BLUE- new ideas

Roberto Chierici (CNRS)

TOPLHCWG open session, 28th-29th November 2013

Introduction

- Part I: new ideas and their implementation
  - Weights and information in a BLUE combination
  - Unknown correlations and conservativeness
    - Work based on arXiv:1307.4003, submitted to EPJC
    - Please also see presentation at the open session in date 29/11/2012
- Part II: towards a common code
  - Proposal from internal discussions
- Iterative BLUE
  - See dedicated talk in agenda

Part I
new ideas

Roberto Chierici (CNRS), Andrea Valassi (CERN)

Reminders and questions

- In a weighted average, the BLUE method finds the parameters λ_i by minimizing the total error
  - The λ_is are not directly related to the impact that a measurement has in the reduction of the total error
    - More so if important correlations enter the game
  - Peculiar features are present when there are high positive correlations
    - Some of the λ_is can be negative
    - $\mathrm{var}(\hat{Y})$ vanishes as correlations tend to unity ($\mathcal{M}_{ij}$ becomes singular)
    - BLUE is undefined in a regime of full correlation

$$\hat{Y} = \tilde{\boldsymbol{\lambda}}\mathbf{y} = \sum_{i=1}^{n} \lambda_i \, y_i$$

$$\mathrm{var}(\hat{Y}) = \sigma_{\hat{Y}}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \, \mathcal{M}_{ij} \, \lambda_j$$

$$\lambda_i = \frac{(\mathcal{M}^{-1}\mathbf{U})_i}{(\tilde{\mathbf{U}}\mathcal{M}^{-1}\mathbf{U})}$$

- Questions
  - Q1: can I estimate the impact of a measurement in a BLUE combination in a unambiguous way?
  - Q2: how do I realize if I am in a regime of high correlations?
  - Q3: what to do when the correlations cannot be precisely estimated and they are large?

Definition of "weights"

- Q1: can I estimate the impact of a measurement in a combination in an unambiguous way?
  - Answer: if there are significant correlations, no.
    - But we can do much better than what done so far.

- Suggest to quote "weights" determined from the concept of *information* (=1/σ²)
  - IIWs: their interpretation is quite simple:
    - IIWs for the measurements are positive by construction
    - Add one IIW for the ensemble of correlations: this can be negative, zero or positive!

$$\mathrm{IW}_i = \frac{1/\sigma_i^2}{1/\sigma_{\hat{Y}}^2}$$

$$\mathrm{IW}_{\mathrm{corr}} = \frac{1/\sigma_{\hat{Y}}^2 - \sum_i 1/\sigma_i^2}{1/\sigma_{\hat{Y}}^2}$$

  - MIWs: quantify the marginal contribution by measurement I, including correlation
    - MIWs are zero or positive by construction
    - Correlations can make MIW smaller or larger than IIW!

$$\mathrm{MIW}_i = \frac{\Delta I_i}{I_{(n\ \mathrm{meas.})}}$$

- IIWs and MIWs can be quoted together with the λ_is (CVWs)
  - We strongly discourage the further use of absolute values of the λ_is :

$$\mathrm{RI}_i = \frac{|\lambda_i|}{\sum_{j=1}^{n} |\lambda_j|}$$

Measurements		BLUE comb. coeff. [%]	IIW [%]	MIW [%]
ATLAS l+jets	172.31 ± 1.55	22.6	37.3	8.2
ATLAS di-l	173.09 ± 1.63	3.6	33.8	0.2
CMS l+jets	173.49 ± 1.06	60.6	79.2	25.1
CMS di-l	172.50 ± 1.52	-8.4	38.8	0.7
CMS all jets	173.49 ± 1.41	21.6	45.0	4.4
Correlations	—	—	−134.1	—

ATLAS-CONF-2013-102
CMS PAS TOP-13-005

Ranking measurements

- Q1': can I unambiguously rank my measurements according to their "importance" in a BLUE combination?
  - Answer: again, if there are significant correlations, no.

- The weights defined in the previous slide can be used for ranking. The result will however depend on the chosen set of weights (meaning it will depend on the importance of correlations and the way they are treated)

- Example:
  - A, B uncorrelated
  - B1, B2 (→B) are correlated with ρ=0.875
  - B11, B12 (→B1) are correlated at 99.999%
  - → ranking depends on how one considers the (combined effect of) correlations
  - → MIWs can be low for sets of measurements largely correlated among themselves
  - → RIs are different if B is an individual measurement or a combination of two (!)
  - → IIWs is a safe convention, but still arbitrary

Measurements		CVW/%	IIW/%	MIW/%	RI/%
A	103.00 ± 3.87	40.00	40.00	40.00	40.00
B	98.00 ± 3.16	60.00	60.00	60.00	60.00
Correlations	—	—	0.00	—	—
BLUE / Total	100.00 ± 2.45	100.00	100.00	100.00	100.00

Measurements		CVW/%	IIW/%	MIW/%	RI/%
A	103.00 ± 3.87	40.00	40.00	40.00	25.00
B1	99.00 ± 4.00	90.00	37.50	50.63	56.25
B2	101.00 ± 8.00	-30.00	9.38	22.50	18.75
Correlations	—	—	13.13	—	—
BLUE / Total	100.00 ± 2.45	100.00	100.00	113.13	100.00

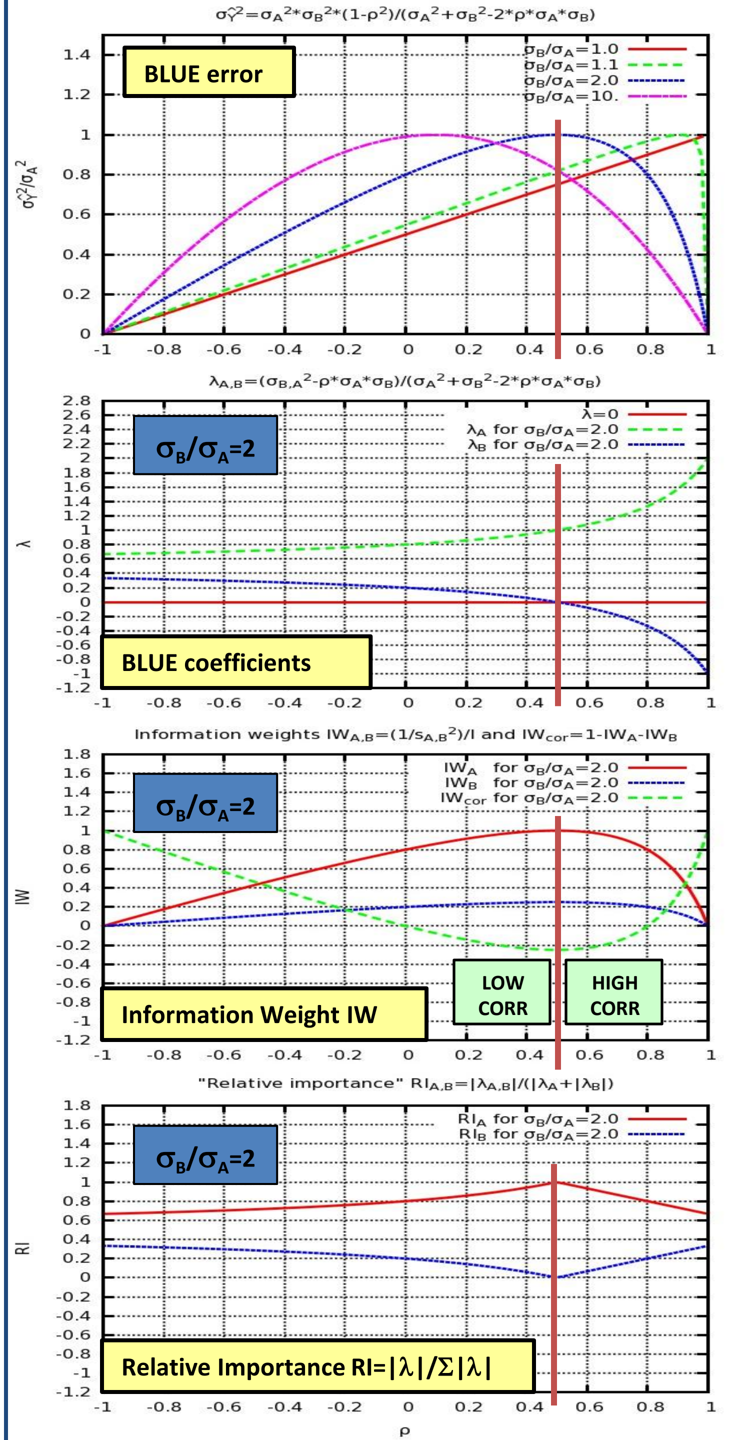
Measurements		CVW/%	IIW/%	MIW/%	RI/%
A	103.00 ± 3.87	40.00	40.00	40.00	25.00
B11	99.01 ± 4.00	45.00	37.50	~0	28.13
B12	98.99 ± 4.00	45.00	37.50	~0	28.13
B2	101.00 ± 8.00	-30.00	9.37	22.50	18.75
Correlations	—	—	-24.37	—	—
BLUE / Total	100.00 ± 2.45	100.00	100.00	62.50	100.00

Weights

- For IIWs the ensemble of the correlation becomes like a measurement per se
  - ➢ Weights and information
  - ➢ Extremely difficult (often not possible) to further split the correlations into sub-components
    - ○ e.g. from different sources, or from just two measurements

- In the very general case of N measurements one can identify two portions of the {ρ} space
  - ➢ Low correlation regime: the error increase with the correlation increasing
  - ➢ High correlation regime: the error decreases with the correlation increasing

- The transition between high and low correlation is invariably identified by one of the following facts:
  - ➢ At least one of the λ_i becomes negative
  - ➢ The total error passes through a maximum
  - ➢ The information from correlations passes through a minimum (meaning dI/d{ρ} changes sign)



Highs and lows

- Q2: how do I realize that I am in a regime of high correlation (for some of the measurements?)
  - ➢ Answer: by using one of the properties from the previous slide:
    - 1. My measurement gets a BLUE weight which is negative
    - 2. The derivative of the information with respect to the correlation of my measurement with at least another measurement in the combination gets positive

$$\frac{\partial I}{\partial \rho_{ij}} = -2I^2 \lambda_i \lambda_j \sigma_i \sigma_j$$

  - ➢ Remember: if you are in there, your error goes down with larger correlations !
    - So you may want to give a second thought to the values of the correlations you put in

- How can I determine the sources/measurements which induce this regime of high correlation (so that I can study them better)?
  - ➢ Check the (normalized) information derivatives with respect to the two-measurement correlations, evaluated at nominal correlation or at ρ=1.

$$\frac{\rho_{ij}^{[s]}}{I}\left(\frac{\partial I}{\partial \rho_{ij}^{[s]}}\right) = -2\frac{\lambda_i \lambda_j \mathscr{M}_{ij}^{[s]}}{\sigma_{\hat{Y}}^2}$$

$\frac{1}{I}@1 \times \frac{\partial I}{\partial f}@1$	ATL10lj	ATL11lj	ATL11aj	CMS10ll	CMS10lj	CMS11ll	CMS11μj
ATL10lj	—	—	—	—	—	TOTAL	
ATL11lj	0.197					0.219	
ATL11aj	-0.004	0.007	—	—	—	—	—
CMS10ll	-0.005	0.009	0.000	—	—	—	—
CMS10lj	-0.001	0.001	0.000	0.000	—	—	—
CMS11ll	-0.010	0.023	0.000	-0.007	-0.001	—	—
CMS11μj	0.149	-0.318	0.005	0.058	0.004	0.111	—

OLD 2012 NUMBERS

Unknown correlations

- Q3: I am in a high correlation regime and I do not really know my correlation. What should I do if I want to be "conservative" rather than wrong?
  - Answer: set your unknown correlation(s) to the value maximizing the final error, or equivalently minimizing the information in the BLUE combination
    - In high correlation regimes this is not 100%

- There are several ways to do so in a pragmatic way, each involving a different degree of arbitrariness. A few techniques are proposed:
  - (Multi-dimensional) minimization of information w.r.t. correlations
    - A minimization as a function of all $N_{sources} \cdot n \cdot (n-1)/2$ would be under-constrained, need to choose the subset of correlations w.r.t. which minimize (for instance by error source or by pair of measurements)

Combination	BLUE
Nominal correlations	173.29 ± 0.95
Minimise by global factor	173.29 ± 0.95
Minimise by error source	173.27 ± 0.95
Minimise by off-diagonal element	173.21 ± 0.95

  - Iterative removal of measurements with negative BLUE coefficients
    - The most conservative choice, even if the least "politically correct"
  - The "onionization" prescription
    - Limit each off-diagonal element of the covariance matrix to be at worse equal to the corresponding diagonal element

$$\begin{cases} (\mathcal{M}')^{[s]}_{ij} \leq (\mathcal{M}')^{[s]}_{ii} = \mathcal{M}^{[s]}_{ii} = (\sigma^{[s]}_i)^2 \\ (\mathcal{M}')^{[s]}_{ij} \leq (\mathcal{M}')^{[s]}_{jj} = \mathcal{M}^{[s]}_{jj} = (\sigma^{[s]}_j)^2 \end{cases}$$

In summary (part I)

- Let us change the way we present the weights in a BLUE combination
  - ➢ Systematic use of IIW, MIW together with the CVW.
  - ➢ They can be used to rank measurements: we should agree if we want to do it, and how.

- Let us not worry any longer about negative CVWs
  - ➢ They are needed and good ! They simply tell us when we are learning from the high correlations between our measurements.
  - ➢ We should worry only when they come in a regime of unknown, high correlations. For this we should always check the behaviour of the information/error as a function of “suspicious“ correlations.

- Let us discuss what is the easiest solution for being “conservative” when in presence of unknown high correlations
  - ➢ Minimization function of ρ is an option: needed only when $dI/d\rho$ becomes positive
  - ➢ Exclude selected measurements from the combination?

Part II
towards a common BLUE code?

Markus, Roberto

Summary of preliminary discussions

Present efforts

- Several independent versions of BLUE have been developed in the combination working groups.
  - The first FORTRAN versions have now started being migrated into C++

- Current versions in use in the WG were reviewed in an internal meeting
  - Mass – ATLAS development (R. Nisius)
  - Top pair cross section – private version+old FORTRAN code (used at the Tevatron)
  - Single top cross section – private version+old FORTRAN code
  - W helicity – BLUE in BAT (K. Kroeninger)
  - New ideas – BlueFin (A. Valassi)

- First discussions and exchange of ideas about the possibility of using a common code, maintained in a more “central” way

Common BLUE code: desiderata

- Combine N measurements:
  - Present results together with various weights: CVWs, IIWs, MIWs
  - Rank measurements according to information weights (in principle a switchable option)
    - IIWs as default?
  - Produce control plots on demand
    - Scan of errors as a function of correlations
    - Information and information derivatives as a function of correlations
  - Warnings if a regime of high correlation is found
    - Rank the worrying correlations

- Treatment of unknown correlation regimes
  - Check the information derivatives with respect to the two-measurement correlations, evaluated at nominal correlation or at $\rho=1$.

- Additional features
  - Possibility for iterative BLUE.
  - Use standard and user friendly conventions for input files (AWA for output).

Implementation of BLUE in BAT (K. Kroeninger)

- Bayesian tools useful when in need of taking priors into account
  - ➢ For instance used for W helicity, with constraints on the sum of the parameters to be combined (the helicity fractions) to unity.

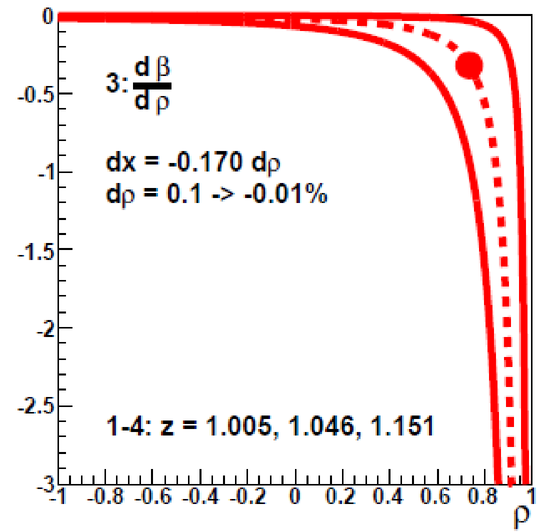
BAT implementation

- Solutions:
  - Analytical BLUE solution:
    - Tested against good-old Fortran version
  - Numerical Maximum Likelihood / Maximum Posterior numerical solution
    - Note: Interpretation Frequentist/Bayesian not an issue here
- Interface:
  - Read in from text file, re-use input from Fortran version
- Requirements:
  - Recent ROOT version
  - Latest BAT version (v0.9.3)

BAT
BAYESIAN ANALYSIS TOOLKIT
http://www.mppmu.mpg.de/bat/

BLUE in C++/Root (R. Nisius)

- Root based package currently used for the LHC top mass combination
  - Many features already implemented (incl. "information")
  - Several crosschecks of public combinations performed



Implementation

- The combination of N estimates for m observables with $N \geq m$ is performed implementing the formulas from NIMA 500 (2003) 391.
- A Root Class *Blue* and one function per input data are used. Both should be compiled using the ACLiC system. This is the Home of Software and Manual (about 30 pages).

Features

- The software has a three step procedure: 1) Fill, 2) Solve, 3) Manipulate estimates.
- It allows to repeat steps 2) and 3) numerous times (without touching the original input) while disabling individual measurements and/or uncertainty sources, and/or by changing the correlation assumptions in a very flexible way.
- It calculates the compatibility of estimates and observables.
- It allows to inspect any pair of estimates to help deciding on their combination.
- It provides a large number of print routines for software control and result inspection.
- It returns the results into local structures for further usage, e.g. displaying.
- Latex and PDF output of estimates used, observables obtained and more is provided.
- Presently 13 example routines *B_Name.cxx* are included. They reproduce published results. At the same time they show how to use the various software features.

BlueFin (A. Valassi)

- Originally developed for testing new ideas about information, weights and minimization procedures:
  - Starting from C++ translation of Fortran code used for LEPEWWG 4f cross sections
  - Now a complete BLUE code (https://svnweb.cern.ch/trac/bluefin), with automatic output in PDF format concerning weights, information derivatives and minimization procedures in case of unknown correlations.

Measurements		CVW/%	IIW/%	MIW/%	RI/%	Unc	Bkgd	Lumi
AXS	95.00 ± 17.92	60.39	50.91	34.69	48.78	10.00	10.00	11.00
BXS	144.00 ± 44.63	-11.90	8.20	8.97	9.61	14.00	40.00	14.00
CXS	115.00 ± 20.81	25.36	37.74	14.63	20.49	18.00	3.00	10.00
DXS	122.00 ± 25.00	26.15	26.15	26.15	21.12	25.00	0	0
Correlations	—	—	-23.01	—	—	—	—	—
BLUE xs	101.30 ± 12.78	100.00	100.00	84.44	100.00	10.14	2.04	7.51

Table 1: BLUE of the combination (χ^2/ndof= 4.23/3). For each input measurement i the following are listed: the central value weight CVW_i or λ_i, the intrinsic information weight IIW_i , the marginal information weight MIW_i, the relative importance RI_i. The intrinsic information weight $\mathrm{IIW}_{\mathrm{corr}}$ of correlations is also shown on a separate row.

OffDiag & ErrSrc	Unc	Bkgd	Lumi	OffDiag
BXS / AXS	0	0.352	0.135	0.487
CXS / AXS	0	-0.056	-0.206	-0.262
CXS / BXS	0	0.044	0.052	0.096
DXS / AXS	0	0	0	0
DXS / BXS	0	0	0	0
DXS / CXS	0	0	0	0
ErrSrc	0	0.340	-0.019	GlobFact 0.321

Table 2: Normalised Fisher information derivatives 1/I*dI/dX for the combination under consideration. The derivatives in the table are computed with respect to scale factors X, representing the ratio of a given correlation to its "current" value in the combination under consideration, and all normalized by the information I for the "current" values of correlations. They are computed for the "current" values of correlations (in this case: nominal correlations). Color boxes indicate normalised derivatives greater than 0.05 (yellow), 0.10 (orange) and 0.15 (red). The last column and last row list information derivatives when the same rescaling factor is used for a given off-diagonal element or error source, which are equal to the sums of individual derivatives in each row and column, respectively.

Summary: towards a “common” version?

- Felt useful to propose a common code in the TOPLHCWG with all the features described above, and maintained “centrally”
  - Publicly accessible on common svn area.
  - Including most, or all, of the desiderata
  - This common version should not prevent independent developments, especially for issues not addressed by the proposed code

- Work is ongoing to define how this goal can be achieved
  - Willingness of the authors to help providing a common version
  - Should that be standalone or integrated in some other frame, more centrally maintained at CERN?
    - Contact with L. Moneta from Root

- Stay tuned, any further suggestion/experience is always appreciated