

# A method and tool for combining differential or inclusive measurements obtained with simultaneously constrained uncertainties

Jan Kieseler CERN, 1211 Geneva 23, jan.kieseler@cern.ch

April 26, 2018

## Abstract

A method is discussed that allows combining sets of differential or inclusive measurements. It is assumed that at least one measurement was obtained with simultaneously fitting a set of nuisance parameters, representing sources of systematic uncertainties. As a result of beneficial constraints from the data all such fitted parameters are correlated among each other. The best approach for a combination of these measurements would be the maximization of a combined likelihood, for which the full fit model of each measurement and the original data are required. However, only in rare cases this information is publicly available. In absence of this information most commonly used combination methods are not able to account for these correlations between uncertainties, which can lead to severe biases as shown in this article. The method discussed here provides a solution for this problem. It relies on the public result and its covariance or Hessian, only, and is validated against the combined-likelihood approach. A dedicated software package implementing this method is also presented. It provides a text-based user interface alongside a C++ interface. The latter also interfaces to ROOT classes for simple combination of binned measurements such as differential cross sections.

## Change notes

The description below corresponds to release 1.2 of Convinvo. This release implements the possibility to define asymmetric uncertainties for orthogonal uncertainty sources using the text-based interface, only. As a consequence, some adaptations to the method needed to be performed with respect to the published version [1]. They are described in Section 2 and the changes to the software interface are described in Section 5.1.

## Contents

<b>1</b>	<b>Introduction</b>
<b>2</b>	<b>Combination method</b>
<b>3</b>	<b>Validation</b>
<b>4</b>	<b>Neglecting correlations</b>
<b>5</b>	<b>Program Installation and User Interface</b>
<b>6</b>	<b>Summary</b>

## 1 Introduction

A common technique to reduce the impact of systematic uncertainties, in particular in precision measurements in high energy physics, is to constrain the range of their variations by the data. A simultaneous fit of these variations and the parameters to be measured can be performed based on prior knowledge of the uncertainties and suitable distributions to constrain them. This technique can reduce the total uncertainty in many cases significantly, as in Refs. [2, 3, 4], and can be used to measure several parameters simultaneously (see e.g. Refs. [5, 6]). However, it leads to non-negligible correlations between all fitted parameters. These are particularly important when at least one measurement that uses this technique contributes to a combination.

1 The most consistent approach to such combination would  
be to define a *combined likelihood* based on the original  
2 models, including all systematic variations, and the orig-  
inal data the models were fit to. Also other commonly  
3 used combination techniques and the corresponding soft-  
4 ware tools would require this information [7, 8, 9, 10], which  
is publicly available only in very rare cases and often un-  
recoverable. This poses a serious problem for a consistent  
5 combination involving measurements obtained with simul-  
taneous nuisance parameter fits.

8 The method described here provides a solution for this  
problem, since it is based on the central results and their co-  
variance or Hessians, only. It allows separating constraints  
12

and correlations imposed by the previously fitted data from those that stem from prior knowledge of the systematic variations.

Therefore, the combination can be performed accounting for correlations between the measurements as well as for correlations and constraints within each individual measurement.

The dedicated software tool “Convino” is also presented in this article. It is specifically developed to perform combinations based on the method described here and provides a simple text-based user interface that can be used without knowledge of any programming language. Assumptions on correlations can be varied in an automated way. Moreover, partially correlated measurements of different quantities (e.g. bins of a differential distribution) can be combined simultaneously accounting for all correlations. In addition to the text-based interface, a C++ interface is provided to define the input to the combination. This interface can either read basic C++ standard library data types or ROOT [11] histogram and graph classes, which are commonly used in high-energy-physics analyses.

The combination method is described in Section 2. It is validated against the combined likelihood approach in Section 3. The effect of neglecting correlations within the same measurement is studied in Section 4. The installation and the user interface of the Convino program is described in Section 5.

## 2 Combination method

The combination is performed using a  $\chi^2$  minimisation. The  $\chi^2$  is defined as

$$\chi^2 = \sum_{\alpha} (\chi_{s,\alpha}^2 + \chi_{u,\alpha}^2) + \chi_p^2 \quad (1)$$

It is composed of three terms: the term  $\chi_{s,\alpha}^2$  represents the results of each measurement  $\alpha$  and its statistical uncertainties. It follows a Neyman or Pearson  $\chi^2$  definition, with the statistical uncertainty being fixed for each measurement or being scaled with the combined value, respectively. A measurement can aim to determine a set of quantities, e.g. bins of a differential cross section, where the Pearson definition is more applicable. Alternatively a single quantity can be measured, e.g. the mass of a particle from a fit of an invariant mass peak position, where the Neyman definition is presumably better suited to describe the measurement. In both cases, the measured quantities are referred to as estimates in the following. The additional term  $\chi_{u,\alpha}^2$  describes the correlations between the systematic uncertainties and constraints on them from the data for each measurement  $\alpha$ . The last term,  $\chi_p^2$ , incorporates prior knowledge of the systematic uncertainties and correlation assumptions between uncertainties of the measurements to be combined.

In a typical measurement that exploits simultaneous constraints on the uncertainties from the data, the sources of uncertainties are uncorrelated prior to the fit to the data and the knowledge about their variations is modelled by independent penalty terms in the original likelihood. The goal of the method described here is to find an approximation for this likelihood that allows disentangling the independent penalty terms from the constraints and correlations between them, which are typically introduced by the data<sup>1</sup>. Therefore, the first central assumption is that the original likelihood for a measurement  $\alpha$  can be approximated by:

$$\chi_{\alpha}^2 = (\chi_{s,\alpha}^2 + \chi_{u,\alpha}^2) + \sum_i (P_i^{\alpha}(\lambda_i))^2, \quad (2)$$

where  $(P_i^{\alpha}(\lambda_i))^2$  represents the penalty term for a systematic uncertainty  $i$  parametrised by a continuous parameter  $\lambda_i$ , such that  $\lambda_i = 1$  corresponds to a  $1\sigma$  variation. The terms  $\chi_{s,\alpha}^2$  and  $\chi_{u,\alpha}^2$  are defined as:

$$\chi_{s,\alpha}^2 = \sum_{\mu\nu} M_{\mu\nu}^{\alpha} \frac{\xi_{\mu}^{\alpha} \xi_{\nu}^{\alpha}}{\tau_{\mu}^{\alpha} \tau_{\nu}^{\alpha}} \text{ and} \quad (3)$$

$$\chi_{u,\alpha}^2 = \sum_{ij} \lambda_i D_{ij}^{\alpha} \lambda_j, \text{ with} \quad (4)$$

$$\xi_{\mu}^{\alpha} = x_{\mu}^{\alpha} - \bar{X}_{\mu} \text{ and} \quad (5)$$

$$\bar{X}_{\mu} = \bar{x}_{\mu} \prod_i (1 - \lambda_i K_{\mu i}^{\alpha} / x_{\mu}^{\alpha}) - \sum_i \lambda_i k_{\mu i}^{\alpha}. \quad (6)$$

Here,  $x_{\mu}^{\alpha}$  is the estimate  $\mu$  obtained in measurement  $\alpha$  and  $\bar{x}_{\mu}$  the combined value to be determined. The relation between both is given by  $\tau_{\mu}^{\alpha} = (\bar{X}_{\mu} / x_{\mu}^{\alpha})^{0.5}$  for the Pearson  $\chi^2$  definition. In case of the Neyman  $\chi^2$ , all  $\tau_{\mu}^{\alpha} = 1$ . The matrix  $M$  represents the inverted statistical covariance of the estimates. The parameters  $k_{\mu i}^{\alpha}$  and  $K_{\mu i}^{\alpha}$  model the effect of the systematic variations on the estimates for absolute or relative uncertainties, respectively. Here, a relative uncertainty is an uncertainty that scales with the measured value such as e.g. the luminosity uncertainty in a cross-section measurement, while absolute uncertainties have a constant value irrespective of the central result. In principle, also other classes of dependencies can be incorporated through additional terms in  $\xi_{\mu}^{\alpha}$ . The correlations between the uncertainties and the constraints that stem from the fit to the data are described by the matrix  $D^{\alpha}$ . Throughout this paper, indices  $\mu$  and  $\nu$  are used for estimates, while systematic uncertainties are denoted with indices  $i$  and  $j$ .

The procedure to obtain the parameters of  $\chi_{\alpha}^2$  from the measurements to be combined are discussed in the following - firstly for results obtained through a simultaneous nuisance parameter fit and secondly for the specific case of orthogonal uncertainties.

<sup>1</sup>The same procedure can be applied if correlations are present prior to the fit to the data

## 2.1 Measurements obtained with simultaneous fits

The second central assumption of the method is that the parameters of  $\chi_\alpha^2$  can be determined from the Hessian of measurement  $\alpha$  evaluated at the best-fit values,  $\vec{H}_{\text{in}}^\alpha$ . The entries of the Hessian can be ordered such that the matrix can be split in the following sub-matrices:

$$\vec{H}_{\text{in}}^\alpha = \begin{pmatrix} \tilde{D} & \kappa^T \\ \kappa & \tilde{M} \end{pmatrix}^\alpha, \quad (7)$$

where  $\tilde{M}$  describes the relation between the estimates  $x_\mu^\alpha$  and  $x_\nu^\alpha$ . The relation between systematic variations and the estimates is described by  $\kappa$ . The matrix  $\tilde{D}$  quantifies the relation between the systematic variations.

All parameters of  $\chi_\alpha^2$  are determined by calculating analytically the Hessian of  $\chi_\alpha^2$ ,  $H^\alpha(\vec{0})$ , and identifying the resulting terms with their counterparts of the input  $\vec{H}_{\text{in}}^\alpha$ . Here  $\vec{0}$  means  $\lambda_i = 0$  and  $x_\mu^\alpha - \bar{x}_\mu = 0 \forall i, \mu$ . The components are calculated as follows:

$$H_{\mu\nu}^\alpha(\vec{0}) = \frac{1}{2} \left( \frac{\partial^2}{\partial \Delta x_\mu^\alpha \partial \Delta x_\nu^\alpha} \chi_\alpha^2 \right) \Big|_{\vec{0}} = M_{\mu\nu}, \quad (8)$$

$$H_{\mu i}^\alpha(\vec{0}) = \frac{1}{2} \left( \frac{\partial^2}{\partial \Delta x_\mu^\alpha \partial \lambda_i} \chi_\alpha^2 \right) \Big|_{\vec{0}} = \sum_\nu M_{\mu\nu} (-\hat{k}_{\nu i}^\alpha), \quad (9)$$

with  $\hat{k}_{\nu i}^\alpha = K_{\nu i}^\alpha + k_{\nu i}^\alpha$ . The matrix  $M$  can be directly identified with  $\tilde{M}$ . Since  $\tilde{M}$  stems from a measurement of a physics quantity,  $\tilde{M}$  is positive definite and therefore invertible. Thus, the parameters  $\hat{k}_{\nu i}^\alpha$  can be determined as:

$$\hat{k}_{\nu i}^\alpha = - \sum_\mu ((M^\alpha)^{-1})_{\mu\nu} \kappa_{\mu i}^\alpha. \quad (10)$$

Since a variation  $i$  is either relative or absolute,  $\hat{k}_{\nu i}^\alpha$  equals either  $K_{\nu i}^\alpha$  or  $k_{\nu i}^\alpha$ , with the other parameter being 0. The terms describing the analytic relations between the systematic uncertainties are calculated as:

$$\begin{aligned} H_{ij}^\alpha(\vec{0}) &= \frac{1}{2} \left( \frac{\partial^2}{\partial \lambda_i \partial \lambda_j} \chi_\alpha^2 \right) \Big|_{\vec{0}} \\ &= D_{ij}^\alpha + \delta_{ij} \frac{1}{2} \frac{\partial^2}{\partial \lambda_i^2} (P_i^\alpha)^2 \Big|_{\lambda_i=0} + \sum_{\mu\nu} M_{\mu\nu}^\alpha \hat{k}_{\nu i}^\alpha \hat{k}_{\mu j}^\alpha. \end{aligned} \quad (11)$$

Only Gaussian penalty terms describing the prior knowledge of the uncertainties are considered<sup>2</sup> ( $P_i^\alpha(\lambda_i) = \lambda_i$ ), which simplify the equations, since in this case

$$\frac{1}{2} \frac{\partial^2}{\partial \lambda_i^2} (P_i^\alpha)^2 \Big|_{\lambda_i=0} = 1. \quad (12)$$

<sup>2</sup>In principle also other penalty terms (e.g. log-normal priors) can be accounted for. However, these lead to a non-constant  $D^\alpha$ .

In consequence  $D^\alpha$  becomes:

$$D_{ij}^\alpha = \tilde{D}_{ij}^\alpha - \delta_{ij} - \sum_{\mu\nu} M_{\mu\nu}^\alpha \hat{k}_{\nu i}^\alpha \hat{k}_{\mu j}^\alpha, \quad (13)$$

such that all parameters of Eq. 2 are defined.

## 2.2 Measurements with orthogonal uncertainties

For orthogonal uncertainties, the same initial  $\chi^2$  described in Eq. 2 is used. However, the calculation of its parameters does not necessarily require the full Hessian. Instead, the calculation of the parameters simplifies to:

$$\hat{k}_{\mu i}^\alpha = \frac{\sigma_{\mu i}^\alpha}{\sigma_{\mu \text{ total}}^\alpha}, \quad (14)$$

with  $\sigma_{\mu i}^\alpha$  being the contribution of uncertainty  $i$  to the total uncertainty  $\sigma_{\mu \text{ total}}^\alpha$  of estimate  $x_\mu^\alpha$ . In case of asymmetric uncertainties, described by an upward variation which results in an uncertainty  $\delta_{\mu,u}^\alpha$  and a downward variation resulting in  $\delta_{\mu,d}^\alpha$ ,  $\hat{k}_{\mu i}^\alpha$  becomes a function of  $\lambda_i$ , such that

$$\hat{k}_{\mu i}^\alpha(\lambda_i) = \begin{cases} \delta_{\mu,u}^\alpha, & \text{if } \lambda_i \geq 0 \\ \delta_{\mu,d}^\alpha, & \text{otherwise} \end{cases} \quad (15)$$

These changes propagate back to Eq. 6. The matrix  $D^\alpha$  is 0, the terms of  $M$  are calculated as:

$$M_{\mu\nu} = \frac{\rho_{\mu\nu}}{\sigma_\mu \sigma_\nu}. \quad (16)$$

Here,  $\rho_{\mu\nu}$  is the statistical correlation between estimate  $\mu$  and  $\nu$ , and  $\sigma_\mu$  and  $\sigma_\nu$  are the corresponding statistical uncertainties.

For the orthogonal uncertainties as well as for measurements obtained by simultaneous fits, the constraints from the prior knowledge of the uncertainties are implemented in Eq. 1 through the term:

$$\chi_p^2 = \sum_{ij} P_i(\lambda_i) (C^{-1})_{ij} P_j(\lambda_j), \quad (17)$$

with  $C$  being the matrix describing the correlation assumptions between the systematic uncertainties. In case no correlations are assumed, the term simplifies to:

$$\chi_p^2(\text{no corr}) = \sum_i P_i^2(\lambda_i). \quad (18)$$

Only Gaussian penalty terms are considered in the following, such that  $P_i(\lambda_i) = \lambda_i$ . For a combination,  $C$  will be of the structure

$$C = \begin{pmatrix} \mathbb{1} & A & \cdots \\ A & \mathbb{1} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}, \quad (19)$$

with matrices  $A$  describing the correlation assumptions, and  $\mathbb{1}$  being the identity matrix.

## 2.3 Technical implementation

The final minimisation of Eq. 1 is performed using the Minuit algorithms [12]. The total uncertainty on each combined value is determined by scanning  $\chi^2 = \chi_{\min}^2 + 1$  using the Minos algorithm. These algorithms as implemented in ROOT 6 as “TMinuit2” are employed.

The correlations that are assumed between systematic uncertainties can vary between -1 and 1. These extremes are special cases for which the correlation matrix  $C$  becomes non-invertible. In practice, a correlation of  $C_{ij} = \pm 1$  means that parameters  $i$  and  $j$  describe the same variation. In such cases, an entry  $C_{ij} = \pm 1$  is replaced by  $C_{ij} = \pm(1 - 10^{-3})$ . The difference to  $\pm 1$  is almost negligible.

For illustration,  $2 \times 2$  parameters are chosen with  $C_{ij} \approx \pm 1$ . The affected part of the  $\chi^2$ ,  $\chi_F^2$ , can be simplified to

$$\chi_F^2 = \frac{1}{1 - C_{ij}^2} (\lambda_i^2 + \lambda_j^2 \mp 2C_{ij}\lambda_i\lambda_j) \quad (20)$$

$$\approx \frac{1}{1 - C_{ij}^2} (\lambda_i \mp \lambda_j)^2 \quad (21)$$

and corresponds to  $(\lambda_i \mp \lambda_j)^2 \cdot 10^6$  for  $C_{ij} = \pm(1 - 10^{-3})$ . Given that a variation of  $\lambda = \pm 1$  corresponds to only a fraction of the total uncertainty on each estimate, the effect of the approximation  $C_{ij} = \pm(1 - 10^{-3})$  is negligible.

## 3 Validation

The validation is based on *pseudo-measurements*. Each pseudo-measurement is a binned likelihood fit with steerable central value and bin-wise uncertainties. This has the advantage, that the full likelihood of each pseudo-measurement is known and can be adjusted to different scenarios. Therefore, it is possible to compare the results obtained with the method proposed here to the ones obtained using the combined likelihood as reference. Since the latter in principle contains arbitrarily more parameters, small deviations are expected.

The validation is first performed with respect to the statistical bias, only. Secondly, the modelling of systematic uncertainties is tested with respect to correlations between the uncertainties of the pseudo-measurements, and the modelling of relative uncertainties.

For each pseudo-measurement a Poisson likelihood is chosen to determine the central result  $\bar{x}_\mu$ . The bin-wise uncertainties are randomly generated and modelled by the parameters  $\lambda_i$ . In the case that an uncertainty corresponds to an absolute variation, its effect on each bin is generated independently. For more than one bin ( $N_{\text{bins}} > 1$ ) this results in correlations between the uncertainties after the fit, as well as in constraints on their variations.

The likelihood for pseudo-measurement  $\alpha$  is defined as:

$$L^\alpha = \prod_\mu \prod_i^{N_{\text{bins}}^\alpha} \mathcal{P}(X_\mu^\alpha, \bar{X}_{\mu i}^\alpha) \cdot \prod_i \tilde{P}_i^\alpha(\lambda_i), \quad (22)$$

with  $\mathcal{P}$  being the Poisson likelihood and  $\tilde{P}_i^\alpha(\lambda_i)$  the Gaussian penalty terms modelling the prior knowledge of each uncertainty. The parameters  $X_\mu^\alpha$  and  $\bar{X}_{\mu i}^\alpha$  are given as:

$$X_\mu^\alpha = \frac{x_\mu^\alpha}{N_{\text{bins}}^\alpha} \text{ and} \quad (23)$$

$$\bar{X}_{\mu i}^\alpha = \frac{\bar{x}_\nu}{N_{\text{bins}}^\alpha} \prod_j \left( \frac{K_{\nu j}^\alpha \lambda_j}{x_\nu} + 1 \right) + \sum_j k_{\nu ij}^\alpha \lambda_j, \quad (24)$$

where  $K_{\nu j}^\alpha$  describes the magnitude of global relative variations and  $k_{\nu ij}^\alpha$  absolute shape variations, different for each bin  $i$ . The value of  $x_\mu^\alpha$  is the input to each pseudo-measurement and corresponds to the number of events that would be observed in a real measurement. The elements of the matrices  $K^\alpha$  and  $k_i^\alpha$  are chosen to describe different validation scenarios. Finally, the fit to determine  $\bar{x}_\mu^\alpha$  is performed and the resulting Hessian is recorded.

The combined likelihood for several pseudo-measurements is given by:

$$L_{\text{comb}} = \left( \prod_\alpha \frac{L^\alpha}{\prod_i \tilde{P}_i^\alpha} \right) \cdot \phi(\lambda_0, \dots, \lambda_N), \quad (25)$$

where  $\phi$  models the prior knowledge of  $N$  systematic uncertainties and the correlation assumptions between them, analogue to  $\chi_p^2$  in Eq. 17. For every validation step, the difference  $\Delta\bar{x}$  between the result obtained with the method proposed in this document and using the combined likelihood is recorded. The compatibility of both approaches is quantified by  $\Delta\bar{x}/\sigma_{\bar{x}}$ , which is the difference between their central results normalised to the total uncertainty of the combined-likelihood combination.

### 3.1 Statistical bias

To evaluate the statistical bias, the impact of systematic uncertainties on each pseudo-measurement is set to 0, corresponding to  $K = 0$  and  $k = 0$ . Only one quantity,  $\bar{x}$ , is determined from two estimates  $x^a$  and  $x^b$ , chosen as:

$$x^a = s \cdot 100, \quad (26)$$

$$x^b = x^a + \gamma\sqrt{x^a}, \quad (27)$$

with  $s$  being a scaling factor and  $\gamma$  describing the compatibility between the estimates. The latter is chosen to be either  $\gamma = 10$  to describe two very incompatible measurements or  $\gamma = 3$  for a more realistic scenario where both estimates still agree well enough to enter a combination.

Two pseudo-measurements are generated for each choice of  $s$  and  $\gamma$  and are combined either using a Pearson or Neyman  $\chi^2$  definition. For both choices, the uncertainties on the combined results agree very well with the ones obtained using the direct combination based on  $L_{\text{comb}}$ . The bias of the central value is shown in Figure 1 relative to the uncertainty of the combined value. It behaves as expected: it is smaller but of opposite sign for the Pearson  $\chi^2$  definition and is reduced with smaller statistical uncertainties and better compatibility between the results.

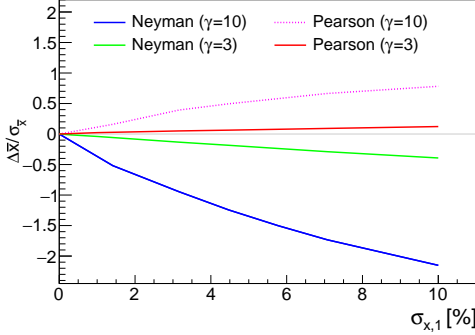


Figure 1: Difference between the combined values using a direct Poisson-likelihood combination and the method proposed here with Neyman and Pearson  $\chi^2$  definition relative to the total uncertainty. The estimates to be combined differ by about  $\gamma\sigma$  and are displayed as a function of the first estimate's relative statistical uncertainty. The second estimate's statistical uncertainty scales accordingly.

### 3.2 Systematic uncertainties

The effect of absolute systematic uncertainties is evaluated by combining two pseudo-measurements, with randomly chosen elements of the matrices  $k_\nu^\alpha$ . An upper threshold  $t$  is defined, such that for each element  $i, j$ :

$$|k_{\nu ij}^\alpha| \leq t \cdot X_\mu^\alpha, \quad (28)$$

limiting the contribution of systematic uncertainties. Two bins, two systematic uncertainties, and one  $x_\mu^\alpha$  per pseudo-measurement are considered. The sign of  $k_{\nu ij}^\alpha$  is chosen to be constant for each systematic uncertainty. The estimates  $x^a$  and  $x^b$  for measurement  $a$  and  $b$  are set to:

$$x^a = 30000 \text{ and} \quad (29)$$

$$x^b = 30600 \quad (30)$$

to reduce the effect of statistical uncertainties. The resulting statistical uncertainty of 0.6% does not account for the

difference of 2% between both values, such that the modelling of the systematic uncertainties will affect the combination significantly.

For large systematic variations the maximisation of Eq. 22 with Minuit can become numerically unstable. This is the case when the variation becomes as large as the nominal entry,  $X_\mu^\alpha$ , in at least one of the bins. Therefore, the Poisson likelihood is approximated with a Gaussian form, which is valid for low statistical uncertainties such as in this test. Thus,  $L^\alpha$  becomes:

$$L^\alpha = \prod_{\mu\nu} \prod_i^{N_{\text{bins}}^\alpha} \exp \left[ -S_{\mu\nu}^\alpha \frac{(\bar{X}_{\mu i}^\alpha - X_\mu^\alpha)(\bar{X}_{\nu i}^\alpha - X_\nu^\alpha)}{2(\bar{X}_{\mu i}^\alpha \bar{X}_{\nu i}^\alpha)^{1/2}} \right]. \quad (31)$$

The matrix  $S^\alpha$  allows modelling direct statistical correlations between  $\bar{X}_{\mu i}^\alpha$  and  $\bar{X}_{\nu i}^\alpha$ . Here,  $S$  is set to  $\mathbf{1}$ .

In total  $2 \times 20,000$  pseudo-measurements are generated, each with a different random choice of the uncertainties. The total relative uncertainty,  $\sigma_x/x$ , on the estimate of pseudo-measurement  $a$  is shown in Figure 2 for different values of the threshold  $t$ . Depending on  $t$ , the uncertainty varies from moderate values to more than 100%. The same applies to pseudo-measurement  $b$  (not displayed). The average constraints on the systematic uncertainties reach from about 90% ( $t = 0.01$ ) to 50% ( $t = 1.0$ ) with respect to their initial  $1\sigma$  variation.

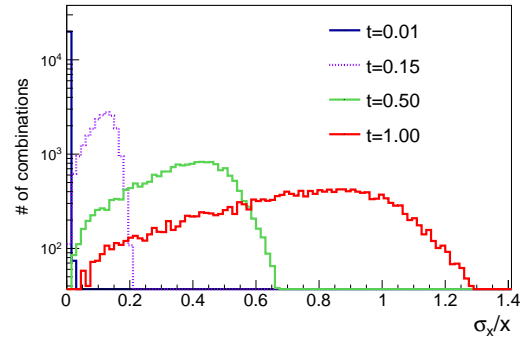


Figure 2: Relative total uncertainty of the pseudo-measurement  $a$  for different values of the threshold  $t$ .

In a first validation step, each uncertainty of one pseudo-measurement is assumed to be highly correlated with exactly one uncertainty of the other pseudo-measurement by assigning a correlation factor  $c = 0.99$ . A total of 20,000 combinations are performed. The ratio  $\Delta^r\sigma$  between the uncertainty on the combined value obtained with the method proposed here and by maximising  $L_{\text{comb}}$  is shown in Figure 3 as a function of  $t$ . Asymmetric uncertainties on the combined value are accounted for and are equally well described. With an increasing contribution of the systematic uncertainties, the  $\Delta^r\sigma$ -distribution becomes slightly

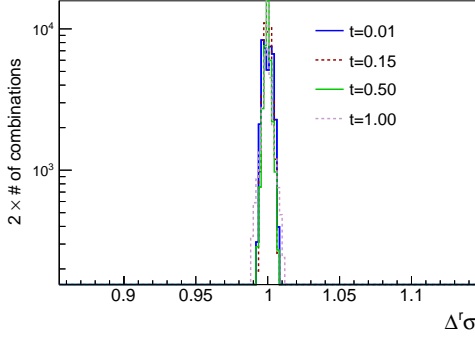


Figure 3: Ratio of the uncertainties on the combined value obtained using the method proposed here and the direct likelihood combination, shown for different values of the upper threshold for systematic uncertainties  $t$ .

broader, but does not indicate any numerically relevant mis-modelling. The resulting values for  $\Delta\bar{x}/\sigma_{\bar{x}}$  are illustrated in Figure 4. For  $t = 0.01$ , the statistical uncertainties are non negligible. This leads to a small bias towards lower values introduced by the choice of Eq. 31 and discussed in the Section 3.1. However, for all choices of  $t$  and all pseudo experiments, the differences between the direct likelihood approach and the method described here are well below 5% of the total uncertainty and can therefore be considered negligible.

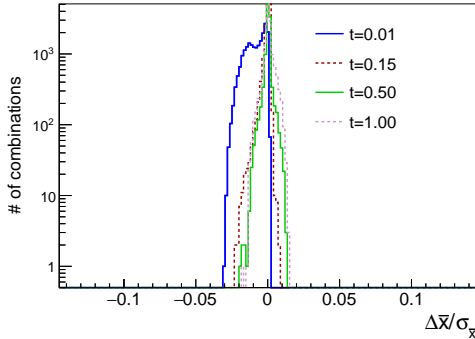


Figure 4: Difference between the combined result  $\bar{x}$  obtained with the method proposed here and using a direct likelihood combination relative to the total uncertainty on  $\bar{x}$ , shown for different values of the upper threshold for systematic uncertainties  $t$ .

Moreover, the dependence on the assumed correlation between the uncertainties of both pseudo-measurements is studied, as well as possible biases with respect to the number of bins in each pseudo-measurement. Figure 5 shows

the dependence of  $\Delta\bar{x}/\sigma_{\bar{x}}$  on the choice for the correlation coefficients  $c$  for  $t = 1$ . The modelling worsens slightly when  $|c|$  decreases, but is below about 3% with respect to the total uncertainty on the combined value for all 20,000 pseudo experiments. Also, the total uncertainty remains well mod-

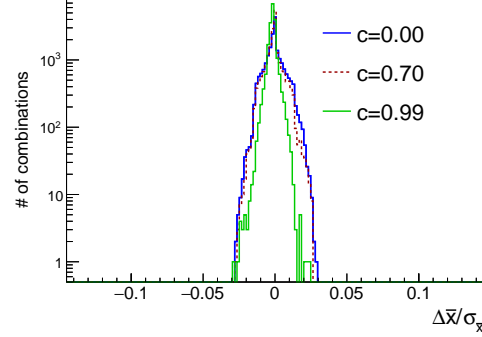


Figure 5: Difference between the combined results  $\bar{x}$  from the method proposed here and using a direct likelihood combination relative to the total uncertainty on  $\bar{x}$ . The distribution is shown for different values of the correlation  $c$  between the systematic uncertainties of both pseudo-measurements.

eled with only a very moderate increase of combinations with  $|\Delta^r\sigma|$  slightly different from 1, as shown in Figure 6. The same conclusion can be drawn when the procedure described here is repeated for a different number of bins in each pseudo-measurement (not shown here). All results for 2, 4, 20, and 100 bins show a good modelling of the combined likelihood approach with respect to the central values and the total uncertainties.

In general, the central result and its uncertainty are very well modelled for a large range of relative contributions from systematic uncertainties, correlations among them, and the chosen number of bins in each pseudo-measurement. Very small deviations of the order of a few per-cent with respect to the total uncertainty are observed. These are expected as a result of reducing the binned information of the initial measurement likelihood to one or many estimates and a corresponding Hessian. The method described here shows a similar stability for different choices of  $x^a$  and  $x^b$ , and of the number of uncertainties. Moreover, it is also valid for multiple estimates within one pseudo-measurement without statistical correlations between them. The case of statistical correlations is discussed separately in Section 3.3.

### 3.3 Modeling of statistical correlations

The correct modelling of statistical correlations between the estimates within a measurement is tested by generating two pseudo-measurements  $a$  and  $b$  similar to Section 3.2, each

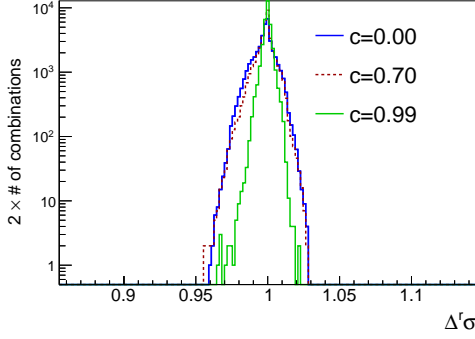


Figure 6: Ratio of the uncertainties on the combined value obtained using the method proposed here and the direct likelihood combination. The distribution is shown for different values of the correlation  $c$  between the systematic uncertainties of both pseudo-measurements.

with two estimates  $x_1^a$  and  $x_2^a$  or  $x_1^b$  and  $x_2^b$ , respectively. The corresponding correlation matrices  $S^a$  and  $S^b$  are randomly chosen to have off-diagonal elements with an absolute value of  $d \pm 0.1$ . In total, 10000 combinations are performed for each choice of  $d = \{0, 0.3, 0.9\}$ ,  $t = \{0.01, 0.50\}$ , and  $c = \{0.00, 0.99\}$ . The values for  $x_\mu^\alpha$  are chosen to be  $x_1^a = 30000$ ,  $x_1^b = 30600$ ,  $x_2^a = 20000$ , and  $x_2^b = 20500$ . The resulting values for  $\Delta\bar{x}_1/\sigma_{\bar{x},1}$  and  $\Delta^r\sigma_1$  are illustrated in Figure 7 and 8 for  $c = 0$ .

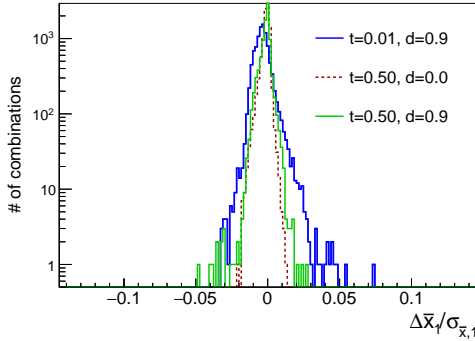


Figure 7: Difference between the combined value of  $\bar{x}_1$  obtained with the method described here and using a direct likelihood combination relative to the total uncertainty of  $\bar{x}_1$ . The distribution is shown for different values of the scale  $t$  for systematic uncertainties and the statistical correlation between the estimates  $d$ . The combination assumes no correlation between the uncertainties of both pseudo-measurements.

No significant mismodelling of the statistical correlation

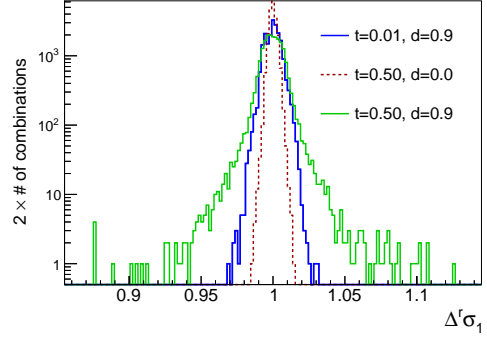


Figure 8: Ratio of the uncertainties on  $\bar{x}_1$  obtained with the method proposed in this document and using a direct likelihood combination. The distribution is shown for different values of the scale for systematic uncertainties  $t$  and the statistical correlation between the estimates  $d$ . The combination assumes no correlation between the uncertainties of both pseudo-measurements.

between estimates of the same measurement can be observed. The dependence on  $c$  is similar to the one discussed in Section 3.2 and not shown here. Also the result of the combination of  $x_2$  shows identical behaviour and is therefore not depicted either. Different choices for  $x_\mu^\alpha$  were tested and confirm a good modelling with respect to  $d$ .

### 3.4 Relative uncertainties

The modelling of relative uncertainties is studied by generating two pseudo-measurements, each of them with one parameter to be combined, one relative uncertainty, and two absolute uncertainties. The relative uncertainty applies to all bins in the same way and will therefore not receive constraints. In consequence, it will be dominant. Thus, the total uncertainty of each pseudo-measurement will differ from the dependence on  $t$  previously illustrated in Figure 4. A total of  $2 \times 5000$  pseudo-measurements are generated. Figure 9 shows the relative uncertainty of pseudo-measurement  $a$ , including one relative uncertainty, as a function of  $t$ . For  $t$  larger than 0.15, the direct likelihood combination shows instabilities in some cases, likely related to the Gaussian penalty terms, while log-normal terms would be more suitable for large relative uncertainties. As shown in Figures 10 and 11, also when combining pseudo-measurements with contributions from relative uncertainties, central values and uncertainties are well modelled, assuming the uncertainties of one pseudo-measurement to be uncorrelated with the uncertainties of the other. The same holds true for high correlations between the pseudo-measurements.

Additionally, the method is validated using exactly one

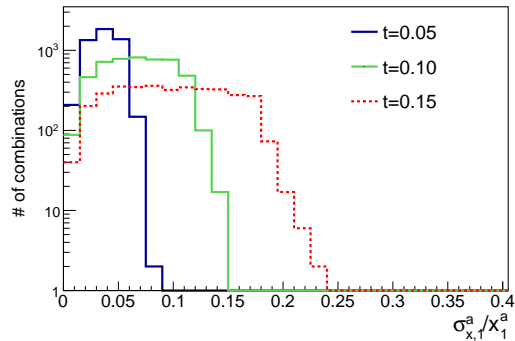


Figure 9: Relative total uncertainty of the pseudo-measurement  $a$  for different values of the threshold  $t$  for systematic uncertainties. All pseudo-measurements comprise one relative and two absolute uncertainties.

estimate per pseudo-measurement and one large relative uncertainty of +15%. The input estimates are set to  $30000 + \beta$ , where  $\beta$  is a randomly generated value between 0 and 750. The uncertainty is assumed to be fully correlated between the pseudo-measurements. This results in asymmetric uncertainties on each pseudo-measurement and the combined value. Moreover, for this particular choice of uncertainties, the combined value can be larger than the highest input estimate. When comparing the direct likelihood combination to the method proposed here, also in this case no bias with respect to the central value or its uncertainties can be observed.

In summary, the combination method described here does not require the original data and the full fit model, but sufficiently describes the initial measurement for a large variety of possible central values, binning choices and uncertainties. In consequence, the combination results are numerically equivalent to a using the full likelihood information, in particular in case of dominant systematic uncertainties.

## 4 Neglecting correlations

For comparison, two pseudo-measurements  $a$  and  $b$  with two bins, two systematic uncertainties, and the same parameters described in Section 3.2 are combined neglecting correlations between uncertainties within the same pseudo-measurement, but still considering strong correlations between pseudo-measurements  $a$  and  $b$ . This approximates the situation in which the BLUE method [7, 8, 9] can be used for the combination. The correlations within one pseudo-measurement are removed by inverting  $(D^\alpha + \mathbb{1})$  in Eq. 2, removing the off-diagonal elements of the resulting covariance matrix, and replacing  $D^\alpha$  by the inverse of this covariance matrix

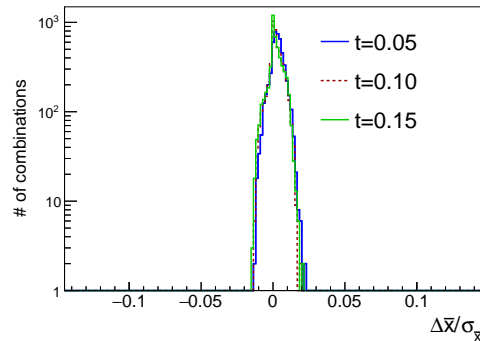


Figure 10: Difference between the combined value of  $\bar{x}$  obtained with the method proposed here and using a direct likelihood combination relative to the total uncertainty of  $\bar{x}$ , shown for different values of the threshold  $t$  for systematic uncertainties. All pseudo-measurements comprise one relative and two absolute uncertainties.

minus 1. By choosing the Neyman  $\chi^2$  definition in addition, this makes this test equivalent to the BLUE method. As shown in Figure 12, this approximation can lead to wrong individual combination results with respect to the central value when the contribution of systematic uncertainties becomes non-negligible. Also, the uncertainty on the combined value can be severely mismodelled if the correlations within one measurement are neglected, as displayed in Figure 13. The total uncertainty can be underestimated or strongly overestimated, in particular if it is dominated by systematic uncertainties. Therefore, it is crucial to model these correlations consistently when performing a combination of results obtained in simultaneous fits of systematic uncertainties and the quantity to be determined.

## 5 Program Installation and User Interface

The method described in Section 2 is implemented in the dedicated Convino program for the combination of experimental results. The source code can be found at <https://github.com/jkieseles/Convino/releases>. It can be compiled using `make` with gcc version 4.9 or newer, or clang 8.0.0 or newer (OSX) and ROOT 6 installed on the system. Other versions might be sufficient but are not tested.

The measurements and the configuration for the combination are contained in human-readable text files. Alternatively, a C++ library is provided with the software package, providing an interface to C++ standard-library or ROOT



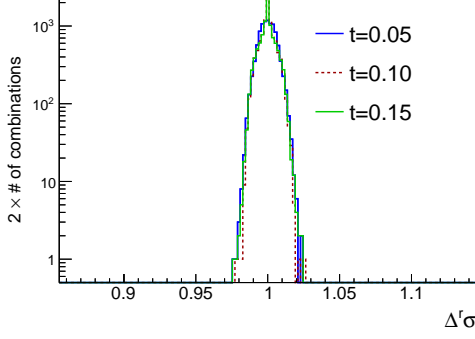


Figure 11: Ratio of the uncertainties on  $\bar{x}$  obtained with the method described in this document and using a direct likelihood combination, shown for different values of the threshold  $t$  for systematic uncertainties. All pseudo-measurements comprise one relative and two absolute uncertainties.

classes, the latter commonly used in high-energy physics. Both interfaces are described in the following, starting with the text-based interface. The discussion of the text-based interface serves as reference for the description of the C++ interface.

## 5.1 Text-based Interface

The “convino” executable can be found in the base directory after compiling. It prints usage information and a list of options if the `-h` option is specified. Other options are:

- `-s` perform correlation scan
- `-p` save scan plots as .pdf in addition to a .root file
- `-d` switch on debug printout
- `--neyman` uses a Neyman  $\chi^2$  instead of the Pearson  $\chi^2$
- `--prefix` defines a prefix for all output files and directories

In addition to the options, a text file is passed to the executable. It is referred to as *base file* in the following and is described in Section 5.1.2. Each measurement comprising one or a set of estimates is described in a *measurement file*. Well documented examples for both types of files are provided in the `examples` directory and should be consulted alongside this manual.

### 5.1.1 Measurement File

Each measurement file consists of blocks. Each block describes estimates or uncertainties. They are defined by a Hessian, a correlation matrix together with constraints, or a set of orthogonal uncertainties. The latter should be provided in the following format:

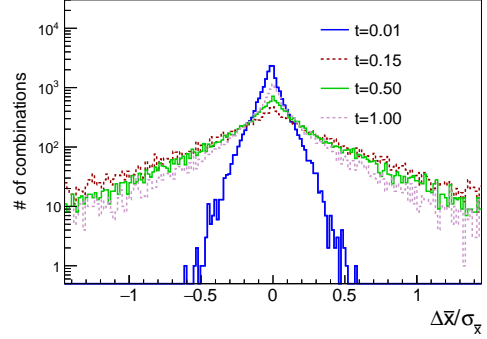


Figure 12: Difference between the combined results neglecting correlations between uncertainties within a measurement and using a direct likelihood combination relative to the total uncertainty on  $\bar{x}$ . The distribution is shown for different values of the upper threshold for systematic uncertainties  $t$ . For comparison with the method proposed here, see Figs. 3 and 4.

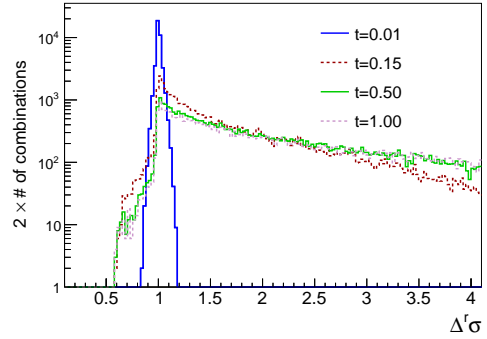


Figure 13: Ratio of the uncertainties on the combined value obtained neglecting correlations between uncertainties within a measurement and using a direct likelihood combination. The distribution is shown for different values of the upper threshold for systematic uncertainties  $t$ . For comparison with the method proposed here, see Figs. 3 and 4.

```
[not fitted]
      sys_a1 sys_b1 sys_c1 stat
estimate_a1 (+5-3) 6.1 2 6
estimate_b1 (-3+2) 1 4 2
[end not fitted]
```

The uncertainties `sys_XX` on the estimates `estimate_XX` are given in absolute values. The keyword `stat` is reserved for the statistical uncertainty. Asymmetric uncertainties are given as a pair surrounded by parentheses. The uncertainties and their effect on the estimates in a measurement using a simultaneous nuisance parameter fit technique are described either by a Hessian or a correlation matrix. The Hessian must be written in the following form:

```
[hessian]
      sys_a3      1944.6
      sys_b3     -1349. 1154.4
estimate_a3 -0.525 0.5398 3.25e-4
estimate_b3 -0.708 0.2706 0 4.89e-4
[end hessian]
```

while the correlation matrix has to include additional information about the constraints on the parameters. These constraints are given in units of  $1\sigma$  variations for the systematic uncertainties, such that a value of 1 corresponds to no reduction and lower values indicate a constraint from the fit to the data. For estimates, the constraints are given in absolute units and correspond to the total uncertainties. In both cases, they are defined in parentheses, such that the correlation matrix is of the format:

```
[correlation matrix]
      sys_a2      (1)      1
      sys_b2      (1)     -0.2      1
estimate_a2 (10.6) 0.547945 0.1059 1
estimate_b2 (12.8) 0.147945 0.4305 0 1
[end correlation matrix]
```

If uncertainties have been described in form of a Hessian or correlation matrix, additional contributions from orthogonal uncertainties can be provided in the `[not fitted]` block. These uncertainties must not have any correlation with the uncertainties defined in the Hessian or the correlation matrix. The next block of the measurement file describes the type of each uncertainty.

```
[systematics]
      sys_a2 = absolute
      sys_b2 = relative
[end systematics]
```

The type can be either `absolute` or `relative`. The default is `absolute` and does not need to be specified explicitly. The last block defines which of the parameters are estimates, and their nominal values:

```
[estimates]
      n_estimates = 2
      name_0      = estimate_a2
      value_0     = 780
      name_1      = estimate_b2
      value_1     = 280
[end measurements]
```

Here, `n_estimates` gives the number of estimates.

### 5.1.2 Base File

The first block of the base file defines the number of measurement files (`nFiles`) to be considered for the combination and the corresponding file names. An example is given below:

```
[input]
      nFiles = 2
      file0  = exampleMeasurement1.txt
      file1  = exampleMeasurement2.txt
[end input]
```

The files must be in the same directory as the base file. The second block defines the observables, the estimates should be combined to:

```
[observables]
      combined_a = estimate_a1 + estimate_a2
      combined_b = estimate_b1 + estimate_b2
[end observables]
```

Here, `estimate_a1` and `estimate_a2` should be combined to `combined_a`, and similarly for `estimate_b1` and `estimate_b2`. The number of estimates that should be combined to a single quantity is not limited, as well as the number of combined values. This makes it possible to combine simultaneously e.g. a large amount of bins from differential cross sections from various channels and experiments. However, in this case, the C++ interface is probably more practical.

The last block describes the correlations that should be assigned using the following syntax.

```
[correlations]
      sys_b1 = (0.2) sys_c2
      sys_c1 = (-0.3) sys_d2
[end correlations]
```

Here, a correlation coefficient of 0.2 is assigned between `sys_b1` and `sys_c2` and -0.3 between `sys_c1` and `sys_d2`. The correlation assumptions between the parameters can be scanned in an automated way. In this case, the following syntax is used to define the scan ranges:

```
[correlations]
      sys_b1 = (0.2 & -0.1 : 0.4) sys_c2
[end correlations]
```

Here, `sys.b1` has a nominal correlation of 0.2 to `sys.c2`. The correlation is scanned from -0.1 to 0.4. If several correlation coefficients should be scanned simultaneously, they have to be specified in a single line:

```
[correlations]
sys_b1=(0.2&-0.1:0.4)sys_c2+(-0.3&0.2:-0.3)sys_d2
[end correlations]
```

In this case, the scan range for a single coefficient can start from positive values to negative values to allow accounting for anti-correlations between the parameters that are scanned simultaneously.

Correlation matrices are positive definite by definition, a correlation matrix  $C$  with large off-diagonal entries might lose this property if ill-posed assumptions are made, such as:

$$C = \begin{pmatrix} 1 & .99 & 0 \\ .99 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix}. \quad (32)$$

In this case, the program exits and it is strongly advised to revise the plausibility of the correlation assumptions.

The results of the combination are saved in the output file `result.txt`, or `<prefix>_result.txt` in case a prefix is specified. The output file contains the original input correlations, the combination results, the minimum  $\chi^2$ , and pulls and constraints on all parameters. The output of the scan, including all correlation matrices, is saved in the file `scan_result.txt`. The corresponding figures are saved as `TGraphAsymmErrors` classes in the file `scanPlots.root`. If pdf-file output was enabled, the resulting Figures can be found in the directory `scan_results`. Examples of such Figures obtained with the example configuration are shown in Figure 14 and 15.

## 5.2 C++ Interface

The C++ interface is optimized for the combination of differential distributions and provides three basic classes which will be described in the following: the class `measurement`, which is analogous to a measurement file discussed in the previous Section, the class `combiner` to perform the combination, and a class `combinationResult` that collects the output of the combination. The `measurement` class and the `combinationResult` class provide interfaces to C++ standard library `std::vector<double>` or alternatively to ROOT histograms and graphs. An example of the usage is provided in `bin/differentialExample.cpp`. Any cpp file that will be placed in the `bin` directory will be compiled automatically when running `make`. Alternatively, the compilation of the Convino package will create the library `libconvino.so` that can be linked against. The header files can be found in the `include` directory.

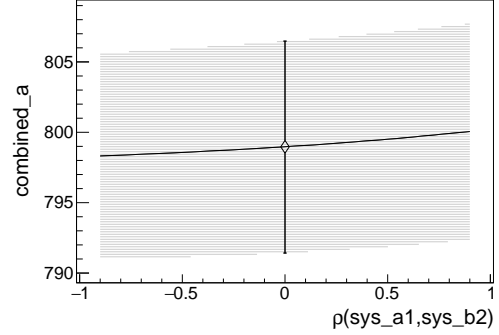


Figure 14: Combined value for `combined_a` for a scan of the correlation coefficient for `sys_a1` and `sys_b2`. The open marker shows the result obtained with the nominal assumption and its uncertainty. The shaded area represents the uncertainty associated to the scanned dependence, indicated by a continuous line. All values are obtained with the example configuration.

Each class is documented in the corresponding header file. Therefore, the documentation here is limited to the general usage.

### 5.2.1 Measurement class

The measurement class provides the possibility to define a set of estimates, their statistical correlations and systematic uncertainties. Each object can only contain one set of estimates at once. In case the information is read from a ROOT TH1 histogram, each measurement class object can contain only one nominal histogram.

For a measurement with orthogonal uncertainties, the following procedure should be applied: the nominal values are set using the function `setMeasured`. Systematic uncertainties can be added in a second step to the measurement object with `addSystematics`. The type of each uncertainty is defined using the function `setParameterType` after all uncertainties have been added. Here, it is recommended to use the parameter name to identify the correct uncertainty. In a last step, statistical correlations between the estimates can be set using the method `setEstimateCorrelation`.

If a measurement comprises correlated uncertainties, the corresponding measurement object should be configured using the function `setHessian`, which defines the uncertainties and estimates at once. Additional orthogonal uncertainties can be added in a subsequent step using `addSystematics`.

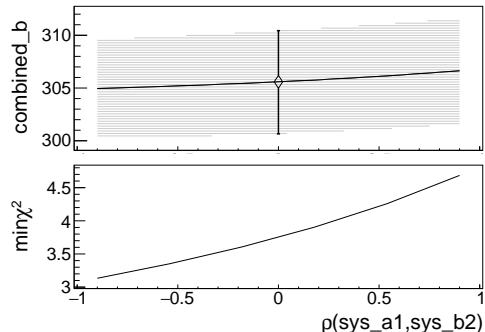


Figure 15: Combined values for `combined_b` (upper panel), and minimum  $\chi^2$  (lower panel) for a scan of the correlation coefficient for `sys_a1` and `sys_b2`. The open marker shows the result obtained with the nominal assumption and its uncertainty. The shaded area represents the uncertainty associated to the scanned dependence, indicated by a continuous line. All values are obtained with the example configuration.

### 5.2.2 Combiner class

Once the individual `measurement` objects are defined, they are added to a `combiner` object using the function `addMeasurement`. For the following combination, it is assumed that the entries of each measurement in the same bin or with the same vector index should be combined. It is not possible to combine a number of estimates from one measurement object with a different number of estimates from another. The correlation assumptions are defined with `setSystCorrelation`. It is advised to use the uncertainties names as input for unambiguous identification.

The combination is initiated by calling the method `combine`, which returns a `combinationResult` class object.

### 5.2.3 CombinationResult class

The `combinationResult` class is a container for all information regarding the inputs to the combination, the correlation matrices, and the combined values as well as the post-combination correlation matrices, pulls and constraints. If differential distributions are combined, the result can be fed back to a ROOT TH1 object or a `TGraphAsymmErrors` using the functions `fillTH1` or `fillTGraphAsymmErrors`.

## 6 Summary

The combination method presented in this document allows combining measurements obtained with simultaneous nuisance parameter fits consistently, taking into account the constraints from the data as well as correlations between

systematic uncertainties within each measurement. In contrast to the optimal case of a direct likelihood combination, based on the product of the individual likelihoods of each measurement, the method does not require the full knowledge of the original data and the fit models. This information would also be required by other commonly used combination methods, however, it is publicly available only in rare cases. It is shown that not accounting for correlations between uncertainties within the same measurement can lead to non-negligible deviations from the combined likelihood approach with respect to the combined value and its uncertainty.

The method described here does not introduce such deviations and relies on the central results and their covariances or Hessians, only, which makes it applicable to a significantly larger variety of combinations. An extensive validation is performed using pseudo-measurement with varying contributions of statistical and systematic uncertainties, correlation assumptions, binning choices, and pri-or statistical correlations. All obtained results and uncertainties are numerically equivalent to a direct likelihood combination. Only for measurements strongly limited by statistical precision, the same known caveats as in other  $\chi^2$  or least-squares-based approaches (e.g. the BLUE method) apply. In addition, the Convino program is presented. It is developed to perform combinations using the method described here and provides a text-based and a C++ user interface. The text-based user interface provides an automatic scan of correlation assumptions and creates the corresponding figures for graphical representation.

## Acknowledgements

I would like to thank O. Behnke for fruitful discussions and comments alongside M. Aldaya Martin, C. Diez Pardos, M. Mulders, and A. Giammanco for editorial suggestions.

## References

- [1] J. Kieseler, “A method and tool for combining differential or inclusive measurements obtained with simultaneously constrained uncertainties,” *EPJC*, vol. 77, no. 11, p. 792, 2017.
- [2] CMS Collaboration, “Measurement of the  $t\bar{t}$  production cross section in the  $e\mu$  channel in proton-proton collisions at  $\sqrt{s} = 7$  and 8 TeV,” *JHEP*, vol. 08, no. 029, 2016.
- [3] CMS Collaboration, “Measurement of the  $t\bar{t}$  production cross section using events with one lepton and at least one jet in pp collisions at  $\sqrt{s}=13$  TeV,” *Submitted to JHEP*, 2017.

- [4] CMS Collaboration, “Measurements of the  $t\bar{t}$  production cross section in lepton+jets final states in pp collisions at 8 TeV and ratio of 8 to 7 TeV cross sections,” *EPJC*, vol. 77, no. 1, 2017.
- [5] J. Kieseler, K. Lipka, and S.-O. Moch, “Calibration of the Top-Quark Monte Carlo Mass,” *PRL*, vol. 116, no. 16, 2016.
- [6] ATLAS Collaboration, “Simultaneous measurements of the  $t\bar{t}$ ,  $W^+W^-$ , and  $Z/\gamma^* \rightarrow \tau\tau$  production cross-sections in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector,” *PRD*, vol. 91, no. 5, 2015.
- [7] L. Lyons, D. Gibaut, and P. Clifford, “How to combine correlated estimates of a single physical quantity,” *NIM A*, vol. 270, no. 1, pp. 110 – 117, 1988.
- [8] R. Nisius, “BLUE: a software package to combine correlated estimates of physics observables within ROOT using the Best Linear Unbiased Estimate method - Program manual, Version 2.1.0,” <http://blue.hepforge.org>, 2016.
- [9] R. Nisius, “On the combination of correlated estimates of a physics observable,” *EPJC*, vol. 74, no. 8, 2014.
- [10] S. Alekhin *et al.*, “HERAFitter,” *EPJC*, vol. 75, no. 7, 2015.
- [11] R. Brun and F. Rademakers, “ROOT: An object oriented data analysis framework,” *NIM A*, vol. 389, 1997.
- [12] F. James and M. Roos, “Minuit: a system for function minimization and analysis of the parameter errors and correlations,” *Comput. Phys. Commun.*, vol. 10, 1975.