A method and program for the combination of measurements obtained with simultaneous nuisance parameter fits.

Jan Kieseler, jan.kieseler@cern.ch

CERN

February 7, 2017

Abstract

A method is proposed that allows to combine measurements obtained with simultaneous nuisance parameter fits. These are performed by fitting a model as a function of the parameters of interest and its uncertainties to the measured data. As a result of such a measurement, the uncertainties are correlated among each other and receive constraints from the data as well as from prior assumptions. The best approach for a combination of these measurement would be the maximization of a combined likelihood. To define this likelihood, the full fit model used for each measurement and the original data is required. Only in rare cases, this information is publicly available. The method described here can be used to model a likelihood equivalent to the combined likelihood based on the public result and its covariance or Hessian, only. The method is incorporated in the Convino program that provides a text-based user interface alongside a C++ interface. The latter also integrates ROOT histogram and graph classes for simple combination of binned measurements such as differential cross sections. The program is not limited to measurements obtained with simultaneous nuisance parameter fits.

Contents

1 Introduction

A common technique to reduce limitations in precision measurements due to systematic uncertainties is to constrain the range of their variations by the properties of the data. A simultaneous fit of these variations and the parameters of interest can be performed based on prior knowledge of the uncertainties and suitable distributions. This technique can reduce the total uncertainty in many cases significantly, as in Refs. [1, 2, 3], and can be used to measure several parameters simultaneously (see e.g. Refs. [4, 5]). However, it leads to non-negligible correlations between all fitted parameters. These are particularly important for a combination of at least one of such measurements with other measurements.

The most consistent way for this combination is to define a combined likelihood based on the original models, including all systematic variations, and the original data the models were fit to. However, the data and the models are publicly available only in very rare cases. The method described here provides a solution to this problem. It can be used to create a likelihood, which is equivalent to the combined likelihood, based on the central results and their covariance or Hessians, only. It allows to separate constraints 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. This cannot be modeled by other frequently used combination programs [6, 7, 8].

The method presented here is incorporated in the "Convino" combination program, which is also described in this note. It 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. points of a differential distribution) can be combined simultaneously accounting for all correlations. In addition to a text-based interface, a C++ interface is provided to define the input to the combination. This interface can either read basic data types, or the program can be configured using ROOT [9] histogram and graph classes, which are commonly used in high-energy-physics analyses.

The method is based on a χ^2 minimization and is described in Section 2. Technical details of the implementation are given in Section 3. A thorough validation of the method with respect to using a combined likelihood is discussed in Section 4. The installation of the Convino program is described in Section 5 and the user interface of the program is presented in Section 6. Section 7 concludes this document.

2 Mathematical framework

The combination is performed using a χ^2 minimization. The corresponding $\chi^2$is defined as:

$$\chi^2 = \sum_\alpha \left(\chi^2_{s,\alpha} + \chi^2_{u,\alpha}\right) + \chi^2_p. \tag{1}$$

It is composed of three terms: a term $\chi^2_{s,\alpha}$ represents the results of each measurement α. It follows a Neyman or Pearson χ^2 definition. A measurement can aim to determine a single quantity, e.g. the mass of a particle, or of a set of quantities, e.g. bins of a differential distribution. In both cases, these quantities are referred to as estimates in the following. The additional term $\chi^2_{u,\alpha}$ describes the correlations between the systematic uncertainties and constraints on them from the data for each measurement α. The last term, χ^2_p incorporates prior knowledge of the systematic uncertainties and correlations between uncertainties of the measurements to be combined. The parametrisation of the terms are discussed in the following - first for a measurement obtained through a simultaneous fit followed by a description for a measurement with orthogonal uncertainties.

2.1 Measurements obtained with simultaneous fits

For each measurement α, its Hessian evaluated at the best-fit values, H^α_{in}, is input to the method. It does not need to include uncertainties that were externalized from the initial fit. The treatment of these uncertainties will be discussed at the end of this Subsection. The Hessian can be split into the following components:

$$H^\alpha_{\text{in}} = \begin{pmatrix} \tilde{C} & \kappa^T \\ \kappa & M \end{pmatrix}^\alpha, \tag{2}$$

where M describes the relation between the estimates x^α_μ and x^α_ν, with μ and ν being indices for estimates. The relation between systematic variations and the estimates is described by κ. The matrix $\tilde{C}$ quantifies the relation between the systematic variations for which indices i and j will be used.

The central assumption of the method presented here is that an individual measurement α can be described by the following χ^2:

$$\chi^2_\alpha = \sum_{\mu\nu} \tilde{M}^\alpha_{\mu\nu} \frac{\xi^\alpha_\mu \xi^\alpha_\nu}{\tau^\alpha_\mu \tau^\alpha_\nu} + \sum_{ij} \lambda_i D^\alpha_{ij} \lambda_j + \sum_i P^2_i(\lambda_i), \text{ with} \tag{3}$$

$$\xi^\alpha_\mu = x^\alpha_\mu - \left(\bar{x}_\mu \prod_i (\lambda_i K^\alpha_{\mu i}/x^\alpha_\mu + 1) + \sum_i \lambda_i k^\alpha_{\mu i} \right). \tag{4}$$

Here, x^α_μ is the the estimate μ obtained in measurement α and $\bar{x}_\mu$ the combined value to be determined. The relation between both is given by $\tau^\alpha_\mu = \bar{x}_\mu / x^\alpha_\mu$ for the Pearson χ^2 definition. In case of the Neyman χ^2, all $\tau^\alpha_\mu = 1$. The systematic variations are modeled by continuous parameters λ_i and their effect on the estimates by $k^\alpha_{\mu i}$ or $K^\alpha_{\mu i}$ for absolute uncertainties and relative uncertainties, respectively. The matrix D^α describes the correlations between the uncertainties and the constraints that stem from the fit to the data, while $P_i(\lambda_i)$ represents terms to implement constraints on each λ_i from the prior knowledge of the uncertainties, uncorrelated with each other.

All parameters of χ^2_α are determined by calculating the Hessian of χ^2_α, $\tilde{H}^\alpha(0)$, analytically and comparing to the input H^α_{in}. The components are calculated as follows:

$$\tilde{H}^\alpha_{\mu\nu}(0) = \frac{1}{2} \left(\frac{\partial^2}{\partial \Delta x^\alpha_\mu \partial \Delta x^\alpha_\nu} \chi^2_\alpha \right) \Bigg|_{\lambda_i = 0,\ \Delta x^\alpha_\mu = 0\ \forall\ i,\ \mu} = \tilde{M}_{\mu\nu}, \tag{5}$$

$$\tilde{H}^\alpha_{\mu i}(0) = \frac{1}{2} \left(\frac{\partial^2}{\partial \Delta x^\alpha_\mu \partial \lambda_i} \chi^2_\alpha \right) \Bigg|_{\lambda_i = 0,\ \Delta x^\alpha_\mu = 0\ \forall\ i,\ \mu} = \sum_\nu \tilde{M}_{\mu\nu}(-\tilde{k}^\alpha_{\nu i}), \tag{6}$$

$$\tag{7}$$

with $\tilde{k}^\alpha_{\nu i} = K^\alpha_{\nu i} + k^\alpha_{\nu i}$ and $\Delta x^\alpha_\mu = x^\alpha_\mu - \bar{x}_\mu$. The matrix $\tilde{M}$ can be directly identified with M. Since M stems from a measurement of a physics quantity, M is positive definite and therefore invertible. Thus, the parameters $\tilde{k}^\alpha_{\nu i}$ can be determined as:

$$\tilde{k}^\alpha_{\nu i} = -\sum_\mu ((M^\alpha)^{-1})_{\mu\nu}\ \kappa^\alpha_{\mu i}. \tag{8}$$

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

$$\tilde{H}^\alpha_{ij}(0) = \frac{1}{2} \left(\frac{\partial^2}{\partial \lambda_i \partial \lambda_j} \chi^2_\alpha \right) \Bigg|_{\lambda_i = 0,\ \Delta x^\alpha_\mu = 0\ \forall\ i,\ \mu} = D^\alpha_{ij} + \delta_{ij} \frac{1}{2} \frac{\partial^2}{\partial \lambda_i^2} P^2_i \Bigg|_{\lambda_i = 0} + \sum_{\mu\nu} M^\alpha_{\mu\nu} \tilde{k}^\alpha_{\nu i} \tilde{k}^\alpha_{\mu j}. \tag{9}$$

Only Gaussian penalty terms describing the prior knowledge of the uncertainties are considered ($P_i(\lambda_i) = \lambda_i$), such that:

$$\frac{1}{2} \frac{\partial^2}{\partial \lambda_i^2} P^2_i \Bigg|_{\lambda_i = 0} = 1. \tag{10}$$

In consequence D^α becomes:

$$D^\alpha_{ij} = \tilde{C}^\alpha_{ij} - \delta_{ij} - \sum_{\mu\nu} M^\alpha_{\mu\nu} \tilde{k}^\alpha_{\nu i} \tilde{k}^\alpha_{\mu j}. \tag{11}$$

Finally, for each measurement the term $\chi^2_{s,\alpha}$ and $\chi^2_{u,\alpha}$ in Eq. 1 can be identified as:

$$\chi^2_{s,\alpha} = \sum_{\mu\nu} M^\alpha_{\mu\nu} \frac{\xi^\alpha_\mu \xi^\alpha_\nu}{\tau^\alpha_\mu \tau^\alpha_\nu} \text{ and} \tag{12}$$

$$\chi^2_{u,\alpha} = \sum_{ij} \lambda_i D^\alpha_{ij} \lambda_j. \tag{13}$$

$$\tag{14}$$

Uncertainties that were externalized from the simultaneous fit do not necessarily have to be accounted for by the Hessian. They can be incorporated through additional parameters λ_i, and corresponding $\tilde{k}^{\alpha}_{\mu i}$. These parameters have no contribution to D^{α}. The calculation of $\tilde{k}^{\alpha}_{\mu i}$ follows the procedure described in the next Subsection.

2.2 Measurements with orthogonal uncertainties

For orthogonal uncertainties, the same likelihood described in Eq. 3 is used. However, the calculation of its parameters does not require the full Hessian. Instead, the parameters can be calculated as:

$$\tilde{k}^{\alpha}_{\mu i} = \frac{\sigma^{\alpha}_{\mu i}}{\sigma^{\alpha}_{\mu\ \mathrm{total}}}, \tag{15}$$

with $\sigma^{\alpha}_{\mu i}$ being the contribution of uncertainty i to the total uncertainty $\sigma^{\alpha}_{\mu\ \mathrm{total}}$ of estimate x^{α}_{μ}. The matrix D^{α} is 0, the terms of M are calculated as:

$$M_{\mu\nu} = \frac{\rho_{\mu\nu}}{\sigma_{\mu}\sigma_{\nu}}. \tag{16}$$

Here, $\rho_{\mu\nu}$ is the statistical correlation between estimate μ and ν, and σ_{μ} and σ_{ν} the corresponding statistical uncertainties.

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

$$\chi^2_p = \sum_{ij} P_i(\lambda_i)(C^{-1})_{ij}P_i(\lambda_j), \tag{17}$$

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

$$\chi^2_p(\text{no corr}) = \sum_{i} P^2_i(\lambda_i). \tag{18}$$

For a combination, C will be of the structure

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

with matrices A describing the correlation assumptions.

3 Technical implementation

The final minimisation of χ^2 is performed using the Minuit [10] algorithms, for the scan of $\chi^2(\vec{\bar{x}}, \vec{\lambda}) = \chi^2_{\mathrm{min}}+1$, the implemented Minos algorithm is used. These algorithms implemented in ROOT6 as "TMinuit2" are employed for the program.

The correlations that are assumed between systematic uncertainties can vary between -1 and 1. The latter extremes are special cases, where 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. However, they are often defined as different parameters on input level and the correlation coefficient is assigned afterwards based on certain assumptions. In such cases, an entry $C_{ij} = \pm 1$ is replaced by $C_{ij} = \pm(1 - 10^{-3})$. The difference to ± 1 is almost negligible, such that any significant deviation would result in a significant increase of χ^2. For 2×2 parameters and $C_{ij} \approx \pm 1$, the corresponding part of χ^2, χ^2_F, can be simplified to

$$\chi^2_F = \frac{1}{1 - C^2_{ij}}(\lambda^2_i + \lambda^2_j \mp 2C_{ij}\lambda_i\lambda_j) \approx \frac{1}{1 - C^2_{ij}}(\lambda_i \mp \lambda_j)^2 \tag{20}$$

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.

4 Validation

The validation is based on pseudo measurements. In contrast to the majority of public results, 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 a combined likelihood, which uses the full information from each pseudo measurement. The validation is performed with respect to the statistical bias, correlations between the uncertainties of the pseudo measurements, and the modeling of relative uncertainties.

Each pseudo measurement is generated using a simultaneous binned Poisson-likelihood fit of the quantities to be determined ($\bar{x}_\mu$) and randomly generated uncertainties with variations modeled by parameters λ_i. In case 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 pseudo measurement α is performed using the likelihood

$$L^\alpha = \prod_\mu \prod_i^{N^\alpha_{\text{bins}}} \mathcal{P}\left(X^\alpha_\mu, \bar{X}^\alpha_{\mu i}\right) \cdot \prod_i \tilde{P}^\alpha_i(\lambda_i), \tag{21}$$

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

$$X^\alpha_\mu = \frac{x^\alpha_\mu}{N^\alpha_{\text{bins}}} \text{ and} \tag{22}$$

$$\bar{X}^\alpha_{\mu i} = \frac{\bar{x}_\nu}{N^\alpha_{\text{bins}}} \prod_j \left(\frac{K^\alpha_{\nu j} \lambda_j{}^\alpha}{x_\nu} + 1 \right) + \sum_j k^\alpha_{\nu i j} \lambda_j, \tag{23}$$

$$\tag{24}$$

with $K^\alpha_{\nu j}$ describing the magnitude of global relative variations and $k^\alpha_{\nu i j}$ absolute shape variations, different for ich bin i. The value of x^α_μ is defined for each pseudo measurement and would correspond to the number of events observed in data in a real measurement. The matrices K^α and k^α_i are generated, with their contribution being adjustable, and the fit to determine $\bar{x}^\alpha_\mu$ is performed.

The combined likelihood for several pseudo measurements in given by:

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

where ϕ models prior knowledge of N systematic uncertainties and the correlation assumptions between them, analogue to χ^2_p 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. This difference is normalised to the uncertainty on the combined value ($\Delta\bar{x}/\sigma_{\bar{x}}$) to quantify the compatibility of both approaches.

4.1 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 to be determined from the input estimates x^a and x^b, given as:

$$x^a = s \cdot 100, \tag{26}$$

$$x^b = x^a + 10\sqrt{x^a}, \tag{27}$$

with s being a scaling factor. The compatibility between x^a and x^b is approximately constant for different values of s. Two pseudo measurements are generated for each choice of s and combined with Convino either using a Pearson or Neyman χ^2 definition. For both choices, the uncertainties on the combined results agree very well with the one obtained using the direct combination based on L_{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 a factor 2 smaller but of opposite sign for the Pearson χ^2 definition and is reduced with smaller statistical uncertainties.

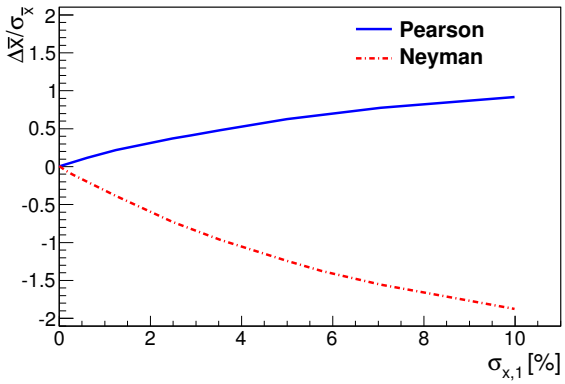

Figure 1: Difference between the combined values using a direct Poisson-likelihood combination and the method proposed here with Neyman and Pearson χ^2 definition relative to the total uncertainty. The estimates to be combined differ by about 10σ and are displayed as a function of the first estimate's statistical uncertainty. The second estimate's statistical uncertainty scales accordingly.

4.2 Systematic uncertainties

The effect of absolute systematic uncertainties is evaluated by combining two pseudo measurements. The matrices k^α_ν are randomly filled with entries. An upper threshold t is chosen, such that for each entry $i,\ j$:

$$|k^\alpha_{\nu ij}| \leq t \cdot X^\alpha_\mu. \tag{28}$$

Two bins, two systematic uncertainties and one x^α_μ per pseudo measurement are considered. The estimates x^a and x^b for measurement a and b are set to:

$$x^a = 30000 \text{ and} \tag{29}$$
$$x^b = 30600 \tag{30}$$

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 modeling of the systematic uncertainties affects the combination significantly.

For large systematic variations, the maximization with Minuit of Eq. 21 can become numerically unstable. This is the case, when the variation becomes as large as the entry X^α_μ 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^α becomes:

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

The matrix S^α allows to model direct statistical correlations between $\bar{X}^\alpha_{\mu i}$ and $\bar{X}^\alpha_{\nu i}$. Here, S is set to $\mathbb{1}$.

In total $2 \times 20,000$ pseudo measurements are generated, each with a different random choice for the uncertainties. The total relative uncertainty, σ_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 reaches from moderate values to being dominant. The same applies to pseudo measurement b, which is not displayed.

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 $c = 0.99$. In total 20,000 pseudo experiments are performed, in each combining a different instance of pseudo measurements a and b. The relative difference $\Delta^r\sigma$ between the uncertainty on the combined value obtained with the method proposed here and by maximizing L_{comb} is also shown in Figure 2 as a function of t. The total uncertainty on the combined value can be asymmetric. Both, the lower and upper uncertainty agree well when comparing both approaches. With increasing contribution of the systematic uncertainties, the difference $\Delta^r\sigma$ becomes slightly broader, but does not show any significant bias.

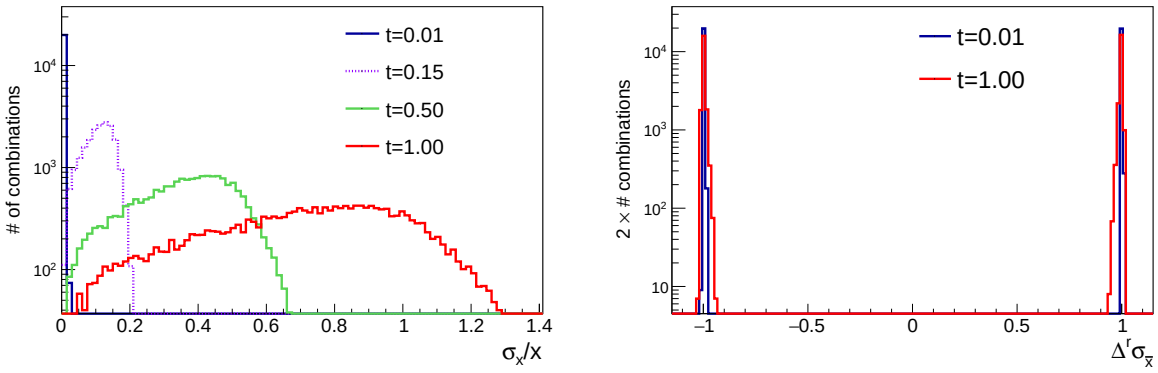

Figure 2: Left: relative total uncertainty of the pseudo measurement a for different values of the threshold t. Right: ratio of the uncertainties on the combined value, $\bar{x}$, obtained using the method proposed here and a direct likelihood combination. The ratio of the lower uncertainties is multiplied with -1.

The resulting values for the compatibility, $\Delta\bar{x}/\sigma_{\bar{x}}$, are illustrated in Figure 3. The distribution broadens slightly with increasing t, but the effects are very small compared to the total uncertainty on the combined value.

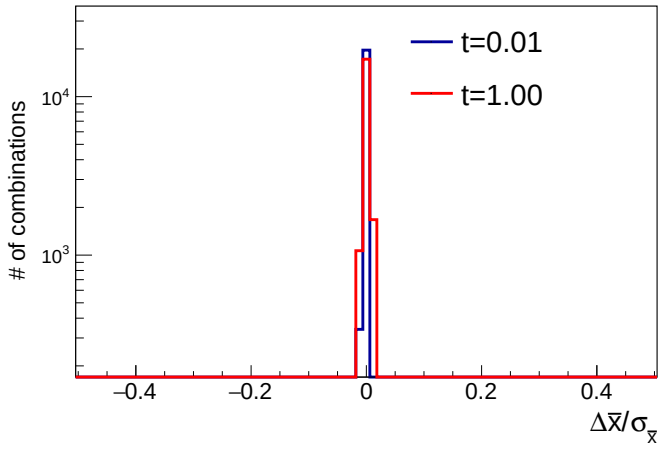

Figure 3: Difference between the combined result $\bar{x}$ obtained with the method proposed here and using a direct likelihood combination relative to the total uncertainty on $\bar{x}$, shown for different values of the upper threshold for systematic uncertainties t.

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 4 shows the dependence of $\Delta\bar{x}/\sigma_{\bar{x}}$ and $\Delta^r\sigma$ on the choice for the correlation coefficients c. For lower c, the relative bias increases slightly, 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 modeled with only a very moderate increase of combinations with $|\Delta^r\sigma|$ slightly different from 1. The same can be observed when repeating the procedure described here for a different number of bins in each pseudo measurement (not shown here). All results for 2, 4, 20, and 100 bins show no bias with respect to the central values as well as the total uncertainty.

In general, there is no bias with respect to the central result as well as its uncertainty 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 effects can be observed, but are expected as a result of neglecting

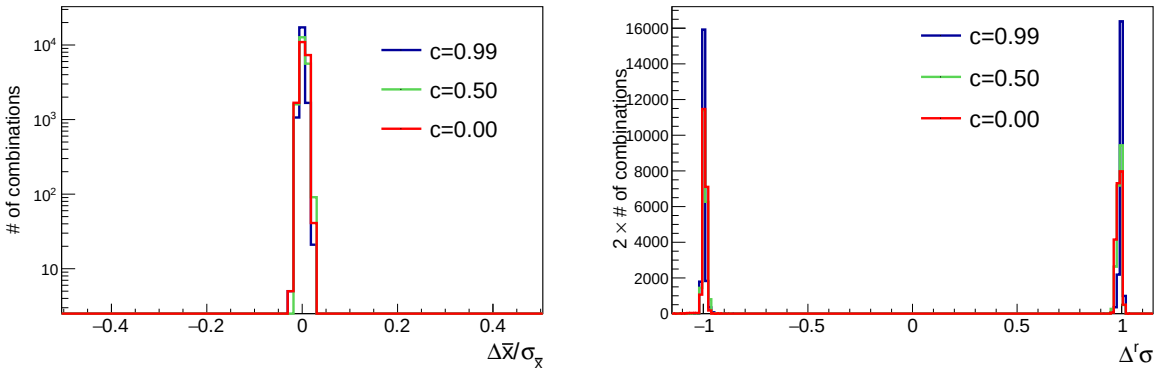

Figure 4: Left: difference between the combined results $\bar{x}$ from the method proposed here and using a direct likelihood combination relative to the total uncertainty on $\bar{x}$. Right: ratio of the uncertainties on $\bar{x}$ obtained with the method described in this document and using a direct likelihood combination. The ratio of the lower uncertainties is multiplied with -1. Both distributions are shown for different values of the correlation c between the systematic uncertainties of both pseudo measurements.

all binned information in any combination approach using unbinned numbers as input parameters. Both observations also hold true for different choices for x^a and x^b, and the number of uncertainties. Moreover, they are also valid for multiple estimates within one pseudo measurement without statistical correlations between them. The case of these correlations is discussed separately in Section 4.3.

For comparison, the combination of the pseudo measurements is repeated, neglecting correlations between systematic uncertainties within the same pseudo measurement, but still considering strong correlations between pseudo measurement a and b. This is achieved by inverting $(D^\alpha+\mathbb{1})$ in Eq. 3, removing the off-diagonal elements of the resulting covariance matrix, and replacing D^α by the inverse of this covariance matrix minus $\mathbb{1}$. This makes the method equivalent to e.g. using BLUE, with the exception of the Pearson χ^2. The latter difference is negligible for the low statistical uncertainties of x^a and x^b. As shown in Figure 5, this approximation can lead to biases with respect to the central value and severely wrong uncertainty estimates when the contribution of systematic uncertainties becomes non-negligible compared to the statistical uncertainty of about 0.6%. Therefore, it is crucial to model these correlations properly when performing a combination of results obtained in simultaneous fits of systematic uncertainties and the quantity to be determined.

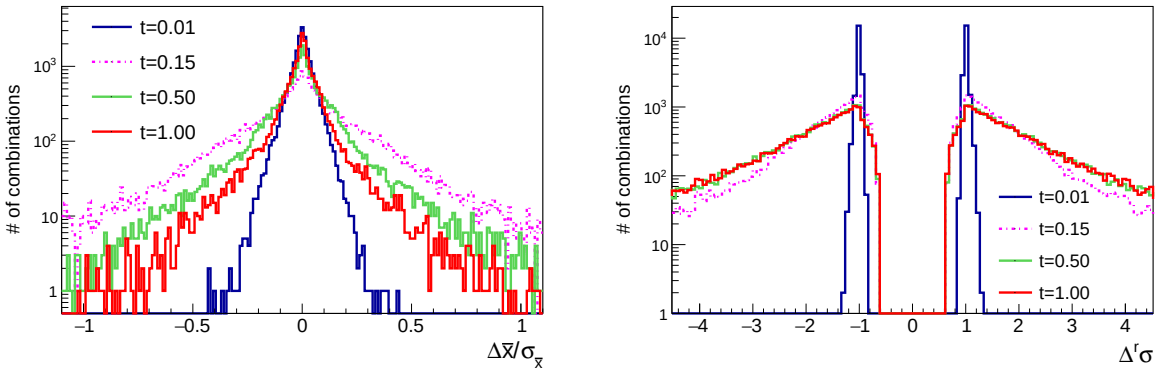

Figure 5: Left: difference between the combined results $\bar{x}$ neglecting correlations between uncertainties within a measurement and using a direct likelihood combination relative to the total uncertainty on $\bar{x}$. Right: ratio of the uncertainties on $\bar{x}$ obtained neglecting correlations between uncertainties within a measurement and using a direct likelihood combination. The ratio of the lower uncertainties is multiplied with -1. Both distributions are shown for different values of the upper threshold for systematic uncertainties t.

4.3 Modeling of direct correlations

The correct modeling of statistical correlations between the estimates within a measurement is tested by generating two pseudo measurements a and b similar to Section 4.2, each with two estimates x_1^a and x_2^a or x_1^b and x_2^b, respectively. The corresponding correlations 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 5000 combinations are performed for each choice of $d = \{0,\ 0.3,\ 0.9\}$, $t = \{0.0,\ 0.5\}$, and $c = \{0.00, 0.99\}$. The values for x_μ^α are chosen to be:

$$x_1^a = 30000, \tag{32}$$
$$x_1^b = 30600, \tag{33}$$
$$x_2^a = 20000, \text{ and} \tag{34}$$
$$x_2^b = 20500. \tag{35}$$

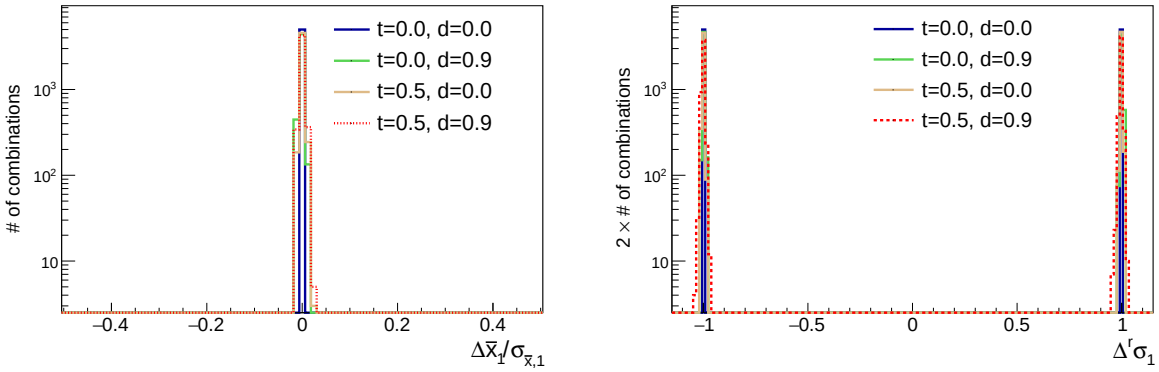

Figure 6: Left: difference between the combined value of $\bar{x}_1$ obtained with the method described here and using a direct likelihood combination relative to the total uncertainty of $\bar{x}_1$. Right: ratio of the uncertainties on $\bar{x}_1$ obtained with method proposed in this document and using a direct likelihood combination. The ratio of the lower uncertainties is multiplied with -1. Both distributions are shown for different values of the scale of systematic uncertainties t and the direct correlation between the estimates d. The combination assumes no correlation between the uncertainties of both pseudo measurements.

The resulting values for $\Delta\bar{x}_1/\sigma_{\bar{x}_1}$ and $\Delta^r\sigma_1$ are illustrated in Figure 6 for $c = 0$, and in Figure 7 for $c = 0.99$. No bias with respect to the modeling of the statistical correlation between estimates of the same measurement can be observed. The dependence on t and c is similar to the one discussed in the previous Section. The result of the combination of x_2 shows identical behavior and is therefore not depicted in addition. Also different choices for x_μ^α were tested and confirm that there is no bias with respect to d.

4.4 Relative uncertainties

The modeling of relative uncertainties is tested by generating two pseudo measurements with two parameters to be fitted, one relative uncertainty, and two absolute uncertainties, each. 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 relative total uncertainty of each pseudo measurement will differ from the dependence on t previously illustrated in Figure 3. In total 2×5000 pseudo measurements are generated. Figure 8 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.

Also when combining the pseudo measurements with contributions from relative uncertainties, there is neither a bias in the central value nor one in the estimation of the total uncertainty, 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 estimate per pseudo measurement and one large relative uncertainty of +15%. The input estimates are set to $30000 + \beta$, where β is a randomly generated value

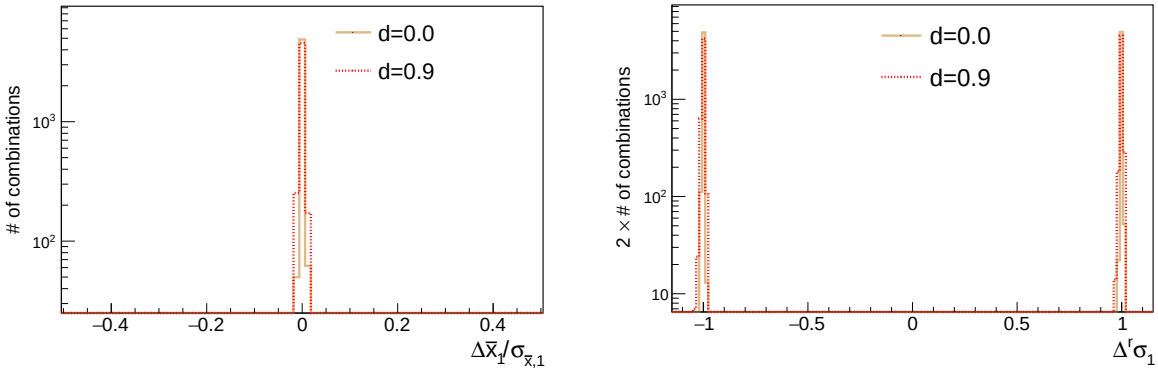

Figure 7: Left: difference between the combined value of $\bar{x}_1$ obtained with the method proposed here and using a direct likelihood combination relative to the total uncertainty of $\bar{x}_1$. Right: ratio of the uncertainties on $\bar{x}_1$ obtained with the method described in this document and using a direct likelihood combination. The ratio of the lower uncertainties is multiplied with -1. Both distributions are shown for different values of the scale of systematic uncertainties t and the direct correlation between the estimates d. The combination assumes a correlation of $c = 0.99$ between each uncertainty of one pseudo measurement with exactly one uncertainty of the other pseudo measurement.

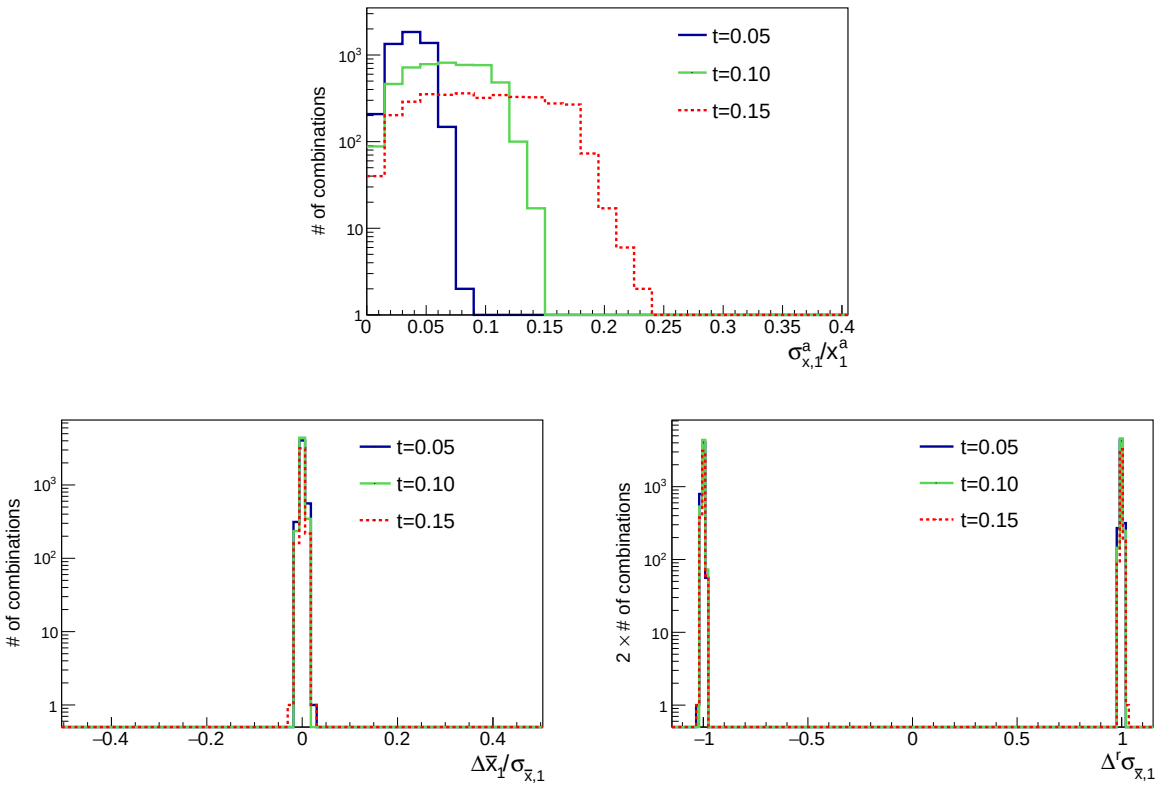

Figure 8: Top: relative total uncertainty of the pseudo measurement a, estimate 1 for different values of the threshold t. Left: difference between the combined value of $\bar{x}_1$ obtained with the method proposed here and using a direct likelihood combination relative to the total uncertainty of $\bar{x}_1$. Right: ratio of the uncertainties on $\bar{x}_1$ obtained with the method described in this document and using a direct likelihood combination. The ratio of the lower uncertainties is multiplied with -1. Both distributions are shown for different values of the scale of systematic uncertainties t and the correlation of c between each uncertainty of one pseudo measurement with exactly one uncertainty of the other pseudo measurement. All pseudo measurements comprise one relative and two absolute uncertainties.

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, the combined value can be larger than the highest input estimate. Figure 9 shows the relative difference between the highest input estimate and the combined value, as well as the possible biases with respect to the combined value and the uncertainty on it. Also in this scenario, no bias can be observed.

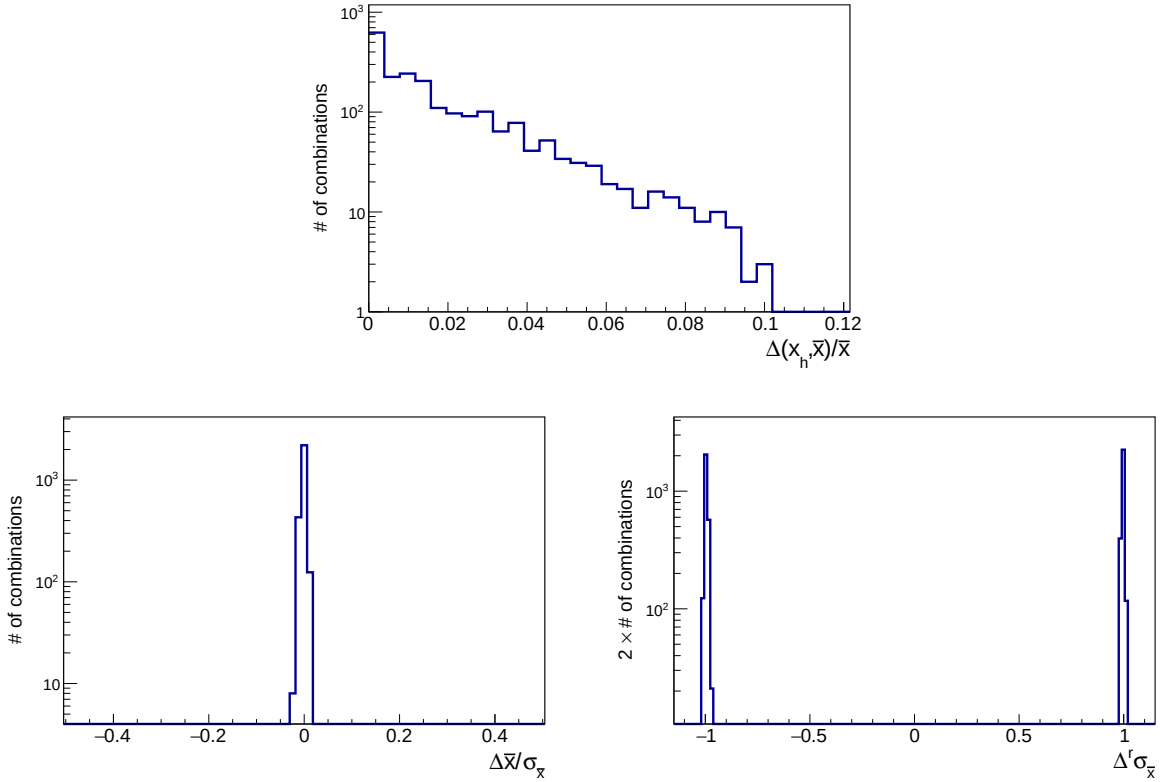

Figure 9: Top: relative difference between highest input estimate and combined result. Left: difference between the combined value of $\bar{x}_1$ obtained with the method described here and using a direct likelihood combination relative to the total uncertainty of $\bar{x}_1$. Right: ratio of the uncertainties on $\bar{x}_1$ obtained with the method proposed in this document and using a direct likelihood combination. The ratio of the lower uncertainties is multiplied with -1. Both distributions are shown for different values of the scale of systematic uncertainties t and the correlation of c between each uncertainty of one pseudo measurement with exactly one uncertainty of the other pseudo measurement. All pseudo measurements comprise one relative and two absolute uncertainties.

5 Installation

The method described in Section 2 is implemented in the Convino program for the combination of experimental results. The program is pre-installed on CERN lxplus (SLC6). For text-based usage, the environment can be set up by `source /afs/cern.ch/user/j/jkiesele/public/Convino/lxplus_env.sh` (sh, bash or zsh, only). The examples can be copied from the `examples` directory and adapted to the user needs.

To use the C++ interface, it is recommended to compile the full source code. It can be found at `https://github.com/jkiesele/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. Also other versions might be sufficient but are not tested.

6 User Interface

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, allowing to interface directly to C++ standard-library or ROOT 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.

6.1 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 χ^2 instead of the Pearson χ^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 *base file* in the following and is described in Section 6.1.2. 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.

6.1.1 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:

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

where the uncertainties `sys_XX` on the estimates `estimate_XX` are given in absolute values. The keyword `stat` is reserved for the statistical uncertainty. 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
  sys_c3       1030.3 -638.6   593.45
  estimate_a3 -0.525   0.5398 -0.231     3.25e-4
  estimate_b3 -0.708   0.2706 -0.448     0          4.89e-4
[end hessian]
```

while the correlation matrix has to include additional information about the constraints, the parameters received in the fit. These constraints are given in parentheses, such that the correlation matrix is of the format:

```
[correlation matrix]
  sys_a2       (1)     1
  sys_b2       (1)     0            1
  sys_c2       (.9)    0           -0.2          1
  sys_d2       (1)    -.2           0.1          0            1
  estimate_a2  (10.6)  0.1826484018 0.0652968037 0.5479452055 0.105936073  1
  estimate_b2  (12.8)  0.0826484018 0.3652968037 0.1479452055 0.4305936073 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. Often, these uncertainties have been externalized from the nuisance parameter fit and must not have any correlation with the ones defined in either 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 defined before 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.

6.1.2 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 filenames indicated with `file0=`, `file1=` etc. 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 + (0 & -0.5 : 0.5) sys_d2
[end correlations]
```

Here, `sys_b1` has a nominal correlation of 0.2 to `sys_c2` and 0 to `sys_d2`. The former correlation is scanned from -0.1 to 0.4, the latter from -0.5 to 0.5.

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

$$C = \begin{pmatrix} 1 & .99 & 0 \\ .99 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix}. \tag{36}$$

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

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 χ^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 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 10.

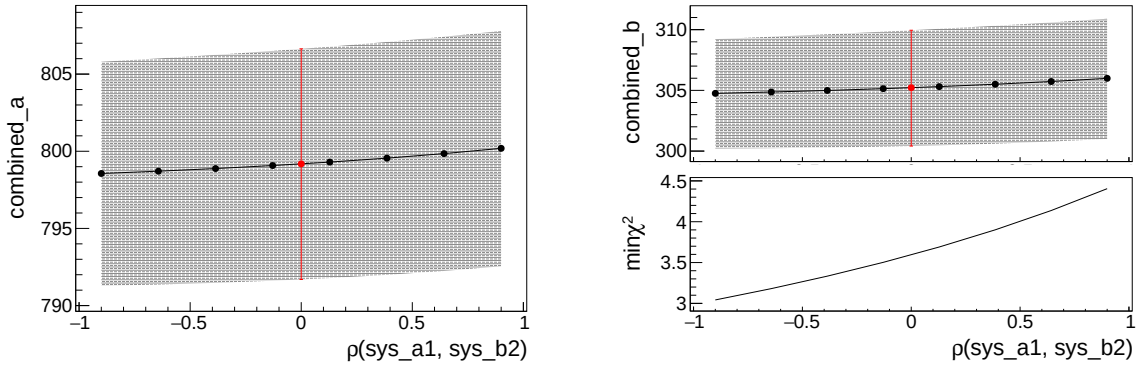

Figure 10: Combined values for `combined_a` (left) and `combined_b` (right) for a scan of the correlation coefficient for `sys_a1` and `sys_b2`. The red overlay shows the result obtained with the nominal assumption with uncertainties and the shaded area the uncertainty associated to each scan point, indicated by black markers. The lower panel on the right shows the dependence on the minimum χ^2 on the correlation coefficient. All values are obtained with the example configuration.

6.2 C++ Interface

The C++ interface is optimized for the combination of differential distributions and provides three basic classes which will be discussed in the following: the class `measurement`, which is analogue 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.

Each class is documented in the corresponding header file. Therefore, the documentation here describes the general strategy, only.

6.2.1 Measurement class

The measurement class provides the possibility to set 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.

In case of 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 using `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 function `setEstimateCorrelation`.

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

6.2.2 Combiner class

Once the individual objects of the `measurement` class 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 here 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 gives back a `combinationResult` class object.

6.2.3 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. It also gives access to pulls and constraints on the parameters after the combination. If differential distributions were combined, the result can be fed back to a ROOT `TH1` object or a `TGraphAsymmErrors` using the functions `fillTH1` or `fillTGraphAsymmErrors`.

7 Summary

The combination method presented in this document allows to consistently combine measurements obtained with simultaneous nuisance parameter fits. 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 likelihood, which is publicly available only in rare cases. Instead, the it relies on the central results and their covariances or Hessians, only. An extensive validation is performed, showing that the results and uncertainties are numerically equivalent to a direct likelihood combination.

The method is implemented in the Convino program, that provides a text-based and a C++ user interface. For both, each measurement is described individually, and then selected to be combined in a second step. The

text-based user interface provides an automatic scan of correlation assumptions and creates the corresponding Figures for graphical representation.

References

[1] CMS Collaboration. Measurement of the $t\bar{t}$ production cross section in the eμ channel in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV. *J. High Energy Phys. 08*, (arXiv:1603.02303), 2016.

[2] 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. (arXiv:1701.06228), 2017.

[3] 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. *Eur. Phys. J.*, C77(1):15, 2017.

[4] Jan Kieseler, Katerina Lipka, and Sven-Olaf Moch. Calibration of the Top-Quark Monte Carlo Mass. *Phys. Rev. Lett.*, 116(16):162001, 2016.

[5] ATLAS Collaboration. Simultaneous measurements of the $t\bar{t}$, W^+W^-, and $Z/\gamma^* \to \tau\tau$ production cross-sections in *pp* collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys. Rev.*, D91(5):052005, 2015.

[6] Richard 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.

[7] Richard Nisius. On the combination of correlated estimates of a physics observable. *Eur.Phys.J. C*, 74, 2014.

[8] S. Alekhin et al. HERAFitter. *Eur. Phys. J.*, C75(7):304, 2015.

[9] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997.

[10] F. James and M. Roos. Minuit: a system for function minimization and analysis of the parameter errors and correlations. *Comput. Phys. Commun.*, 10:343, 1975.