The Abdus Salam
International Centre
for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2358-5

Joint ICTP-IAEA Workshop on Nuclear Structure Decay Data: Theory and Evaluation

6 - 17 August 2012

Averaging Techniques

B. Singh
McMaster University
Canada

Strada Costiera, 11 - 34151 - Trieste - Italy • Tel. +39 0402240111 • Fax. +39 040224163 • sci_info@ictp.it • www.ictp.it
ICTP is governed by UNESCO, IAEA, and Italy, and it is a UNESCO Category 1 Institute

Averaging Techniques for Experimental Measurements

Balraj Singh (McMaster University)

(Prepared by Michael Birch and Balraj Singh, McMaster Univ.)

ICTP-IAEA workshop, Trieste, Aug 6-17 (2012)

Averaging methods:

- While weighted and un-weighted averaging methods are most common, they do not give satisfactory answers all the time.
- Modified weighted averaging methods: LWM, NRM, RT
- Other averaging methods: Mandel-Paule method, Method of best representation (MBR); and median method Bootstrap.
- Methods of finding outliers (for possible rejection) in a dataset also exist to help the evaluator to average the appropriate data since a poor measurement may skew the average
- Here we description of these methods; handling asymmetric uncertainties
- Display of a computer code incorporating above methods

Need for averaging methods?

- Often measurements of a single quantity are made independently by different experimenters using different methods/apparatus.
- How does an evaluator handle a data set to obtain best recommended value?

- Example: Mean lifetime of Free Neutron ($n \rightarrow p + e^- + \bar{\nu}$)

Neutron mean lifetime data

Author	Original Value (sec)	Reanalyzed value (sec)
2000Ar07	885.4(10)	881.6(21)*: 2012Ar05
1989Ma34	887.6(30) *	882.5(21) * ; 2012St10
2010Pi07	880.7(18) *	
2005Se01	878.5(8) **	
2003De25	886.8(34)	886.3(39) *; 2005Ni13
1990By05	893.6(53)	889.2(48) *; 1996By03
1993Ma62	882.6(27) *	
1992Ne11	888.4(33)	Withdrawn by 2005Se01

F. E. Wietfeldt and G.L. Greene: The neutron lifetime: RMP 83, 1117 (Oct-Dec 2011)

Options for a data evaluator

- Ideally one will prepare a critical compilation by reading every paper, evaluating methodology of measurement and assignment of both statistical and systematic uncertainties.
- Sufficient details not being available in a publication could be a reason for not including the measurement in the analysis. Could contact authors if possible to obtain the details and policy of uncertainty assignments. But when encountering large data sets and/or quite old measurements, this may become impossible or cumbersome.
- Select a dataset from available experimental results which you believe represents reliable set of measurements and realistic uncertainties. (Watch out very low and unrealistic uncertainties).
- Sometimes a single measurements merits recommendation due to superior technique used and uncertainty budget accounted for properly.
- Often resort to statistical methods since such a single measurement rarely exists.

Statistical Procedures: conditions

- Each measurement is made using **accurate** techniques
  - By examining the methods of each experiment, result(s) not satisfying this assumption should be discarded
  - For example in a half-life measurement using integral beta counting, if the impurities also present but were not well known then it is possible the result is skewed. One may select data based on discrete gamma rays.
- Each measurement is **independent** and uncorrelated with other measurements
  - The data set should only include results which are obtained by different groups or by the same group using different methods
- The standard deviation of each measurement is correctly estimated (i.e. the **precision** is reasonable)
  - The experimenter, when quoting the final value, properly analyzed and accounted for both statistical and systematic uncertainties
  - If it is clear the quoted uncertainty is unrealistically low, it may be necessary to inflate the uncertainty at the discretion of the evaluator

Statistical Procedures

- Check if the dataset you have selected is discrepant: poor agreement between different measurements, i.e. deviations of >3 or so quoted standard deviations.
- Take a weighted average. If reduced χ^2 > critical χ^2 at 95-99% confidence level, then the data are deemed discrepant. If data are not discrepant then various methods described later will most likely converge.
- If data seem discrepant, look for outliers. Over the years several methods have been proposed, but these should be used with extreme caution. It may happen that the outlier is actually closest to the true value!

Outliers in a data set

- Identifying and possibly omitting outliers is not a process universally agreed upon and often discouraged. Finally it comes down to the discretion of an evaluator.
- Two prescriptions of finding such data points are Chauvenet's and Peirce's criteria, both circa 1860, former being more popular, although, latter is more rigorous statistically.

  In Chauvenet's words, " For the general case….. when there are several unknown quantities and several doubtful observations, the modifications which the rule [*Chauvenet's criterion*] requires renders it more troublesome than Peirce's formula……What I have given may serve the purpose of giving the reader greater confidence in the correctness and value of Peirce's Criterion". William Chauvenet, A Manual of Spherical and Practical Astronomy V.II , Lippincott, Philadelphia, 1st Ed (1863)

Chauvenet's Criterion (in manual of practical astronomy)

- William Chauvenet decided (circa 1860) an "outlier" should be defined as a value in a set of n measurements for which the deviation from the mean, $d_i = |x_i - \bar{x}|$, would be observed with probability less than $1/2n$ assuming the data are distributed according to a normal distribution with the sample mean, $\bar{x}$, (unweighted average) and variance, s^2, given by the unbiased sample variance (a quantity defined in any statistics text). Iterative approach with one outlier picked up at a time

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad\qquad s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

- Note that the uncertainties of the individual data points are not taken into account
- A formula for the criterion is thus obtained by the following calculation

$$\Pr(X \geq \bar{x} + d_i) + \Pr(X \leq \bar{x} - d_i) < \frac{1}{2n}$$

$$\int_{\bar{x}+d}^{\infty} \mathcal{N}(x;\bar{x},s)dx + \int_{-\infty}^{\bar{x}-d} \mathcal{N}(x;\bar{x},s)dx < \frac{1}{2n}$$

$$1 - \operatorname{erf}\left(\frac{d_i}{\sqrt{2}\,s}\right) < \frac{1}{2n}$$

$$n \cdot \operatorname{erfc}\left(\frac{d_i}{\sqrt{2}\,s}\right) < \frac{1}{2}$$

where $\operatorname{erf}(x)$ is the "error function" defined by

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

and $\operatorname{erfc}(x)$ is the complimentary error function defined by $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$

Peirce's Criterion

- Benjamin Peirce developed a criterion for finding outliers a few years before Chauvenet and his work is more mathematically rigorous
- He assumes the data are distributed according to the same normal distribution as Chauvenet, however the principle used to identify outliers is very different
- A subset of m points are called outliers if

  (likelihood of the complete data set) < (likelihood of the remainder of the data set)(Probability of the existence of m outliers)
- The principle includes the identification of more than one outlier hence the procedure for identifying outliers need not be iterated as with Chauvenet's criterion
- It is difficult to obtain an exact, closed form solution to the inequality above using the appropriate likelihood functions; however an iterative procedure can be used to find the maximum deviation from the mean, above which the measurements can be considered outliers by the above principle

Peirce's Criterion

- After working with the mathematical formulation for Peirce's principle the following four equations are derived to obtain the ratio of the maximum deviation from the unweighted mean, d_{max}, to the square root of the sample variance, s, as defined for Chauvenet's Criterion: $r_{max}=d_{max}/s$.
  - Suppose in a set of n measurements m are suspected as outliers

$$Q^n = \frac{m^m (n-m)^{n-m}}{n^n} \qquad (1)$$

$$\lambda^{n-m} R^n = Q^n \qquad (2)$$

$$r_{\max}^2 = \lambda^2 + \frac{n-1}{m}(1-\lambda^2) \qquad (3)$$

$$R = e^{\frac{1}{2}\left(r_{\max}^2 - 1\right)} \operatorname{erfc}\left(\frac{r_{\max}}{\sqrt{2}}\right) \qquad (4)$$

  - These lend themselves to the iterative procedure to find r_{max}
    - 1. Calculate Q using equation (1)
    - 2. Begin with an approximate value for R
    - 3. Use Q and R to calculate λ by equation (2)
    - 4. Use λ to calculate r_{max} using equation (3)
    - 5. Use r_{max} to refine the estimate on R using equation (4)
    - Repeat steps 3-5 until R converges to one value, the r_{max} which gives that R is the required maximum ratio

Peirce's Criterion

- To apply Peirce's criterion:
  - First assume one point is an outlier ($m=1$), then check if that is true by checking if any points exceed the maximum deviation from the unweighted mean calculated as on the previous slide
  - If there are any outliers then assume there is one more (for example if 3 points exceed the maximum deviation then try the calculation with $m=4$) and repeat the calculation until no more outliers are identified
- Note that even though Peirce's criterion in more rigorous than Chauvenet's and does not arbitrarily choose a probability which indicates outliers, this formulation still does not include the uncertainties of the respective data points as they are not included in the likelihood functions

Method to including uncertainties by M. Birch

- It is proposed here that a criterion for identifying outliers which takes into account the uncertainties on each data point may be defined as follows:
  - A measurement $x_i \pm \sigma_i$ is outlier with respect to a supposed mean $\mu \pm \sigma_\mu$ if the difference

  $d = x_i - \mu$ is "inconsistent with zero" at a given confidence level, α.
- It can then be proven that the random variable

  $D = X_i - M$ will be normally distributed about d with variance $\sigma_d^2 = \sigma_i^2 + \sigma_\mu^2$ where X_i and M are normally distributed random variables with their respective peak values at the measurement x_i and supposed mean μ
- We can say D is inconsistent with zero at a confidence level α if $\Pr(0<D<2d) > \alpha$ when $d > 0$ or $\Pr(2d<D<0) > \alpha$ if $d < 0$, since these intervals correspond to the set of values more likely to occur than zero.
- This results in the formula

$$\operatorname{erf}\left(\frac{|x_i - \mu|}{\sqrt{2\left(\sigma_i^2 + \sigma_\mu^2\right)}} \right) > \alpha$$

Outlier Identification Including Uncertainties

- This criterion should be applied by simply checking each measurement individually; should not be iterated!
- It should always be kept in mind that this criterion identifies outliers with respect to the given mean which should be the evaluator's best estimate of the true value, generally a weighted average. This may be chosen using any of the averaging techniques to be discussed.

- Evaluator's choice whether to keep or omit outliers prior to using averaging procedures

Coming back to neutron mean lifetime: Average (recommended) value

- Particle Data Group 2012 evaluation: used weighted average by including for the first time, seemingly discrepant value from 2005Se01, but not pointed out an outlier by any method, provided revised value from 2012St10 is used; otherwise the original value from these authors is an outlier according to new proposed procedure.

  PRD 86, 010001 (July 2012): 880.1(11) sec (with inflated σ)

  Reduced χ^2=2.98 compared to critical χ^2=2.80 at 99%CL, 2.10 at 95%CL

  (Inclusion of 2012St10 correction gives 880.0(9) or 880.4(10)-LWM)

  Reduced χ^2=2.15 compared to critical χ^2=2.80 at 99%CL, 2.10 at 95%CL

  In 2006, 2008, 2010 PDG evaluations Adopted value was 885.7(8) sec; 2005Se01 value was not included as it was much too low to give a meaningful average! However, caution was recommended.

Mathematical Definitions

- Probability Density Function (PDF)
  - A function, $f(x)$, of a continuous random variable for which the integral over an interval gives the probability of the value of the random variable lying in that interval.

$$\Pr(a \leq X \leq b) = \int_a^b f(x)dx$$

  - PDFs are also normalized: $\int_{-\infty}^{\infty} f(x)dx = 1$

Mathematical Definitions

- Mean
  - The mean (or expected) value of a continuous random variable X with PDF $f(x)$ is defined to be:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

  - For a discrete random variable, X, with n possibilities $x_1, x_2, \ldots, x_n$ which occur with associated probabilities $p_1, p_2, \ldots, p_n$ such that $\sum_{i=1}^{n} p_i = 1$ the mean value is:

$$E[X] = \sum_{i=1}^{n} p_i x_i$$

Mathematical Definitions

- Variance
  - The variance of a random variable with mean μ is defined to be:

$$Var[X] = E[(X-\mu)^2]$$

$$= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \quad \text{(Continuous)}$$

$$= \sum_{i=1}^{n} (x_i-\mu)^2 p_i \quad \text{(Discrete)}$$

Mathematical Definitions

- The Normal (Gaussian) Distribution
  - A random variable is normally distributed if it has a PDF of the form

$$\mathcal{N}(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

  - One can then show the mean of a normally distributed random variable is μ and the variance is σ^2

Mathematical Definitions

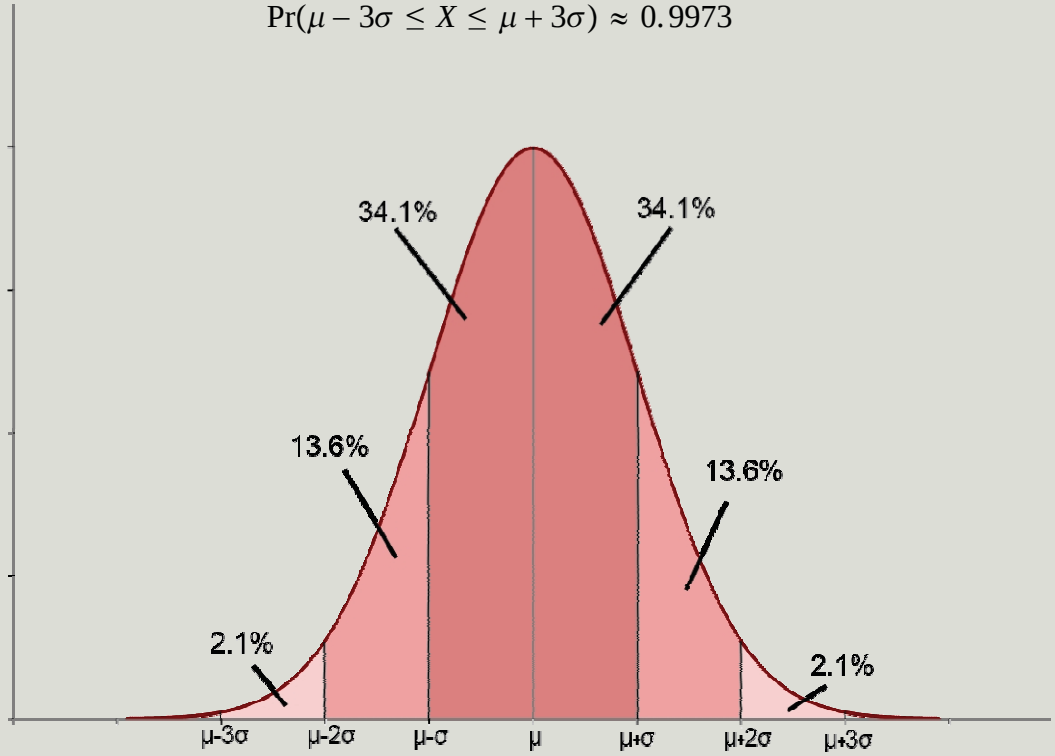
- Standard Deviation
  - Standard deviation is defined to be the square root of the variance.
  - Hence, for a normal distribution the standard deviation is σ.
  - This is an important measure for a normally distributed random variable, X, since the probability of X lying within 1σ, 2σ and 3σ of the mean is 68%, 95% and 99.7% respectively

Mathematical Definitions

$$\Pr(\mu - \sigma \leq X \leq \mu + \sigma) = \int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \approx 0.6827$$

$$\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.9545$$

$$\Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.9973$$

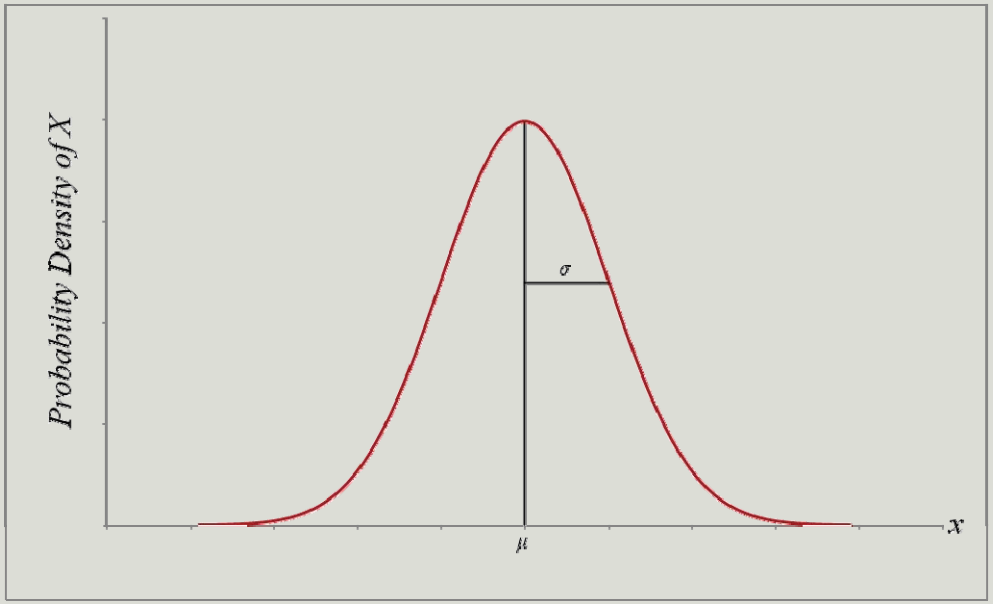


Importance of the Normal Distribution

- The Central Limit Theorem
  - For a set of n independent and identically distributed random variables $X_1, X_2, \ldots, X_n$ with mean μ and variance σ^2, the quantity $Y = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$ tends to be normally distributed with mean 0 and variance σ^2 as n tends to infinity.
  - In the context of experiment one can think of the n random variables as realizations of many sources for error in a measurement (e.g. In various electronic devices), the central limit theorem then says the total error in each measurement can be expected to follow a normal distribution
  - It can also be argued that the use of a normal distribution for error frequency is the best assignment based on the available information without making additional assumptions

Definition of Uncertainty

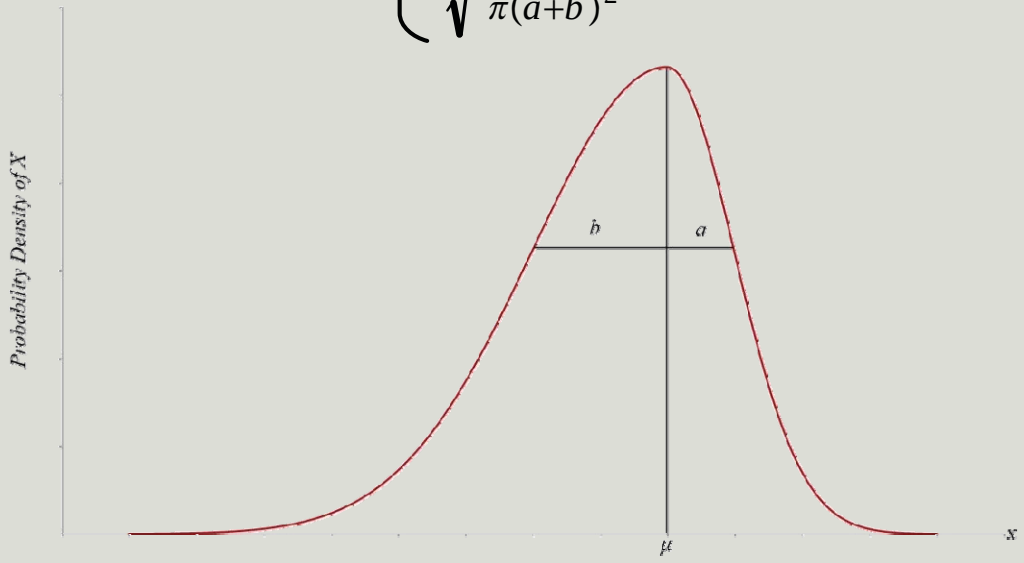
- Consistent with the Central Limit Theorem, a quoted measurement with uncertainty of a quantity X, $\mu \pm \sigma$, represents a normal distribution with mean μ and standard deviation σ



Extension to Asymmetric Uncertainty

- For a quoted measurement with asymmetric uncertainty of a quantity X, μ^{+a}_{-b} , it is assumed the values μ, a, and b correspond to the parameters of the PDF

$$\mathcal{N}_A(x;\mu,a,b) = \begin{cases} \sqrt{\frac{2}{\pi(a+b)^2}}\, e^{-\frac{(x-\mu)^2}{2b}}, & x \leq \mu \\ \sqrt{\frac{2}{\pi(a+b)^2}}\, e^{-\frac{(x-\mu)^2}{2a}}, & x > \mu \end{cases}$$

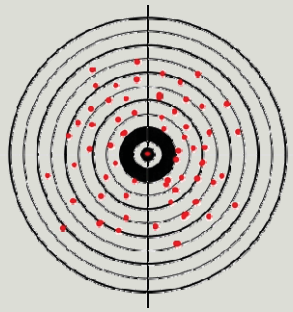


Associated Definitions

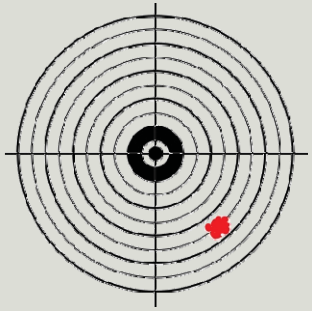
- Reproducibility
  - A measurement which is reproducible would have its result repeated if the experiment were re-performed
  - If the uncertainty is correctly quoted the result should reflect reproducibility of 68% within its error
- Precision
  - A measurement which is precise is one for which the uncertainty is low
- Accuracy
  - A measurement which is accurate is one for which the measured value is close to the “true value”

Associated Definitions

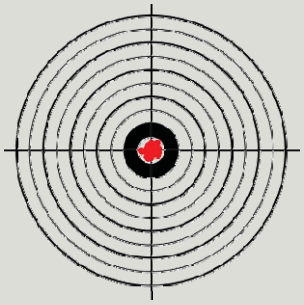
- Clearly, measurements can be:
  - Neither Precise nor accurate
  - Accurate but not precise (first figure)



  - Precise but not accurate (second figure)

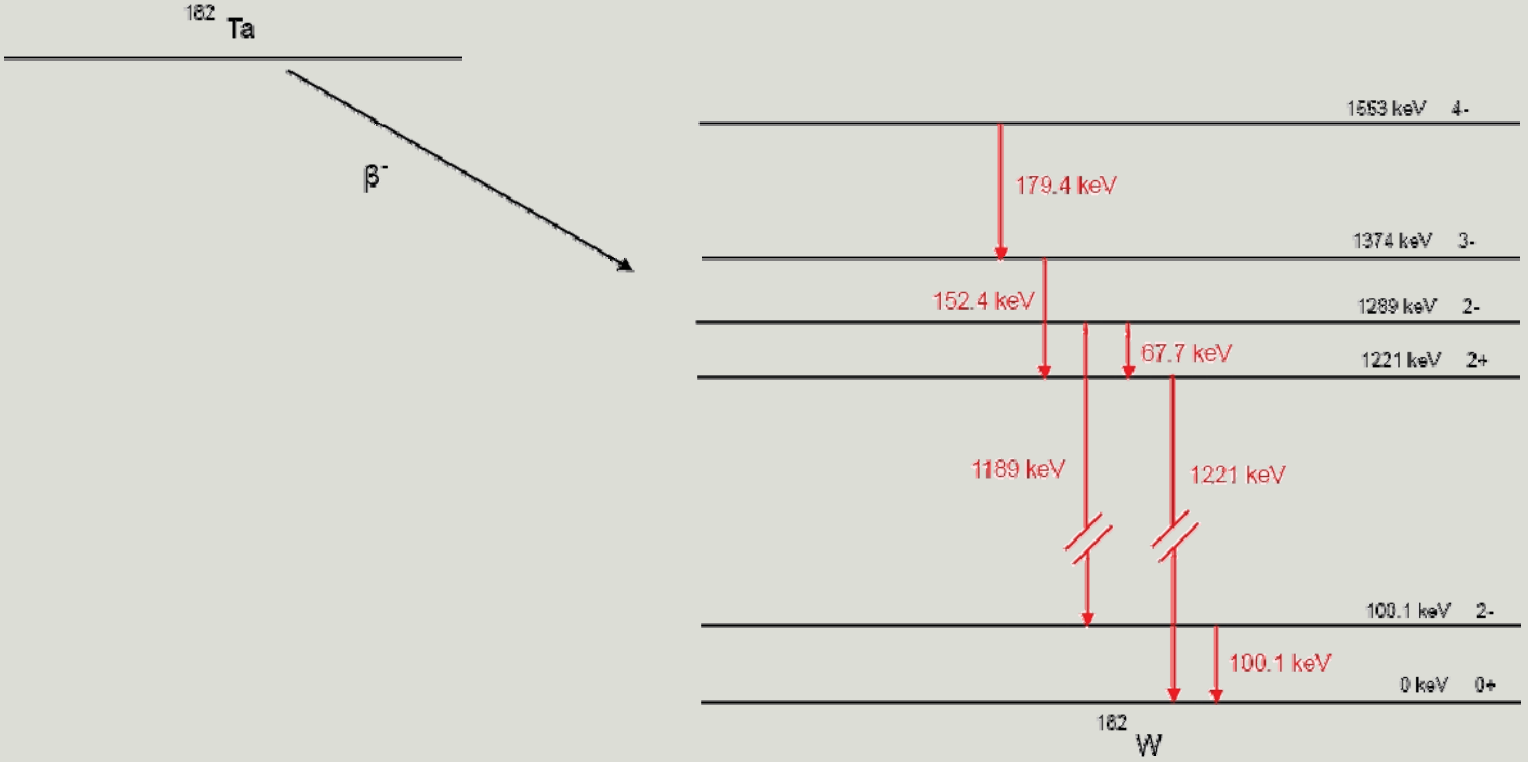


  - Both Precise and Accurate (ideal) (third figure)



Example: ^{182}Ta

- Frequently used in gamma ray detector calibration



Evaluation of ^{182}Ta Half-life

Reference	Measurement (days)	Method
1980Sc07	114.43(4)	4π ion. chamber
1973Vi13	114.740(24)	Well-type NaI(Tl)
1972Em01	115.0(2)	4π ion. chamber
1967Wa29	117.3(10)	Diff. ion. chamber
1958Sp17	118.4(5)	GM counter
1958Ke26	114.80(12)	Single-ion chamber
1957Wr37	115.05(25)	Single-ion chamber
1951Ei12	111.2(5)	Single-ion chamber
1951Si25	111(1)	Single-ion chamber
Meitner	117.5(18)	-
1947Se33	117(3)	-
Zumstein et al.	117(3)	-

Cannot identify methodology problems in experiments; all seem equally valid.

⇒Reasonable to assume accuracy

Evaluation of ^{182}Ta Half-life

Reference	Measurement (days)	Method
1980Sc07	114.43(4)	4π ion. chamber
1973Vi13	114.740(24)	Well-type NaI(Tl)
1972Em01	115.0(2)	4π ion. chamber
1967Wa29	117.3(10)	Diff. ion. chamber
1958Sp17	118.4(5)	GM counter
1958Ke26	114.80(12)	Single-ion chamber
1957Wr37	115.05(25)	Single-ion chamber
1951Ei12	111.2(5)	Single-ion chamber
1951Si25	111(1)	Single-ion chamber
1948Me29	117.5(18)	-
1947Se33	117(3)	-
Zumstein et al.	117(3)	-

Same group; different methods

Same method; different groups

⇒ Reasonable to assume independence

Evaluation of ^{182}Ta Half-life

Reference	Measurement (days)	Method
1980Sc07	114.43(4)	4π ion. chamber
1973Vi13	114.740(24)	Well-type NaI(Tl)
1972Em01	115.0(2)	4π ion. chamber
1967Wa29	117.3(10)	Diff. ion. chamber
1958Sp17	118.4(5)	GM counter
1958Ke26	114.80(12)	Single-ion chamber
1957Wr37	115.05(25)	Single-ion chamber
1951Ei12	111.2(5)	Single-ion chamber
1951Si25	111(1)	Single-ion chamber
1948Me29	117.5(18)	-
1947Se33	117(3)	-
Zumstein et al.	117(3)	-

No single data point seems unrealistically precise compared with others

⇒Reasonable to assume correct precision estimation

Evaluation of ^{182}Ta Half-life

- With the assumptions of statistical methods reasonably justified a recommended value may be obtained via these procedures
- Many Possibilities:
  - Unweighted average
  - Weighted average
  - Limitation of Relative Statistical Weights Methods (LRSW/LWM)
  - Normalized Residuals Method (NRM)
  - Rajeval Techique (RT)
  - Bootstrap Method
  - Mandel-Paule Method (MP)
  - Method of Best Representation (MBR)

Unweighted Average

- Origin
  - Maximum likelihood estimator for the mean of a normal distribution from which a sample was taken
  - Extra implied assumption: the data set is a sample from a single normal distribution
- Formula for a set of measurements $\{x_1, x_2, \ldots, x_n\}$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\text{(uncertainty estimate)} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n(n-1)}}$$

Unweighted Average

- Result for the ^{182}Ta half-life data set
  - 115.29(68) days
- Unweighted average treats all measurements equally as if all from the same distribution, however different experiments have different precisions and therefore different standard deviations and probability distributions
- Does not make use of the third assumption: the standard deviation of each measurement is well estimated

Weighted Average

- Origin
  - Maximum Likelihood estimator for the common mean of a set of normal distributions with known variances
  - Extra implicit assumption: the data set is well suited to a single mean about which each result is independently normally distributed with the standard deviation quoted in the uncertainty
- Formula for a set of measurements $\{x_1, x_2, \ldots, x_n\}$ with associated uncertainties $\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$

$$x_w = \frac{\sum_{i=1}^{n} w_i x_i}{W}, \text{ where } W = \sum_{i=1}^{n} w_i \quad \text{and} \quad w_i = \frac{1}{\sigma_i^2}$$

$$\text{(uncertainty estimate)} = \frac{1}{\sqrt{W}}$$

Note about Extension to Asymmetric Uncertainties

- Since uncertainties are to represent the standard deviation, which is the square-root of the variance, the weights of weighted averaging and all other quantities which use the uncertainties in their formulation will instead use the standard deviations calculated from the PDF defined previously, $g(x;\mu,a,b)$
- One can compute the variance of that distribution to obtain the following formula for the standard deviation

$$\text{(Standard deviation)} = \sqrt{(1-\tfrac{2}{\pi})(a-b)^2 + ab}$$

Weighted Average – Chi-Squared Test

- The weighted average makes use of all three original assumptions as well an additional proposition, the Chi-Squared test gives an indication to the validity of this extra assumption
- Theorem:
  - If $X_1, X_2, \ldots, X_n$ are normally distributed continuous random variables with means μ_i and standard deviations σ_i ($i=1, \ldots, n$) then the following quantity has a Chi-Squared distribution with n degrees of freedom

$$Y = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2$$

Weighted Average – Chi-Squared Test

- In the case of the weighted average $\mu_1=\mu_2=\ldots=\mu_n=x_w$, which is deduced from the data. The "uses up" one degree of freedom so by the previous theorem the quantity (called chi-squared after its expected distribution) should have n-1 degrees of freedom

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{x_i - x_w}{\sigma_i}\right)^2$$

- The number of degrees of freedom is simply a parameter of the chi-squared distribution which determines its mean and variance
- For a chi-squared distribution with v degrees of freedom, the mean is v and the variance is 2v
- Therefore the reduced chi-squared can be defined which should be close to unity

$$\chi^2_R = \frac{\chi^2}{n-1}$$

Weighted Average – Chi-Squared Test

- A confidence level α can be used to assign a critical value of chi-squared which, if exceeded, indicates it is reasonable to reject the assumption of weighted averaging

$$\alpha = \Pr(\chi^2 \leq \chi^2_C)$$

$$= \int_0^{\chi^2_C} \chi^2_\nu(x)dx$$, where $\chi^2_\nu(x)$ is the PDF of the chi-squared distribution with v degrees of freedom

- For example the critical reduced chi-square for five degrees of freedom at a 95% confidence level is approximately 11.05/5=2.21

Weighted Average – External Uncertainty

- The uncertainty estimate $\frac{1}{\sqrt{W}}$ is based purely upon the standard deviations intrinsic to each measurement, hence it is an internal uncertainty
- An external uncertainty based on the spread of the values can be calculated by multiplying the internal uncertainty by the square-root of the reduced chi-squared (the Birge Ratio)

$$\text{(external uncertainty estimate)} = \sqrt{\frac{\chi^2}{W(n-1)}}$$

- It is recommended for data evaluation that the higher of the internal and external uncertainties be used as the standard deviation to accompany the recommended weighted mean, although, it is unrealistic to assume that each input uncertainty is underestimated by this factor (also called scaling or inflation factor)

Weighted Average

- Result for the ^{182}Ta half-life data set
  - 114.668(81) days; reduced chi-squared=16.24
- Reduced chi-squared is very much greater than critical chi-squared.
- Indicates a problem with one or more of the assumptions about the data set
- Other methods may be attempted which try to resolve the discrepancy

Limitation of Relative Statistical Weights (LRSW/LWM)

- Origin
  - Adopted by the IAEA during a CRP on gamma and X-ray standards
- Formulation
  - A “Relative Statistical Weight” is defined to be the ratio of the individual weight of a data point to the sum of all weights.
  - Searches for outliers (original version: Chauvenet’s criterion)
  - If the data are deemed discrepant and If any data point has a relative weight greater than 50% its weight is reduced to be 50% by increasing the uncertainty of the measurement, an ordinary weighted average is then calculated.
  - If the critical chi-squared is still exceeded after this adjstment then the unweighted average is also calculated, if the weighted and unweighted averages overlap within their uncertainties then the weighted average is adopted, otherwise the unweighted value is adopted
  - If necessary, the uncertainty of the adopted result is then increased to overlap the uncertainty of the most precise value in the data set

Limitation of Relative Statistical Weights (LRSW/LWM)

- This procedure addresses the third assumption regarding estimation of standard deviation
- If one value has greatly under-estimated its uncertainty as to get more than 50% of the weighting it is corrected
- The final adjustment of uncertainty also ensures a somewhat conservative estimate of the adopted standard deviation
- Since ultimately a weighted average is still performed, the same assumptions apply but to a modified data set in which some of the uncertainties may be greater. Hence a chi-squared test can still be used to determine if one should reject the weighted average assumptions

Limitation of Relative Statistical Weights (LRSW/LWM)

- Result for the ^{182}Ta half-life data set
  - 114.62(10) days; reduced chi-squared=15.47
  - The uncertainty on 1973Vi13 was increased from 0.024 to 0.037 to lower its weight to 50%
- Increasing the uncertainty of the most precise data point raised the final standard deviation estimate and lowered the reduced chi-squared, however it is still unacceptably high
- Perhaps try another method

Normalized Residuals Method (NRM)

- Origin
  - NIM Paper by M.F. James et al. (1992)
- Formulation
  - For each data point a “normalized residual,” r_i, is calculated

$$r_i = \sqrt{\frac{w_i W}{W - w_i}}\left(x_i - x_w\right)$$, where w_i, W and x_w are as before

  - If $|r_i|$ is greater than some critical value R then the uncertainty of that point is adjusted such that $|r_i| = R$
  - Once all the required adjustments have been made an ordinary weighted average is calculated with the adjusted data set

Normalized Residuals Method (NRM)

- This again addresses the third assumption and adjusts uncertainties which may have been under estimated based on how far the point lies from the bulk of the data
- The critical value R can be approximated based on the probability, p (in percent), of one point out of n in the data set having a normalized residual greater than the critical value

$$R \approx \sqrt{1.8\left(\frac{n}{p}\right) + 2.6}$$

- Once again the chi-squared test can be applied to the modified data set since a weighted average is performed

Normalized Residuals Method (NRM)

- Result for the ^{182}Ta half-life data set
  - 114.738(44) days; reduced chi-squared=3.78
  - Uncertainties increased for 1980Sc07, 1958Sp17, 1951Ei12, 1951Si25 (with R from p=1)
- Reduced chi-squared far improved, but still greater than the critical reduced chi-squared of 2.25 for a distribution with n-1=11 degrees of freedom at a 99% confidence level

Rajeval Technique

- Origin
  - NIM Paper by M.U. Rajput and T.D. Mac Mahon (1992)
- Formulation
  - Done in three stages:
    - Population Test – checks for outliers and excludes them from the remainder of the analysis
    - Consistency Test – check the remainder of points for consistency
    - Adjustment – the uncertainty on points which appear inconsistent with the rest of the data have their uncertainties increased until the whole data set is consistent

Rajeval Technique – Population Test

- The quantity y_i is calculated for each data point $y_i = \frac{x_i - \mu_i}{\sqrt{\sigma_i^2 + \sigma_{\mu i}^2}}$, where μ_i is the unweighted mean excluding the ith point and $\sigma_{\mu i}$ is the associated standard deviation
- If $| y_i |$ is greater than the critical value of 1.96 the data point is an outlier at a 95% confidence level and is excluded
- The test can be made less severe by using a critical value of 2×1.96 (99% CL) or 3×1.96 (99.99% CL)

Rajeval Technique – Consistency Test

- The quantity z_i is calculated, which is normally distributed with mean 0 and unit variance, thus the probability of attaining values less than z_i can also be computed

$$z_i = \frac{x_i - x_w}{\sqrt{\sigma_i^2 + \sigma_w^2}}$$, where x_w is as before and σ_w is the uncertainty estimate on the weighted average

$$\Pr(Z \le z_i) = \int_{-\infty}^{z_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

- The absolute difference of this probability from 0.5 is a measure of the central deviation of the measurement, if it exceeds the critical value $0.5^{n/(n-1)}$ then the data point is considered inconsistent

Rajeval Technique – Adjustment

- Any points which were deemed inconsistent have their uncertainties incremented by adding the weighted average uncertainty in quadratutre

$$\sigma_i' = \sqrt{\sigma_i^2 + \sigma_w^2}$$

- Steps two and three are repeated until no data point is considered inconsistent
- Once the iteration is finished an ordinary weighted average is calculated on the modified data set

Rajeval Technique

- This procedure first attempts to verify our first original assumption that all measurements are accurate by looking for outliers and removing them
- It also tries to validate the third assumption as LWM and NRM did by increasing uncertainties on “inconsistent” data points
- Since it too is based on a weighted average in the end the chi-squared test can be applied
- Result for the ^{182}Ta half-life data set
  - 1958Sp17, 1951Ei12, and 1951Si25 were marked as outliers in the first stage at 99% confidence, if these points are included anyways the result is **14.761(72) days; reduced chi-squared=2.14** and half the data points have uncertainty adjustments
  - If the outliers are allowed to be excluded the result is **114.766(61); reduced chi-squared=1.50** with the four most recent measurements receiving uncertainty adjustment
- Both reduced chi-squared values are acceptable at a 99% confidence level, however major modifications were made to the uncertainties in data points to attain the final result. This is a common feature with the Rajeval Technique

Bootstrap Method

- Origin
  - Commonly employed in data analysis in medicine and social sciences
- Formulation for a set of n measurements
  - If the three original assumptions are satisfied then a Monte Carlo approach can be taken in which n points are randomly sampled from the normal distributions defined by the measurements and the median is taken of the sample
  - The median of a discrete sample $x_1, x_2, \ldots, x_n$ is the central value $x_{n/2}$ when the sample is sorted in increasing order and the number of elements is odd; and is the unweighted average of the two central elements $x_{n/2}, x_{n/2+1}$ of the sorted sample when n is even
  - This sampling procedure is repeated many times (800,000 is the default for the present implementation) and finally an unweighted average is taken of the medians
  - The uncertainty is estimated using the unbiased sample variance

$$\text{(uncertainty estimate)} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

Bootstrap Method

- An advantage to the Bootstrap Method is it has little sensitivity to outliers or very precise data points.
- The Bootstrap method does not return the mean of any probability distribution, therefore the chi-squared test does not apply here since the test determines whether one can reasonably reject the proposed common mean to a set of normal distributions
- The numeric value of the reduced chi-squared can serve as a general indicator of the consistency of the data set at the discretion of the evaluator
- Result for the ^{182}Ta half-life data set
  - 115.15(70) days; reduced chi-squared=68.57

Mandel-Paule Method

- Origin
  - Simplified approximation to the maximum likelihood estimator of the parameter μ in the measurement model of inter-laboratory experiments: $x_{ij} = \mu + b_i + e_{ij}$, where x_{ij} is the jth measurement in the ith laboratory, b_i is the error contribution for the laboratory and e_{ij} is the error contribution for the particular measurement
  - Developed by Mandel and Paule at NIST (1982)
  - Used by NIST (USA) for adopted values of standard references
- Formulation
  - The result is again a weighted average, however the weights are of the form

$w_i = \frac{1}{y + \sigma_i^2}$, where y is found as the solution to the equation $\sum_{i=1}^{n} w_i (x_i - x_m)^2 = n - 1$

with x_m being the Mandel-Paule mean, $x_m = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$

  - The square-root of y also serves as the uncertainty estimate for the method

Mandel-Paule Method

- Again the weighted average chi-squared test does not apply here because the measurement model used is different than that of weighted averaging, but the value is still as general indicator of consistency
- Result for the ^{182}Ta half-life data set
  - 115.0(21) days; reduced chi-squared=41.28
  - Note the large error and closeness to the unweighted average value (115.29 days), this is common in the MP method since y effectively evens out the weighting by being included in all values

Method of Best Representation (MBR)

- Origin
  - Developed as an alternative to other averaging techniques
- Formulation
  - The MBR builds a “Mean Probability Density Function”, $M(x)$, to represent the entire data set by calculating the unweighted mean of the individual PDFs
  - The value of the Mean PDF evaluated at a measured value x_i is interpreted as being proportional to the frequency with which that measurement is expected to occur
  - Weights are then assigned to each measurement according to its expected relative frequency and a weighted average is computed

Method of Best Representation (MBR)

The Mean PDF, $M(x) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{N}(x;x_i,\sigma_i)$ (where the ordinary normal distribution is substituted with the asymmetric normal distribution defined previously for asymmetric uncertainties) is used to define the weights

$w_i = \frac{M(x_i)}{\sum_{i=1}^{n} M(x_i)}$, which then define the mean $x_B = \sum_{i=1}^{n} w_i x_i$ (note analogy with statistical expected value). The internal uncertainty is estimated according to

$\sqrt{\sum_{i=1}^{n} w_i^2 \sigma_i^2}$, which follows from a theorem about linear combinations of normally distributed random variables, and the external uncertainty is estimated by

$\sqrt{\sum_{i=1}^{n} w_i (x_i - x_B)^2}$ (note analogy with statistical variance). As with the weighted average, the higher of the internal and external uncertainties should be used as the uncertainty estimate for the final result

Method of Best Representation (MBR)

- The MBR has the advantage that it does not modify the data set in any way, but still does not rely heavily on the first of the original assumptions: that all measurements are accurate
- Measurements which are close together will build up the Mean PDF in that region, giving a higher weight, whereas an apparent outlier would receive less weight because of its low expected frequency (but it is still not discounted entirely, which is important since later experiments may show the “outlier” was actually the only accurate measurement of the set)
- The final assumption is also still considered since the height of the peak of an individual normal distribution is inversely proportional to the standard deviation, hence the maximum contribution a measurement can make to the Mean PDF also depends on the uncertainty
- However, this assumption also plays less critical of a role since the dependence of the weights on the uncertainties is not as great as with the weighted average because in MBR both the central value and uncertainty of a measurement contribute to the weight.

Method of Best Representation (MBR) – Test of Mean PDF Model

- The MBR does rely on the Mean PDF model of the data set being an accurate description, which is not necessarily true *a priori*
- To test the model a variation on the Chi-Squared test for the weighted average is used
- The expected number of measurements above, n_e^+, and below, n_e^-, the mean is calculated by

$$n_e^- = n\Pr(X \le x_B) = n\int_0^{x_B} M(x)dx$$

$$n_e^+ = n\Pr(X > x_B) = n(1 - \Pr(X \le x_B)) = n - n_e^-$$

- This expectation is compared with the actual number of measurements above, n_a^+ and below, n_a^-, using the statistic, Q, which should have an approximate chi-squared distribution with one degree of freedom if the model is valid

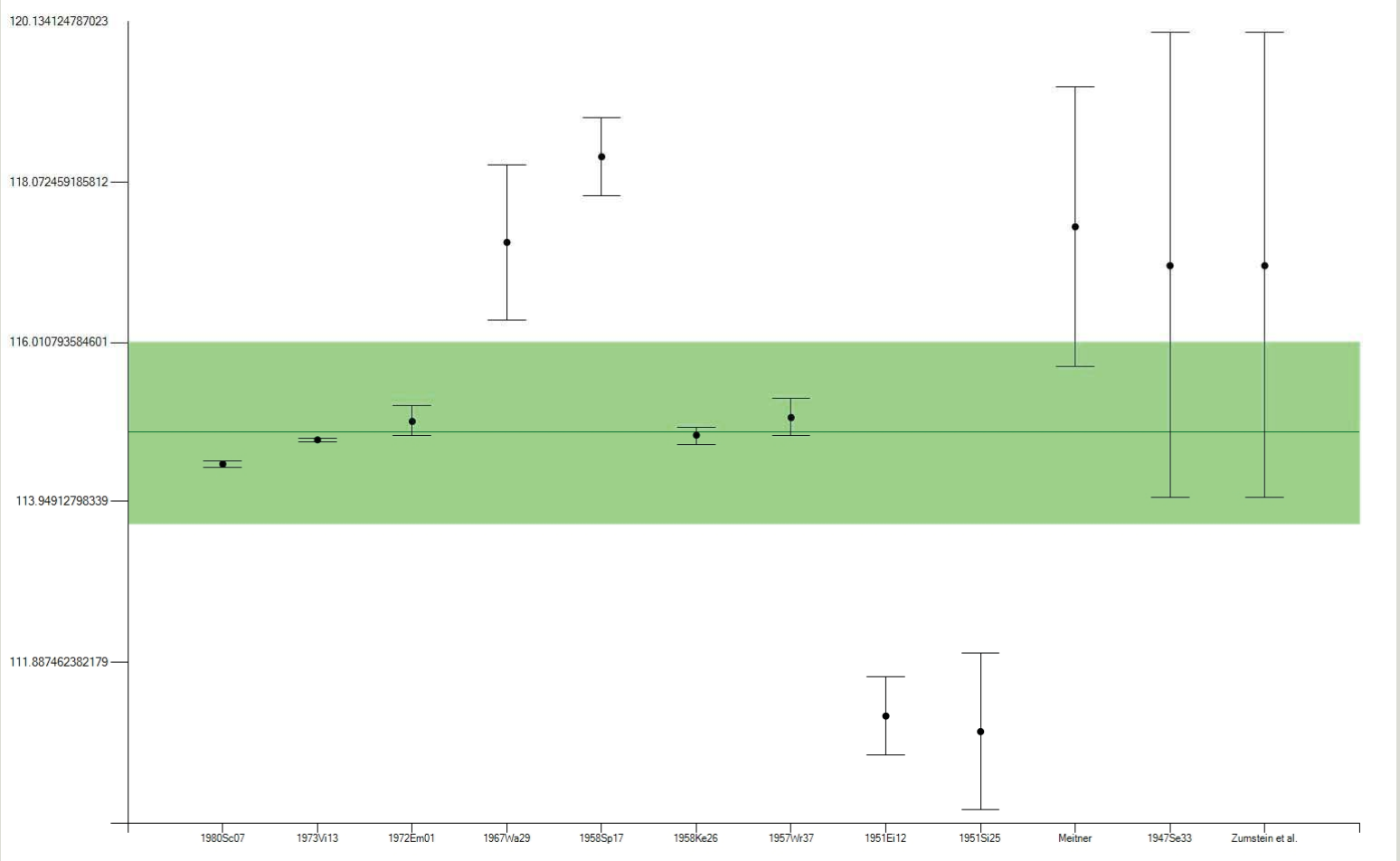
$$Q = \frac{(n_a^+ - n_e^+)^2}{n_e^+} + \frac{(n_a^- - n_e^-)^2}{n_e^-}$$

- Therefore the confidence level of the test which could reject the mean PDF model is $\Pr(X<Q)$ and thus the confidence we can hold in the model is $1 - \Pr(X<Q)$

Method of Best Representation (MBR)

- Result for the ^{182}Ta half-life data set
  - 114.8(12) days; Confidence Level=72.9%
- This result overlaps all the measurements except for the two highest and two lowest, the value also lies close to the five most precise measurements (see figure)
- The confidence level also indicates the model can be reasonably accepted
- Therefore this result could be used as a recommended value for the ^{182}Ta half-life

Method of Best Representation (MBR)

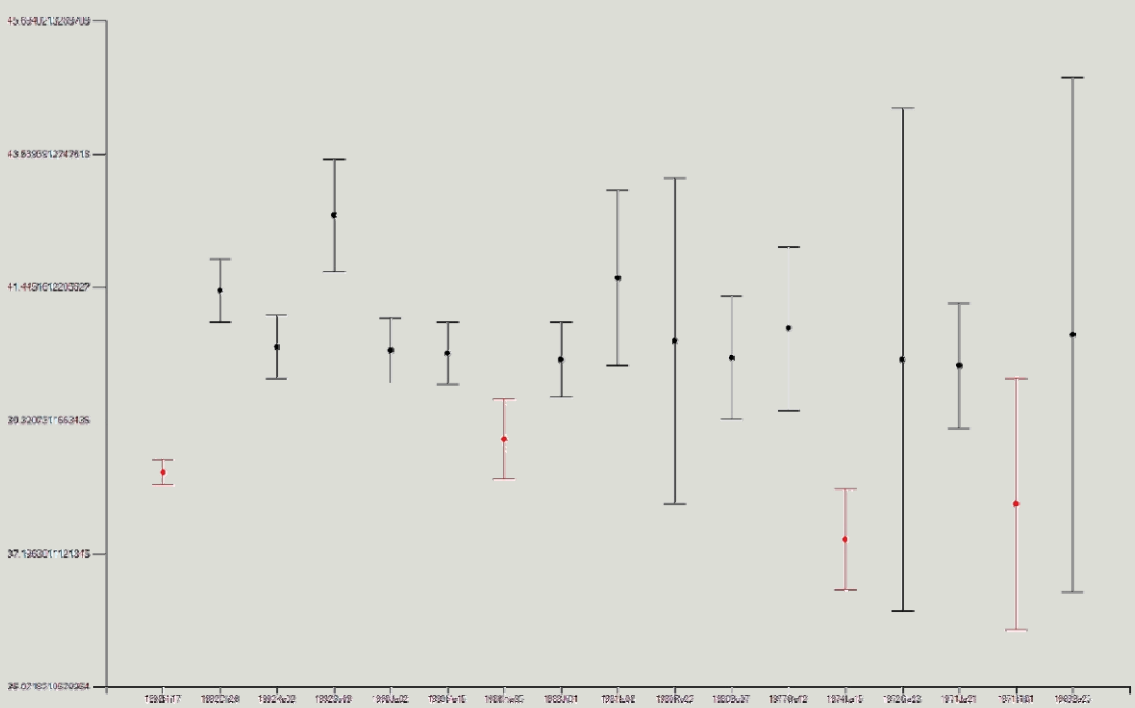


Summary of Results for ^{182}Ta Half-Life

Method	Result	Test Statistic
Unweighted Average	115.26(68)	-
Weighted Average	114.668(81)	$\frac{\chi^2}{n-1}$ =16.24
LWM	114.62(10)	$\frac{\chi^2}{n-1}$ = 15.47
NRM	114.738(44)	$\frac{\chi^2}{n-1}$ = 3.78
RT (outliers used)	114.761(71)	$\frac{\chi^2}{n-1}$ = 2.14
RT (outliers excluded)	114.766(61)	$\frac{\chi^2}{n-1}$ = 1.50 (<critical at 95%)
Bootstrap	115.15(70)	-
Mandel-Paule	115.0(21)	-
MBR	114.8(12)	72.9% acceptance confidence level
ENSDF DDEP	114.74(12) 114.61(13)	NRM for 5 most precise values. χ^2=2.2 LWM , χ^2=16!

^{182}Ta Gamma Intensity – 100.1 keV

Reference	Measurement
1998Mi17	38.5(2)
1992Ch26	41.4(5)
1992Ke02	40.5(5)
1992Su09	42.6(9)
1990Ja02	40.45(51)
1990Me15	40.4(5)
1986Wa35	39.03(64)
1983Ji01	40.3(6)
1981Is08	41.6(14)
1980Ro22	40.6(26)
1980Sc07	40.33(98)
1977Ge12	40.8(13)
1974La15	37.43(80)
1972Ga23	40.3(40)
1971Ja21	40.2(10)
1971Ml01	38(2)
1969Sa25	40.7(41)



^{182}Ta Gamma Intensity – 100.1 keV

Method	Result	Test Statistic
Unweighted Average	40.18(31)	-
Weighted Average	39.48(30)	$\frac{\chi^2}{n-1}$ =4.77; critical: 2.0, 99%
LWM	39.48(78)	$\frac{\chi^2}{n-1}$ =4.77
NRM	40.29(26)	$\frac{\chi^2}{n-1}$ =2.04
RT (outlier used)	40.28(23)	$\frac{\chi^2}{n-1}$ =1.27
RT (outlier excluded)	40.46(21)	$\frac{\chi^2}{n-1}$ =1.08 (<critical at 95%)
Bootstrap	40.28(47)	-
Mandel-Paule	40.10(93)	-
MBR	40.34(85)	73.9% :acceptance CL
ENSDF DDEP	40.3(3) 40.42(24)	NRM LWM

^{100}Pd: First 2^+ Level at 665.5 keV: Mean-lifetime measurement by RDDS

- No lifetime currently given in 2008 update in ENSDF
- New measurements:
  - 2009Ra28 – PRC 80, 044331: 9.0(4) ps
    - Cologne Plunger
  - App. Rad. & Iso. 70, 1321 (July 2012), also 2011An04: Acta Phys.Pol. B42, 807 and Thesis by V.Anagnostatou (U. of Surrey): 13.3(9) ps
    - New Yale Plunger device (NYPD)
    - Authors note statistics not as good as 2009 work, however experiment done in inverse kinematics
    - One common author (Radeck, first author of 2009 work)

^{100}Pd: First 2^+ State Lifetime

Method	Result	Comment
Unweighted Average	11.2(22)	
Weighted Average	9.7(16)	Reduced Chi-Squared=19.1 too large
MBR	10.3(20)	CL=100%
LWM	11.2(22)	Reduces to unweighted average for two points since max. weight=50%, and the data are discrepant
NRM/RT	-	Not to be performed on less than three points (recommendation by original authors)
Bootstrap	11.1(16)	Very close to unweighted average
Mandel-Paule	11.1(30)	Very close to unweighted average

- Decision to make: MBR seems the best choice, or one of the points individually?

^{222}Th Alpha Decay Half-Life

- Measurements:
  - 1970Va13: 2.8(3) ms
    - Exclude : first observation of ^{222}Th, half-life does not seem reliable
  - 1970To07: 4(1) ms
    - Exclude: stated in paper that the ^{222}Th alpha peak was very weak
  - 1990AnZu: 2.6(6) ms
    - Exclude: same experiment as 1991AuZZ
  - 1991AuZZ: **2.2(2)** ms *
  - 1999Ho28: 4.2(5) ms
    - Exclude: same group as 1999Gr28
  - 1999Gr28: **2.2(3)** ms and **2.1(1)** ms
  - 2000He17: **2.0(1)** ms
  - 2001Ku07: **2.237(13)** ms
  - 2005Li17: **2.4(3)** ms
- Could take an average of values not excluded, however 2001Ku07 is the only paper to give a decay curve which shows good statistics and measurement of decay curve for 40 half-lives and the fragment-alpha correlation method is superior to other methods.
- Only drawback about 2001Ku07: from conference proceedings!
- One can adopt 2001Ku07, increasing the uncertainty to 1% if one feels it is too precisely quoted. ENSDF value revised June 2012 based on above: 2.24(3) ms.

Conclusion and recommendations

- Many averaging procedures exist for analyzing data which satisfy three major assumptions, some rely on each assumption more or less than others and some add extra assumptions
- The evaluator should be aware of the assumptions being made when employing these techniques
- Which method returns the most acceptable result is chosen at the discretion of the evaluator, guided by available statistical tests for the methods
- For difficult data sets, methods may need to be combined to produce an acceptable result
- Averaging may not be necessary if careful analysis of the data set show adopting one value is a reasonable choice. Such analysis should be done on every data set before averaging
- A computer code by Michael Birch determines outliers and deduces averages using all the methods described here.