Combination and checks of highly correlated measurements of the muon precession frequency in magnetic field for the FNAL measurement of Muon g-2

for the FNAL Muon $g-2$ collaboration,

Alberto Lusiani

https://orcid.org/0000-0002-6876-3288

2025 European Physical Society Conference on High Energy Physics (EPS-HEP 2025), Marseille, 7-11 July 2025

Muon $g-2$ measurement relies on measurement of muon precession frequency

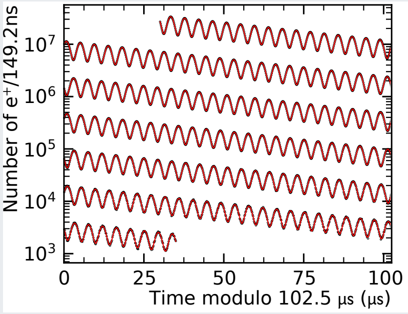
$$\frac{g_\mu - 2}{2} = a_\mu = \left[\frac{\omega_a}{\tilde{\omega}'_p(T_r)}\right] \cdot \left[\frac{\mu'_p(T)}{\mu_B}\right] \left[\frac{m_\mu}{m_e}\right]$$

- ω_a: 114 ppb stat., 51 ppb syst.
- $\tilde{\omega}'_p(T_r)$: 5 ppb stat., 56 ppb syst.
- $\mu'_p(T)/\mu_B$: 4 ppb
- m_μ/m_e: 22 ppb

$$\omega_a = \omega_a^{\text{wiggle plot fit}} + \omega_a^{\text{beam dynamics corrections}}$$

- $\omega_a^{\text{wiggle plot fit}}$: 114 ppb stat., 30 ppb syst.
- $\omega_a^{\text{beam dynamics corrections}}$: 1 ppb stat., 42 ppb syst.

FNAL-E989 muon $g-2$ measurement released on 3-Jun-2025 (Run 4+5+6 samples only)



ω_a^m wiggle plot fit

- ~220 billion muon decays for Run 4+5+6
- main trend: cosine-modulated decreasing exponential
- corrections: ~25 additional parameters for beam dynamics effects, muon loss, rate-dependent detector gain effects
- 20 different fits by 8 groups
  - (4 out of 8 groups shared same blinding offsets)
- all fits performed on 4 datasets

Run 4,5,6 ω_a measurements, 8 groups, 8 methods, 20 analyses

groups

g
BU
RE1
RE2
RE3
RE4
EU
Ky
SJTU

reconstructions

r
RW
RE
RI2
RQ

analyses

g	m	r	rq
BU	T, A, RT, RA	RW	K
RE1	T, A	RE	R
RE2	T, A	RE	
RE3	RT, RA	RE	R
RE4	ST, SA	RE	
EU	T, A, RT, RA	RI2	R
UK	Q, RQ	RQ	R
SJTU	T, A	RW	

measurement methods

m	method
T	Threshold
A	Asymmetry weighted
RT	Ratio T
RA	Ratio A
Q	Charge
RQ	Ratio Charge
ST	Stroboscopic T
SA	Stroboscopic A

pileup subtraction

ps
RW-BU
RE
RW-EU
RW-SJTU

ratio quartering methods

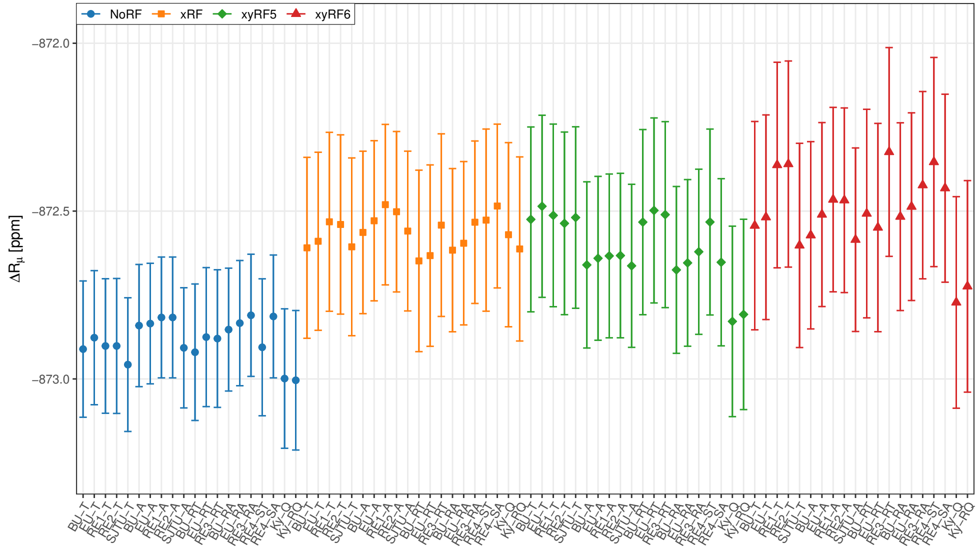
rq	ratio quartering methods
R	random
K	kernel

envelope modeling

em
analytic function
spline
Gaussian Process Regression

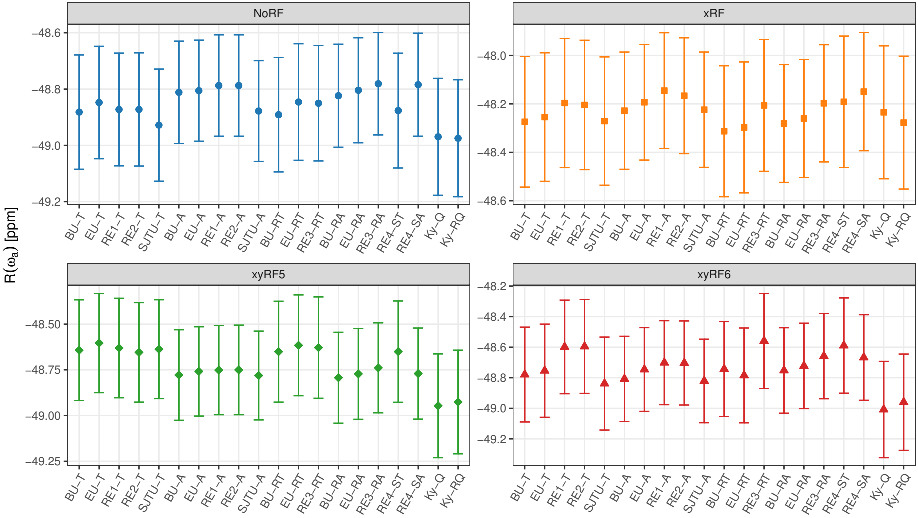
$\frac{\omega_a^m}{\tilde{\omega}'_p(T_r)} = R_\mu \propto a_\mu$ consistent across datasets, as expected

R_μ for 20 analyses and for 4 Run 4+5+6 datasets, stat. unc. only, no corrections



ω_a^m measurements must be consistent within each dataset

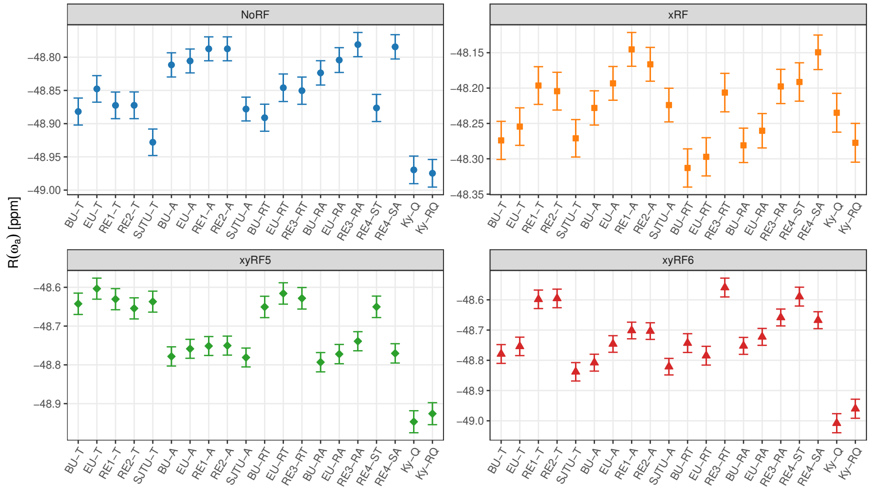
20 ω_a^m measurements on 4 Run 4+5+6 datasets



- however, statistical uncertainties within same datasets are highly correlated

ω_a^m measurements with just their systematic uncertainties' estimates

20 ω_a^m measurements on 4 Run 4+5+6 datasets



- even assuming uncorrelated systematics, ω_a^m measurements are inconsistent
- part of statistical uncertainties are uncorrelated between different fits and measurements methods

Use data bootstrap samples to estimate ω_a^m measurements correlation

- for each dataset, build 200 bootstrap samples of same size by ramdomly selecting DAQ subruns, with repetitions [B. Efron, Annals Statist. 7 (1979)]
  (subruns are DAQ files, each containing around 37 K muon decays with energetic positrons in the calorimeters)
- for each analisys i, perform fits on 200 bootstrap samples $\Rightarrow$ ω_a^m fit measurements $m_{i,b}$
- estimate ω_a^m covariance: $\text{cov}(m_1, m_2) = \frac{\sum_b (m_{1,b} - \overline{m}_1)(m_{2,b} - \overline{m}_2)}{N_b - 1}$ with $\overline{m}_i = \frac{\sum_b m_{i,b}}{N_b}$
- expected residual uncertainty $\sigma^2(m_2 - m_1) = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$

bootstrap samples produced for most analyses

group	method	RF datasets
BU	T, A, RT, RA	No, x, xy5, xy6
EU	T, A, RT, RA	No, x, xy5, xy6
RE1	T, A	No, x, xy5, xy6
RE2	T, A	No, x, xy5, xy6
RE3	RA	No, x, xy5, xy6
RE4	SA	No, x, xy5, xy6
SJTU	T, A	No, x, xy5, xy6

correlations between Q-x vs. A-x & T-x analyses

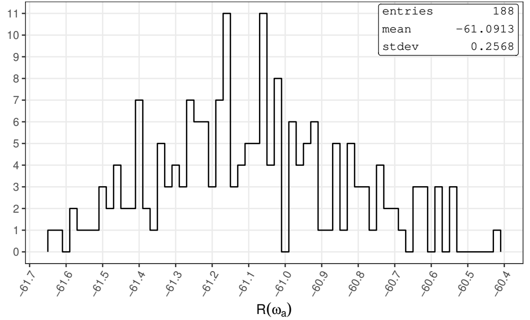
- bootstrap samples fits not done for Q-x analyses
- reliable correlation evaluations using fits on 34 sub-samples of Run 4+5+6 sample

past evaluations of analyses' correlations

- used for checks and addition some alternative evaluations of analyses' correlations from past dedicated studies done, for Run 2+3 measurement (partly outdated)

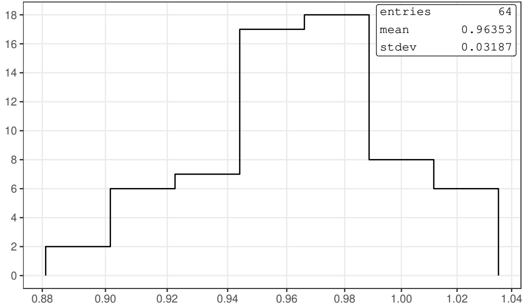
Bootstrap ω_a^m distribution width consistent with fit uncertainty

trimmed bootstrap ω_a^m distribution



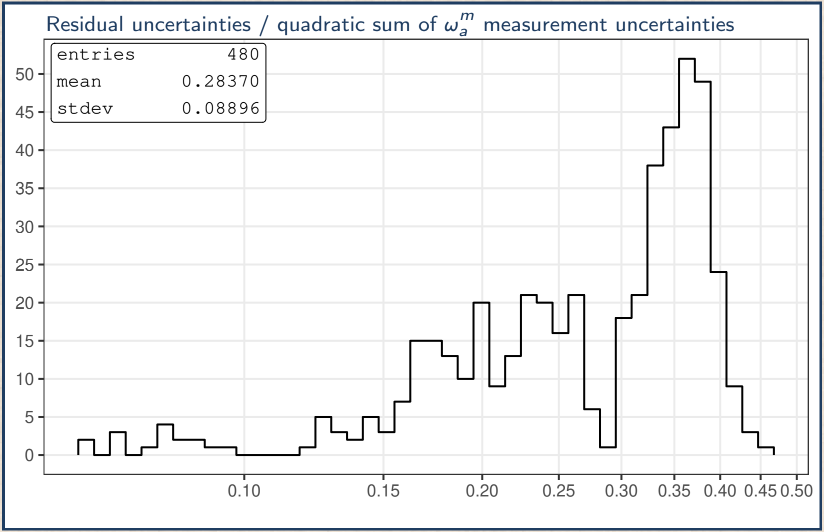
- removed $>3\sigma$ outliers of $m_{i,b}$ distribution

bootstrap RMS width / ω_a^m fit uncertainty



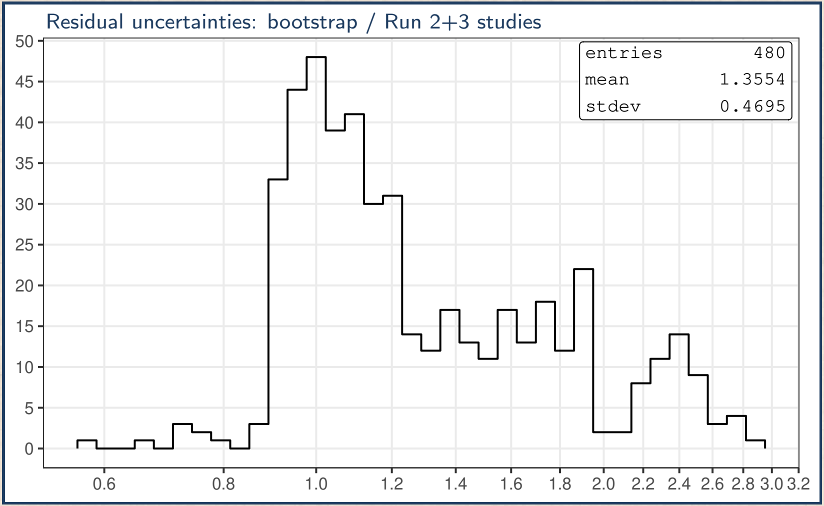
- bootstrap ω_a^m RMS width consistent with ω_a^m fit uncertainty when taking into account 3σ trimming

Residual uncertainties



- large uncertainties: different methods (Q-A, Q-T, A-T)
- small uncertainties: ratio vs. non-ratio, different reconstructions

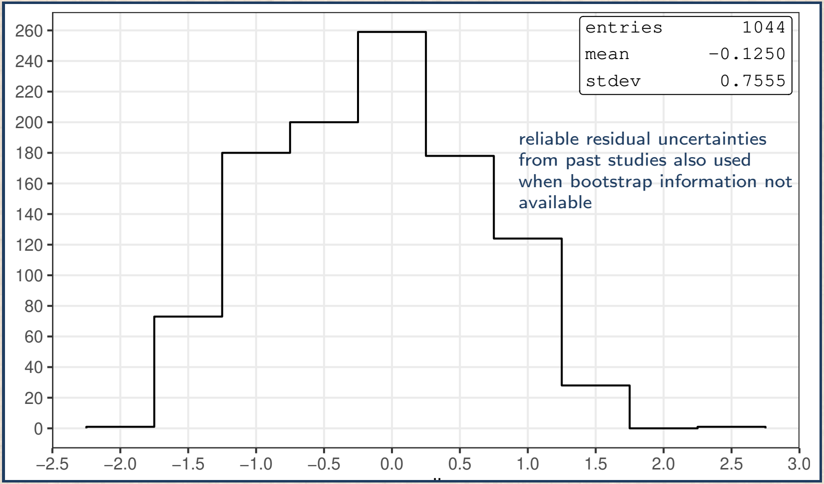
Past residual uncertainties' estimates ∼consistent with bootstrap ones



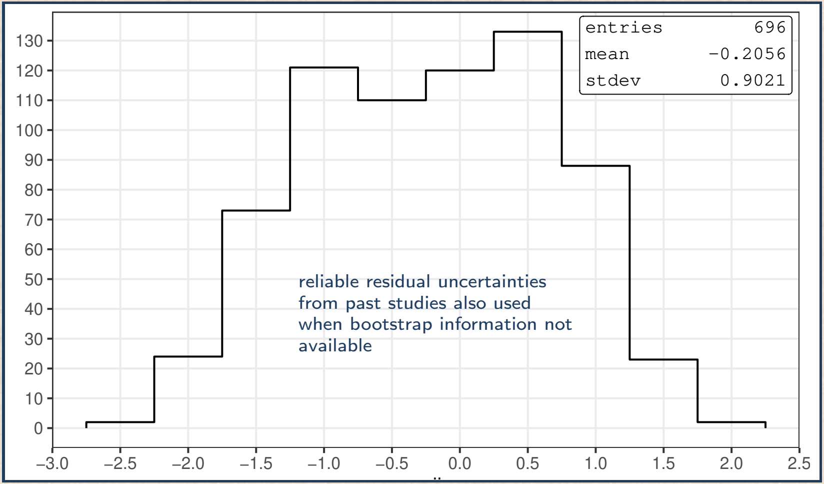
- note: past Run 2+3 estimates not fully updated to Run 4+5+6 analysis features

Pulls – dealing with analysis-group-dependent blinding offsets

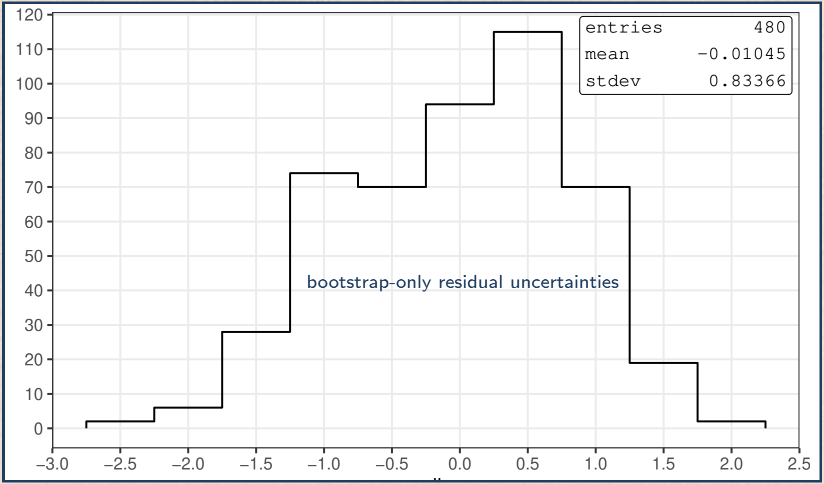
- consistency checks for group-dependent blinding offsets possible relying on different datasets
- $\text{pull} = \dfrac{(\omega_a^{\text{dataset 2, analysis 2}} - \omega_a^{\text{dataset 2, analysis 1}}) - (\omega_a^{\text{dataset 1, analysis 2}} - \omega_a^{\text{dataset 1, analysis 1}})}{\sqrt{\sigma^2(\omega_a^{\text{dataset 2, analysis 2}} - \omega_a^{\text{dataset 2, analysis 1}}) + \sigma^2(\omega_a^{\text{dataset 1, analysis 2}} - \omega_a^{\text{dataset 1, analysis 1}})}}$



Pulls after removing group-dependent blinding offsets / 1



Pulls after removing group-dependent blinding offsets / 2



- in several occasions, inconsistencies prompted debugging and improving analyses procedures and fitting

Combination of highly correlated ω_a^m measurements is same dataset

minimum χ^2 / minimal uncertainty / optimal combination

- $\chi^2 = \sum_{ij} (m_i - \overline{m})\, V_{ij}^{-1}\, (m_j - \overline{m}) = \sum_{ij} (m_i - A_{i1}^t \overline{m})\, V_{ij}^{-1}\, (m_j - A_{j1}\overline{m})$; $\qquad A_{ki} = \begin{pmatrix} 1 \\ 1 \\ \vdots \end{pmatrix}$
- $\overline{m} = \sum_i w_i m_i$ optimal combination is linear combination of measurements
- $w_i = (A^t V^{-1} A)^{-1} A^t V^{-1}$ optimal weights; $\qquad \sigma_{\text{av}}^2 = w_i^t V_{ij} w_i$ uncertainty of linear combination

- optimal combination of highly correlated measurements is unstable when covariance is uncertain [G. Cowan, Statistical data analysis, 1998] and can result in
  - negative weights
  - unrealistically small uncertainty of average
- use even average of most precise and most independent analyses
  - 1 analysis for each of 7 groups that used A/RA/SA methods
  - optimal combination for reliability of independent systematic uncertainties
  - note that all other analyses remain useful for consistency checks

Combination of highly correlated ω_a^m measurements is same dataset / 2

even average of 7 A-x analyses with conservative uncertainty

- $\bar{\omega}_a^{m,\ \text{dataset}} = w_i \omega_a^{m,\ \text{dataset, analysis-}i}$ $\qquad w_i = 1/7$
- $\sigma\left(\bar{\omega}_a^{m,\ \text{dataset}}\right) = w_i^t V_{ij} w_i$ $\qquad$ (continues to hold as it is linear combination)
- since estimated V_{ij} from bootstrap is uncertain, use full 100% correlation to be conservative

  $\sigma\left(\bar{\omega}_a^{m,\ \text{dataset}}\right) = w_i^t V_{ij}^{100\%} w_i$ $\qquad$ ($V_{ij}^{100\%}$ = covariance corresponding to 100% correlation)
- estimate that even averaging and conservative correlation increase optimal $\sigma\left(\bar{\omega}_a^{m,\ \text{dataset}}\right)$ by 1.5%

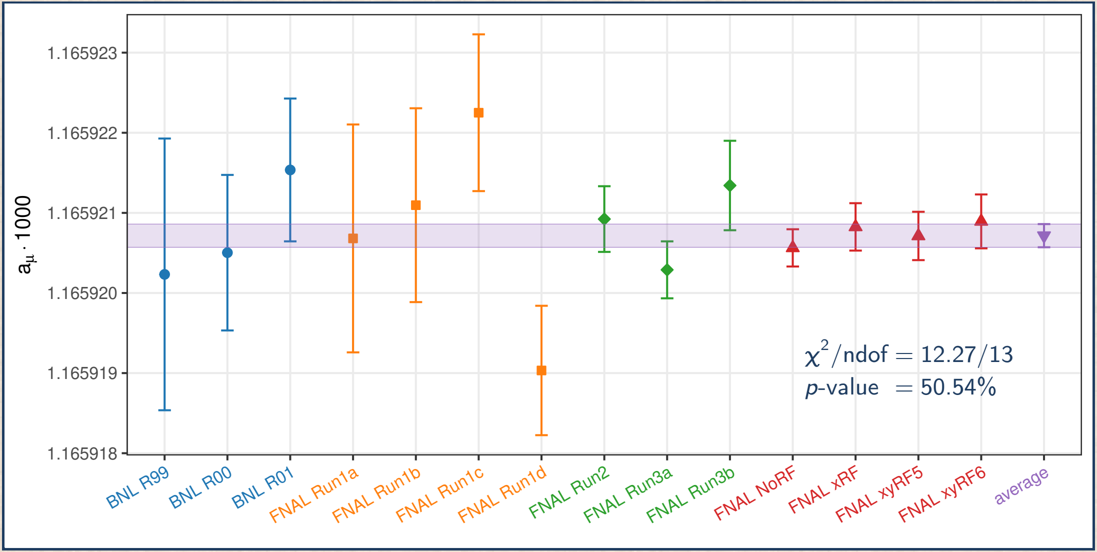
Summary

- used bootstrap technique to estimate correlation of different ω_a^m analyses on the same data in order to compute residuals' uncertainties to perform consistency checks
- resilient combination of highly correlated analyses, with minor 1.5% inflation of uncertainty
  - even average of independent most precise analyses
  - using 100% conservative correlation to compute uncertainty of combination
  - remaining analyses remain useful for consistency checks

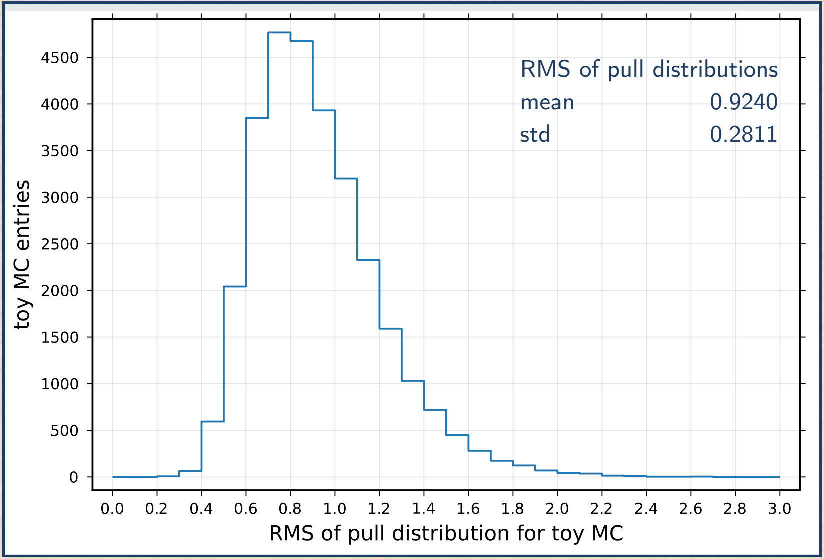
End

Backup Slides

Consistency of a_μ on datasets



Toy MC expected distribution of pulls standard deviation



Critical correlation

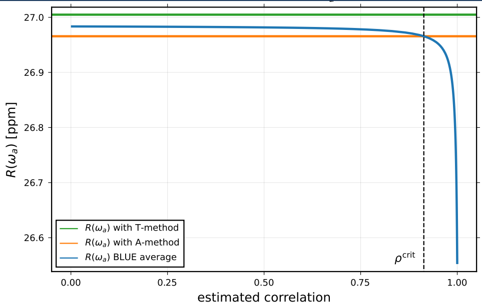
- when averaging two measurements A, B with $\sigma_B > \sigma_A$ the more uncertain measurement has weight
  - $w_B > 0$ when $\rho < \sigma_A/\sigma_B$
  - $w_B = 0$ when $\rho = \sigma_A/\sigma_B$ critical correlation
  - $w_B < 0$ when $\rho > \sigma_A/\sigma_B$

  [G. Cowan, Statistical data analysis, 1998]
- when two measurements have the same uncertainty, the critical correlation is 100%

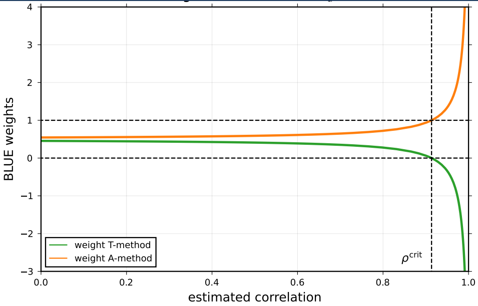
BLUE average vs. assumed correlation

- average of 1 A and 1 T method measurements
- T uncertainty 10% larger than A
- true correlation $\rho = \sigma_A/\sigma_T$, $\Rightarrow$ w_A=1, w_T=0
- unstable outcomes for assumed $\rho > \sigma_A/\sigma_T$

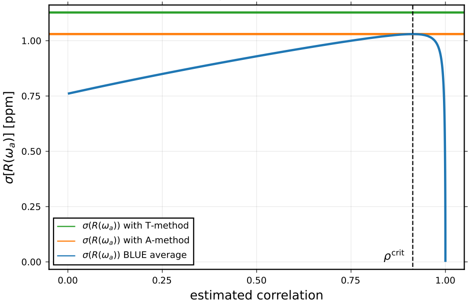
BLUE average value vs. estimated correlation



BLUE average weights



BLUE average uncertainty vs. est...d correlation



Averaging highly correlated measurements with imperfect correlation

- minimum χ^2 average of correlated measurement using imprecise correlation can be "'"un-physical"
  - especially when conservatively estimating fully correlated systematic uncertainties