Evaluation of Discrepant Data, I

Desmond MacMahon
National Physical Laboratory
Teddington, United Kingdom

Evaluation of Discrepant Data

- What is the half-life of ^{137}Cs?

- Look at the published data from experimental measurements

- For greater detail: T. D. MacMahon, A. Pearce, P. Herns, Convergence of Techniques for the evaluation of Discrepant Data,
  Appl. Radiat. Isot. **60 (2004)275-281**

Measured Half-lives of Cs-137

Authors		Measured half-lives	
		in days	
		$t_{1/2}$	σ
Wiles & Tomlinson (1955a)		9715	146
Brown et al. (1955)		10957	146
Farrar et al. (1961)		11103	146
Fleishman et al. (1962)		10994	256
Gorbics et al. (1963)		10840	18
Rider et al. (1963)		10665	110
Lewis et al. (1965)		11220	47
Flynn et al. (1965)		10921	183
Flynn et al. (1965)		11286	256
Harbottle (1970)		11191	157
Emery et al. (1972)		11023	37
Dietz & Pachucki (1973)		11020.8	4.1
Corbett (1973)		11034	29
Gries & Steyn (1978)		10906	33
Houtermans et al. (1980)		11009	11
Martin & Taylor (1980)		10967.8	4.5
Gostely (1992)		10940.8	6.9
Unterweger (2002)		11018.3	9.5
Schrader (2004)		10970	20

Evaluation of Discrepant Data

- The measured data range from 9715 to 11286 days.
- What value are we going to use for practical applications?
- Simplest procedure is to take the unweighted mean:

x

- If x_i (for i = 1 to N) are the individual values of the half-life, the unweighted mean, x_u, and associated standard deviation (σ_u) are given by:

Unweighted Mean

$$x_u = \frac{\sum x_i}{N}$$

$$\sigma_u = \sqrt{\frac{\sum \left(x_i - x_u\right)^2}{N\left(N-1\right)}}$$

Unweighted Mean

- Gives the result: 10936 ± 75 days
- However, the unweighted mean is influenced by outliers in the data, in particular the first low value of 9715 days
- Secondly, the unweighted mean takes no account of different authors making measurements of different precision, so we effectively lose some of the information content of the listed data

Weighted Mean

- We can take into account the authors' quoted uncertainties σ_i, i = 1 to N, by weighting each value, using weights w_i to give the weighted mean, (x_w):

$$w_i = \frac{1}{\sigma_i^2}$$

$$x_w = \frac{\sum x_i w_i}{\sum w_i}$$

Weighted Mean

- Standard deviation of the weighted mean (σ_w) is given by:

$$\sigma_w = \sqrt{\frac{1}{\sum w_i}}$$

- And for the half-life of Cs-137, a value of 10988 ± 3 days results

Weighted Mean

- This result has a small uncertainty, but how reliable is the value ?
- How do we know that all the data are consistent?
- We can look at the deviations of the individual data from the mean, compared to their individual uncertainties
- We can define a quantity 'chi-squared'

$$\chi_i^2 = \frac{(x_i - x_w)^2}{\sigma_i^2}$$

Weighted Mean

- We can also define a 'total chi-squared':

$$\chi^2 = \sum_i \chi_i^2$$

- 'Total chi-squared' should be equal to the number of degrees of freedom (i.e. to the number of data points minus one) in an ideal consistent data set

Weighted Mean

- So, we can define a 'reduced chi-squared':

$$\chi_R^2 = \frac{\chi^2}{N-1}$$

- which should be close to unity for a consistent data set.

Weighted Mean

- For the Cs-137 data under consideration, ‘reduced chi-squared’ is 18.6, indicating significant inconsistencies in the data
- We need to look at the data again
- Can we identify the most discrepant data?

Measured Half-lives of Cs-137

Authors		Measured half-lives	
		in days	
		$t_{1/2}$	σ
Wiles & Tomlinson (1955a)		**9715**	**146**
Brown et al. (1955)		10957	146
Farrar et al. (1961)		11103	146
Fleishman et al. (1962)		10994	256
Gorbics et al. (1963)		10840	18
Rider et al. (1963)		10665	110
Lewis et al. (1965)		11220	47
Flynn et al. (1965)		10921	183
Flynn et al. (1965)		11286	256
Harbottle (1970)		11191	157
Emery et al. (1972)		11023	37
Dietz & Pachucki (1973)		11020.8	4.1
Corbett (1973)		11034	29
Gries & Steyn (1978)		10906	33
Houtermans et al. (1980)		11009	11
Martin & Taylor (1980)		10967.8	4.5
Gostely (1992)		10940.8	6.9

Weighted Mean

- Highlighted values are the more discrepant
- Their values are far from the mean and their uncertainties are small
- In cases such as the Cs-137 half-life, the uncertainty, (σ_w) ascribed to the weighted mean is much too small
- One way of taking into account the inconsistencies is to multiply the uncertainty of the weighted mean by the Birge ratio:

Weighted Mean

- Birge Ratio:

$$\sqrt{\frac{\chi^2}{N-1}} = \sqrt{\chi_R^2}$$

- This would increase the uncertainty of the weighted mean from 3 days to 13 for Cs-137, which would be more realistic.

Limitation of Relative Statistical Weights (LRSW)

- This procedure has been adopted by the IAEA in the Coordinated Research Programme on X-ray and gamma-ray standards
- A Relative Statistical Weight is defined as

$$\frac{w_i}{\sum w_i}$$

- If the most precise value in a data set (value with the smallest uncertainty) has a relative weight greater than 0.5, the uncertainty is increased until its relative weight has dropped to 0.5

Limitation of Relative Statistical Weights (LRSW)

- Avoids any single value having too much influence in determining the weighted mean although for Cs-137, there is no such value)
- LRSW procedure compares the unweighted mean with the new weighted mean; if they overlap, i.e.

$$\left| x_u - x_w \right| \leq \sigma_u + \sigma_w$$

the weighted mean is the adopted value

Limitation of Relative Statistical Weights (LRSW)

- If the weighted mean and the unweighted mean do not overlap, the data are inconsistent , and the unweighted mean is adopted

- Whichever mean is adopted, the associated uncertainty is increased if necessary, to cover the most precise value in the data set

Limitation of Relative Statistical Weights (LRSW)

- Cs-137 half-life:
- Unweighted Mean is 10936 ± 75 days
- Weighted Mean is 10988 ± 3 days
- These two means do overlap, so the weighted mean is adopted
- Most precise value in the data set is that of Dietz and Pachucki (1973) of 11020.8 ± 4.1 days
- Therefore, the uncertainty in the weighted mean is increased to 33 days to give 10988 ± 33 days

Median

- Individual values in a data set are listed in order of magnitude
- If there is an odd number of values, the middle value is the median
- If there is an even number of values, the median is the average of the two middle values
- Median has the advantage that this approach is very insensitive to outliers

- See also: J. W. Müller, Possible Advantages of a Robust Evaluation of Comparisons, J. Res. Nat. Inst. Stand. Technol., **105** (2000) 551-555; Erratum, *ibid*., **105** (2000) 781.

Median

- We now need some way of attributing an uncertainty to the median
- First we have to determine the quantity ‘median of the absolute deviations’ or ‘MAD’

$$MAD \quad = \quad med\ \{ | \, x_i - \tilde{m} \, | \} \quad for\ i = 1, 2, 3, \ldots.. N$$

$where\ \tilde{m}\ is\ the\ median\ value$

Median

- Uncertainty in the median can be expressed as:

$$\frac{1.9 \times MAD}{\sqrt{(N-1)}}$$

Median

- Median is 10970 ± 23 days for the Cs-137 half-life data already presented
- As for the unweighted mean, the median does not use the uncertainties assigned by the authors, so again some information is lost
- However, the median is much less influenced by outliers than is the unweighted mean

Evaluation of Discrepant Data

- In summary, we have:

- Unweighted Mean: 10936 ± 75 days
- Weighted Mean: 10988 ± 3 days
- LRSW: 10988 ± 33 days
- Median: 10970 ± 23 days

Evaluation of Discrepant Data, II

Desmond MacMahon
National Physical Laboratory
Teddington, United Kingdom

Evaluation of Discrepant Data

Cs-137 half-life:

- Unweighted Mean: 10936 ± 75 days
- Unweighted mean can be influenced by outliers and has a large uncertainty
- Weighted Mean: 10988 ± 3 days
- Weighted mean has an unrealistically low uncertainty due to the high quoted precision of one or two measurements; value of 'chi-squared' is very high, indicating inconsistencies in the data

Evaluation of Discrepant Data

Cs-137 half-life:

- LRSW: 10988 ± 33 days
- Limitation of Relative Statistical Weights has not increased the uncertainty of any value in the case of Cs-137, but has increased the overall uncertainty to include the most precise value
- Median: 10970 ± 23 days
- Median is not influenced by outliers, nor by particularly precise values; however this approach ignores all the uncertainty information supplied with the measurements

Evaluation of Discrepant Data

- There are two other statistical procedures which attempt to:

  (i) identify the more discrepant data, and

  (ii) decrease the influence of these data by increasing their uncertainties

- These procedures are known as the Normalised Residuals technique and the Rajeval technique

- See also: M.V. Rajput, T. D. MacMahon, Techniques for Evaluating Discrepant Data, Nucl. Instrum. Meth. Phys. Res., **A312** (1992) 289-295.

Evaluation of Discrepant Data

- Normalised Residuals technique:

  A normalised residual for each value in a data set is defined as follows:

$$R_i = \sqrt{\frac{w_i W}{(W - w_i)}} \times (x_i - x_w)$$

$$where \quad x_w = \frac{\sum x_i w_i}{W}; \quad w_i = \frac{1}{\sigma_i^2}; \quad W = \sum w_i$$

Evaluation of Discrepant Data

- A limiting value (R_0) of the normalised residual for a set of N values is defined as:

$$R_0 = \sqrt{1.8 \ln N + 2.6} \quad \mathit{for} \quad 2 \le N \le 100$$

- If any value in the data set has $|R_i| > R_0$, the weight of the value with the largest R_i is reduced until the normalised residual is reduced to R_0
- This procedure is repeated until no normalised residual is greater than R_0.

Evaluation of Discrepant Data

- Weighted mean is then re-calculated with the adjusted weights
- Results of applying this method to the Cs-137 data is shown on the next table, which shows only those values whose uncertainties have been adjusted

Author	Half-life (days)	Original Uncertainty	R_i **R_0 = 2.8**	Adjusted Uncertainty
Wiles 1955	9715	146	- 8.7	453
Gorbics 1963	10840	18	- 8.3	52
Rider 1963	10665	110	- 2.9	114
Lewis 1965	11220	47	4.9	88
Dietz 1973	11020.8	4.1	10.1	18.4
Martin 1980	10967.8	4.5	- 5.4	8.7
Gostely 1992	10940.8	6.9	- 7.4	16.4
Unterweger 2002	11018.3	9.5	3.3	15.5
New Weighted Mean	**10985**	**10**		

Rajeval Technique

- This technique is similar to the normalised residuals technique: only inflate the uncertainties of the more discrepant data, although a different statistical recipe is used
- Also has a preliminary population test which allows the rejection of highly discrepant data
- Normally makes more adjustments than the normalised residuals method, but the outcomes are usually very similar

Rajeval Technique

Initial Population Test:

outliers in the data set are detected by calculating the quantity y_i:

$$y_i = \frac{x_i - x_{ui}}{\sqrt{\sigma_i^2 + \sigma_{ui}^2}}$$

where x_{ui} is the unweighted mean of the whole data set excluding x_i, and σ_{ui} is the standard deviation associated with x_{ui}

Rajeval Technique

- Critical value of $|y_i|$ at 5% significance is 1.96
- At this stage, only values with $|y_i| > 3 \times 1.96 = 5.88$ are rejected

- Cs-137 half-life data: only the first value of 9715 ± 146 days is rejected, with a value of $|y_i| = 8.61$

Rajeval Technique

Standardised deviates, Z_i, are calculated in the next stage of the procedure

$$Z_i = \frac{x_i - x_w}{\sqrt{\sigma_i^2 - \sigma_w^2}} \quad where \quad \sigma_w = \sqrt{\frac{1}{W}}$$

Rajeval Technique

Probability integral for each Z_i

$$P(Z) \quad = \quad \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) dt$$

is determined.

Rajeval Technique

- Absolute difference between P(Z) and 0.5 is a measure of the ‘central deviation’ (CD)

- Critical value of the central deviation (cv) can be determined by the expression:

$$cv = \left[(0.5)^{\frac{N}{N-1}} \right] \quad for\ N > 1$$

Rajeval Technique

If the central deviation (CD) of any value is greater than the critical value (cv), that value is regarded as discrepant, and the uncertainties of the discrepant values are adjusted to

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

Rajeval Technique

- An iteration procedure is adopted in which σ_w is recalculated each time and added in quadrature to the uncertainties of those values with CD > cv

- Iteration process is terminated when all CD < cv

- Cs-137 half-life data: one value is rejected by the initial population test, and 8 of the remaining 18 values have their uncertainties adjusted as shown on the next table

Author	Half-life (days)	Original Uncertainty	CD **cv = 0.480**	Adjusted Uncertainty
Gorbics 1963	10840	18	0.500	74
Rider 1963	10665	110	0.498	159
Lewis 1965	11220	47	0.500	125
Dietz 1973	11020.8	4.1	0.500	28
Corbett 1973	11034	29	0.443	34
Houtermans 1980	11009	11	0.473	22
Gostely 1992	10940.8	6.9	0.500	15
Unterweger 2002	11018.3	9.5	0.499	27
New Weighted Mean	**10970**	**4**		

Rajeval Technique

Compare Rajeval technique table with that for the Normalised Residuals technique; differences are seen to be:

1. Rajeval technique has rejected the Wiles and Tomlinson value
2. Normally the Rajeval technique makes larger adjustments to the uncertainties of discrepant data than does the Normalised Residuals technique, and has a lower final uncertainty

Evaluation of Discrepant Data

- We now have 6 methods of extracting a half-life from the measured data:

Evaluation Method	Half-life (days)	Uncertainty
Unweighted Mean	10936	75
Weighted Mean	10988	3
LRSW	10988	33
Median	10970	23
Normalised Residuals	10985	10
Rajeval	10970	4

Evaluation of Discrepant Data

- Already pointed out that the unweighted mean can be influenced by outliers, and therefore is to be avoided if possible.

- Weighted mean can be heavily influenced by discrepant data that have small quoted uncertainties, and would only be acceptable if the reduced chi-squared is small, i.e. close to unity. This criterion is certainly not the case for Cs-137 half-life with a reduced chi-squared of 18.6

Evaluation of Discrepant Data

- Limitation of Relative Statistical Weights (LRSW) for Cs-137 half-life data still chooses the weighted mean, but inflates the associated uncertainty to cover the most precise value
- Therefore, both the LRSW value and associated uncertainty are heavily influenced by the most precise value of Dietz and Pachucki, which is identified as the most discrepant value in the data set by the Normalised Residuals and Rajeval techniques

Evaluation of Discrepant Data

- Median is a more reliable estimator - very insensitive to outliers and to discrepant data
- However, by not using the experimental uncertainties, the median approach is not making use of all the information available
- Normalised Residuals and Rajeval techniques have been developed to address the problems of other techniques and to maximise the use of all the experimental information available

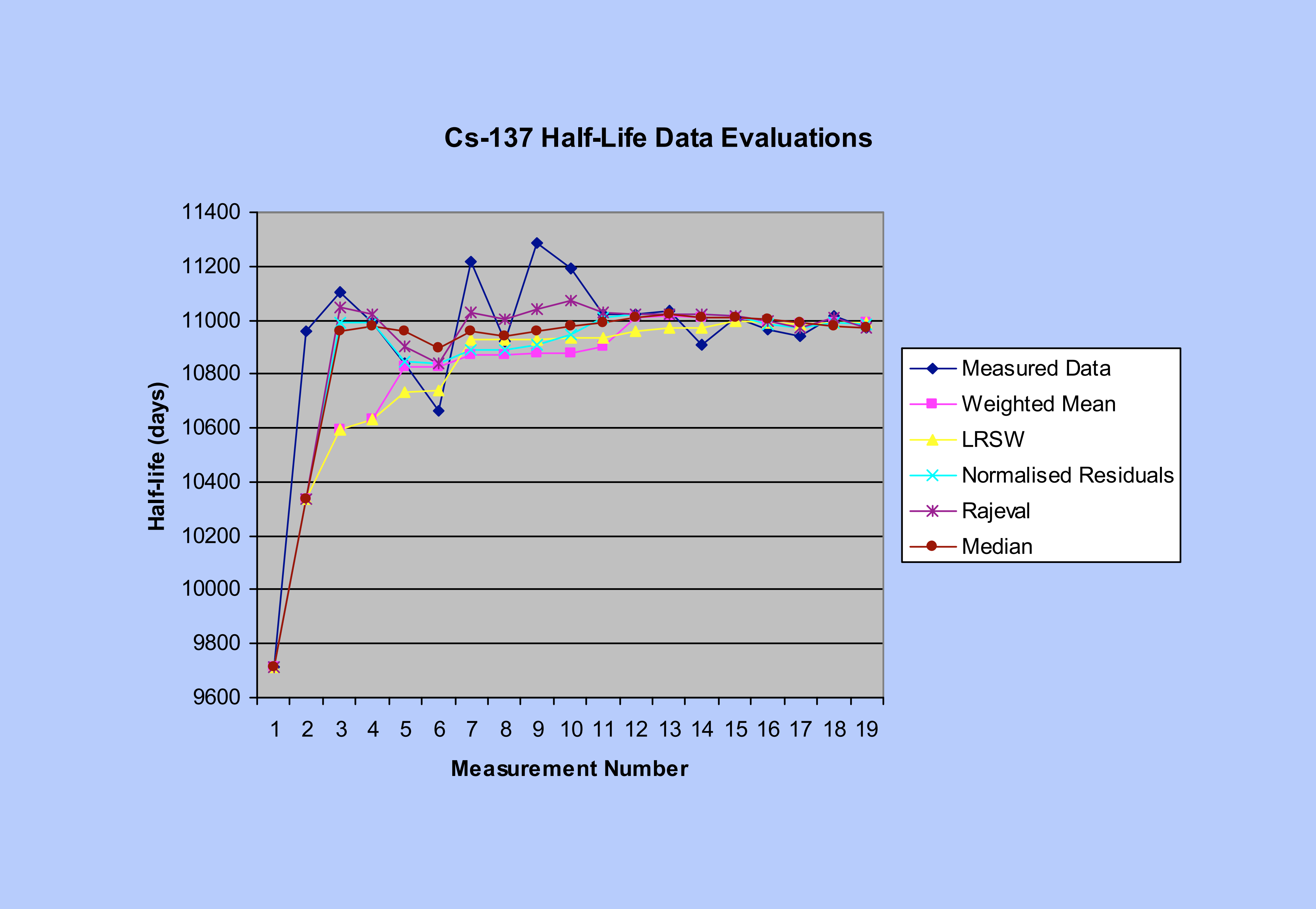
Evaluation of Discrepant Data

- Normalised Residuals and Rajeval techniques use different statistical methods to reach the same objective: to identify discrepant data and to increase the uncertainties of such data to reduce their influence on the final weighted mean
- Author's opinion: best value for the half-life of Cs-137 would be that obtained by taking the mean of the Normalised Residuals and Rajeval values, together with the larger of the two uncertainties

Evaluation of Discrepant Data

Adopted half-life of Cs-137 would be

10977 ± 10 days



Cs-137 Half-Life Data Evaluations
Half-life (days)
11400
11200
11000
10800
10600
10400
10200
10000
9800
9600
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Measurement Number
Measured Data
Weighted Mean
LRSW
Normalised Residuals
Rajeval
Median

Evaluation of Discrepant Data

- The previous image shows how the evaluation techniques behave as each new data point is added to the data set
- Left-hand portion of the plot shows that the weighted mean and LRSW values take much longer to recover from the first low and discrepant value than do the other techniques

Evaluation of Discrepant Data

- Next image shows an expanded version of the second half of the previous image, revealing in more detail how the different techniques behave as the number of data points reaches 19
- Taking into account the 19^{th} point, the overall spread in the evaluation techniques is only 18 days or 0.16%

Cs-137 half-life data - expanded version of the end of the previous plot