Statistical Correlations Between Run 1 ω_a Analyses

Nicholas Kinnaird

Boston University – nickkinn@bu.edu

August 19, 2020

Abstract

This note presents calculated statistical correlation coefficients between the different Run 1 ω_a analyses. The correlation coefficients were calculated with a Monte Carlo simulation taking real data as input, for which the implementation is described. The statistical correlations presented here include the effects from the different reconstructions, different analysis methods (TARQ), and different analyzer-specific analysis parameters. Depending on the combination methodology adopted, these statistical correlations may be needed in order to properly combine the results. Various versions of the statistical correlations were provided to the combination effort, and results for the combination will be detailed by a forthcoming note. These statistical correlations are the most accurate to date for Run 1, and can be used as a starting point going forward for combinations of future ω_a analyses.

Contents

1 Introduction

The Run 1 ω_a analysis effort consisted of six independent groups conducting a total of eleven different analyses for each of the four Run 1 datasets[1]. These groups included Boston University (BU), Cornell University (CU), University of Washington (UW), a collaborative group from several European institutions under the moniker 'Europa' (EU), Shanghai Jiao Tong University (SJTU), and the University of Kentucky (UK). Each analysis was carried out independently of the rest. The groups from CU, UW, EU, and SJTU all performed T and A-Method analyses, while the group from BU performed T and R-Method analyses. The CU group utilized the ReconEast positron reconstruction, while the rest (barring the UK group) utilized the ReconWest reconstruction. The group from UK performed a Q-Method analysis, and therefore did not use either the ReconEast or ReconWest positron reconstructions. Details on the ReconEast and ReconWest positron reconstructions can be found in the theses of D. Sweigart [2] and A. Fienberg [3] respectively. Within each of the analyses, parameters were chosen according to preference or analysis specifics. These parameters correspond to energy thresholds for the various methods, binning parameters, and fit start and end times. Table 1 gives a summary of the different analyses for the different groups and the associated parameters[2]. References for the specifics of the analyses are included in the table.

The four datasets for the Run 1 analysis consisted of the 60h, HighKick, 9d, and Endgame datasets. These are sometimes abbreviated in different manners, such as 60h, HK, 9d, EG, or 1a, 1b, 1c, 1d respectively. The former is used at times in this document. The final, commonly blinded, best-fit R values for all eleven analyses of all four datasets are given in Table 2, where the equation

$$\omega_a = 2\pi \cdot 0.2291\,\mathrm{MHz} \cdot (1 + (R - \Delta R) \times 10^{-6}), \tag{1}$$

gives the relationship between ω_a and R, and ΔR is the common blinding offset. In order to provide a single ω_a number per dataset that will feed into the final determination of a_μ, these R values need to be combined in some way. Depending on the combination methodology adopted, correlation coefficients between the different analyses may be necessary in order to properly combine the results. These correlation coefficients will depend upon the specific analysis types, reconstructions, and parameters used in the various analyses, as given in Table 1. Preliminary versions of the correlation coefficients and results using one combination approach were determined by A. Keshavarzi in early 2020 [4]. The correlation coefficients therein were determined from a Monte Carlo simulation built from first principles. Since that work, a better Monte Carlo simulation has been developed which includes relevant information from real data and the individual analysis parameters. Using the updated Monte Carlo, the correlation coefficients were re-estimated. Those new correlation coefficients are presented here, along with details about the Monte Carlo.

[1] Reference [1] provides a nice readers guide and synopsis summary for the Run 1 analysis, and includes many references.

[2] Different pileup methods were used by the groups with varying parameters, however those details are not included here.

Run 1 Analyses and Parameters						
Analysis Effort	Cornell U.	U. Washington	Europa	Shanghai Jiao Tong U.	Boston U.	U. Kentucky
Abbreviation	CU	UW	EU	SJTU	BU	UK
Lead Analyzer	D. Sweigart	A. Fienberg	M. Sorbara	B. Li	N. Kinnaird	T. Gorringe
Analysis References	[2, 5]	[3, 6]	[7]	[8]	[9, 10]	[11]
Parameter						
Analysis Methods (TARQ)	TA	TA	TA	TA	TR	Q
Reconstruction	East	West	West	West	West	Q
Bin Width (ns)	149.2	149.19	149.19	149.2	149.2	150
Bin Edge (ns)	0	53.62	0	0	0	0
T-Method Threshold (MeV)	1700–	1700–6000	1680–7020	1700–9300	1700–	-
A-Method Threshold (MeV)	1000–3000	1000–3020	1080–3020	1000–3100	-	-
R-Method Threshold (MeV)	-	-	-	-	1700–	-
Q-Method Threshold (MeV)	-	-	-	-	-	300–
Fit Start Time (μs)	30.2876	30.19	30.1364	30.2876	30.2876	30
Fit Start Time (EndGame) (μs)	49.982	49.88308	49.8295	49.982	49.982	49.9762
Fit End Time (μs)	649.9152	649.92526	650.021	671.4	650.0644	215.5

Table 1: Individual analysis groups, the lead analyzers in those groups, and the methods and parameters in the respective analyses. Parameters provided via private communication and through discussion in the Run 1 Combination Task Force [12]. Note that the EndGame dataset had a different fit start time compared to the rest, hence the additional row in the table.

Run 1 Commonly Blinded Results								
Analysis	60h		HK		9d		EG	
	R	σ_R	R	σ_R	R	σ_R	R	σ_R
BU T	−28.8023	1.3582	−27.0442	1.1561	−27.9171	0.9301	−27.7020	0.7584
BU R	−28.9668	1.3598	−27.2093	1.1574	−27.9218	0.9327	−27.7654	0.7576
CU T	−28.2111	1.3377	−27.2093	1.1336	−28.0202	0.9126	−27.7152	0.7474
CU A	−28.2288	1.2079	−26.9466	1.0234	−27.5532	0.8240	−27.5902	0.6758
UW T	−28.6199	1.3308	−27.0049	1.1277	−27.8989	0.9079	−27.7144	0.7437
UW A	−28.6373	1.2184	−26.9657	1.0302	−27.5723	0.8305	−27.6694	0.6799
EU T	−28.8848	1.3327	−27.0806	1.1203	−27.8890	0.9067	−27.8772	0.7435
EU A	−28.4813	1.1938	−27.0213	1.0120	−27.5998	0.8146	−27.7276	0.6679
SJTU T	−28.7398	1.3314	−27.0019	1.1281	−27.8935	0.9084	−27.6658	0.7441
SJTU A	−28.4228	1.2061	−27.0910	1.0223	−27.7440	0.8224	−27.6945	0.6729
UK Q	−29.2062	2.0585	−24.9464	1.7478	−26.2794	1.4032	−27.9905	1.2690

Table 2: Best-fit results for R and the respective statistical errors for the different analyses of the Run 1 datasets, in units of ppm. R is commonly blinded, and contains both a software and hardware blinding offset.

2 Monte Carlo Details

In order to determine the correlation coefficients between the various analyses, a Monte Carlo simulation was developed which includes information from real data and incorporates the various analyzer parameters as given in Table 1. This Monte Carlo generated pseudo-data according to the different analysis types, reconstructions, and datasets, filled histograms defined by the various parameters, and then fit the resulting samples. The fit parameters from the samples were then plotted against one another and the correlation coefficients calculated for the different datasets.

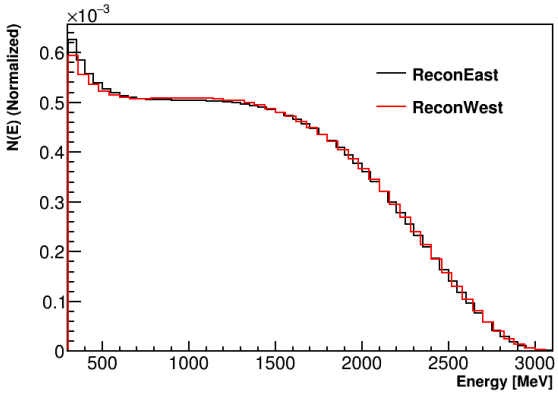
2.1 Generating Function

Pseudo-data was generated using `ROOT`'s `TF2->GetRandom2()` method. The 2D function used to generate the data is given by a five parameter function with energy dependence on the number, asymmetry, and phase terms,

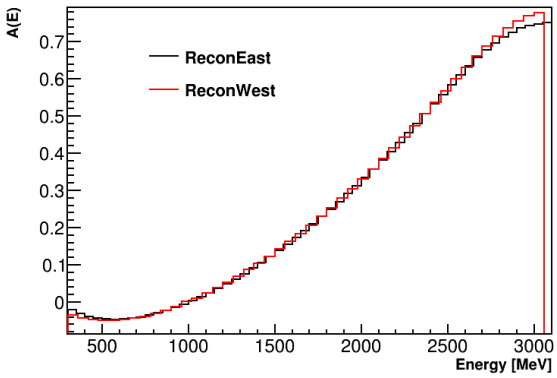
$$N(t, E) = N_0(E) \cdot e^{-t/\tau_\mu} \cdot (1 + A(E) \cos(\omega_a t + \phi(E))), \tag{2}$$

with the time-dilated muon lifetime τ_μ equal to 64 440 ns and the $g - 2$ frequency ω_a set as in Equation 1, with R set to 0.

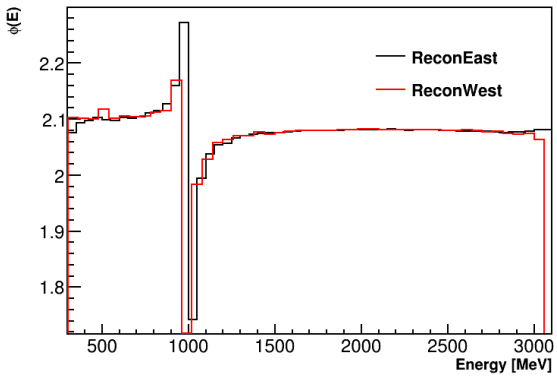
The energy dependent terms in Equation 2 were determined from energy-binned fits to the data provided by D. Sweigart for ReconEast and M. Sorbara for ReconWest reconstructions respectively, for each of the four datasets in Run 1. Figure 1 shows the respective histograms for the 9d dataset, where it can be seen that energy bin fits to the two reconstructions produce very similar histograms.



(a) $N(E)$



(b) $A(E)$



(c) $\phi(E)$

Figure 1: N, A, and ϕ as functions of energy from energy bin fits to ReconEast (black) and ReconWest (red) data for the 9d dataset. These histograms are representative of the quantities for all datasets, though there are slight differences. The ReconEast energy bin fits ranged from 300 MeV to 3100 MeV in bins of 50 MeV for a total of 56 energy bins, while the ReconWest energy bin fits ranged from 300 MeV to 3060 MeV in bins of 60 MeV for a total of 46 energy bins, hence the different bin edges in the plot. There are some slight differences at low and high energies within the various plots, but for the most part energy bin fits to the two reconstructions behave very similarly.

The ReconEast energy bin fits ranged from 300 MeV to 3100 MeV in bins of 50 MeV for a total of 56 energy bins, while the ReconWest energy bin fits ranged from 300 MeV to 3060 MeV in bins of 60 MeV for a total of 46 energy bins. Two `TF2`s were constructed using either the ReconEast or ReconWest histograms as input. The number of Y points in the 2D functions were set as the number of energy bins in the respective histograms, and the function was defined such that the energy binned histograms would be linearly interpolated between the values at the bin centers. The number of X points in the 2D functions were set such that the function ranged from 0 ns to 699 971.8 ns in steps of $149.2/2 = 74.6$ ns for a total of 9383 points. This 'point-width' was chosen for a few reasons. The first is that the maximum number of allowed X or Y points in a `TF2` is 10000, and this point-width results in a number of points close to but not above that value. The second is that the chosen point-width was either exactly or very close to a multiple of the bin widths chosen by the analyzers. It was found that using the maximum number of points such that the point-width was 70 ns (for a range up to 700 000 ns) resulted in aliasing frequencies appearing in the some of the FFTs of the fit residuals to the generated data. With a point-width of 74.6 ns these aliasing frequencies disappeared entirely except in some Q-Method fits. In order to remove this issue entirely a different implementation besides that described here would be needed, with finer granularity or better interpolation. The 2D function parameters are summarized in Table 3.

2D Function Parameters		
	ReconEast Input	ReconWest Input
Energy Range (MeV)	300–3100	300–3060
Energy Points	56	46
Energy Point-Width (MeV)	50	60
Time Range (ns)	0–699971.8	0–699971.8
Time Points	9383	9383
Time Point-Width (ns)	74.6	74.6

Table 3: Parameters used to define the `TF2`s in the Monte Carlo pseudo-data generation.

2.2 ReconEast vs ReconWest

The two 2D `TF2`s generate pseudo-data corresponding to either the ReconEast or the ReconWest reconstruction. In order to properly estimate correlation coefficients between reconstructions, generated hits need to be converted between the two. For this requirement, a comparison between ReconEast and ReconWest is necessary. J. LaBounty studied the different reconstructions of clusters in detail [13]. In doing so he determined that the cluster times between the two reconstructions were the same to high precision, but that there were differences in the reconstructed cluster energies. He took D. Sweigart's ReconEast and A. Fienberg's ReconWest analyses respectively, isolated clusters produced from the same waveforms, and compared the hit energies to each other. The calorimeter sum of this comparison for the 9d dataset is shown in Figure 2. As shown most hits lie along the unit slope line. Some hits lie in different bands due to reconstruction differences in how

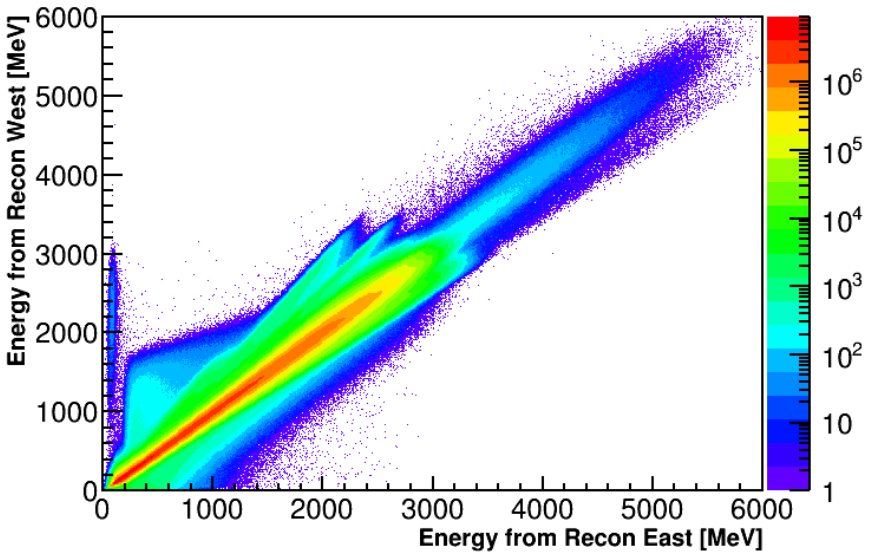


Figure 2: Log-scale plot of ReconEast energies vs ReconWest energies for the same produced clusters for the 9d dataset. The majority of hits form a Gaussian around the unit slope line, however there are bands of hits outside this region due to reconstruction differences in how certain hits and energies are treated. This pattern was determined to be stable throughout the fill. ReconEast vs ReconWest comparison performed by J. LaBounty [13].

certain hits and energies are treated[3]. Taking 1D projections of this distribution, like in Figure 3, energies can be converted between ReconEast and ReconWest by sampling those projections.

Before describing the event generation, a couple of points should be mentioned. First is that the bin width used by J. LaBounty was 10 MeV, discrete enough for the purposes of this Monte Carlo and also finer than the energy bin widths used in the generation of the `TF2`s. Second is that the comparison between reconstructions was built upon clusters before they had been corrected for pileup, hence the presence of counts at energies beyond the magic momentum of 3.094 GeV (plus the energy resolution). It is expected that this should be a fine approximation to use since pileup is a minimal effect after it has been corrected. Third is that this comparison was built upon D. Sweigart's and A. Fienberg's analyses. D. Sweigart is the only ReconEast analyzer, while A. Fienberg was one of several ReconWest analyzers. Since A. Fienberg used different clustering parameters than the rest of the ReconWest analyzers, (`timeCutoffLow = 2, timeCutoffHigh = 3` vs `timeCutoffLow = 3, timeCutoffHigh = 5`), the comparison isn't perfect. For the most technically correct results the ReconEast vs ReconWest comparison should be done individually for all analyzers, though the final results are not expected to change significantly.

[3]One example is that some individual crystals in some calorimeters have poor relative energy calibration between the two reconstructions.

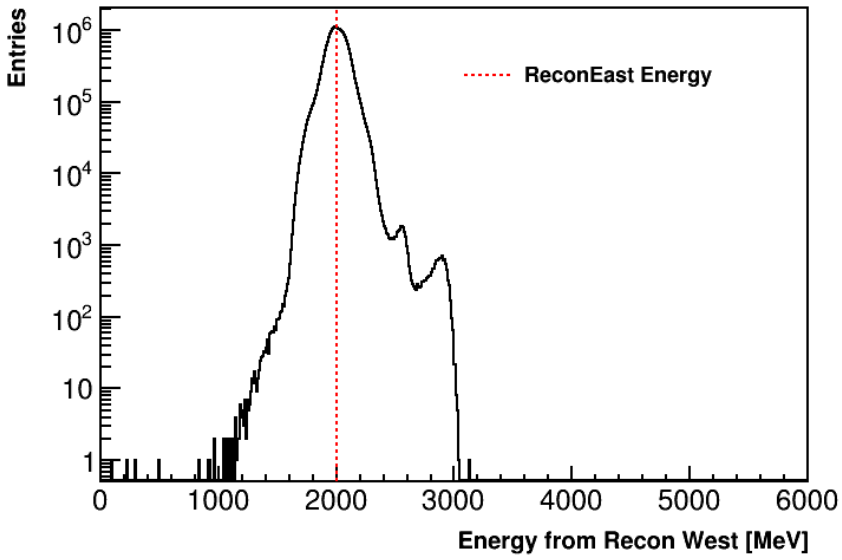


Figure 3: A Y-axis or ReconWest energy projection for a chosen ReconEast energy of 2 GeV. The plot is on a log-scale.

2.3 Event Generation and Randomization

Many samples of pseudo-data were generated via the same procedure. These samples were generated by submitting 1000 jobs each to the grid for the East-To-West and West-To-East variants[4], for each dataset. The number of jobs which completed varied only slightly, except for the EG dataset which had less completed jobs due to timing out on the grid; see Table 4.

	Jobs Completed	
Dataset	East-To-West	West-To-East
60h	1000	999
HK	1000	999
9d	1000	1000
EG	851	856

Table 4: Number of jobs completed out of 1000 after submission to the grid for the East-To-West and West-To-East variants. When calculating the statistical errors on the calculated correlation coefficients, the smaller of the two numbers corresponding to the larger error is used.

The number of hits per dataset was determined empirically such that the error on the fitted R values were comparable to the Run 1 results shown in Table 2. The approximate level of statistics

[4]'East-To-West' means that the 2D function with input from the ReconEast energy binned functions was used to generate the data, and vice versa.

for the 60h, HK, 9d, and EG datasets are given by the counts {5.04e9, 7.01e9, 1.08e10, 2.21e10} respectively. Note that these numbers depend on the energy and time ranges of the `TF2`s and hence do not correspond directly to the real dataset statistics; what was important and thus replicated was the number of hits which ended up in the wiggle plot within the fit ranges. The errors on the fitted R values were always within 10 ppb to 20 ppb of the actual real data analysis values (and many times within a few ppb).

Note that in the event generation no time-randomization was included, as is done by the analyzers in order to remove effects of the fast rotation. It is expected that in the Monte Carlo with no fast rotation effect, with enough random seeds, the mean fitted R value is equivalent to the non-time-randomized R value. While the spread in the R values will be slightly less without the time-randomization, this is expected to negligibly affect the final correlation coefficients. This effect could in the future be included at the expense of more processing time, where it is expected that the event generation would take significantly longer.

The procedure for the event generation was as follows:

1. A number of hits per sample and per dataset was determined via `PoissonD()` method calls on the static number of statistics per dataset such that each sample had a very slightly different amount of statistics.

2. For each hit or entry, `TF2->GetRandom2()` is called on one of the two 2D input functions, with either ReconEast or ReconWest input, to generate a time and energy for a hit.

3. For either variant, East-To-West or West-To-East, the corresponding ReconWest or ReconEast energy was determined by randomly sampling the corresponding energy projection as determined from J. LaBounty's studies. This projection was sampled via `TH1->GetRandom()` method calls. The first generated energy, designated either as ReconEast or ReconWest, was also designated as the Q-Method energy.

4. The generated time-energy pairs were then used to fill histograms defined by the analyzer parameters given in Table 1, with the corresponding weights set according to the specific analysis types. In the A-Method histograms, the asymmetry weights were determined from a linear interpolation of the asymmetry energy binned histogram, while in the Q-Method histograms the energies themselves were used as the weights.

5. The R-Method has intrinsic randomization in how the counts get put into the four ratio histograms, and so ratio histograms were produced for 100 different random seeds. The randomization of counts was performed via `GetUniform()` method calls in order to properly distribute the data.

In all instances of the randomization, including the randomization on the number of events, sampling the `TF2`s, sampling the ReconEast vs ReconWest energy projections, and the ratio randomization, the `ROOT TRandomMixMax` class is used. The standard `TRandom3` class was initially used and found to be inadequate for this high precision statistical study. `TRandomMixMax` has the advantage of producing better random numbers with high periodicity, and also being fast enough to produce the high statistics in a reasonable time frame[5][14].

[5]This generator is the default used in Geant4.

2.4 Fitting the Pseudo-Data

Once the pseudo-data had been generated, the data was fit either with a simple five parameter function in the TAQ-Methods,

$$N(t) = N_0 \cdot e^{-t/\tau_\mu} \cdot (1 + A\cos(\omega_a t + \phi)), \tag{3}$$

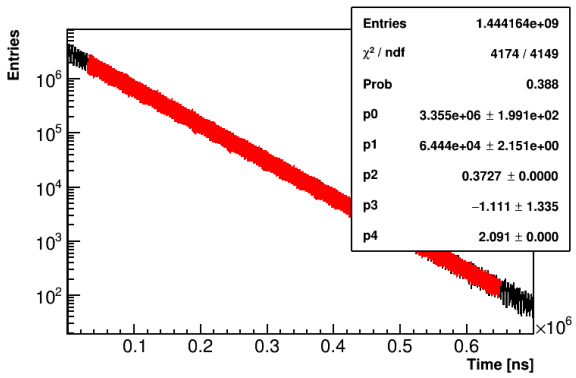
or a simple three parameter function in the case of the R-Method,

$$R(t) = A\cos(\omega_a t + \phi), \tag{4}$$

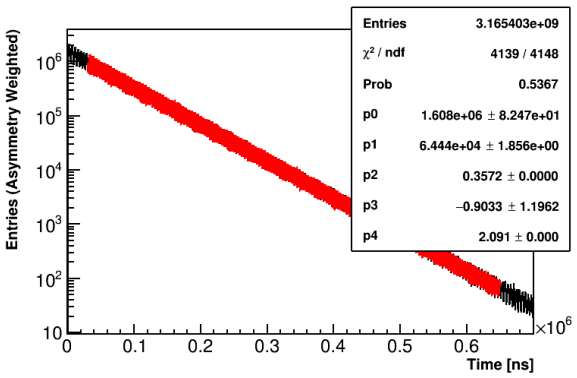
with fit ranges defined by the associated parameters in Table 1. In both cases the energy dependent pieces from Equation 2 are gone, and the fitted constants in general converge to their average integrated values. Figure 4 shows examples of fits to the four different methods, for the same generated sample. In general the fit parameters are very similar to one another between the different methods, however due to the different analyzer specific parameters like energy threshold, the fit parameters can be slightly, but significantly, different. Figure 5 shows the FFTs of the fit residuals from Figure 4. What sometimes occurred, though isn't shown here, is that aliasing peaks can appear at beat frequencies with ω_a, due to differences in the point-width frequency of the input 2D function and the bin width of the histogram.

Once all samples and seeds were fitted, the fit parameters were stored into `ROOT` trees such that the parameters for the different methods and samples could be compared. Similarly, measures of the goodness-of-fit were also stored, in order to verify that the pseudo-data was being fit correctly. P value distributions for the four methods for one set of samples is shown in Figure 6. As shown the distributions are flat for the T and A-Methods, and are not flat for the R and Q-Methods. The Q-Method is understood to be non-flat due to the difference in bin width and point-width mentioned previously, while the origin of the non-flatness of the R-Method is unknown at this time[6]. Figure 7 show the pulls on R for the different samples in the same set of fitted samples. The distributions in general are very close to unit-Gaussians, with some slight deviations in the R and Q-Method cases leading to slight pulls on R. This should have a very minimal effect on the calculated correlation coefficients, but it should be kept in mind.

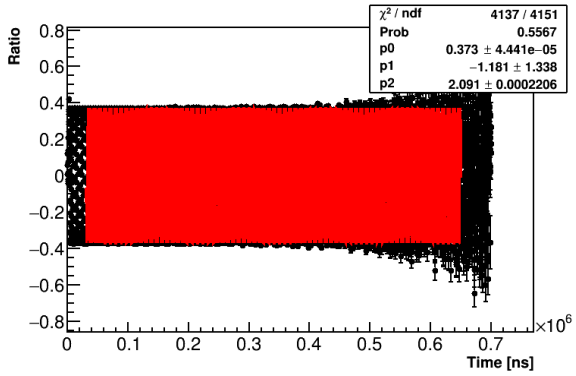
[6]It may be that the random generator class `TRandomMixMax` is insufficient for the ratio method randomization, which performs many more randomization calls than the other methods.



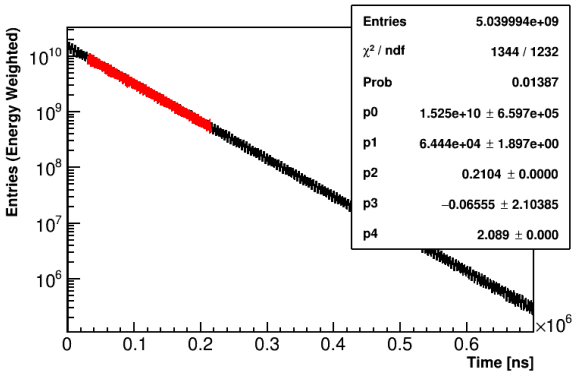
(a) T-Method



(b) A-Method

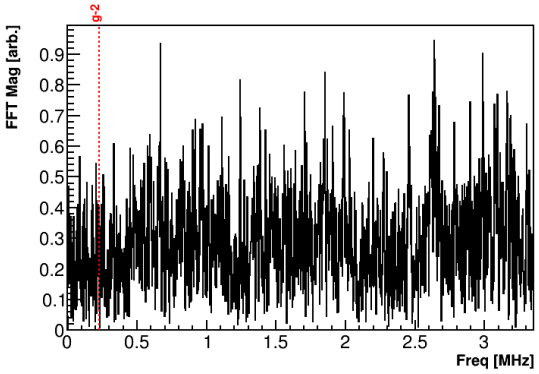


(c) R-Method

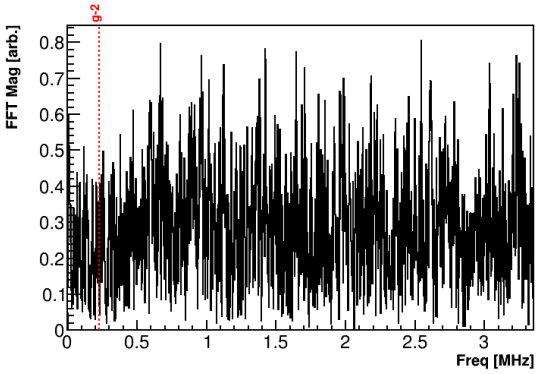


(d) Q-Method

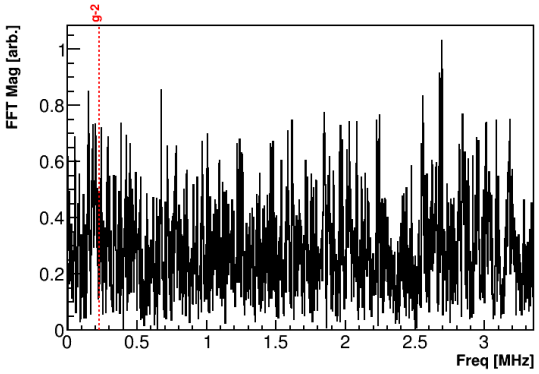
Figure 4: TARQ-Method fits to a sample generated from the 60h East-To-West Monte Carlo data.



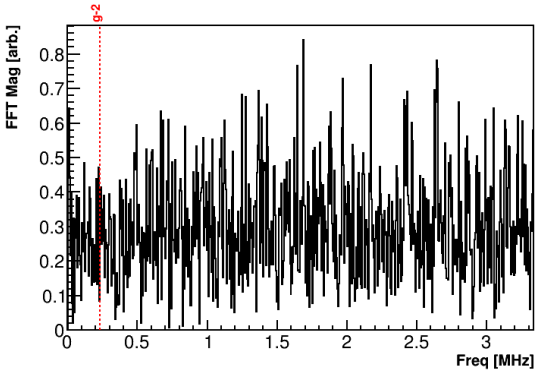
(a) T-Method



(b) A-Method

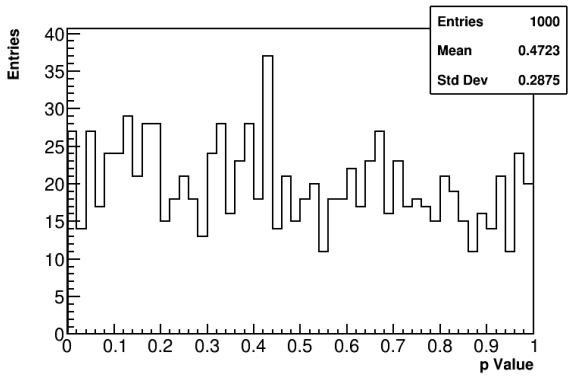


(c) R-Method

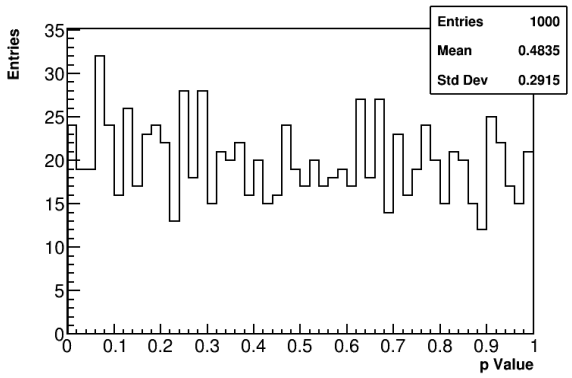


(d) Q-Method

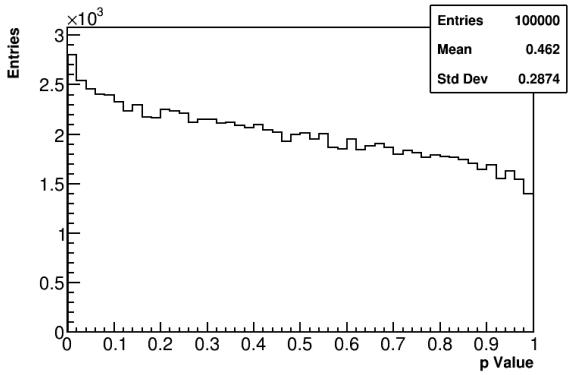
Figure 5: FFTs of the fit residuals for those fits shown in Figure 4. No peaks rise above the noise, though in some Q-Method fits an aliasing frequency between $g-2$ and the point-width of the generating function can be seen.



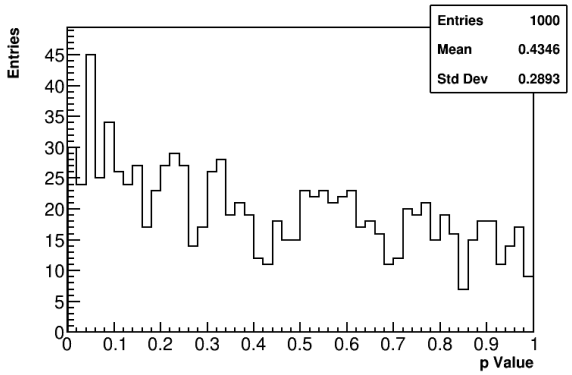
(a) T-Method



(b) A-Method

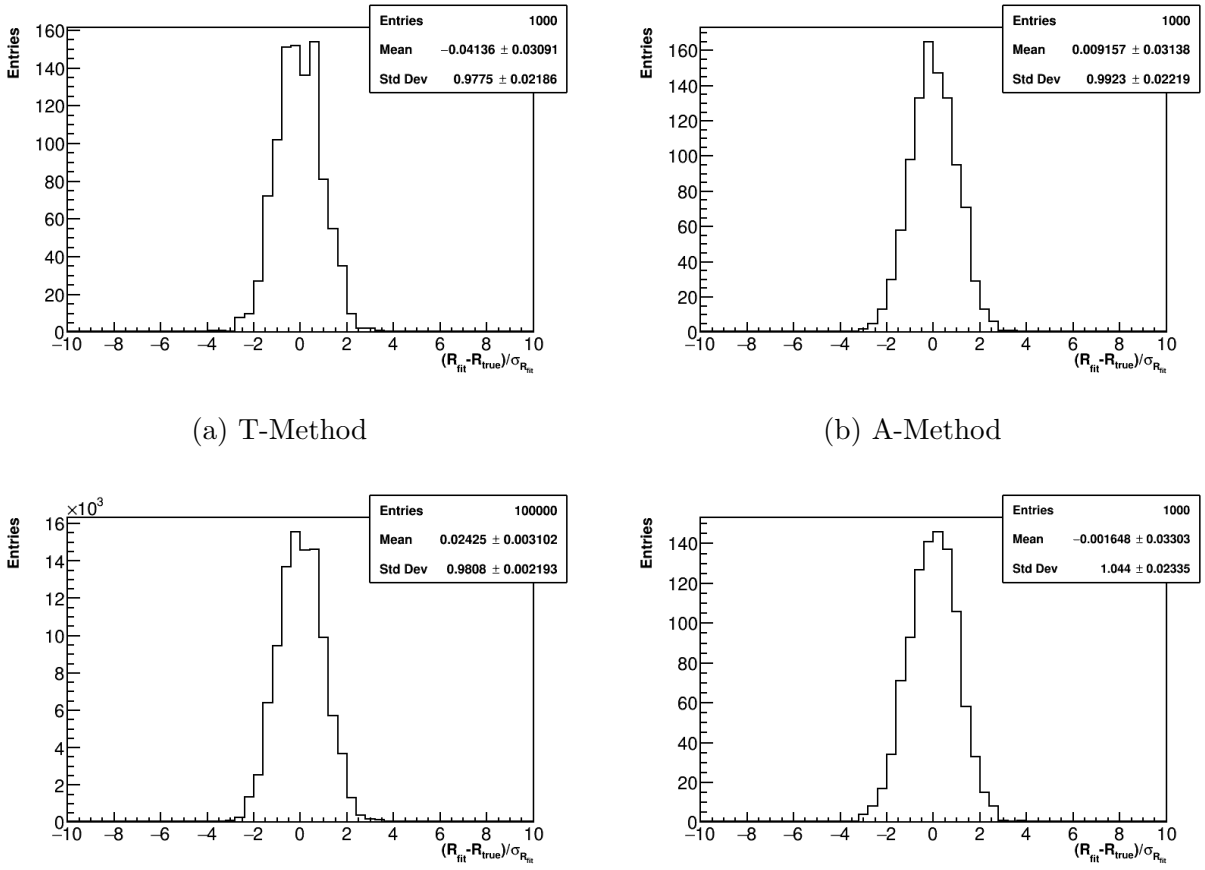


(c) R-Method



(d) Q-Method

Figure 6: P value distributions for fits to 1000 samples, for the 60h East-To-West Monte Carlo pseudo-data. The T and A-Method p values are flat, while the R and Q-Method p values are not. In the case of the Q-Method, the non-flat shape can be attributed to the difference in bin width and 2D function point-width. The origin of the non-flat shape for the R-Method is unknown at this time.



(a) T-Method

(b) A-Method

(c) R-Method

(d) Q-Method

Figure 7: Pull distributions for fits to 1000 samples, for the 60h East-To-West Monte Carlo pseudo-data. The various distributions are close to unit-Gaussians, with differences in some R and Q-Method cases which is attributed to imperfect data-generation.

2.5 Calculating the Correlations and Errors

Once the generated pseudo-data was fitted, the R values and other parameters were plotted against one another for different analyses and methods to determine the correlation coefficients. An example of such a scatter plot is shown in Figure 8. The final correlation coefficients are calculated as the average of the correlation coefficients produced with the East-To-West and West-To-East variants. The systematic errors on the coefficients are determined as the difference between the average correlation coefficients and either of the two variants. By construction this systematic error will include it's own statistical error since the two variants come from different grid job submissions, however the extra error is simply conservatively included in the systematic number.

The statistical error on the determined correlation coefficients is dependent on the number of points or samples used to calculate the coefficients as well as the coefficients themselves. These statistical errors are calculated via the Fisher-z transform of the correlation coefficients r,

$$z = \tanh^{-1} r, \tag{5}$$

where the variance on z goes as

$$V(z) = \frac{1}{n-3}, \tag{6}$$

and n is the number of samples. By converting the correlation coefficients r to z, finding the upper and lower ranges when adding or subtracting the square root of the variance, and then converting back to r, the statistical errors are determined. This is described more fully in Section 9.5 of G. Cowan's *Statistical Data Analysis* [15]. The positive and negative statistical errors as a function of the correlation for 851 samples is shown in Figure 9. As shown the statistical error decreases as the correlation increases. To quote a single statistical error for each correlation coefficient, the average of the positive and negative errors was taken. As shown the error is $\mathcal{O}(10^{-2} - 10^{-4})$ depending on the correlation coefficient. At correlation coefficients of 90%, 95%, and 99% the statistical errors are approximately 0.6%, 0.3% and less than 0.1% respectively. The statistical errors are sometimes comparable to the systematic errors, and sometimes noticeably less. The total error is then the quadrature sum of the statistical and systematic errors as usual.

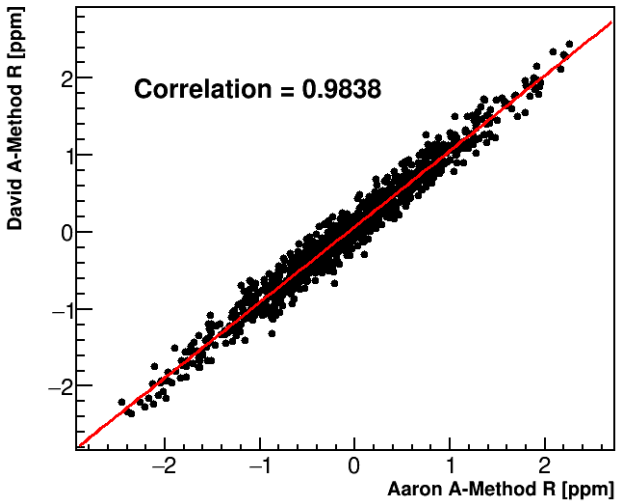


Figure 8: Scatter plot between R values for two different analyses, from jobs submitted for the 9d dataset East-To-West variant. The correlation coefficient between these two analyses is determined from this scatter plot, and the red line simply guides the eye.

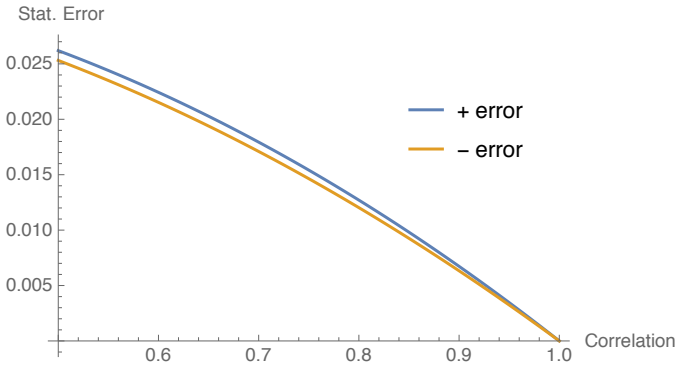


Figure 9: Positive and negative statistical errors on the correlation coefficient as a function of the correlation, for 851 samples.

3 Results

Correlation coefficients were calculated among all analyzers, all analyses, for all datasets, and between all parameters. The correlation coefficients between the R values are presented here, since they are of the most importance[7]. The correlation matrix between different analyzers for the EG dataset is shown in Figure 10. In one of the proposed combination approaches, a staged averaging method [12, 16], the correlation coefficients between the different reconstructions are needed. In order to determine these the R values between the ReconWest T and A-Methods were averaged among the four and three different analyses respectively, before populating the scatter plots and determining the correlation coefficients[8]. The correlation matrix for the EG dataset at this reconstruction level is shown in Figure 11. Tables for the correlation coefficients at the analyzer level and their errors for all four datasets are given in Tables 5 through 8. Tables for the correlation coefficients at the reconstruction level and their errors for all four datasets are given in Tables 9 through 12.

There are quite a few numbers listed throughout the tables. In order to facilitate the processing of all the information, some general summary points are as follows:

- Correlations across the four datasets are very similar, barring some slight differences in random table entries which are usually within error of each other. An exception is for the Q-Method correlations for the 60h dataset which are around 2-3% higher than in the rest of the datasets.
- The T-A-Methods correlation is typically around 90%, T-Q around 50%, T-R around 99.6%, A-R around 90%, and A-Q around 57%, regardless of reconstruction.
- Errors on the analyzer level correlation coefficients are typically around 0.5% depending on the specific correlation. In general the larger the correlation the smaller the error, just as dictated by the statistical error. Some exceptions include the errors on the Q-Method correlations which are at the 2–3% level, and the errors on the T-R-Method correlations which are typically sub 0.1%. In general most of the errors are at the same order as the pure statistical errors, though there are some cases where the total error is around 2 times larger.
- The T-Method correlations between ReconWest analyzers is typically around 98–99%, the A-Method correlations between ReconWest analyzers is typically around 99%.
- The ReconEast-ReconWest correlations are around 95% for the T-Methods, and around 98.9% for the A-Methods. The errors on these correlations are around 0.4 and 0.1% respectively.

As described in D. Sweigart's collaboration talk [16], two analyses enter the "high-correlation" regime when

$$r_{ij} \geq \sigma_i/\sigma_j, \tag{7}$$

[7]Correlation coefficients between other fit parameters are not particularly useful. One case of interest is the R $-\phi$ correlations for a potential multi-parameter combination.

[8]It should be noted as well that other combinations of various methods were made, one example being a weighted average of ReconWest AR-Methods, in order to provide a variety of correlation coefficients for corresponding types of combinations.

where r_{ij} is the correlation coefficient between two analyses i and j, σ_i and σ_j are the errors of the two analyses, and $\sigma_i < \sigma_j$. In this regime, the χ^2 combination approach (detailed in Reference [16]) entirely favors analysis j and the combination result can even be outside the individual values of the two analyses. When comparing the correlation coefficients to these cut-offs, something like a third of so of the correlation coefficients are beyond this regime. The amount beyond the cut-off for these correlations is anywhere from small fractions of a percent to a few percent. In a good amount of these cases, this difference is within the calculated error on the coefficients. In other cases the difference cannot solely be explained by the error. If the Monte Carlo was made more like real data, with systematic effects included, then the correlations would decrease and perhaps all coefficients would drop below the cut-off (with high enough stats), though it is not known for sure.

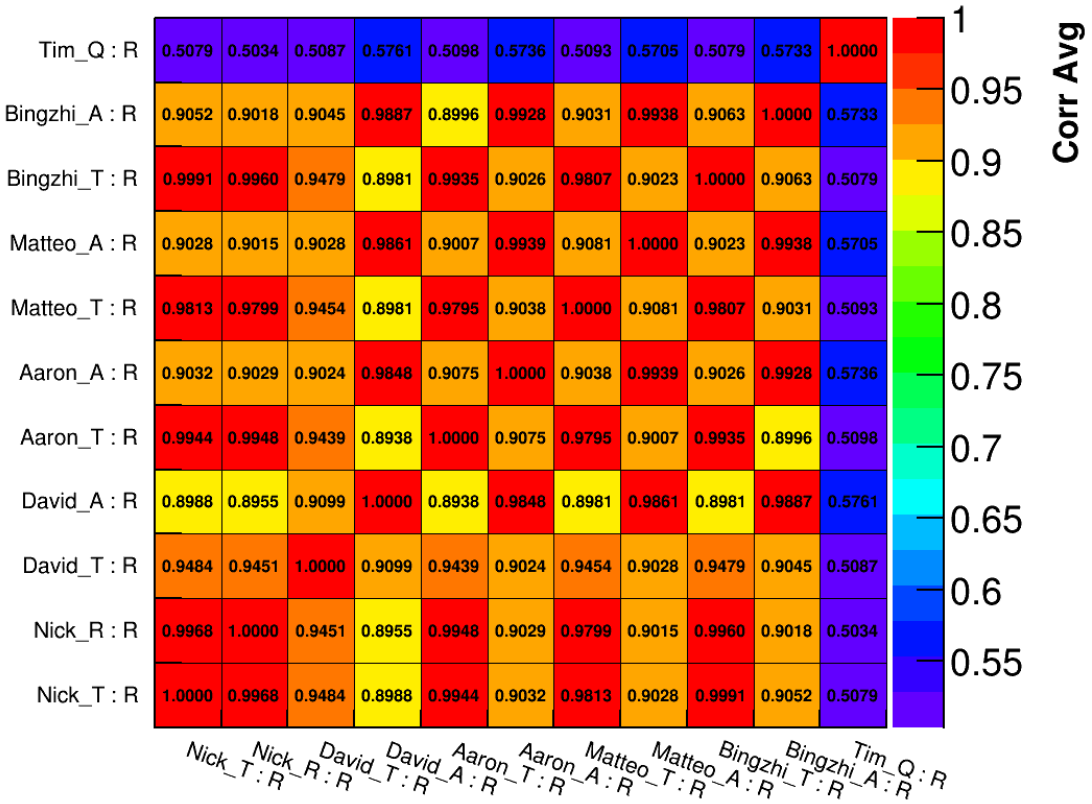


Figure 10: Correlation coefficient matrix for the R values between different analyzers and methods for the EG dataset. Along the axes are the first names of the analyzers, the method used to fit the data, and then the fit parameter after the colon. The names here {Nick, David, Aaron, Matteo, Bingzhi, Tim} are the first names of the analyzers for the BU, CU, UW, EU, SJTU, and UK groups respectively. The Z axis label "Corr Avg" stands for the fact that the correlations are averages of the East-To-West and West-To-East variants of the Monte Carlo.

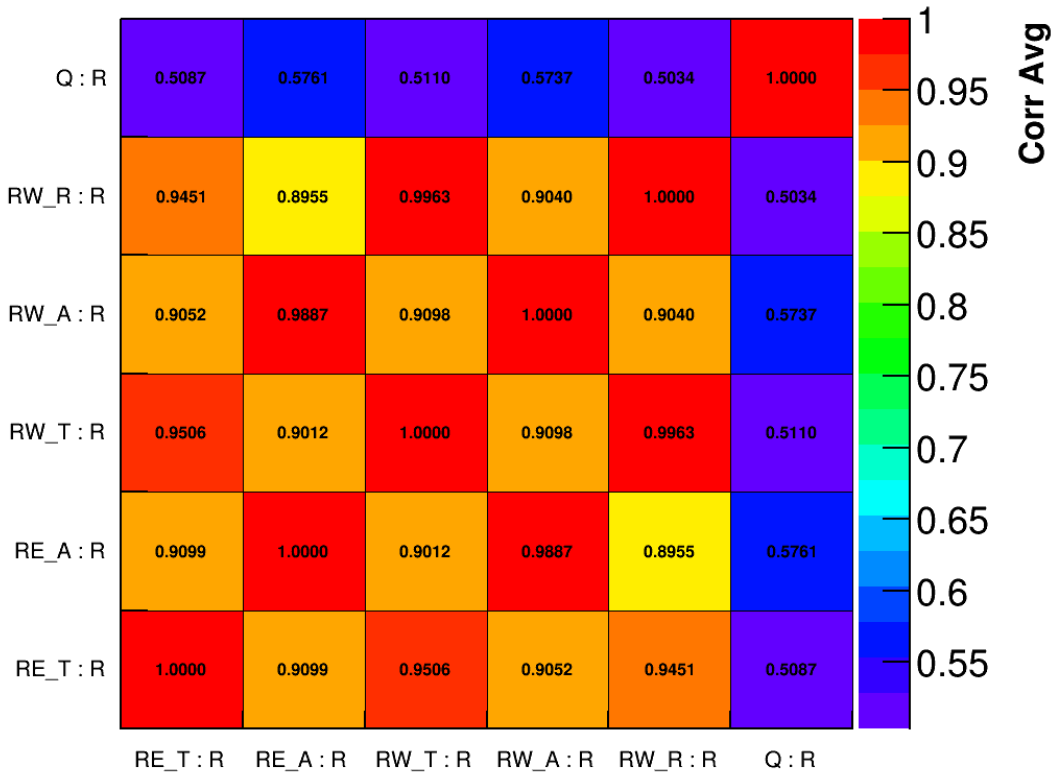


Figure 11: Correlation coefficient matrix between different reconstructions and methods for the EG dataset. Along the axes are the reconstructions, the method used to fit the data, and then the fit parameter after the colon. The ReconEast, R-Method, and Q-Method analyses consist of the single respective analyses. The ReconWest T and A-Method analysis results were averaged among the four and three different analyses respectively before calculating the coefficients. The Z axis label "Corr Avg" stands for the fact that the correlations are averages of the East-To-West and West-To-East variants of the Monte Carlo.

60h Correlation Coefficients – Analyzer Level											
	BU T	BU R	CU T	CU A	UW T	UW A	EU T	EU A	SJTU T	SJTU A	UK Q
BU T	1.0000	0.9964	0.9445	0.8992	0.9935	0.8977	0.9783	0.9014	0.9992	0.9050	0.5279
	0.0000	0.0002	0.0037	0.0063	0.0004	0.0062	0.0014	0.0059	0.0001	0.0057	0.0309
BU R	0.9964	1.0000	0.9418	0.8969	0.9943	0.8987	0.9776	0.9008	0.9956	0.9023	0.5256
	0.0002	0.0000	0.0040	0.0065	0.0004	0.0061	0.0015	0.0060	0.0003	0.0060	0.0306
CU T	0.9445	0.9418	1.0000	0.9043	0.9395	0.8933	0.9399	0.8963	0.9436	0.8992	0.5248
	0.0037	0.0040	0.0000	0.0059	0.0039	0.0068	0.0042	0.0066	0.0038	0.0064	0.0289
CU A	0.8992	0.8969	0.9043	1.0000	0.8972	0.9840	0.8999	0.9855	0.8985	0.9892	0.5926
	0.0063	0.0065	0.0059	0.0000	0.0062	0.0010	0.0064	0.0010	0.0065	0.0008	0.0306
UW T	0.9935	0.9943	0.9395	0.8972	1.0000	0.9065	0.9771	0.9025	0.9928	0.9022	0.5324
	0.0004	0.0004	0.0039	0.0062	0.0000	0.0057	0.0014	0.0060	0.0005	0.0060	0.0322
UW A	0.8977	0.8987	0.8933	0.9840	0.9065	1.0000	0.9016	0.9933	0.8971	0.9916	0.5927
	0.0062	0.0061	0.0068	0.0010	0.0057	0.0000	0.0059	0.0005	0.0062	0.0006	0.0364
EU T	0.9783	0.9776	0.9399	0.8999	0.9771	0.9016	1.0000	0.9091	0.9775	0.9041	0.5406
	0.0014	0.0015	0.0042	0.0064	0.0014	0.0059	0.0000	0.0055	0.0014	0.0058	0.0345
EU A	0.9014	0.9008	0.8963	0.9855	0.9025	0.9933	0.9091	1.0000	0.9006	0.9934	0.5928
	0.0059	0.0060	0.0066	0.0010	0.0060	0.0005	0.0055	0.0000	0.0060	0.0004	0.0369
SJTU T	0.9992	0.9956	0.9436	0.8985	0.9928	0.8971	0.9775	0.9006	1.0000	0.9058	0.5271
	0.0001	0.0003	0.0038	0.0065	0.0005	0.0062	0.0014	0.0060	0.0000	0.0057	0.0305
SJTU A	0.9050	0.9023	0.8992	0.9892	0.9022	0.9916	0.9041	0.9934	0.9058	1.0000	0.5893
	0.0057	0.0060	0.0064	0.0008	0.0060	0.0006	0.0058	0.0004	0.0057	0.0000	0.0343
UK Q	0.5279	0.5256	0.5248	0.5926	0.5324	0.5927	0.5406	0.5928	0.5271	0.5893	1.0000
	0.0309	0.0306	0.0289	0.0306	0.0322	0.0364	0.0345	0.0369	0.0305	0.0343	0.0000

Table 5: Correlation coefficients between R values for the 60h dataset, at the analyzer level. In each table cell, the top number is the correlation coefficient and the bottom number is the error on the coefficient.

HK Correlation Coefficients – Analyzer Level											
	BU T	BU R	CU T	CU A	UW T	UW A	EU T	EU A	SJTU T	SJTU A	UK Q
BU T	1.0000 0.0000	0.9967 0.0002	0.9469 0.0040	0.8924 0.0077	0.9939 0.0005	0.8938 0.0065	0.9799 0.0017	0.8946 0.0071	0.9992 0.0001	0.8978 0.0070	0.4982 0.0245
BU R	0.9967 0.0002	1.0000 0.0000	0.9439 0.0042	0.8907 0.0081	0.9943 0.0005	0.8955 0.0065	0.9785 0.0017	0.8948 0.0072	0.9958 0.0003	0.8959 0.0072	0.4959 0.0249
CU T	0.9469 0.0040	0.9439 0.0042	1.0000 0.0000	0.8968 0.0068	0.9408 0.0046	0.8858 0.0068	0.9409 0.0038	0.8865 0.0068	0.9458 0.0042	0.8900 0.0067	0.4913 0.0266
CU A	0.8924 0.0077	0.8907 0.0081	0.8968 0.0068	1.0000 0.0000	0.8872 0.0089	0.9839 0.0014	0.8900 0.0071	0.9846 0.0011	0.8913 0.0079	0.9892 0.0008	0.5635 0.0256
UW T	0.9939 0.0005	0.9943 0.0005	0.9408 0.0046	0.8872 0.0089	1.0000 0.0000	0.8993 0.0064	0.9776 0.0025	0.8928 0.0077	0.9930 0.0005	0.8920 0.0080	0.5015 0.0238
UW A	0.8938 0.0065	0.8955 0.0065	0.8858 0.0068	0.9839 0.0014	0.8993 0.0064	1.0000 0.0000	0.8943 0.0064	0.9935 0.0007	0.8928 0.0067	0.9918 0.0007	0.5642 0.0245
EU T	0.9799 0.0017	0.9785 0.0017	0.9409 0.0038	0.8900 0.0071	0.9776 0.0025	0.8943 0.0064	1.0000 0.0000	0.8998 0.0062	0.9791 0.0018	0.8949 0.0066	0.5066 0.0236
EU A	0.8946 0.0071	0.8948 0.0072	0.8865 0.0068	0.9846 0.0011	0.8928 0.0077	0.9935 0.0007	0.8998 0.0062	1.0000 0.0000	0.8936 0.0073	0.9935 0.0004	0.5637 0.0247
SJTU T	0.9992 0.0001	0.9958 0.0003	0.9458 0.0042	0.8913 0.0079	0.9930 0.0005	0.8928 0.0067	0.9791 0.0018	0.8936 0.0073	1.0000 0.0000	0.8984 0.0071	0.4987 0.0245
SJTU A	0.8978 0.0070	0.8959 0.0072	0.8900 0.0067	0.9892 0.0008	0.8920 0.0080	0.9918 0.0007	0.8949 0.0066	0.9935 0.0004	0.8984 0.0071	1.0000 0.0000	0.5612 0.0259
UK Q	0.4982 0.0245	0.4959 0.0249	0.4913 0.0266	0.5635 0.0256	0.5015 0.0238	0.5642 0.0245	0.5066 0.0236	0.5637 0.0247	0.4987 0.0245	0.5612 0.0259	1.0000 0.0000

Table 6: Correlation coefficients between R values for the HK dataset, at the analyzer level. In each table cell, the top number is the correlation coefficient and the bottom number is the error on the coefficient.

9d Correlation Coefficients – Analyzer Level											
	BU T	BU R	CU T	CU A	UW T	UW A	EU T	EU A	SJTU T	SJTU A	UK Q
BU T	1.0000 0.0000	0.9965 0.0004	0.9445 0.0037	0.8912 0.0109	0.9939 0.0006	0.8935 0.0086	0.9787 0.0023	0.8952 0.0103	0.9993 0.0001	0.8983 0.0094	0.4936 0.0261
BU R	0.9965 0.0004	1.0000 0.0000	0.9414 0.0040	0.8885 0.0118	0.9944 0.0005	0.8944 0.0088	0.9782 0.0017	0.8945 0.0105	0.9958 0.0004	0.8957 0.0099	0.4888 0.0265
CU T	0.9445 0.0037	0.9414 0.0040	1.0000 0.0000	0.9002 0.0064	0.9392 0.0045	0.8905 0.0066	0.9419 0.0037	0.8906 0.0068	0.9438 0.0039	0.8944 0.0066	0.5000 0.0248
CU A	0.8912 0.0109	0.8885 0.0118	0.9002 0.0064	1.0000 0.0000	0.8861 0.0122	0.9833 0.0011	0.8904 0.0088	0.9840 0.0011	0.8903 0.0111	0.9886 0.0009	0.5710 0.0213
UW T	0.9939 0.0006	0.9944 0.0005	0.9392 0.0045	0.8861 0.0122	1.0000 0.0000	0.8990 0.0091	0.9770 0.0018	0.8931 0.0117	0.9931 0.0007	0.8929 0.0109	0.4999 0.0270
UW A	0.8935 0.0086	0.8944 0.0088	0.8905 0.0066	0.9833 0.0011	0.8990 0.0091	1.0000 0.0000	0.8966 0.0070	0.9932 0.0005	0.8926 0.0088	0.9923 0.0005	0.5750 0.0214
EU T	0.9787 0.0023	0.9782 0.0017	0.9419 0.0037	0.8904 0.0088	0.9770 0.0018	0.8966 0.0070	1.0000 0.0000	0.9021 0.0082	0.9778 0.0023	0.8974 0.0081	0.4992 0.0241
EU A	0.8952 0.0103	0.8945 0.0105	0.8906 0.0068	0.9840 0.0011	0.8931 0.0117	0.9932 0.0005	0.9021 0.0082	1.0000 0.0000	0.8944 0.0104	0.9940 0.0004	0.5698 0.0214
SJTU T	0.9993 0.0001	0.9958 0.0004	0.9438 0.0039	0.8903 0.0111	0.9931 0.0007	0.8926 0.0088	0.9778 0.0023	0.8944 0.0104	1.0000 0.0000	0.8988 0.0095	0.4913 0.0263
SJTU A	0.8983 0.0094	0.8957 0.0099	0.8944 0.0066	0.9886 0.0009	0.8929 0.0109	0.9923 0.0005	0.8974 0.0081	0.9940 0.0004	0.8988 0.0095	1.0000 0.0000	0.5651 0.0216
UK Q	0.4936 0.0261	0.4888 0.0265	0.5000 0.0248	0.5710 0.0213	0.4999 0.0270	0.5750 0.0214	0.4992 0.0241	0.5698 0.0214	0.4913 0.0263	0.5651 0.0216	1.0000 0.0000

Table 7: Correlation coefficients between R values for the 9d dataset, at the analyzer level. In each table cell, the top number is the correlation coefficient and the bottom number is the error on the coefficient.

EG Correlation Coefficients – Analyzer Level											
	BU T	BU R	CU T	CU A	UW T	UW A	EU T	EU A	SJTU T	SJTU A	UK Q
BU T	1.0000	0.9968	0.9484	0.8988	0.9944	0.9032	0.9813	0.9028	0.9991	0.9052	0.5079
	0.0000	0.0002	0.0040	0.0113	0.0006	0.0097	0.0013	0.0107	0.0001	0.0108	0.0275
BU R	0.9968	1.0000	0.9451	0.8955	0.9948	0.9029	0.9799	0.9015	0.9960	0.9018	0.5034
	0.0002	0.0000	0.0045	0.0122	0.0007	0.0107	0.0015	0.0115	0.0003	0.0118	0.0275
CU T	0.9484	0.9451	1.0000	0.9099	0.9439	0.9024	0.9454	0.9028	0.9479	0.9045	0.5087
	0.0040	0.0045	0.0000	0.0064	0.0042	0.0064	0.0049	0.0064	0.0038	0.0063	0.0270
CU A	0.8988	0.8955	0.9099	1.0000	0.8938	0.9848	0.8981	0.9861	0.8981	0.9887	0.5761
	0.0113	0.0122	0.0064	0.0000	0.0121	0.0011	0.0124	0.0010	0.0113	0.0010	0.0230
UW T	0.9944	0.9948	0.9439	0.8938	1.0000	0.9075	0.9795	0.9007	0.9935	0.8996	0.5098
	0.0006	0.0007	0.0042	0.0121	0.0000	0.0099	0.0017	0.0115	0.0007	0.0120	0.0274
UW A	0.9032	0.9029	0.9024	0.9848	0.9075	1.0000	0.9038	0.9939	0.9026	0.9928	0.5736
	0.0097	0.0107	0.0064	0.0011	0.0099	0.0000	0.0108	0.0006	0.0096	0.0006	0.0233
EU T	0.9813	0.9799	0.9454	0.8981	0.9795	0.9038	1.0000	0.9081	0.9807	0.9031	0.5093
	0.0013	0.0015	0.0049	0.0124	0.0017	0.0108	0.0000	0.0113	0.0014	0.0115	0.0267
EU A	0.9028	0.9015	0.9028	0.9861	0.9007	0.9939	0.9081	1.0000	0.9023	0.9938	0.5705
	0.0107	0.0115	0.0064	0.0010	0.0115	0.0006	0.0113	0.0000	0.0107	0.0005	0.0234
SJTU T	0.9991	0.9960	0.9479	0.8981	0.9935	0.9026	0.9807	0.9023	1.0000	0.9063	0.5079
	0.0001	0.0003	0.0038	0.0113	0.0007	0.0096	0.0014	0.0107	0.0000	0.0106	0.0278
SJTU A	0.9052	0.9018	0.9045	0.9887	0.8996	0.9928	0.9031	0.9938	0.9063	1.0000	0.5733
	0.0108	0.0118	0.0063	0.0010	0.0120	0.0006	0.0115	0.0005	0.0106	0.0000	0.0231
UK Q	0.5079	0.5034	0.5087	0.5761	0.5098	0.5736	0.5093	0.5705	0.5079	0.5733	1.0000
	0.0275	0.0275	0.0270	0.0230	0.0274	0.0233	0.0267	0.0234	0.0278	0.0231	0.0000

Table 8: Correlation coefficients between R values for the EG dataset, at the analyzer level. In each table cell, the top number is the correlation coefficient and the bottom number is the error on the coefficient.

60h Correlation Coefficients – Recon. Level						
	RE T	RE A	RW T	RW A	RW R	Q
RE T	1.0000	0.9043	0.9467	0.8984	0.9418	0.5248
	0.0000	0.0059	0.0036	0.0065	0.0040	0.0289
RE A	0.9043	1.0000	0.9033	0.9886	0.8969	0.5926
	0.0059	0.0000	0.0060	0.0008	0.0065	0.0306
RW T	0.9467	0.9033	1.0000	0.9097	0.9961	0.5348
	0.0036	0.0060	0.0000	0.0055	0.0003	0.0320
RW A	0.8984	0.9886	0.9097	1.0000	0.9028	0.5930
	0.0065	0.0008	0.0055	0.0000	0.0059	0.0359
RW R	0.9418	0.8969	0.9961	0.9028	1.0000	0.5256
	0.0040	0.0065	0.0003	0.0059	0.0000	0.0306
Q	0.5248	0.5926	0.5348	0.5930	0.5256	1.0000
	0.0289	0.0306	0.0320	0.0359	0.0306	0.0000

Table 9: Correlation coefficients between R values for the 60h dataset, at the reconstruction level, after the ReconWest T-Method and A-Method R values were averaged among the different analyzers. In each table cell, the top number is the correlation coefficient and the bottom number is the error on the coefficient.

HK Correlation Coefficients – Recon. Level						
	RE T	RE A	RW T	RW A	RW R	Q
RE T	1.0000	0.8968	0.9482	0.8896	0.9439	0.4913
	0.0000	0.0068	0.0037	0.0066	0.0042	0.0266
RE A	0.8968	1.0000	0.8946	0.9883	0.8907	0.5635
	0.0068	0.0000	0.0075	0.0009	0.0081	0.0256
RW T	0.9482	0.8946	1.0000	0.9018	0.9962	0.5037
	0.0037	0.0075	0.0000	0.0063	0.0002	0.0238
RW A	0.8896	0.9883	0.9018	1.0000	0.8975	0.5643
	0.0066	0.0009	0.0063	0.0000	0.0068	0.0250
RW R	0.9439	0.8907	0.9962	0.8975	1.0000	0.4959
	0.0042	0.0081	0.0002	0.0068	0.0000	0.0249
Q	0.4913	0.5635	0.5037	0.5643	0.4959	1.0000
	0.0266	0.0256	0.0238	0.0250	0.0249	0.0000

Table 10: Correlation coefficients between R values for the HK dataset, at the reconstruction level, after the ReconWest T-Method and A-Method R values were averaged among the different analyzers. In each table cell, the top number is the correlation coefficient and the bottom number is the error on the coefficient.

9d Correlation Coefficients – Recon. Level						
	RE T	RE A	RW T	RW A	RW R	Q
RE T	1.0000	0.9002	0.9471	0.8939	0.9414	0.5000
	0.0000	0.0064	0.0035	0.0065	0.0040	0.0248
RE A	0.9002	1.0000	0.8940	0.9876	0.8885	0.5710
	0.0064	0.0000	0.0103	0.0009	0.0118	0.0213
RW T	0.9471	0.8940	1.0000	0.9027	0.9963	0.4985
	0.0035	0.0103	0.0000	0.0088	0.0003	0.0249
RW A	0.8939	0.9876	0.9027	1.0000	0.8969	0.5713
	0.0065	0.0009	0.0088	0.0000	0.0097	0.0214
RW R	0.9414	0.8885	0.9963	0.8969	1.0000	0.4888
	0.0040	0.0118	0.0003	0.0097	0.0000	0.0265
Q	0.5000	0.5710	0.4985	0.5713	0.4888	1.0000
	0.0248	0.0213	0.0249	0.0214	0.0265	0.0000

Table 11: Correlation coefficients between R values for the 9d dataset, at the reconstruction level, after the ReconWest T-Method and A-Method R values were averaged among the different analyzers. In each table cell, the top number is the correlation coefficient and the bottom number is the error on the coefficient.

EG Correlation Coefficients – Recon. Level						
	RE T	RE A	RW T	RW A	RW R	Q
RE T	1.0000	0.9099	0.9506	0.9052	0.9451	0.5087
	0.0000	0.0064	0.0039	0.0062	0.0045	0.0270
RE A	0.9099	1.0000	0.9012	0.9887	0.8955	0.5761
	0.0064	0.0000	0.0115	0.0008	0.0122	0.0230
RW T	0.9506	0.9012	1.0000	0.9098	0.9963	0.5110
	0.0039	0.0115	0.0000	0.0104	0.0003	0.0273
RW A	0.9052	0.9887	0.9098	1.0000	0.9040	0.5737
	0.0062	0.0008	0.0104	0.0000	0.0112	0.0232
RW R	0.9451	0.8955	0.9963	0.9040	1.0000	0.5034
	0.0045	0.0122	0.0003	0.0112	0.0000	0.0275
Q	0.5087	0.5761	0.5110	0.5737	0.5034	1.0000
	0.0270	0.0230	0.0273	0.0232	0.0275	0.0000

Table 12: Correlation coefficients between R values for the EG dataset, at the reconstruction level, after the ReconWest T-Method and A-Method R values were averaged among the different analyzers. In each table cell, the top number is the correlation coefficient and the bottom number is the error on the coefficient.

3.1 Result Deviations

Of interest beyond the statistical correlations themselves are the deviations between the best-fit R results presented in Table 2. The 1σ allowed deviation between two analysis results is given by

$$\sigma_{\text{allowed}} = \sqrt{\sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j}, \tag{8}$$

where again σ_i and σ_j are the errors of the two analyses, and r_{ij} is the correlation coefficient between them. Using this equation, the determined correlation coefficients, and the values in Table 2, the allowed differences and subsequent deviations between all of the analyses for the four datasets were calculated. Tables 13 through 16 display this information, with the allowed differences in ppb. In general the deviations between the different analyses are spread around 0 with a few deviations rising into the positive and negative 2σ range. Some specifics to note are that the Q-Method best-fit R values for the HK and 9d datasets are something like 1.35σ or so above the rest depending on the exact correlation, and the R-Method best-fit R values are consistently below the rest barring a few entries, up to around 2σ in some cases. The former is most likely statistical in nature, whereas the latter is most likely systematic, as the R-Method has been shown to be less susceptible to systematic effects [9].

In order to get a sense of the spread of deviations, all deviation entries from the top-side of the table diagonals were taken for each of the datasets and put into a histogram, as shown in Figure 12. The resulting distribution, while not a perfect Gaussian, shows a nice symmetry around 0 with only a few entries close to 2σ or higher[9]. This histogram provides some confidence that the observed differences in the fits to data are natural and to be expected. Lastly, the calculation of these allowed differences and associated deviations are purely statistical. It is expected that with systematic components included, the deviations would get smaller as the correlations would decrease and the errors on the best-fit R values would increase.

[9]Note that correlation entries in the tables are themselves highly correlated, so a perfect Gaussian is not expected.

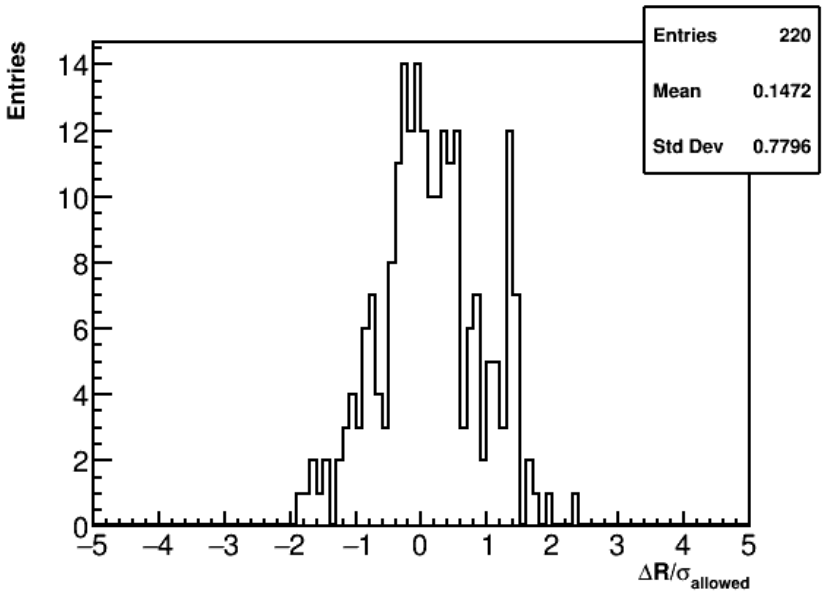


Figure 12: Deviations between best-fit R values for all datasets for the individual analysis comparisons. The entries from the top-right of the diagonal of Tables 13 through 16 are taken and filled into this histogram. In general such a pull distribution should be a unit Gaussian, however entries from the same row or column will be correlated, so a perfect Gaussian is not expected. This is particularly apparent with the spike around 1.35σ, which comes from some of the Q-Method correlations in the HK and 9d datasets, in both of which the Q-Method fits returned a greater R-value than the rest of the methods.

60h Analysis Differences											
	BU T	BU R	CU T	CU A	UW T	UW A	EU T	EU A	SJTU T	SJTU A	UK Q
BU T	0.0000 +0.00	0.1154 −1.42	0.4495 +1.32	0.5943 +0.96	0.1555 +1.17	0.5983 +0.28	0.2817 −0.29	0.5890 +0.55	0.0609 +1.03	0.5781 +0.66	1.7692 −0.23
BU R	0.1154 +1.42	0.0000 +0.00	0.4609 +1.64	0.6015 +1.23	0.1464 +2.37	0.5963 +0.55	0.2862 +0.29	0.5912 +0.82	0.1288 +1.76	0.5867 +0.93	1.7731 −0.14
CU T	0.4495 −1.32	0.4609 −1.64	0.0000 +0.00	0.5710 −0.03	0.4642 −0.88	0.6017 −0.71	0.4628 −1.46	0.5933 −0.46	0.4484 −1.18	0.5854 −0.36	1.7710 −0.56
CU A	0.5943 −0.96	0.6015 −1.23	0.5710 +0.03	0.0000 +0.00	0.5879 −0.67	0.2172 −1.88	0.5813 −1.13	0.2050 −1.23	0.5846 −0.87	0.1774 −1.09	1.6581 −0.59
UW T	0.1555 −1.17	0.1464 −2.37	0.4642 +0.88	0.5879 +0.67	0.0000 +0.00	0.5619 −0.03	0.2851 −0.93	0.5731 +0.24	0.1600 −0.75	0.5741 +0.34	1.7583 −0.33
UW A	0.5983 −0.28	0.5963 −0.55	0.6017 +0.71	0.2172 +1.88	0.5619 +0.03	0.0000 +0.00	0.5767 −0.43	0.1419 +1.10	0.5888 −0.17	0.1572 +1.36	1.6580 −0.34
EU T	0.2817 +0.29	0.2862 −0.29	0.4628 +1.46	0.5813 +1.13	0.2851 +0.93	0.5767 +0.43	0.0000 +0.00	0.5553 +0.73	0.2827 +0.51	0.5696 +0.81	1.7456 −0.18
EU A	0.5890 −0.55	0.5912 −0.82	0.5933 +0.46	0.2050 +1.23	0.5731 −0.24	0.1419 −1.10	0.5553 −0.73	0.0000 +0.00	0.5786 −0.45	0.1389 +0.42	1.6580 −0.44
SJTU T	0.0609 −1.03	0.1288 −1.76	0.4484 +1.18	0.5846 +0.87	0.1600 +0.75	0.5888 +0.17	0.2827 −0.51	0.5786 +0.45	0.0000 +0.00	0.5641 +0.56	1.7665 −0.26
SJTU A	0.5781 −0.66	0.5867 −0.93	0.5854 +0.36	0.1774 +1.09	0.5741 −0.34	0.1572 −1.36	0.5696 −0.81	0.1389 −0.42	0.5641 −0.56	0.0000 +0.00	1.6632 −0.47
UK Q	1.7692 +0.23	1.7731 +0.14	1.7710 +0.56	1.6581 +0.59	1.7583 +0.33	1.6580 +0.34	1.7456 +0.18	1.6580 +0.44	1.7665 +0.26	1.6632 +0.47	0.0000 +0.00

Table 13: Differences in results for the 60h dataset between the different analyses. The top number is the allowed ppb difference in R, σ_{allowed}, as calculated from the correlation coefficients and analysis errors. The bottom number is the calculated deviation between analyses, calculated as $\sigma_{\text{dev}} = (R_{\text{column}} - R_{\text{row}})/\sigma_{\text{allowed}}$, where R_{column} and R_{row} come from Table 2 for the respective analyses.

HK Analysis Differences											
	BU T	BU R	CU T	CU A	UW T	UW A	EU T	EU A	SJTU T	SJTU A	UK Q
BU T	0.0000 +0.00	0.0943 −1.75	0.3736 −0.44	0.5218 +0.19	0.1293 +0.30	0.5184 +0.15	0.2310 −0.16	0.5172 +0.04	0.0544 +0.78	0.5094 −0.09	1.5421 +1.36
BU R	0.0943 +1.75	0.0000 +0.00	0.3846 −0.00	0.5262 +0.50	0.1256 +1.63	0.5153 +0.47	0.2389 +0.54	0.5172 +0.36	0.1087 +1.91	0.5143 +0.23	1.5454 +1.46
CU T	0.3736 +0.44	0.3846 +0.00	0.0000 +0.00	0.5015 +0.52	0.3891 +0.53	0.5267 +0.46	0.3878 +0.33	0.5246 +0.36	0.3724 +0.56	0.5169 +0.23	1.5470 +1.46
CU A	0.5218 −0.19	0.5262 −0.50	0.5015 −0.52	0.0000 +0.00	0.5208 −0.11	0.1842 −0.10	0.5114 −0.26	0.1788 −0.42	0.5119 −0.11	0.1502 −0.96	1.4443 +1.38
UW T	0.1293 −0.30	0.1256 −1.63	0.3891 −0.53	0.5208 +0.11	0.0000 +0.00	0.4934 +0.08	0.2380 −0.32	0.5080 −0.03	0.1338 +0.02	0.5101 −0.17	1.5328 +1.34
UW A	0.5184 −0.15	0.5153 −0.47	0.5267 −0.46	0.1842 +0.10	0.4934 −0.08	0.0000 +0.00	0.5022 −0.23	0.1181 −0.47	0.5086 −0.07	0.1314 −0.95	1.4438 +1.40
EU T	0.2310 +0.16	0.2389 −0.54	0.3878 −0.33	0.5114 +0.26	0.2380 +0.32	0.5022 +0.23	0.0000 +0.00	0.4888 +0.12	0.2301 +0.34	0.5004 −0.02	1.5251 +1.40
EU A	0.5172 −0.04	0.5172 −0.36	0.5246 −0.36	0.1788 +0.42	0.5080 +0.03	0.1181 +0.47	0.4888 −0.12	0.0000 +0.00	0.5064 +0.04	0.1163 −0.60	1.4439 +1.44
SJTU T	0.0544 −0.78	0.1087 −1.91	0.3724 −0.56	0.5119 +0.11	0.1338 −0.02	0.5086 +0.07	0.2301 −0.34	0.5064 −0.04	0.0000 +0.00	0.4956 −0.18	1.5365 +1.34
SJTU A	0.5094 +0.09	0.5143 −0.23	0.5169 −0.23	0.1502 +0.96	0.5101 +0.17	0.1314 +0.95	0.5004 +0.02	0.1163 +0.60	0.4956 +0.18	0.0000 +0.00	1.4472 +1.48
UK Q	1.5421 −1.36	1.5454 −1.46	1.5470 −1.46	1.4443 −1.38	1.5328 −1.34	1.4438 −1.40	1.5251 −1.40	1.4439 −1.44	1.5365 −1.34	1.4472 −1.48	0.0000 +0.00

Table 14: Differences in results for the HK dataset between the different analyses. The top number is the allowed ppb difference in R, σ_{allowed}, as calculated from the correlation coefficients and analysis errors. The bottom number is the calculated deviation between analyses, calculated as $\sigma_{\text{dev}} = (R_{\text{column}} - R_{\text{row}})/\sigma_{\text{allowed}}$, where R_{column} and R_{row} come from Table 2 for the respective analyses.

9d Analysis Differences											
	BU T	BU R	CU T	CU A	UW T	UW A	EU T	EU A	SJTU T	SJTU A	UK Q
BU T	0.0000 +0.00	0.0775 −0.06	0.3076 −0.34	0.4220 +0.86	0.1040 +0.17	0.4177 +0.83	0.1911 +0.15	0.4148 +0.76	0.0415 +0.57	0.4088 +0.42	1.2432 +1.32
BU R	0.0775 +0.06	0.0000 +0.00	0.3164 −0.31	0.4281 +0.86	0.1007 +0.23	0.4171 +0.84	0.1940 +0.17	0.4174 +0.77	0.0874 +0.32	0.4149 +0.43	1.2487 +1.32
CU T	0.3076 +0.34	0.3164 +0.31	0.0000 +0.00	0.3974 +1.18	0.3175 +0.38	0.4156 +1.08	0.3100 +0.42	0.4151 +1.01	0.3052 +0.42	0.4083 +0.68	1.2334 +1.41
CU A	0.4220 −0.86	0.4281 −0.86	0.3974 −1.18	0.0000 +0.00	0.4213 −0.82	0.1513 −0.13	0.4130 −0.81	0.1469 −0.32	0.4139 −0.82	0.1241 −1.54	1.1522 +1.11
UW T	0.1040 −0.17	0.1007 −0.23	0.3175 −0.38	0.4213 +0.82	0.0000 +0.00	0.3978 +0.82	0.1948 +0.05	0.4085 +0.73	0.1067 +0.05	0.4089 +0.38	1.2327 +1.31
UW A	0.4177 −0.83	0.4171 −0.84	0.4156 −1.08	0.1513 +0.13	0.3978 −0.82	0.0000 +0.00	0.4019 −0.79	0.0973 −0.28	0.4100 −0.78	0.1031 −1.67	1.1483 +1.13
EU T	0.1911 −0.15	0.1940 −0.17	0.3100 −0.42	0.4130 +0.81	0.1948 −0.05	0.4019 +0.79	0.0000 +0.00	0.3912 +0.74	0.1911 −0.02	0.4002 +0.36	1.2332 +1.31
EU A	0.4148 −0.76	0.4174 −0.77	0.4151 −1.01	0.1469 +0.32	0.4085 −0.73	0.0973 +0.28	0.3912 −0.74	0.0000 +0.00	0.4062 −0.72	0.0900 −1.60	1.1532 +1.14
SJTU T	0.0415 −0.57	0.0874 −0.32	0.3052 −0.42	0.4139 +0.82	0.1067 −0.05	0.4100 +0.78	0.1911 +0.02	0.4062 +0.72	0.0000 +0.00	0.3982 +0.38	1.2417 +1.30
SJTU A	0.4088 −0.42	0.4149 −0.43	0.4083 −0.68	0.1241 +1.54	0.4089 −0.38	0.1031 +1.67	0.4002 −0.36	0.0900 +1.60	0.3982 −0.38	0.0000 +0.00	1.1581 +1.26
UK Q	1.2432 −1.32	1.2487 −1.32	1.2334 −1.41	1.1522 −1.11	1.2327 −1.31	1.1483 −1.13	1.2332 −1.31	1.1532 −1.14	1.2417 −1.30	1.1581 −1.26	0.0000 +0.00

Table 15: Differences in results for the 9d dataset between the different analyses. The top number is the allowed ppb difference in R, σ_{allowed}, as calculated from the correlation coefficients and analysis errors. The bottom number is the calculated deviation between analyses, calculated as $\sigma_{\text{dev}} = (R_{\text{column}} - R_{\text{row}})/\sigma_{\text{allowed}}$, where R_{column} and R_{row} come from Table 2 for the respective analyses.

EG Analysis Differences											
	BU T	BU R	CU T	CU A	UW T	UW A	EU T	EU A	SJTU T	SJTU A	UK Q
BU T	0.0000 +0.00	0.0602 −1.05	0.2421 −0.05	0.3325 +0.34	0.0806 −0.15	0.3256 +0.10	0.1461 −1.20	0.3266 −0.08	0.0346 +1.04	0.3226 +0.02	1.0990 −0.26
BU R	0.0602 +1.05	0.0000 +0.00	0.2495 +0.20	0.3372 +0.52	0.0781 +0.65	0.3257 +0.29	0.1513 −0.74	0.3282 +0.12	0.0687 +1.45	0.3276 +0.22	1.1029 −0.20
CU T	0.2421 +0.05	0.2495 −0.20	0.0000 +0.00	0.3100 +0.40	0.2497 +0.00	0.3221 +0.14	0.2464 −0.66	0.3215 −0.04	0.2409 +0.21	0.3188 +0.07	1.0972 −0.25
CU A	0.3325 −0.34	0.3372 −0.52	0.3100 −0.40	0.0000 +0.00	0.3337 −0.37	0.1183 −0.67	0.3270 −0.88	0.1122 −1.22	0.3274 −0.23	0.1012 −1.03	1.0387 −0.39
UW T	0.0806 +0.15	0.0781 −0.65	0.2497 −0.00	0.3337 +0.37	0.0000 +0.00	0.3125 +0.14	0.1505 −1.08	0.3230 −0.04	0.0851 +0.57	0.3249 +0.06	1.0960 −0.25
UW A	0.3256 −0.10	0.3257 −0.29	0.3221 −0.14	0.1183 +0.67	0.3125 −0.14	0.0000 +0.00	0.3182 −0.65	0.0751 −0.77	0.3204 +0.01	0.0816 −0.31	1.0406 −0.31
EU T	0.1461 +1.20	0.1513 +0.74	0.2464 +0.66	0.3270 +0.88	0.1505 +1.08	0.3182 +0.65	0.0000 +0.00	0.3115 +0.48	0.1460 +1.45	0.3193 +0.57	1.0964 −0.10
EU A	0.3266 +0.08	0.3282 −0.12	0.3215 +0.04	0.1122 +1.22	0.3230 +0.04	0.0751 +0.77	0.3115 −0.48	0.0000 +0.00	0.3208 +0.19	0.0746 +0.44	1.0437 −0.25
SJTU T	0.0346 −1.04	0.0687 −1.45	0.2409 −0.21	0.3274 +0.23	0.0851 −0.57	0.3204 −0.01	0.1460 −1.45	0.3208 −0.19	0.0000 +0.00	0.3145 −0.09	1.0977 −0.30
SJTU A	0.3226 −0.02	0.3276 −0.22	0.3188 −0.07	0.1012 +1.03	0.3249 −0.06	0.0816 +0.31	0.3193 −0.57	0.0746 −0.44	0.3145 +0.09	0.0000 +0.00	1.0411 −0.28
UK Q	1.0990 +0.26	1.1029 +0.20	1.0972 +0.25	1.0387 +0.39	1.0960 +0.25	1.0406 +0.31	1.0964 +0.10	1.0437 +0.25	1.0977 +0.30	1.0411 +0.28	0.0000 +0.00

Table 16: Differences in results for the EG dataset between the different analyses. The top number is the allowed ppb difference in R, σ_{allowed}, as calculated from the correlation coefficients and analysis errors. The bottom number is the calculated deviation between analyses, calculated as $\sigma_{\text{dev}} = (R_{\text{column}} - R_{\text{row}})/\sigma_{\text{allowed}}$, where R_{column} and R_{row} come from Table 2 for the respective analyses.

4 Conclusions

Correlation coefficients were determined between analyses and methods for the different Run 1 ω_a analyzers and datasets, using a Monte Carlo with real data input. These correlation coefficients and variations of them were used in different approaches to the ω_a combination, and associated results will be detailed in a forthcoming note by D. Sweigart [12][10]. These correlations and the Monte Carlo provide a starting point for any combination effort of future analyses of Run 2 and beyond.

While some determined correlations did exceed the conditions for the high-correlation regime, in general the correlations were very consistent across datasets and methods. When comparing the Monte Carlo used to generate the pseudo-data to real data, there are a number of facets where differences might arise from, as well as places where the Monte Carlo could be improved:

- There are systematic differences between analyses, both in how various systematics affect the best-fit R values (eg. less effect in the R-Method), and how analyses correct and estimate the effects. The errors from the systematic effects are not included in the calculation of the correlations here.

- One of the main systematic differences in the analyses is the use of different pileup methods. Including the effect of pileup should be sub-leading since it is the corrected time spectrum that is of most importance, and different energy thresholds have a greater effect on the number of counts. It would be quite an undertaking to implement multiple pileup methods into the Monte Carlo, and probably not worth the effort.

- The Q-Method is a different analysis method completely, and it's differences won't be adequately modeled here. Since the correlations are so much less than the other datasets it's not so much of a concern, but it should be kept in mind. The larger errors on the Q-Method correlations in part reflect this.

- The `ROOT TF2`s used to generate the data could use finer time and energy points (with a different implementation), so that the hit generation wouldn't introduce aliasing frequencies in some of the Q-Method fits.

- The ReconEast vs ReconWest comparison was based solely on clusters from a single ReconWest analysis, A. Fienberg's. The comparison could be done with the hits from other ReconWest analyzers, though the differences are expected to be small. The comparison was also made on non-pileup corrected counts.

- Energy bin functions for ReconWest were provided by a single ReconWest analyzer, M. Sorbara. Other analyzers most likely have small differences.

- The analyses to data randomize out the fast rotation by randomizing times at the cyclotron period, and this effect is not included in the Monte Carlo. Different analyses then use different numbers of random seeds, and some analyzers quote mean R values from from fits to the seeds while others quote the closest R value to the mean. Differences in results due to

[10]The favored approach is a staged averaging method, using no correlation coefficients in the combination. An error on the combination value is then determined using different combination methods using the provided correlation coefficients in a variety of ways.

these choices would increase the amount of expected statistical deviation slightly. Generating data corresponding to a number of random seeds specific to every analyzer would make the computation time needed increase drastically, and probably wouldn't be feasible.

- The BU analysis effort randomized out the vertical waist effect. This increases the error on the best-fit R values by 10 ppb to 30 ppb depending on dataset, and would result in slightly different (perhaps negligibly) correlation coefficients.

- The `TRandomMixMax` randomization class may be insufficient as shown by the non-flat R-Method p value distribution. A better random number generator may be used, such as `Ranlux64`, however there is the downside that the computation time would increase significantly.

The effects of each of these points would in general decrease the correlations between the analyses, as the differences between them increase. It is possible though not guaranteed that implementing the respective changes would push the correlation coefficients below the high-correlation regime cut-off. Some of the points may be straightforwardly improved at the cost of computation time, while others would take significant work for most likely relatively little gain. Whether it is worth improving the correlation coefficients or not for future runs will depend on the combination approaches adopted. For now the correlation coefficients presented here are the most accurate ones available for the experiment.

References

[1] T. Gorringe. *Readers guide and synopsis for run 1, ω_a analysis*. Muon $g-2$ Note 243 DocDB 23555. Aug. 2020.

[2] David Allen Sweigart. "A measurement of the anomalous precession frequency of the positive muon". PhD thesis. Cornell University, May 2020.

[3] Aaron Fienberg. "Measuring the Precession Frequency in the E989 Muon $g-2$ Experiment". PhD thesis. University of Washington, Jan. 2019.

[4] A. Keshavarzi. *A statistical combination of ω_a for Run 1*. Muon $g-2$ DocDB 21494. Feb. 2020.

[5] D. Sweigart. *T-/A-Method Analysis of the Run 1 Data Sets*. Muon $g-2$ DocDB 21323. Feb. 2020.

[6] A. Fienberg. *UW ω_a Analysis*. Muon $g-2$ DocDB 21338. Feb. 2020.

[7] J. Price and M. Sorbara. *Europa*. Muon $g-2$ DocDB 21326. Feb. 2020.

[8] B. Li. *SJTU Run-1 ω_a Blinded Analysis*. Muon $g-2$ DocDB 21329. Feb. 2020.

[9] Nicholas Kinnaird. "Muon spin precession frequency extraction and decay positron track fitting in Run 1 of the Fermilab Muon $g-2$ Experiment". PhD thesis. Boston University, Jan. 2020.

[10] N. Kinnaird. *BU Analysis Report*. Muon $g-2$ DocDB 21320. Feb. 2020.

[11] T. Gorringe. *Analysis Type: Q*. Muon $g-2$ DocDB 21302. Feb. 2020.

[12] D. Sweigart et al. Run 1 Combination Task Force Discussion. 2020.

[13] J. LaBounty. *Run 1 Recon East vs. Recon West Comparisons*. Muon $g-2$ DocDB 21567. Feb. 2020.

[14] *TRandom Class Reference*. https://root.cern.ch/doc/master/classTRandom.html.

[15] G. Cowan. *Statistical Data Analysis*. Clarendon Press Oxford, 1998.

[16] D. Sweigart. *Towards the ω_a Combination for Run 1*. Muon $g-2$ DocDB 23308. July 2020.