e. for $\beta = 0$. In this situation, the value of $x_2$ is irrelevant in the linear combination of Eq. 3, and consequently $x = x_1$ and $\sigma_x = \sigma_1$. Finally, $\sigma_x/\sigma_1$ is exactly zero if $\rho = \pm 1$ in accordance with the variance of $X_2$ for the conditional PDF given $x_1$ and $\rho$, shown in Section 2. This means that for the fully correlated or fully anti-correlated case of two estimators, given $x_1$, the result is known for sure, and the outcome of the second estimate has to be $x_2 = \rho x_1$. For combinations of experimental results, for which for all pairs of estimates there are also uncorrelated components of the uncertainty, this situation never happens.

The typical situation is that both $\rho$ and $z$ are only known with some precision. In this situation it is essential to analyse the sensitivity of the combination to this imperfect knowledge that is encoded in the respective derivatives. The derivatives of $\beta$ and $\sigma_x/\sigma_1$ with respect to the parameters $\rho$ and $z$ have been derived in this paper and are given in Eqs. 12–15.

$$
\frac{d\beta}{d\rho} = \frac{z(1 - z^2)}{(1 - 2\rho z + z^2)^2}
$$

$$
\frac{d\sigma_x}{d\rho} = z(z - \rho)(1 - \rho z)^{-1/2} \left(1 - \rho^2(1 - 2\rho z + z^2)^3\right)^{1/2}
$$

$$
\frac{d\beta}{dz} = \frac{\rho(1 + z^2) - 2z}{(1 - 2\rho z + z^2)^2}
$$

$$
\frac{d\sigma_x}{dz} = (1 - \rho z)^{-1/2} \left(1 - \rho^2(1 - 2\rho z + z^2)^3\right)^{1/2}
$$

The resulting variations of the combined value, Eq. 3, are given in Eqs. 16–17.

$$
\frac{dx}{d\rho} = (x_2 - x_1) \frac{d\beta}{d\rho}
$$

$$
\frac{dx}{dz} = (x_2 - x_1) \frac{d\beta}{dz}
$$

The derivatives of $\beta$ and $\sigma_x/\sigma_1$ with respect to $\rho$ as functions of $\rho$, and for various $z$ values, Eq. 12 and Eq. 13, are shown in Figures 2(c) and 2(d). The equations for $\beta$ and $\sigma_x/\sigma_1$, this time as a function of $z$ and for various $\rho$ values, are shown in Figures 3(a) and 3(b). Finally, the derivatives of $\beta$ and $\sigma_x/\sigma_1$ with respect to $z$ as functions of $z$, and for various $\rho$ values, Eq. 14 and Eq. 15, are shown in Figures 3(c) and 3(d). These derivatives can be used to visualise the sensitivity of the combined result to the imperfect knowledge on both the correlation $\rho$ and the uncertainty ratio $z$ of the individual estimators. With this information the stability of the combined result can be assessed and a decision can be taken on whether to refrain from combining. This decision should only be based on the parameters of the combination but not on the outcome for a particular pair of estimates $x_1$ and $x_2$. This is because these parameters are features of the underlying two-dimensional pdf of the estimators, whereas the two specific values are just a pair of estimates, i.e. a single possible likely or unlikely outcome of results. A suggestion for how to proceed is given in Section 9.

4 Estimator properties

In general, in experimental analyses an estimator is constructed by studying Monte Carlo simulated events that are taken as data substitutes. Using those events it is verified that the estimator is unbiased. By applying the method to data, the measured value of the estimator, i.e. the estimate, e.g. $x_1$, is obtained together with its statistical uncertainty. Subsequently, individual systematic uncertainties are obtained for the estimator and assigned to the estimate. For example, in top quark mass measurements like Ref. [10], this is achieved, e.g. by changing the reconstructed objects like leptons and jets within their uncertainties, by altering the underlying Monte Carlo model for the signal, and by varying the background evaluations from data or simulations. In these procedures, the systematic variations per source $k$ of uncertainty are chosen to be performed in an uncorrelated way, and the actual values of the uncertainties are considered one standard deviation Gaussian uncertainties. Consequently, the total systematic uncertainty is calculated as the square root of the quadratic sum of the contributions from the individual sources. Finally, the result is quoted as:

$$
x_1 = \text{value } \pm \text{stat } \pm \sqrt{\sum_k \text{syst}^2}.
$$

To enable their combination, the breakdown of systematic uncertainties is provided. Consequently, the features of the estimates are:

1) they are unbiased,
2) their uncertainties are assumed to be Gaussian,
3) the uncertainty sources are constructed to be uncorrelated,
When performing the combination of a pair $ij$ of estimates, for each source $k$ of uncertainty a correlation $\rho_{ijk}$ has to be assigned for that pair. The statistical uncertainties are either uncorrelated, or, for the case of two estimates obtained from overlapping or even the same data events, their correlation can be obtained within the analysis by means of pseudo-experiments, as described e.g. in Ref. [10]. For the systematic uncertainties, the value of the assigned correlation always is a physics motivated choice that can only be made with some uncertainty. The easiest case occurs if the uncertainties of the estimators have been determined in exactly the same way, e.g. within one experiment while using the identical procedure for all estimates. In this case, the assumption of $\rho_{ijk} = 1$ is justified, and any observed difference in the size of uncertainty $\sigma_{ik} \neq \sigma_{jk}$ is likely caused by the different sensitivities of the estimators to that particular source of uncertainty. The uncertainty of this correlation assumption can be assessed by varying the value of $\rho_{ijk}$ within bounds to be chosen. Given the estimator property 3) for each source $k$ this should be performed independently from all other sources.

A more complicate situation arises however, when combining estimates obtained by different experiments, which even may have partly been derived without knowledge on the procedure applied for the respective other result. Given the difference in strategy, there may be a smaller correlation. In addition, even for $\rho_{ijk} = 1$ differences in the size of the uncertainty can originate from a different size of variation performed for the two estimators. As an example, one experiment may perform larger variations of Monte Carlo parameters than another, an example of which can be found in Ref. [11]. In this situation given the different dependences of $\beta$ and $\sigma_f/\sigma_1$ on $\rho$ and $\beta$, the difference can not be accounted for by changes in $\rho_{ijk}$, but the most appropriate choice is to vary $\sigma_{ik}$ and/or $\sigma_{jk}$.

Given the above, an individual assessment of the situation per source $k$ is strongly preferred. In contrast, any automated procedure very likely can not properly account for the specific situations of all sources $k$. In any case, all systematic variations on the assumptions should be performed obeying the features of the estimators listed above.

Frequently, the question arises whether a pair $(1, 2)$ of estimates is consistent. This can be decided upon using a $\chi^2$ that is defined as the squared ratio of the difference of the estimates, $\Delta$, and its uncertainty, $\sigma_\Delta$:

$$\chi^2_{12} = \left( \frac{\Delta}{\sigma_\Delta} \right)^2 = \frac{(x_1 - x_2)^2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$  \hspace{1cm} (19)$$

Which is the significance of the difference of the estimates of being inconsistent with zero. Alternatively, the related $\chi^2$ probability for one degree of freedom, $P(\chi^2, 1)$, can be used to evaluate the probability for an even larger $\chi^2$ to occur for any other pair [6]. Finally, only consistent estimates should be combined, otherwise the combined result is not trustworthy. An example of such a combination is shown in the next section.

5 Relative uncertainties

The formulas described above assume that the estimators have absolute uncertainties $\sigma_{ik}$. Here, the term absolute uncertainty means that the value of the uncertainty is identical for all possible values of the estimator pdf, i.e. it is independent of the actual value of the estimate. This means it is the same for the actual estimate, any combined value, and the true value. Therefore, irrespectively of whether it was calculated for the estimate, it also applies to the combined value. In contrast, a relative uncertainty (e.g. of some percent) varies across the pdf and depends on the actual value of the estimate. Consequently, in this situation the uncertainty assigned to the estimate is formally incorrect, since it should correspond to the uncertainty of the estimator pdf, which is a constant.

Within the BLUE method this can be accounted for approximately by performing the combination in an iterative way, in which, starting from the initially assigned value, after each iteration the uncertainty is replaced by the expected uncertainty of the true value $x_T$, approximated by the one of the combined value $x$. For most applications, for a given source $k$ of systematic uncertainty, a linear dependence of the uncertainty $\sigma_{ik}$ on $x$ is assumed3, however, there also exist more complicate cases like the one discussed in Ref. [12].

It is worth noticing that during the iterations the originally assigned uncertainties of the estimates are altered, albeit at unchanged correlation assumptions. For example, when using the same linear dependence for all estimates $i$ and a given source of uncertainty $k$, this means that at convergence the uncertainties from this source are identical for all estimates, i.e. they amount to a given fraction of the combined result. Assuming this behaviour for all uncertainties of a pair of estimates leads to $z = 1$, $\beta = 0.5$ for all possible values of $\rho$, and the combination reduces to a simple averaging of the estimates, i.e. $x = (x_1 + x_2)/2$, irrespectively of their

3 Sometimes in the literature the terms additive (absolute) and multiplicative (relative) uncertainties are used instead.
4 Typically, e.g. for counting experiments, the estimate is proportional to the observed number of events $N$, whereas the statistical uncertainty scales with $\sqrt{N}$, i.e. it is not linear in the estimate.
Table 1 Comparison of the combinations for Peelle’s Pertinent Puzzle for the BLUE method with absolute and relative uncertainties using various scenarios for the estimates and their correlation. The five scenarios analysed are: A the original values for the estimates, k = 0, 1 and correlations ρ_{12k}, B = A but with all uncertainties scaled by a factor two, C/D = A but with a changed value for the second estimate and with the original/rescaled uncertainties, and E = A but with a decreased value of the assumed correlation for the systematic uncertainty. The estimates are listed together with their uncertainties. In addition given are the parameters and results of the combination.

| Estimates | Value | Stat | Syst | Full |
|-----------|-------|------|------|------|
| x         | σ₀    | σ₁   | σ    | Comment    |
| A         | 1.00  | 0.10 | 0.20 | 0.22 |
|           | 1.50  | 0.15 | 0.30 | 0.34 |
| B         | 1.00  | 0.20 | 0.40 | 0.45 |
|           | 1.50  | 0.30 | 0.60 | 0.67 |
| C         | 1.00  | 0.10 | 0.20 | 0.22 |
|           | 1.25  | 0.15 | 0.30 | 0.34 |
| D         | 1.00  | 0.10 | 0.20 | 0.22 |
|           | 1.25  | 0.13 | 0.25 | 0.28 |
| E         | 1.00  | 0.10 | 0.20 | 0.22 |
|           | 1.50  | 0.15 | 0.30 | 0.34 |

|           | x     | σ₀    | σ₁   | σ     | χ²   | ρ    | z    | β    | ς₂    | dβ/ς₁ | ρσ₂   | dρ/ς₁ | ρς₂   | dρ/ς₁ |
|-----------|-------|-------|------|-------|------|------|------|------|-------|--------|--------|--------|--------|--------|
| BLUE      |       |       |      |       |      |      |      |      |       |        |        |        |        |
| A         | 0.88  | 0.13  | 0.18 | 0.22  | 5.9  | 0.80 | 1.50 | -0.24| 0.98  | -2.60  | -0.45  | -0.55  | -0.15  |
| B         | 0.88  | 0.26  | 0.35 | 0.44  | 1.5  | 0.80 | 1.50 | -0.24| 0.98  | -2.60  | -0.45  | -0.55  | -0.15  |
| C         | 0.94  | 0.13  | 0.18 | 0.22  | 1.5  | 0.80 | 1.50 | -0.24| 0.98  | -2.60  | -0.45  | -0.55  | -0.15  |
| D         | 1.00  | 0.10  | 0.20 | 0.22  | 2.2  | 0.80 | 1.25 | 0.00 | 1.00  | -2.22  | 0.00   | -1.42  | 0.00   |
| E         | 1.15  | 0.08  | 0.17 | 0.19  | 1.6  | 0.04 | 1.50 | 0.30 | 0.85  | -0.19  | 0.37   | -0.29  | 0.17   |

| Relative uncertainties |       |       |      |       |      |      |      |      |       |        |        |        |        |
| A         | 1.25  | 0.09  | 0.25 | 0.27  | 8.0  | 0.80 | 1.00 | 0.50 | 0.95  | 0.00   | 0.26   | -2.50  | 0.47   |
| B         | 1.25  | 0.18  | 0.50 | 0.53  | 2.0  | 0.80 | 1.00 | 0.50 | 0.95  | 0.00   | 0.26   | -2.50  | 0.47   |
| C         | 1.02  | 0.10  | 0.21 | 0.23  | 2.3  | 0.80 | 1.20 | 0.08 | 1.00  | -1.95  | 0.09   | -1.66  | 0.06   |
| D         | 1.12  | 0.08  | 0.23 | 0.24  | 2.5  | 0.80 | 1.00 | 0.50 | 0.95  | 0.00   | 0.26   | -2.50  | 0.47   |
| E         | 1.25  | 0.09  | 0.21 | 0.20  | 1.7  | 0.04 | 1.00 | 0.50 | 0.72  | 0.00   | 0.35   | -0.52  | 0.36   |

correlation. Only the uncertainty σₓ depends on the value of the correlation.

Numerically, the difference of using absolute or relative uncertainties rarely is of importance, especially so when combining precision measurements. This is because a difference of n% between the estimates and the combined value only results in a relative change of n% in σ_{ik}. Given that σ_{ik} in itself is small, this likely ends up in very small differences in x and σₓ, in any case well below the size of the respective uncertainty.

At first sight a counter example is the original formulation of Peelle’s Pertinent Puzzle [8, 9], for which the estimates (labelled scenario A) are given in Table 1. The statistical uncertainties are uncorrelated and the systematic uncertainties are fully correlated, which results in ρ = 0.8. Given that a percentage uncertainty of 10% (20%) is quoted for the statistical (systematic) uncertainty, the ratio of the total uncertainty equals the ratio of the estimates, i.e. z = 1.5. The compatibility of the two estimates, calculated from Eq. 19, is bad, i.e. χ²_{12} = 5.9 and P(χ², 1) = 1.5%, which means whatever method is used, a combination of this pair of estimates is questionable.

Given the procedures applied to obtain the systematic uncertainty it should be possible to decide whether this is an absolute or relative uncertainty. Here, the combination is performed for both assumptions, i.e. using either absolute and relative uncertainties for all sources of uncertainty, see also Ref. [9]. The results are listed in Table 1, scenario A. In the case of relative uncertainties, given the combined value, the final statisti-
cal (systematic) uncertainties assigned to the estimates are 0.13 (0.25), i.e. they are equal for both estimates and different from the values quoted in the upper part of the table. Due to the changes in uncertainties, for the BLUE method with relative uncertainties the compatibility of the two estimates is even worse, i.e. $\chi^2_{12} = 8.0$ and $P(\chi^2, 1) = 0.5\%$. As explained above, by construction, the result for $x$ is the mean of the two estimates.

For this example the difference in $x$ obtained with the two methods is significant. Using a $\chi^2$ analogous to Eq. 19 this time defined as the squared difference of the combined results, divided by the larger of the two variances the results are incompatible, i.e. the resulting value is $\chi^2 = (0.37/0.27)^2 = 1.9$. However, this is a mere consequence of the incompatible input and not of the differences of the method. This can be seen by the additional scenarios B-E given in Table 1. They are designed to artificially improve the compatibility of the input while using different aspects of the estimates. The parameters of the combinations depend on $\rho$ and $z$ such that they only change, if one of those changes, in particular, they are independent of the assumptions on the estimates.

The estimates are altered by either changing: (B) the size of the uncertainties, (C, D) the value of the less precise estimate, and (E) the correlation of the systematic uncertainties. The target value of the estimate compatibility for the BLUE method with absolute uncertainties was a $\chi^2$ of about 1.5.

For scenario B the uncertainties are doubled. For none of the methods does this change the relative importance of the estimates, however it improves their compatibility.

For scenarios C, D the value of the less precise estimate $x_2$ is reduced to make it more compatible with $x_1$. The difference of the two scenarios is that in C the changed value for $x_2$ is considered another possible outcome, namely a value consistent with the conditional pdf for $X_2$. Consequently, the originally assigned uncertainties are kept. In contrast, for scenario D the uncertainties are scaled to amount to the same fractional uncertainties as were originally assumed in A. Again the compatibility of the estimates and the combined results are improved. For scenario D, by construction, all parameters of the combined result obtained using relative uncertainties are identical to the ones in scenario A, but for the combined value, because the mean is changed due to the changed estimate $x_2$.

For scenario E the correlation is reduced, yielding a similar level of agreement. Here, again by construction, the combined result obtained using relative uncertainties is identical to the one in scenario A, but for its uncertainty which is reduced due to the smaller correlation of the estimates. Finally, the resulting compatibilities of the combined results are $\chi^2 = 1.9, 0.49, 0.12, 0.25, 0.25$ for scenarios A, B, C, D, E, which means that the differences of the methods diminish when using consistent input. This means the difference observed for scenario A is not caused by the differences in the methods, but by the inconsistency of the input.

The definition of whether a given source of uncertainty is an absolute and relative uncertainty has to be made in view of the actual procedure followed to determine this uncertainty. Nevertheless, as purely numerical examples, and without any physics motivation, for a number of examples of publicly available combinations, to evaluate the numerical importance for real applications, the results for both assumptions are given below. All values quoted follow the convention of Eq. 18. The two examples for which originally relative uncertainties are assigned are the combination of lifetimes of $B$ mesons [12], and of the cross-section for single top quark production at the LHC [13]. In these cases for comparison absolute uncertainties are assumed for all sources. The two examples for which originally absolute uncertainties are assigned are the latest combinations of the measurements of the top quark mass $m_{\text{top}}$, performed at the Tevatron [14] and the LHC [11]. In these cases for comparison relative uncertainties are assumed for all sources of systematic uncertainties.

The corresponding results are for the $B$-lifetime:

$$\tau \ [\text{ps}] = 1.13 \pm 0.09 \pm 0.11 \ (\text{relative})$$

$$\tau \ [\text{ps}] = 1.13 \pm 0.09 \pm 0.09 \ (\text{absolute}), \ (20)$$

for the single top quark production cross-section:

$$\sigma \ [\text{pb}] = 85.3 \pm 4.1 \pm 11.5 \ (\text{relative})$$

$$\sigma \ [\text{pb}] = 83.7 \pm 4.6 \pm 11.2 \ (\text{absolute}), \ (21)$$

for $m_{\text{top}}$ measured at the Tevatron:

$$m_{\text{top}} \ [\text{GeV}] = 173.21 \pm 0.51 \pm 0.71 \ (\text{absolute})$$

$$m_{\text{top}} \ [\text{GeV}] = 173.26 \pm 0.51 \pm 0.71 \ (\text{relative}), \ (22)$$

and finally, for $m_{\text{top}}$ measured at the LHC:

$$m_{\text{top}} \ [\text{GeV}] = 173.29 \pm 0.23 \pm 0.92 \ (\text{absolute})$$

$$m_{\text{top}} \ [\text{GeV}] = 173.29 \pm 0.23 \pm 0.92 \ (\text{relative}). \ (23)$$

In all cases the difference of the pair of results is small compared to their statistical uncertainties. The results on $m_{\text{top}}$ are almost indistinguishable (Tevatron) or, at the quoted precision, identical (LHC).

6 The concept of reduced correlations

Reduced correlations postulate that for each pair of estimates, e.g. the pair (1, 2), and a given source of
uncertainty \( k \), the smaller of the individual uncertainties, e.g. \( \sigma_{1k} < \sigma_{2k} \), is fully correlated, and the remainder is uncorrelated. This replaces the covariance \( \rho_{12k}\sigma_{1k}\sigma_{2k} \) by the square of the smaller of the individual uncertainties, e.g. \( \sigma_{1k}^2 \) for this source, see e.g. [15]. This is equivalent to assuming the correlation to amount to the ratio of the smaller to the larger uncertainty, \( \rho_{12k} = \sigma_{1k}/\sigma_{2k} = 1/z_{12k} \).

The impact of this concept can be seen by analysing the contribution of the source \( k \) to the covariance matrix separated into the postulated uncorrelated (u) and correlated (c) parts that reads:

\[
V_k = \begin{pmatrix}
\sigma_{1k}^2 & \rho_{12k}\sigma_{1k}\sigma_{2k} \\
\rho_{12k}\sigma_{1k}\sigma_{2k} & \sigma_{2k}^2
\end{pmatrix} = \begin{pmatrix}
0 & 0 \\
0 & \sigma_{1k}^2 - \sigma_{1k}^2
\end{pmatrix}_u + \begin{pmatrix}
\sigma_{1k}^2 & \sigma_{2k}^2 \\
\sigma_{2k}^2 & \sigma_{1k}^2
\end{pmatrix}_c
\]

By construction, this replaces one source of uncertainty by two and assigns zero (full) correlation to the first (second) term, i.e. \( \sigma_{1k}^2 = 1 \cdot \sigma_{1k}\sigma_{1k} \). Typically, this concept is applied to sources that initially had a correlation of \( \rho_{12k} > 1/z_{12k} \) or even \( \rho_{12k} = 1 \). In this situation, the correlation is always reduced with respect to the initial value, hence the name.

If this source is the only uncertainty, this will lead to \( \beta = 0 \). For an arbitrary number of sources the covariance with reduced correlations reads:

\[
\rho_{\text{red}}\sigma_{1}\sigma_{2} = \sum_{1<2} \sigma_{1k}^2 + \sum_{2<1} \sigma_{2k}^2 \leq \sum_{1} \sigma_{1k}^2 \leq \sigma_{1}^2
\]

where \( \rho_{\text{red}} \) is the total reduced correlation of the pair of estimates. The first (second) term sums the variances of the sources where \( x_1 \) (\( x_2 \)) has the smaller uncertainty. If the second estimate does not have a smaller uncertainty for any of the sources, for the first inequality the equal sign is realised, otherwise replacing \( \sigma_{2k}^2 \) by \( \sigma_{1k}^2 \) in the second sum will increase the covariance. Finally, if there are also no initially uncorrelated uncertainties, for the second inequality the equal sign is valid. In any case, comparing the first and last terms the result is:

\[
\rho_{\text{red}} \leq \frac{\sigma_{1}}{\sigma_{2}} = \frac{1}{z},
\]

which means \( \beta \geq 0 \) is ensured by the method. As a consequence, by construction \( x \) is always within \( x_1 \) and \( x_2 \). However, as has been shown above, due to the conditional probability, the true value \( x_T \) is outside this interval in the majority of all cases.

Apart from this deficiency, also from physics arguments this procedure is questionable as can be seen from an example. Let's assume there are two estimates of the same experiment, which suffer from the same source of uncertainty (lets say an energy scale uncertainty), but apply different phase space requirements, e.g. on the jet transverse momentum \( p_t \). Typically, the uncertainty on these scales decrease with increasing \( p_t \), such that the estimate with the stronger requirement will have the smaller uncertainty. The method now effectively assigns a correlation to the uncertainty from this source, which is zero (one) for \( p_t < p_{t,\text{min}} \) (\( p_t > p_{t,\text{min}} \)), where \( p_{t,\text{min}} \) is the larger of the two minimum transverse momenta required for the two estimates, see Eq. 24. As a result, firstly, the limit of the correlation for \( p_t = p_{t,\text{min}} \) from above and below is different. Secondly, the uncertainty slightly below \( p_{t,\text{min}} \) is by construction independent of the one slightly above. Given that the value of \( p_{t,\text{min}} \) is arbitrary, and that the facts that lead to the uncertainty of the energy scale do not change across the threshold, an unphysical situation. This is an example, where, for \( \rho_{12k} = 1 \) a difference in \( z_{12k} \) is attempted to be cured by an ad hoc change in \( \rho_{12k} \). However, the dependence of \( \beta \) and \( \sigma_x/\sigma_1 \) on \( p \) and \( z \) are different.

Sometimes this method is advertised as a conservative approach, in the sense that it will increase the variance of the combined result \( \sigma_{x}^2 \). However, this is not true for all situations. If, for a given \( z \), e.g. the value of \( \rho \) is only slightly larger than \( 1/z \), the resulting \( \rho_{\text{red}} \) may be much smaller than \( 1/z \), such that \( \sigma_x/\sigma_1 \) is actually reduced from its initial value, see Eq. 11 and Figure 2(b). Consequently, the uncertainty assigned to the combined value by using reduced correlations may be either larger or smaller, depending on the initial value of \( \rho \) and the size of the reduction. For a specific example the impact is evaluated below.

7 Methods to maximise the variance

On top of the reduced correlations discussed in the previous section, and apart from the proposal to simply ignore estimates with negative BLUE weights, a number of methods have been suggested to arrive at a conservative combined estimate, i.e. to maximise the variance of the combined result \( \sigma_{x}^2 \). All attempts work by reducing the correlation in an artificial, but controlled way. Given Figure 2(b) they will only be active for \( \rho > 1/z \) which means \( \beta < 0 \). The three methods suggested in Ref. [16] multiply the initially assigned correlations per source \( k \) for any pair \( ij \) of estimates by factors \( f_{ijk} \). These factors are either chosen:

i) globally, \( f_{ijk} = f \) for all \( i,j,k \),

ii) per uncertainty source, \( f_{ijk} = f_k \) for all \( i,j \), or

iii) per pair of estimates, \( f_{ijk} = f_{ij} \) for all \( k \).

All methods are not flexible enough and do not obey some of the properties of the estimates outlined in Section 4. As examples, by varying all sources simultane-
ously the method i) does not obey property 3) of the estimates, namely that all sources are assumed to be uncorrelated. It also does not take into account that the knowledge on the correlation may differ from source to source. Method ii) does not take into account that the uncertainties on ρ_{ijk} likely are better known for pairs of estimates from the same experiment, than for pairs of estimates from different experiments, or even obtained at different colliders. Method iii), although calculated per pair, in reality corresponds to specific ρ_{ijk} values. Since the variation is done per pair, e.g. (i, j) or (i', j), where i, i' are assumed to be estimates from the same experiment and j from another experiment, this very likely leads to very different assumptions on the correlation for source k across experiments. Again, the available knowledge on this can not be respected by this automated procedure.

8 A hypothetical example

The impact of the reduced correlations and the three ways to maximise the variance of the combined result are discussed on the basis of a hypothetical example, motivated by typical estimates occurring in top quark mass measurements. For simplicity, only two estimates and three uncertainty sources are used. The extension to more estimates and uncertainty sources is straightforward.

The two estimates are given in Table 2. They are analysed for four different scenarios in which the assumption on either the correlation, or the size of the uncertainty for one of the sources is changed one at a time. Using Eq. 19 and calculating \( P(\chi^2, 1) \), the compatibility of the estimates is assessed for the BLUE method and for all scenarios\(^6\). The values obtained with this procedure are \( P(\chi^2, 1) = 0.33, 0.45, 0.18, 0.18 \), for scenarios A, B, C, D.

Given the assigned correlations per source and \( z = 1.54 \), scenario A corresponds to a situation where \( \rho = 0.78 > 1/z = 0.65 \) and consequently \( \beta = -0.22 < 0 \). This situation is visualised in Figure 4, where the eight sub-figures correspond to Figures 2–3, and the black points to the pair of estimates investigated. Consequently, in Figure 4 (2), the point is to the right of the peak which sits at \( \beta = 0 \).

In Figure 4, the sensitivity of the combination to variations of \( \rho \) and \( z \) is visualised by the three curves per sub-figure. For a given functional dependence of one of the functions, e.g. \( \beta (\rho) \), they show the sensitivity to the respective other parameter, here \( z \), using the actual value (dashed line) and two changed values (full lines). For the pair of changed values, either \( z \) is multiplied by 0.9 or 1.1, Figure 4 (1–4), or \( \rho \) is changed by \( \pm 0.1 \), Figure 4 (5–8). This indicates the impact of 10% uncertainties on their respective initial values. The figure shows that the dependence of \( \beta \) and \( \sigma_x/\sigma_1 \) and their derivatives on one of the parameters strongly depends on the value the respective other parameter has. As an example, the sensitivity to \( \rho \) of the derivative of \( \beta \) with respect to \( z \), visualised by the spread of the three lines in Figure 4 (7), varies strongly with \( z \). For the chosen example it is smallest close to the black point, i.e. to the actual pair of values of \( \rho \) and \( z \). For two estimates with \( z = 1.1 \) the sensitivity to \( \rho \) would be much larger. In contrast, for the derivative of \( \sigma_x/\sigma_1 \) with respect to \( z \), Figure 4 (8), the chosen point in phase space lies close to the region with the largest spread of the curves, signalling a large \( \rho \) dependence. The quoted derivatives of \( \sigma_x/\sigma_1 \) in Figure 4 (4, 8), show that a 10% change in \( \rho \) has a much larger impact on the uncertainty of the combined value than the corresponding change in \( z \), which means that for this particular case, it is more important to correctly determine \( \rho \) rather than \( z \). The values of all parameters and for all scenarios investigated are listed in Table 2.

Given the initial correlation assumptions, the reduced correlations act on both systematic uncertainties and yield \( \rho = 0.44 \). As a result of this strong reduction of the correlation, the resulting value of \( \sigma_x/\sigma_1 \) is lower than for the initially assigned correlations, see Table 2. Because there is a non zero uncorrelated component to the uncertainty for both estimates, the reduced correlations can not switch off \( x_2 \) completely, as it would otherwise do, see Eq. 25.

For this example, the three methods for maximising the variance, at the quoted precision, all give the same combined result, which is achieved for \( f = 0.83 \), \( f_k = 0.34 \) (or \( f_k = 1/0.77 \) for \( k = 1/2 \), and finally, \( f_{ij} = f = 0.83 \), respectively. Consequently, with these algorithms, the second estimate is switched off in different ways, i.e. they all give \( \beta = 0 \) and \( x = x_1 \), as it would be the case if estimates with negative weights would be ignored.

For scenario B the systematic uncertainty \( k = 1 \) is assumed to be uncorrelated rather than fully correlated. By this assumption the correlation is reduced such that the point moves to the left of the peak in Figure 4 (2) and the BLUE combination results in a positive value for \( \beta \). Given that \( \beta \) is very close to zero the estimate \( x_2 \) would improve \( x_1 \) by less than 1%. For the reduced correlations, which now only act on the source \( k = 2 \), the correlation is further decreased, such that the predicted

---

\(^6\)The reduced correlations and the methods to maximise the variance reduce the value of \( \rho \) while keeping \( \rho > 0 \). Consequently, the resulting \( \chi^2 \) values obtained from Eq. 19 are always smaller.
Table 2 Combinations of two correlated estimates using the BLUE method for different scenarios (A–D), and using the different methods described in the text. The two estimates used are given together with their uncertainties. The four scenarios analysed for the estimates 1, 2, and uncertainties, \(k = 0, 1, 2\) with correlations \(\rho_{12k}\) are: A the default values of the uncertainties with two fully correlated systematic uncertainties, B=A but the first systematic uncertainty is assumed to be uncorrelated, C(D)=A but for the second systematic uncertainty the smaller (larger) of the two values is taken for both estimates. For the maximisation of the variance no values are given for scenarios B–D, since they coincide with the BLUE results.

### Estimates

| Value | Stat | Sys1 | Sys2 | Syst | Full | Comment |
|-------|------|------|------|------|------|---------|
| \(x_1\) | \(\sigma_0\) | \(\sigma_1\) | \(\sigma_2\) | \(\sigma_{syst}\) | \(\sigma\) |
| A | 172.10 | 0.60 | 0.50 | 0.70 | 0.86 | 1.05 | Correlations of \(\rho_{120} = 0\), \(\rho_{121} = \rho_{122} = 1\) |
| B | 172.10 | 0.40 | 0.70 | 1.40 | 1.57 | 1.62 |
| C | 172.10 | 0.60 | 0.50 | 0.70 | 0.86 | 1.05 | Changed correlation for the first systematics \(\rho_{121} = 0\) |
| D | 172.10 | 0.40 | 0.70 | 0.70 | 0.99 | 1.07 |

### Reduced correlations

| Value | Stat | Sys1 | Sys2 | Syst | Full | Comment |
|-------|------|------|------|------|------|---------|
| \(x\) | \(\sigma_0\) | \(\sigma_1\) | \(\sigma_2\) | \(\sigma_{syst}\) | \(\sigma\) | \(\rho\) | \(z\) | \(\beta\) | \(\frac{\sigma}{\pi_1}\) | \(\frac{d\beta}{\pi_1}\)| \(\frac{1}{\pi_1}\) | \(\frac{d\beta}{\pi_2}\) | \(\frac{1}{\pi_2}\) | \(\frac{d\beta}{\pi_2}\) |
| A | 171.88 | 0.74 | 0.46 | 0.55 | 0.71 | 1.02 | 0.78 | 1.54 | -0.22 | 0.98 | -2.32 | -0.42 | -0.48 | -0.14 |
| B | 172.17 | 0.56 | 0.47 | 0.75 | 0.88 | 1.04 | 0.58 | 1.54 | 0.07 | 1.00 | -0.84 | 0.10 | -0.45 | 0.04 |
| C | 172.56 | 0.37 | 0.59 | 0.70 | 0.92 | 0.99 | 0.75 | 1.02 | 0.46 | 0.94 | -0.14 | 0.27 | -1.96 | 0.43 |
| D | 171.56 | 0.37 | 0.59 | 1.40 | 1.52 | 1.57 | 0.89 | 1.01 | 0.46 | 0.98 | -0.33 | 0.26 | -4.59 | 0.45 |

### Maximisation of the variance

| Value | Stat | Sys1 | Sys2 | Syst | Full |
|-------|------|------|------|------|------|
| A | 172.10 | 0.60 | 0.50 | 0.70 | 0.86 | 1.05 | 1.00 | -1.13 |
| B | 172.26 | 0.51 | 0.51 | 0.73 | 0.89 | 1.02 | 0.44 | 0.97 | -0.51 |
| C | 172.32 | 0.47 | 0.52 | 0.75 | 0.86 | 0.98 | 0.29 | 1.54 | 0.22 | 0.94 | -0.34 |
| D | 172.57 | 0.37 | 0.55 | 0.70 | 0.89 | 0.96 | 0.66 | 1.02 | 0.47 | 0.92 | -0.08 |

Improvement in precision of 6% is even larger than for scenario A. In contrast, since the maximisation of the variance is only attempted to the right of the peak in Figure 4 (2) none of the algorithms i)–iii) is proposing any change.

The scenarios C and D implement the situation in which for \(\rho_{12k} = 1\) for \(k = 2\) the difference has been caused by the use of different procedures. Either estimate \(x_1\) has a ‘too crude’ procedure assigned such that not all features of this source are accounted for, and the quoted uncertainty is underestimated (C), or for estimate \(x_2\) a ‘too generous’ variation was performed such that the quoted uncertainty is overestimated (D). In these scenarios the BLUE combinations give significant and different improvements. This means it is worth investigating whether the difference in uncertainty is caused by different sensitivities of the estimators used, or by different procedures followed and in the latter case if possible, to harmonise those.

For the reduced correlations, given the assigned identical uncertainties, the source \(k = 2\) is not altered, see Section 6. Because in addition, the uncertainties for \(k = 1\) are much smaller than those for \(k = 2\), the method is almost switched off, i.e. \(\rho_{red} \approx \rho\). It is worth noticing that the results of the BLUE method for scenarios C and D are much different from the result of the reduced correlations for scenario A, exemplifying the different sensitivities to \(\rho\) and \(z\). For the BLUE method and at the quoted precision, the values of \(\beta\) in C and D are identical, and much different from the one for scenario A. They also differ strongly from the value obtained by applying the reduced correlations for scenario A.
9 How to decide on and perform a combination

The proposed procedure is described for the situation of \( m \) estimates of the same observable and fully respects the properties of the estimates given in Section 4. The extension to more than one observable is straightforward. As an example, the procedure is applied to the input of the latest combination of \( m_{\text{op}} \) measurements performed at the Tevatron [14]. Based on the initial input and the default assumptions on the correlations, the following questions are addressed:

I) Are the estimates compatible?

II) Which estimates are worth combining?

III) What are the consequences of varying \( \rho_{ijk} \)?

IV) What are the consequences of varying \( z_{ijk} \)?

Clearly, the outcome will depend on the initial assumptions, see e.g. the results for Peelle’s Pertinent Puzzle listed in Table 1.

For answering I), the compatibility is addressed by the \( \chi^2 \) defined in Eq. 19, and calculating \( P(\chi^2, 1) \). Inconsistent sets of estimates should not be combined, instead the reason for this should be searched for. From the 66 \( \chi^2 \) values of the pairwise compatibility tests for the twelve estimates from Ref. [14], 18 are above one, of which one (one) is above two (three), the smallest value being about \( P(\chi^2, 1) = 8\% \), resulting in a reasonable distribution of \( \chi^2 \) values.

For answering II), starting from the most precise estimate \( i \) it is proposed to rank the estimates \( j \neq i \) by their importance, defined as their potential improvement in the most precise estimate, calculated using Eq. 11 and identifying \( 12 = ij \). This procedure takes into account the correlation and the relative uncertainty of the two estimates, but is deliberately independent from the existence of all other estimates. This sugges-
Table 3 The list of estimates from Ref. [14]. The most precise estimate is CDF(II) l+j. The other estimates are listed according to their importance, defined as the achieved improvement of the combined uncertainty with respect to the most precise estimate, obtained by performing pairwise combinations of each estimate with the most precise one. The correlation ρ and relative uncertainties z are given together with the two main parameters of the combination, β and σz/σ1 and their derivatives with respect to ρ and z. Entries quoted as 0.00 mean that the absolute value of the actual number was below 0.005.

| Estimate        | Value | Stat | Syst | ρ   | z     | β   | dβ | σz/σ1 | dσ | dz² |
|-----------------|-------|------|------|-----|-------|-----|----|-------|----|-----|
| CDF(II) l+j     | 172.85| 0.52 | 0.98 |     |       |     |    |       |    |     |
| D0(II) l+j      | 174.94| 0.83 | 1.24 | 0.30| 1.34  | 0.30| 0.91| -0.27 | 0.31| -0.46| 0.20|
| CDF(I) l+j      | 176.10| 5.10 | 5.31 | 0.49| 6.61  | -0.06| 0.93| -0.19 | -0.44| 0.01 | -0.01|
| CDF(II) Met     | 173.95| 1.26 | 1.37 | 0.32| 1.67  | 0.17| 0.96| -0.41 | 0.25 | -0.29| 0.10|
| CDF(II) had     | 172.47| 1.43 | 1.41 | 0.29| 1.80  | 0.15| 0.96| -0.39 | 0.24 | -0.23| 0.08|
| D0(II) dil      | 174.00| 2.36 | 1.49 | 0.18| 2.51  | 0.09| 0.98| -0.33 | 0.20 | -0.09| 0.03|
| CDF(I) had      | 186.00| 10.00| 5.72 | 0.30| 10.34 | -0.02| 0.98| -0.11 | -0.22| 0.00 | 0.00|
| CDF(II) dil     | 170.28| 1.95 | 3.13 | 0.48| 3.31  | -0.07| 0.98| -0.43 | -0.25| -0.01| -0.02|
| CDF(II) l+j     | 167.40| 10.30| 4.90 | 0.29| 10.24 | -0.02| 0.98| -0.11 | -0.22| 0.00 | 0.00|
| CDF(I) l+j      | 180.10| 3.60 | 3.87 | 0.22| 4.75  | 0.00| 1.00| -0.22 | -0.02| 0.01 | 0.00|
| CDF(I) had      | 186.00| 10.00| 5.72 | 0.30| 10.34 | -0.02| 0.98| -0.11 | -0.22| 0.00 | 0.00|
| D0(II) l+j      | 174.94| 0.83 | 1.24 | 0.30| 1.34  | 0.30| 0.91| -0.27 | 0.31| -0.46| 0.20|
| D0(II) dil      | 174.00| 2.36 | 1.49 | 0.18| 2.51  | 0.09| 0.98| -0.33 | 0.20 | -0.09| 0.03|
| CDF(II) had     | 166.90| 9.00 | 2.82 | 0.08| 8.46  | 0.00| 1.00| -0.12 | 0.04 | 0.00 | 0.00|
| D0(II) l+j      | 178.40| 12.30| 3.61 | 0.11| 11.51 | 0.00| 1.00| -0.09 | -0.03| 0.00 | 0.00|
| CDF(II) dil     | 170.28| 1.95 | 3.13 | 0.48| 3.31  | -0.07| 0.98| -0.43 | -0.25| -0.01| -0.02|

Solved according to importance

| Estimate | Value | Stat | Syst | ρ   | z   | β   | dβ  | σz/σ1 | dσ  | dz²  |
|----------|-------|------|------|-----|-----|-----|-----|-------|-----|------|
| CDF(II) l+j | 172.85 ± 0.52 ± 0.98 |     |      |     |     |     |     |       |     |      |
| D0(II) l+j  | 173.48 ± 0.44 ± 0.91 |     |      |     |     |     |     |       |     |      |
| CDF(I) l+j  | 173.26 ± 0.54 ± 0.77 |     |      |     |     |     |     |       |     |      |
| CDF(II) Met | 173.33 ± 0.52 ± 0.75 |     |      |     |     |     |     |       |     |      |
| CDF(II) had | 173.23 ± 0.51 ± 0.73 |     |      |     |     |     |     |       |     |      |
| D0(II) dil  | 173.23 ± 0.51 ± 0.73 |     |      |     |     |     |     |       |     |      |
| CDF(I) had  | 173.10 ± 0.50 ± 0.73 |     |      |     |     |     |     |       |     |      |
| CDF(II) dil | 173.13 ± 0.50 ± 0.72 |     |      |     |     |     |     |       |     |      |
| CDF(II) l+j | 173.19 ± 0.51 ± 0.71 |     |      |     |     |     |     |       |     |      |
| D0(II) l+j  | 173.17 ± 0.51 ± 0.71 |     |      |     |     |     |     |       |     |      |
| D0(II) had  | 173.18 ± 0.51 ± 0.71 |     |      |     |     |     |     |       |     |      |
| CDF(I) l+j  | 173.21 ± 0.51 ± 0.71 |     |      |     |     |     |     |       |     |      |

Fig. 5 Results of successive combinations according to importance of the estimates of mtop from [14]. The first line shows the result of the most precise estimate. All following lines report the combined result after adding the estimate listed to the previously accumulated list. Combinations, below the line with its mtop value given in red never improve the total uncertainty by more than 1%.

The details of the hypothetical pairwise combinations are listed in Table 3. Looking at the parameters of the combination it is apparent that the importance of the exact knowledge of ρ and z strongly depends on the pair of estimate under consideration. As an example, the derivatives of σz/σ1 with respect to ρ vary by about a factor of 10-20 in absolute size. In addition, they have different signs, such that for some pairs the uncertainty on the combined result is reduced with reducing the correlation, for others is instead increased.

The first line in Figure 5 shows the result of the most precise estimate. All following lines report the results of successive combinations after adding the estimate listed to the previously accumulated list. If, as an example, an improvement in the total uncertainty of at least 1% for each individual remaining estimate to be included is desired, only the first five estimates should be combined.

If instead the estimates were sorted according to their absolute BLUE weights for the combination based on all estimates, which takes into account the correlations of all estimates (and the fact that the uncertainty is reduced on both sides of σz/σ1 = 1, β = 0), the same five estimates would have been chosen. If instead the estimates were sorted by their inverse variance 1/σ1², which ignores all correlations and weights the estimates use, and for which it is not worth to perform the difficult task of finding the appropriate variations in ρijk and zijk for assessing the stability of the combined result. The result of applying this procedure to the input of the latest combination of mtop measured at the Tevatron [14] is shown in Figure 5.
as if $\rho_{ij} = 0$, a slightly different list would be used. In the latter case, as can be seen from the values of $z$ reported in Table 3, the estimate CDF(I) l+j would not be used, but D0(II) dil would be used instead, despite the fact that looking at $\sigma_x/\sigma_1$ its impact is much smaller, demonstrating the large importance of the correlation.

When using the proposed method, the corresponding result of $m_{\text{top}}$ is shown in red. The BLUE weights of the five estimates in the order they appear in Figure 5 are: 0.61, 0.23, −0.06, 0.12, 0.10. No combination below this line improves the total uncertainty by more than 1%.

The situation for the pair containing the most precise estimate and the one with the negative BLUE weight is shown in Figure 6. For all sub-figures and all coordinate axes Figures 4 and 6 are drawn using identical ranges. Compared to Figure 4, there is a very flat behaviour for this pair of estimates, but for the derivative of $\sigma_x/\sigma_1$ with respect to $\rho$, Figure 6(4).

After performing the selection, the combination of all selected estimates is performed to get the central value and the breakdown of uncertainties. The compatibility of the selected estimates is improved, only two $\chi^2$ values exceed one, and the smallest $P(\chi^2, 1)$ value is about 19%. By construction, the result of the combination is very close to the one based on all estimates. Only little information is lost, but it is much more clear which estimates contain the information, and the investigation of the stability of the result is more simple.

As said above, the values of $\rho_{12k}$ and $z_{12k}$ are only known with some uncertainties. The task is to evaluate the consequences of this for the combined value. Looking at the figures of pairwise combinations like Figure 6 the most critical pairs and parameters can easily be identified. To assess its stability, individual uncertainty sources have to be investigated for possible variations of $\rho_{ijk}$ and $z_{ijk}$. This should be done in view of the details of the procedures applied, and it should be decided whether a variation in $\rho_{ijk}$ or $z_{ijk}$ is the appropriate choice.

To investigate III) independent variations per source $k$ are performed in which either $\rho_{ijk}$ is varied within a range determined by analysing the procedures used for the estimates, or, to get an indication of the possible effect, by multiplying the initially assigned correlation by a factor $r$, using the range $r = 1 \rightarrow 0$, and investigating the difference in the uncertainty of the combined result. If found appropriate, the observed differences could be
added quadratically to the uncertainty of the combined result to account for the uncertainties in the assigned correlations.

Since the detailed information on reasonable variations of the correlations is only available to the experiments that actually determined the estimates, for the example presented, the full range of \( r = 1 \rightarrow 0 \) has been used for all sources that remain correlated after the selection of estimates. For this example all variations lead to an increase of the combined value \( x \). The square root of the quadratic sum of the differences between the combined value of the default assignments and the ones obtained with the changed assumption on the correlation for all sources \( k \) amounts to 0.26 GeV. This number is dominated by a single source that contributes with 0.23 GeV. Given this, a correlated variation of the correlation assumption of all sources would result in an only slightly larger value of 0.29 GeV. However, this evaluation is disfavoured, since it violates property 3) of the estimates. In addition, the individual variations reveal which sources are the important ones for the stability.

To investigate IV) an indicative procedure is to assume identical values \( \sigma_{ik} = \sigma_{jk} \) for pairs of estimates and to repeat the combination. If this test results in large variations, it is advisable to understand whether the difference of \( \sigma_{ik} \) and \( \sigma_{jk} \) is due to different sensitivities of the estimators, or caused by different procedures followed in determining the uncertainties. In the latter case one should try to harmonise the procedures. For a numerical example of such a situation see Table 2. Investigating the procedures in detail, likely smaller variations of \( \sigma_{ik} \) turn out to be appropriate. Since this information is only available to the experiments that actually determined the estimates, for the example presented, this has not been investigated here. Depending on the details of the situation this can easily be more important than variations of \( r \), as can been seen from the example of the hadronisation uncertainty for the LHC \( m_{t\bar{t}} \) combination [11].

10 Summary and conclusions

In this paper the combination of correlated estimates has been reviewed using the Best Linear Unbiased Estimate (BLUE) method, mainly concentrating on the special case of two estimates of the same observable.

It has been shown that the underlying conditional probability inevitably leads to the fact that for positively correlated estimators in most of the cases, for a given pair of estimates to be combined, the true value is not within the interval spanned by the estimates. This fact should be respected by any combination method. All combination methods constructed to force the combined value to lie within the interval spanned by the estimates, violate this consequence of the conditional probability, and are wrong by construction. These methods will lead to worse results than the BLUE method that achieves this predicted behaviour by means of negative weights, which occur if they reduce the variance of its unbiased result. This situation is realised if the mean of the conditional probability of the less precise estimator is further away from the true value than the more precise estimate. This is the case whenever the correlation of the estimates \( \rho \) is larger than 1/\( z \), the ratio of the smaller and the larger uncertainty.

For any pair of estimates the dangerousness of their combination is encoded in the derivatives, of the main parameters of the combination, which are the weight of the less precise estimate \( \beta \), and the ratio of the uncertainty of the combined results and the more precise estimate \( \sigma_{x}/\sigma_{1} \). Those derivatives were derived with respect to \( \rho \) and \( z \), which themselves typically are only known with some uncertainty.

A critical assessment of methods proposed to deal with the uncertainty on the correlations has been given. Especially, it has been argued that reduced correlations mix \( \rho \) and \( z \) and act in an unphysical way. Other methods constructed to maximise the variance of the combined result are too general, do not respect all properties of the estimates, and do not reflect the different knowledge on the correlations that likely is available for estimates of the same experiment, or those obtained at the same collider compared to those from different experiments and/or colliders. For all other methods discussed, the uncertainty in the knowledge on the relative size of the uncertainties per source \( k \), is ignored throughout, however, this can be numerically much more important.

A detailed proposal for a procedure to combine a number of estimates and to evaluate the stability of the result has been made. It has been argued that the decision on including a given estimate into the combination should be based on its potential improvement with respect to the most precise estimate, i.e. on the relative gain of uncertainty of the combined value with respect to the most precise one for hypothetical pairwise combinations. The stability of the result should be assessed source by source in view of the uncertainty on the knowledge on \( \rho_{ijk} \) and \( z_{ijk} \), while respecting the properties of the estimates. Given the different dependence of the two parameters \( \beta \) and \( \sigma_{x}/\sigma_{1} \) of the pairwise combination on \( \rho \) and \( z \), it is advisable to assess the impact on a case by case basis performing appropriate changes in \( \rho_{ijk} \) or \( z_{ijk} \). A freely available software package to perform these investigations has been written.
Finally, all ways to assess the uncertainty on the combined result by variations of the \( \rho_{ijk} \) and \( z_{ijk} \) are only indicative of possible sensitivities. If large sensitivities occur, a better understanding and possibly harmonisation of the input, and ways to calculate, rather than postulate the correlations as is frequently done, are much preferred.

Acknowledgements I like to thank Allen Caldwell and Giorgetto Cortiana for fruitful discussions on the subject and useful suggestions to the manuscript.

References

1. L. Lyons and D. Gibaut and P. Clifford, Nucl. Instr. and Meth. A270, 110 (1988)
2. A. Valassi, Nucl. Instr. and Meth. A500, 391 (2003)
3. R. Nisius. BLUE: a ROOT class to combine a number of correlated estimates of one or more observables using the Best Linear Unbiased Estimate method. URL http://blue.hepforge.org
4. R. Brun and F. Rademakers, Nucl. Instr. and Meth. A389, 81 (1997). Proceedings of AIHENP’96 Workshop, Lausanne, Sep. 1996
5. L. Lyons, Statistics for nuclear and particle physicists (Cambridge University Press, 1986). ISBN 0 521 37934 2
6. R.J. Barlow, Statistics: a guide to the use of statistical methods in physical sciences (John Wiley & Sons Ltd., 1989). ISBN 0 471 92295 1
7. M.G. Cox, et al., Metrologia 43, 268 (2006). URL http://stacks.iop.org/43/S268
8. R.W. Peelle, Internal Memorandum, Oak Ridge National Laboratory, Washington DC, USA, unpublished (1987)
9. S. Chiba and D.L. Smith. A suggested procedure for resolving an anomaly in least-squares data analysis known as ‘Peelle’s Pertinent Puzzle’ and the general implications for nuclear data evaluation (1991). URL http://www.osti.gov/scitech/biblio/10121367
10. The ATLAS Collaboration, Eur. Phys. J. C72, 2046 (2012)
11. The ATLAS and CMS Collaborations. Combination of ATLAS and CMS results on the mass of the top-quark mass using up to 4.9 fb\(^{-1}\) of \( \sqrt{s} = 7 \) TeV LHC data (2013). URL http://cdsweb.cern.ch/record/1601811
12. L. Lyons and A.J. Martin and D.H. Saxon, Phys. Rev. D41, 982 (1990)
13. The ATLAS and CMS Collaborations. Combination of single top-quark cross-section measurements in the t-channel at \( \sqrt{s} = 8 \) TeV with the ATLAS and CMS experiments (2013). URL http://cdsweb.cern.ch/record/1601029
14. The Tevatron Electroweak Working Group for the CDF and DØ Collaborations. Combination of CDF and DØ results on the mass of the top quark using up to 8.7 fb\(^{-1}\) at the Tevatron (2013). URL http://arxiv.org/abs/1305.3929v2
15. The OPAL Collaboration, Eur. Phys. J. C40, 287 (2005)
16. A. Valassi and R. Chierici. Information and treatment of unknown correlations in the combination of measurements using the BLUE method (2013). URL http://arxiv.org/abs/1307.4003