Averaging Methods for Experimental Measurements

Balraj Singh and Michael Birch Department of Physics and Astronomy, McMaster University, Hamilton, Canada IAEA-ICTP Workshop Aug 22-Sept 2, 016

Basic Definitions: Normal Distribution

Typically a measurement quoted as $(value) \pm (uncertainty)$ is interpreted as representing a normal distribution with mean given by the value and standard deviation given by the uncertainty.

Basically valid for statistical uncertainties.

Systematic uncertainties should be treated separately.



Basic Definitions: Asymmetric Normal Distribution

- Generalization of normal distribution to have different widths on the left and right
- Used as the interpretation for asymmetric uncertainties $\mu l b \uparrow + a$
- Same as normal distribution if a=b

Basic Definitions: Chi-Squared Distribution

- Definition:
 - Let $Z\downarrow1$, $Z\downarrow2$, ..., $Z\downarrowk$ be independent normally distributed random variables with zero mean and unit variance
 - Then the random variable $Q = \sum_{i=1}^{k} \frac{1}{k} \frac$
- The chi-squared test combines the definition above with the interpretation of experimental results as normal distributions to test the consistency of the data when taking a weighted average
 - The $\chi 12$ statistic is a random variable; we can only say data are inconsistent up to some confidence limit, i.e. $\Pr(\chi 12 \le \chi \downarrow crit 12) = 0.95$ or $\Pr(\chi 12 \le \chi \downarrow crit 12) = 0.99$
 - We recommend choosing a critical chi-squared at 95% (about 2σ)

Basic Definitions: Chi-Squared Distribution



Basic Definitions: Precision and Accuracy

- A measurement is precise if the variance when repeating the experiment (i.e. statistical uncertainty) is low
- A measurement is accurate if the central value is close to the "true value" (i.e. the systematic error is low)
- Ideally need precise and accurate measurement.
- Example: assume true value=15.02
 - Result: 15 ± 2: accurate but not precise
 - 14.55 ± 0.05: precise but not accurate
 - 15.00 ± 0.05: precise as well as accurate



All Evaluations begin with a Compilation of all available data (good and not so good)

- Compilation:
 - Complete (to the best of our ability) record of all experimental measurements of the quantity of interest
 - More than just of list of values; includes experimental methodology and other notes about how the value was determined, any reference standards used
- Evaluation:
 - The process of determining a single recommended result for the quantity of interest from a compilation
 - Compilation must be pruned to include only measurements which the evaluator believes are accurate, mutually independent and given with well-estimated uncertainties

When Do We Average?

- If the pruned dataset has one best measurement we do NOT need to average
 - e.g. best measurement could use a superior experimental technique, or agree with all other results but be more (reliably) precise
- If the pruned dataset has more than one measurement which the evaluator cannot decide between, only then we need to take an average

How Do We Average?

- Lots of ways... (see 2004Mb11: Appl. Rad. & Isot. 60, 275 for brief description)
 - Unweighted average
 - Weighted average
 - Limitation of Relative Statistical Weights Method (LWM or LRSW)
 - Normalized Residuals Method (NRM)
 - Rajeval Technique (RT)
 - Expected Value Method (EVM)
 - Bootstrap
 - Mandel-Paule (MP)
 - Power-Moderated Mean (PMM) (recent technique from EU-JRC-IRMM, Geel)
- One code to perform them all (except PMM): Visual Averaging Library (V.AveLib)

Visual Averaging Library By Michael Birch

- Available from http://www.physics.mcmaster.ca/~birchmd/ codes/V.AveLib_release.zip
- E-mail contacts: birchmd@mcmaster.ca or balraj@mcmaster.ca
- Written in Java (platform independent)
 - Requires Java Runtime Environment (JRE) available from Oracle website
 - Plotting features require GNU plot, freely available from http://www.gnuplot.info/
- Detailed documentation for all averaging and outlier detection methods
- Summary of V.AveLib features follows



Asymmetric Uncertainties in V.AveLib

- V.AveLib handles asymmetric uncertainties in a mathematically consistent way based on notes published in arXiv by R. Barlow (see e.g. arXiv:physics/0401042, Jan 10, 2004 [physics.data-an])
- All inputs are interpreted as describing asymmetric normal distributions
- To compute a weighted average, these distributions are used to construct a log-likelihood function, $\ln L$, for the mean which is then maximized
- The internal uncertainty estimate is found using the $\Delta \ln L = -1/2$ interval; external is found by multiplying by the "Birge ratio" (more on that later)

Unweighted Average

- Formula: $x = 1/N \sum_{i=1}^{\infty} 1/N = x \downarrow_i$; $\sigma \downarrow_{int} = (\sum_{i=1}^{\infty} 1/\sigma \downarrow_i 1/2) = 1/2$; $\sigma \downarrow_{ext} = \sqrt{1/N(N-1)} \sum_{i=1}^{\infty} 1/N = (x \downarrow_i - x)/2$
- Pros:
 - Simple; treats all measurements equally
 - Maximum likelihood estimator for the mean of a normal distribution, given a sample
- Cons:
 - Ignores uncertainties
- Recommended usage:
 - For discrepant data when discrepancy cannot be resolved with confidence by the evaluator

Weighted Average

- Formula: $x \downarrow w = 1 / \sum \sigma \downarrow i \uparrow -2 \sum i = 1 \uparrow N = w \downarrow i x \downarrow i , w \downarrow i = \sigma \downarrow i \uparrow -2 ; \sigma \downarrow int = (\sum i = 1 \uparrow N = 1 / \sigma \downarrow i \uparrow 2) \uparrow -1/2 ; \sigma \downarrow ext = \sigma \downarrow int \sqrt{1/(N-1)} \sum i = 1 \uparrow N = (x \downarrow i x \downarrow w) \uparrow 2 / \sigma \downarrow i \uparrow 2$
- Pros:
 - Maximum likelihood estimator for the common mean of normal distributions with different standard deviations, given a sample
 - Weighted by inverse squares of uncertainties (strictly valid for statistical uncertainties)
 - Well accepted in the scientific community
- Cons:
 - Can be dominated by a single very precise measurement
 - Not suitable for discrepant data (or data with underestimated uncertainties)
- Recommended Usage:
 - Always try this first; accept its result if the χ^2 is smaller than the critical χ^2 ; try another method otherwise

Limitation of Statistical Weights Method (LWM)

- Pros:
 - Same essential methodology as the weighted average
 - Limits maximum weight for a value to 50% in case of discrepant data
- Cons:
 - Arbitrary
 - Recommends unweighted average if the final result does not overlap the most precise measurement (within uncertainty)
- Recommended usage:
 - Sometimes useful in cases of discrepant data. (Note that DDEP group uses this as a general method of averaging)

Normalized Residuals Method (NRM)

- Primary Reference:
 - M.F. James, R.W. Mills, D.R. Weaver, Nucl. Instr. and Meth. in Phys. Res. A313, 277 (1992)
- Pros:
 - Same essential methodology as the weighted average
 - Automatically inflates uncertainties of measurements for which the uncertainty is deemed underestimated, based on a certain criterion; see manual for details
- Cons:
 - · Evaluator may not agree with inflated uncertainties
- Recommended usage:
 - Good alternative to weighted average for weakly discrepant data; again only accept if χ^2 is smaller than the critical χ^2

Rajeval Technique (RT)

- Primary Reference:
 - M.U. Rajput and T.D. MacMahon, Nucl. Instr. and Meth. in Phys. Res. A312, 289 (1992).
- Pros:
 - Same essential methodology as the weighted average
 - Automatically suggests the evaluator remove severe outliers
 - Automatically inflates uncertainties of measurements for which the uncertainty is deemed underestimated, based on a certain criterion
- Cons:
 - Uncertainty inflation can be extreme (factor of 3 or more), difficult to justify
- Recommended usage:
 - Rare. Uncertainty increases are often too severe to justify

Expected Value Method (EVM)

- Primary Reference:
 - M. Birch, B. Singh, Nucl. Data Sheets 120, 106 (2014)
 - Uses weightings proportional to a "mean probability density"
- Pros:
 - Does not alter input data
 - Robust against outliers
 - Consistent results under data transformations (e.g. B(E2) to lifetime)
- Cons:
 - Uncertainty estimate tends to be larger than weighted average (although M. Birch would argue this is a pro and the weighted average uncertainty is often too small)
- Recommended Usage:
 - Alternative to weighted average for discrepant data where the evaluator is not comfortable with uncertainty adjustments

Bootstrap

- Pseudo-Monte-Carlo, creates new "datasets" by sampling from distributions described by input data
- Pros:
 - Commonly used in bio-statistical and epidemiological applications
- Cons:
 - Resampling method, only meaningful when a large number of measurements are available, and ordinary weighted averaging method gives a large χ^2
- Recommended usage:
 - Alternative to weighted average when many measurements (~> 10) have been made

Mandel-Paule (MP)

- Primary Reference:
 - A.L. Rukhin and M.G. Vangel, J. Am. Stat. Assoc. 93 303 (1998)
 - Maximum-likelihood method which assumes additional global uncertainty
- Pros:
 - Used by National Institute of Standards and Technology (NIST)
 - Robust against outliers
- Cons:
 - Essentially increases the uncertainty of each measurement until they are all consistent
- Recommended usage:
 - Sometimes useful in the case of discrepant data, possibly covers unknown systematic errors

A Recent Averaging Method

- Power-moderated mean (PMM)
- Primary reference:
 - S. Pommé and J. Keightley, Metrologia 52, S200-S212 (2015)
 - Download an Excel spreadsheet implementing the method available as supplementary material to the article.
- Pros
 - Based on Mandel-Paule (MP) formalism.
 - Smooth transition between weighted average and unweighted average
- Cons
 - Same limitations as MP method. Has been used in some recent papers.

Internal vs. External Uncertainty

- Internal uncertainty:
 - · Uncertainty in average based on uncertainties in the input measurements
- External uncertainty:
 - Uncertainty in the average based on spread of input values (c.f. variance of a sample)
 - For weighted average and derivative methods (LWM, NRM, RT), calculated using "Birge Ratio" (square root of χ^2 ; see R. T. Birge, Phys. Rev. 40, 207 (1932))
- V.Ave.Lib choses larger of the two, but evaluator may prefer one or the other based on other considerations
- Both are listed in the full report file, which V.AveLib will save upon the user's request

What If My Data Is Inconsistent and I Don't Know Why?

- Sometimes, when there is a large number of measurements, the weighted average can give a large χ^2 even though it is not obvious which measurements are discrepant
- In this case outlier detection methods may help the evaluator decide which measurements should not be included in the average
- V.AveLib offers 3 outlier detection methods:
 - Chauvenet's Criterion
 - Peirce's Criterion
 - Birch's Criterion

Chauvenet's Criterion

- Assumes measurements are sampled from a normal distribution and removes measurements that are on the tails
- Historically used to catch typos in (hand-written) astronomical and marine data
- Cons:
 - Somewhat arbitrary
 - Does not consider uncertainties
- Recommended usage:
 - Popular with DDEP; used in LWM (by default, but can be changed to another method)

Peirce's Criterion

- Primary Reference:
 - B. Peirce, Astronomical Journal vol. 2, issue 45, 161 (1852),
 - Maximizes Prob(dataset) x Prob(outliers) by increasing the number of outliers one point at a time
- Pros:
 - Better mathematical formalism than Chauvenet's
- Cons:
 - Does not consider uncertainties
- Recommended usage:
 - General opinion is that Peirce's method is better than Chauvenet's

Birch's Criterion

- Determines which points differ from a given mean by more than a given confidence limit (default 99%)
- Pros:
 - Considers uncertainties
 - Can be reversed to give the "Consistent Minimum Variance" averaging method
- Cons:
 - Requires input result to compare data to (default is the weighted average)
- Recommended Usage:
 - Can help find outliers in large sets of data; use the EVM result as the input mean to compare data to

Example: ¹³⁷Cs Half-Life

Reference	Measurement (Days)	Comment	Reference	Measurement (Days)	Comment
1951FIAA	12053(1096)	Outlier	1973Co39	11034(29)	
1955Br06	10957(146)		1973Di01	11020.8(41)	
1955Wi21	9715(146)	Outlier	1978Gr08	10906(33)	
1958MoZY	10446(+73-37)	Outlier	1980Ho17	11009(11)	
1961Fa03	11103(146)		1980RuZX	10449(147)	Superseded by 1990Ma15
1961Gl08	10592(365)		1980RuZY	10678(140)	Superseded by 1990Ma15
1962Fl09	10994(256)		1982RuZV	10678(140)	Superseded by 1990Ma15
1963Go03	10840(18)		1982HoZJ	11206(7)	Superseded by 2014Un01
1963Ri02	10665(110)		1983Wa26	10921(19)	
1964Co35	10738(66)		1989KoAA	10941(7)	
1965Fl01[1]	10921(183)		1990Ma15	10967.8(45)	
1965Fl01[2]	11286(256)		1992G024	10940.8(69)	
1965Le25	11220(47)		1992Un02	11015(20)	Superseded by 2014Un01
1966Re13	11030(110)	Superseded by 1972Em01	2002Un02	11018.3(95)	Superseded by 2014Un01
1968Re04	11041(58)	Superseded by 1972Em01	2004Sc04	10970(20)	
1970Ha32	11191(157)		2012Be08,2013Be06	10942(30)	
1970Wa19	10921(16)	Superseded by 1983Wa26	2012Fi12	10915(55)	Superseded by 2014Un01
					Correction of NIST
1972Em01	11023(37)		2014Un01	10900(12)	measurements due to source holder movement

Example Case: ¹³⁷Cs Half-Life

- Unweighted average
 - 10960(33) d
- Weighted average
 - 10976.1(95); χ^2 = 16.05 > 1.54 = χ^2_{crit}
- LWM
 - 10976(41); $\chi^2 = 16.05 > 1.54 = \chi^2_{crit}$
- NRM
 - 10952.3(70); $\chi^2 = 4.02 > 1.54 = \chi^2_{crit}$

- RT
 - 10957.3(73); $\chi^2 = 2.62 > 1.54 = \chi^2_{crit}$
- EVM
 - 10964(71); 95.4% confidence (different goodness of fit test here)
- Bootstrap
 - 10959(26); χ² = 18.45 (not really relevant here)
- Mandel-Paule
 - 10959(97); χ² = 18.44 (not really relevant here)
- PMM
 - 10959(25); χ² = 4.01

Example Case: ¹³⁷Cs Half-Life

- Try identifying outliers using Birch's Criterion with EVM
 - Finds 1964Co35: 10738(66) and 1965Le25: 11220(47)
 - Re-do averages \rightarrow little change
- Unweighted average
 - 10958(33)
- Weighted average
 - 10975.7(94); $\chi^2 = 15.66 > 1.57 = \chi^2_{crit}$
- LWM
 - 10976(41); $\chi^2 = 15.66 > 1.57 = \chi^2_{crit}$
- NRM
 - 10952.3(66); $\chi^2 = 3.57 > 1.57 = \chi^2_{crit}$

- RT
 - 10955.4(74); $\chi^2 = 2.31 > 1.57 = \chi^2_{crit}$
- EVM
 - 10963(59); 99.3% confidence
- Bootstrap
 - 10959(25); χ² = 18.23 (not really relevant here)
- Mandel-Paule
 - 10954(61); χ² = 19.97 (not really relevant here)
- PMM
 - 10954(18); χ² = 3.96

Example Case: ¹³⁷Cs Half-Life

- Chi-squared too high to accept weighted average or NRM
- Unweighted average, NRM, RT, EVM, bootstrap, MP, PMM give similar values, but quite different uncertainties
- Choose to adopt bootstrap result (one might think that the EVM uncertainty is too large to recommend)
- Conclusion: 10959(25) (Bootstrap) or 10954(18) (PMM)
 - ENSDF: 30.08(9) y or 10986(33) (2007 update) (tropical 1y=365.2422 d)
 - DDEP: 10976(30) (Feb 2006)
 - 2004Mb11: 10981(11) d (evaluation by D. MacMahon)



²²²Th Alpha Decay Half-Life

Measurements:

- 1970Va13: 2.8(3) ms
 - Exclude : first observation of ²²²Th, half-life does not seem reliable
- 1970To07: 4(1) ms
 - Exclude: stated in paper that the ²²²Th 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 a weighted average of selected values, however 2001Ku07 is the only paper to give a decay curve showing good statistics and decay curve followed for 40 half-lives. Fragment-alpha correlation method used, superior to other methods.
- Drawback about 2001Ku07: paper in a conference proceedings, not sure whether peer-reviewed!

Example: ⁸²Zn Half-life

- 2012Ma37: PRL 109, 112501: 228(10) ms at ORNL: γ-gated growth and decay curve
- 2014Xu07: PRL 113, 03255: 177.9(25) ms at RIKEN: βγ-coin method
- 2016Al10: PRC 93, 044325: 155(17)(20) ms (155(26) ms) at ORNL (same group as 2012Ma37): growth and decay curves with gammagating. Point out that analysis depends on procedure for subtraction of background. Systematic uncertainty of 20 ms assigned by authors to account for background subtraction issues.
- WA of 2014Xu07 and 2016Al10 gives 177.7(25) ms, χ^2 =0.8

Example: ⁵²Ni Half-life

- 1994Fa06: PRC 49, 2440: 38(5) ms: GANIL
- 2007Do17: NP-A 792, 18: 40.8(2) ms: GANIL
- 2016Or03: PRC 93, 044336: 42.8(3) ms: GANIL

WA: 41.4(7) ms, χ²=16
LWM: 41.8(7) ms, χ²=11
NRM: 40.9(3) ms
UWM: 40.5(14) ms or 41.8(10) ms if 1994Fa06 is omitted.

¹⁰⁰Pd: First 2⁺ level at 665.5 keV: Mean-lifetime measurement by RDDS

- Measurements:
 - 2009Ra28 PRC 80, 044331: 9.0(4) ps
 - ⁹²Mo(¹¹B,2np),E=43 MeV; RDDS method: Cologne Plunger
 - 2012An17: App. Rad. & Iso. 70, 1321, 2011An04: Acta Phys.Pol. B42, 807 and Thesis by <u>V. Anagnostatou</u> (U. of Surrey): 13.3(9) ps
 - ²⁴Mg(⁸⁰Se,4n),E=268 MeV: RDDS method: New Yale Plunger device (NYPD)
 - Authors note: statistics not as good as in the 2009 work
 - Involves inverse kinematics
 - WA=9.7(16) ps; reduced χ^2 =19.1: too large. UWA=11.2(22) ps.
 - In evaluations, we prefer the value from 2009Ra28.

Asymmetric uncertainties: example: ¹¹⁴Ba decay

• Measurements:

- L. Capponi et al.: PRC 94, 024314 (Aug 2016): 380 (+190-110) ms from 17, α-α-α correlated events: at Univ. of Jyvaskyla
- 1997Ja12: Janas et al., NP-A 627, 119: 0.43 (+30-15) s: from decay curve for positrons: at GSI. (Value in ENSDF, 2012 update)

WA: 411 (+146-97) ms, reduced χ^2 =0.04. UWA=410(120) ms Adopt WA but uncertainty may be inflated to +190-110, minimum in the dataset NUBASE-12 (2012Au07) recommended 530(230) ms based on a symmetrization procedure applied to 1997Ja12 value.

480(220) ms from symmetrization by upper and lower bounds.

General Half-Life Evaluation Guidelines

Based on presentation by A.L. Nichols and B. Singh at the IAEA-NSDD meeting, April 2015: INDC(NDS)-0687

- Identify, accumulate and document ALL the published measurements of the half-life of the specified nuclear level(s) i.e. complete compilation of available data.
- Consider any features of each specific measurement for either rejection or increased preference, based on your experience and judgements. Examples include the following:

-acceptance or rejection of *grey* references (publications that have not been fully peer reviewed: laboratory reports; conference proceedings; sometimes the journal issue of a set of conference papers);

-measurement technique (compared with others, the technique is judged/known to be more appropriate for the half-life being addressed),

-recognised difficulties and complications (e.g. impact of impurities, detector limitations, background subtraction, dead-time losses, relative to "standards");

-known reliability or improvements in a particular measurement technique (improvements might make the date of the measurements important);

-regular measurement programme of specific half-lives for applications (normally a policy in national standards laboratories) can result in rejecting all but the most recently reported value;

Half-Life Evaluation Guidelines

-if the same author(s) determine a particular half-life based on the same measurement technique/ apparatus, only consider the most recent value in deducing the recommended value.

 Issues faced by an evaluator to derive a recommended half-life with an uncertainty at the 1o level from a set of data varying widely with measurement techniques, data handling procedures by the measurers, problems with the detail (or lack thereof) provided in a publication, unrealistically low uncertainties, particularly obvious when systematic uncertainties are ignored by the experimenters.

-reject measurements that do not quantify the uncertainty (budgets) at all;

-reject or be cautious of measurements with uncertainties that are judged to be totally unrealistic and/or incorrect;

-reject or be cautious of half-life studies that suffer from insufficient measurement time when determining activity decay as a function of time in order to quantify the slope of such a plot, and which do not provide details of counting losses;

-increase the uncertainty in a particular measurement on the basis of known limitations in the measurement technique, hopefully described adequately in the paper;

-increase uncertainties in the process of weighted-mean calculation, and subsequently recycle until the weighting of any particular half-life measurement does not exceed a prescribed level (one common practice is "no more than 50% weighting").

Half-Life Evaluation Guidelines

- Identify outliers, document and discard, based on the criteria adopted in least-squares analysis codes. V-AVELIB computer code can be used to analyse selected data.
- All acceptable half-life data to be analysed by means of various techniques.
 -define which method is the most appropriate in a certain situation;
 -role of reduced x² in such analyses needs to be discussed;
- As an overall guide:
 - -adopt WM value and uncertainty when measured half-life data are not discrepant;
 - -adopt value from other procedures when measured half-life data exhibit discrepancies;

-the recommended uncertainty should generally be no lower than the lowest uncertainty in sets of experimental half-life data that are not individually defined in terms of separated component uncertainties;

-if the statistical and systematic components of the half-life uncertainty have been quantified as separate entities in the various measurements, the recommended overall uncertainty in the half-life should be the sum of the lowest systematic uncertainty to be found in the data set and the weighted mean of the statistical uncertainty;

-the final uncertainty should not be lower than 0.01%.

 Literature coverage: some articles are in non-nuclear physics journals such as Health Physics, Geochronology and Geochemistry, and Planetary and Earth Sciences; and may not be in NSR database.

JGAMUT: Adopted Levels and Gammas

- JGAMUT: designed to handle gamma-ray energies and intensities
- Available from http://www.physics.mcmaster.ca/~birchmd/codes/ JGAMUT_release.zip

JGAMUT v1.1.2 (Jan 27, 2016)	- • ×					
Input ENS file:	Browse					
Intermediate file:	Browse					
Averages file:	Browse					
Averages ENSDF format file:	Browse					
Use tab separated intermediate file						
Use decay dataset intensity normalization (i.e. normalize single gamma to 100 instead of strongest gamma in each level to 100)						
Limit minimum uncertainty of average to be the minimum uncertainty in the data set (Gamma-by-Gamma only)						
Use non-numeric uncertainties in unweighted average (Gamma-by-Gamma only)						
Include energy shifts in fit (GAMUT only)						
Default value (GAMUT only) for non-numeric uncertainties (e.g. AP, CA), in percent:						
Output Averages in ENSDF format						
Create Intermediate File Calculate Linear Energy Shifts Remove Dataset from Intermediate File Create Averages File (Gamma-by-Gamma) Create	Averages Table (GAMUT)					
Status: Ready!						



JGAMUT: Adopted Levels and Gammas

- Intermediate file
 - Grouping of gamma-ray data from all input datasets into a tabular format
 - Warning: this grouping is not perfect and requires verification by evaluator
- Gamma-by-gamma averages
 - Performs a weighted average (or NRM or unweighted average, depending on the discrepancy of the data) of the measurements for each gamma ray
- GAMUT algorithms (based on procedure by R.B. Firestone, LBL-Report 26024 (1991))
 - Energy algorithm performs a least-squares fit to level scheme (similar to GTOL)
 - Intensity algorithm performs a chi-square minimization

JGAMUT: Additional Features

- Preprocessing of the data
 - Can correct calibration differences between datasets through linear systematic shifts of the measured energies
 - Can remove all measurements from an entire dataset from the intermediate file (allows evaluator to exclude faulty measurements)
- Output can be in the format of an adopted levels and gammas dataset
 - Warning: this output is not perfect and requires verification by the evaluator
- Mathematical detail of all features is given in the user manual