

A NEW APPROACH TO THE ESTIMATION OF ANALYTICAL PRECISION

MICHAEL THOMPSON and RICHARD J. HOWARTH

*Applied Geochemistry Research Group, Imperial College, London SW7 2BP
(Great Britain)*

(Received February 7, 1977; revised and accepted June 6, 1977)

ABSTRACT

Thompson, M. and Howarth, R.J., 1978. A new approach to the estimation of analytical precision. *J. Geochem. Explor.*, 9: 23–30.

Methods used for the estimation of analytical precision commonly suffer from two deficiencies which give rise to misleading results. These are: (1) the methods take no account of changes in absolute or relative error over the concentration range, and (2) they tend for other reasons to produce optimistically-biassed results. The difficulties can be avoided by the correct use of duplicate determinations. One method presented allows precision parameters to be estimated. The other gives rise to a simple control chart for use in geochemical analysis.

INTRODUCTION

In previous work it has been shown that when there is a wide range of concentration in a set of samples, both the absolute and relative errors in analytical determinations can vary significantly over the range (Thompson and Howarth, 1976). In such a case, the standard deviation alone cannot properly describe analytical precision. An equation or plot relating standard deviation to concentration is more appropriate. This approach leads to a realistic estimate in contrast to the usual assumption of either a constant absolute error (by using the standard deviation), or a constant relative error (by using the coefficient of variation). It has been shown that a meaningful estimation of precision over a range of concentrations can be obtained by the duplication of analyses. In addition, many aspects of the methods currently used tend to provide optimistically biassed estimates of analytical precision. Such estimates may be immediately comforting, but are ultimately unhelpful. Proper use of duplicated data can avoid these difficulties.

In this paper working instructions are presented for two simple procedures utilising duplicate analyses. A detailed discussion of the procedures with examples of their application can be found elsewhere (Thompson and Howarth, 1976). In the first method to be used, where 50 or more duplicated

results are available, the variation of standard deviation of the determination (s_c) is expressed as a function of the concentration (c) and the standard deviation at zero concentration (s_0), thus:

$$s_c = s_0 + kc \quad (1)$$

The parameters s_0 and k can be used to quantify the precision (p_c) at concentration c , by means of the definition $p_c = 2s_c/c$.

This gives the expression:

$$p_c = 2s_0/c + 2k \quad (2)$$

Thus, the value $2k$ corresponds to the precision which is observed at concentrations well above the detection limit.

In suitable cases, i.e. when there are observations at low concentrations, the practical detection limit c_d (when $p_c = 1.0$) can also be estimated from the expression:

$$c_d = 2s_0/(1 - 2k) \quad (3)$$

This method thus enables the precision characteristics of an analytical system to be estimated.

The second method is appropriate when the number of duplicate observations available is insufficient for the valid estimation of the parameters s_0 and k to be made. In this case the data are tested against an empirical standard of precision, again expressed in the form of equation (1). The method consists of a rapid graphical procedure which can be used as a useful control chart for within-batch precision when about 10 or more duplicated results are available.

The success of either of these methods depends critically upon the duplication being properly carried out, according to the following conditions:

(1) The duplicate analyses must be performed on splits of all the actual samples, or a random selection of them.

(2) Each of the two sub-samples must be taken through the whole analytical procedure as if it were an independent sample.

(3) The position of the second sub-sample in the analytical sequence must not be systematically related to that of the first, but should be distributed at random in the batch of samples passing through the analytical process. This makes the methods generally easier to use for within-batch precision studies.

(4) The data must not be rounded off too severely. At least one significant figure containing uncertainty must be retained.

(5) Sub-zero measurements when obtained must be recorded and used as such and not set to zero or other arbitrary value. The same applies to values falling below a presumed detection limit.

(6) The sub-samples should be numbered so that it is impossible for them to be identified as such at the time of analysis.

The omission of any of these steps will give optimistically-biased esti-

mates which will not truly reflect the precision behaviour of the system. A computer program for the automatic randomisation of samples and selection of duplicates has been published (Howarth, 1977).

If an assessment of the total effect of field error plus sampling error is required, then duplicate samples taken at the field site should be treated as the two sub-samples. Further splitting of the field samples is not required.

METHOD 1: FOR 50 OR MORE DUPLICATED RESULTS

Procedure

(1) From the ($N > 50$) pairs of results a_i, b_i ($i = 1, 2, \dots, N$), form lists of the pair means $(a_i + b_i)/2$ and the corresponding absolute differences $|a_i - b_i|$ (do *not* logtransform the data).

(2) Sort the list of the means into increasing order and the differences into the corresponding order.

(3) Select the first eleven results and calculate the mean of the pair means and the median (i.e. the central value, *not* the average) of the differences.

(4) Repeat this procedure for successive groups of eleven results, and obtain corresponding lists of means and medians. Ignore any terminal group of less than eleven results.

(5) *Either*: plot the medians as a function of the means and obtain the intercept and slope of the line graphically by eye; *or*: obtain the same parameters by regression. These parameters correspond respectively with s_0 and k in equation (1).

Discussion

The whole procedure is illustrated in Fig. 1, which shows means versus differences, plotted as small squares, for a set of 100 pairs of duplicated results. The data set is divided into groups of eleven individual points by the vertical lines, and the mean concentrations and median differences of each group are shown by the large squares. Regression of the medians on the means gives the values $s_0 = 5.20$ (standard error = 1.53) and $k = 0.046$ (standard error = 0.014). There are a few technical details which need emphasising. For strictly unbiased estimates, the parameters obtained by regression must be multiplied by 1.048, which is a factor derived from the properties of the half-normal distribution. However, omission of this step is unlikely to lead to problems, as the difference is rarely outside the 95% confidence limits of the estimates. The parameters are valid for estimating precision only within the concentration range spanned. Thus for meaningful estimates of s_0 (and therefore detection limit) a reasonable proportion of the samples must have concentrations approaching the detection limit. When estimating the parameters by regression it is good practice to calculate in addition the standard errors, and thereby determine whether s_0 and k are significantly greater than zero. Additional refinements

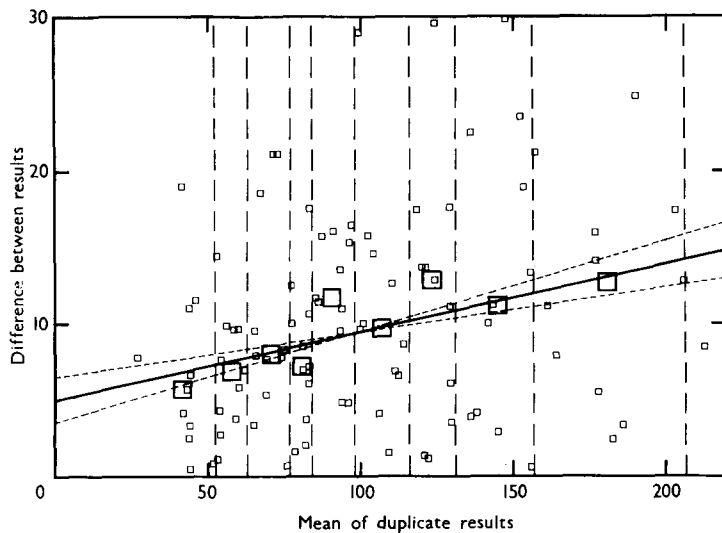


Fig. 1. Procedure for method 1 (>50 duplicated results available). Small squares are individual results. Large squares show the median absolute difference and mean concentration for each range bounded by vertical lines. The diagonal lines are the regression line with confidence limits of ± 1.0 standard error.

consist of testing linearity, and the normal distribution of error. A computer program DUPAN3 for performing all of these operations is available (Thompson, 1978).

METHOD 2: FOR 10 TO 50 DUPLICATED RESULTS

Procedure

- (1) Specify the precision required in the form:

$$s_c = s_0 + kc$$

Either s_0 or k could be zero if appropriate.

- (2) Form two new equations from this, namely:

$$d_{90} = 2.326 (s_0 + kc)$$

$$d_{99} = 3.643 (s_0 + kc)$$

- (3) Plot d_{90} and d_{99} over a suitable range of c , to form the control chart.
- (4) As in Method 1, obtain the pair means $(a_i + b_i)/2$ and absolute differences $|a_i - b_i|$ of the duplicate results.
- (5) Plot these points on the control chart.

Discussion

This procedure is illustrated in Fig. 2 for the specification $s_c = 1.0 + 0.05c$ i.e. a system with a detection limit of 2.2 units (from equation 3) and a precision approaching 10% at high concentrations (from equation 2). The lines d_{90} and d_{99} are respectively the 90th and 99th percentiles of the absolute difference between duplicates as a function of concentration, assuming a normal distribution of error. If the duplicate analytical data comply with the specification, on average 90% of the points will fall below the d_{90} line and 99% below the d_{99} line. If the precision is better, then a higher proportion will tend to fall below the lines. If the precision is worse, the opposite will tend to occur.

Statistical variations in the number of points falling in each area will occur, and it may be desirable to calculate the probability of any particular occurrence by the binomial theory. Tables I and II list the binomial probabilities associated with this type of chart, giving the chance that M or more points from a total of N will fall above the specified percentile, i.e., d_{90} or d_{99} . As an example, in Fig. 2, three of the 14 points plotted on the chart fall above the 90th percentile (d_{90}), and one falls above d_{99} . This distribution can be interpreted by reference to Table I under $N = 14$ and $M = 3$, which gives the value 0.158. This signifies that only in about 16% of cases will this distribution (or worse) occur if the

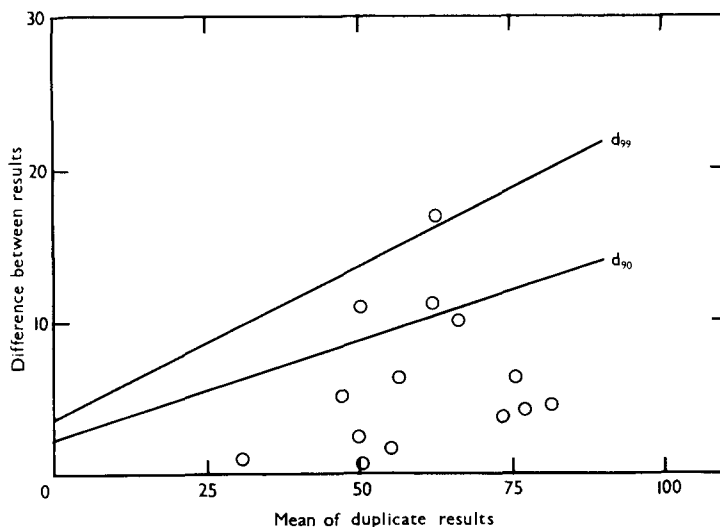


Fig. 2. Procedure for method 2 (<50 duplicated results available). Lines marked d_{90} and d_{99} are percentiles for absolute differences in the specification $s_c = 1.0 + 0.05c$.

analytical data comply with the specification. Likewise the corresponding value obtained from Table II for the single point falling above d_{99} , is about 13%. The implication is that the precision is probably worse than the specification.

For general geochemical purposes we have had printed a control chart for 10% precision (i.e. with percentile lines drawn for the specification $s_c = 0.05c$) on logarithmic axes. This chart is shown in Figure 3, and it has been consistently useful for rapid precision control over the course of several years. The points plotted in Fig. 3 are the same data as plotted in Fig. 2. However, the distribution of the points between the three areas is different, because of the more stringent specification in Fig. 3.

CONCLUDING REMARKS

A full account of the theoretical background to the method, its scope and robustness will be found in Thompson and Howarth (1976). Examples discussed include determination of lead, arsenic and zinc by emission spec-

TABLE I

The probability that M or more points out of N will fall above the 90th percentile of the chart (single event probability = 0.100000)

N	$M=1$	$M=2$	$M=3$	$M=4$	$M=5$	$M=6$	$M=7$	$M=8$	$M=9$	$M=10$	$M=11$	$M=12$
1	.100000											
2	.190000	.010000										
3	.271000	.028000	.001000									
4	.343900	.052300	.003700	.000100								
5	.409510	.081460	.008560	.000460	.000010							
6	.468559	.114265	.015850	.001270	.000055	.000001						
7	.521703	.149834	.025652	.002728	.000177	.000006	.000000					
8	.569633	.186895	.038052	.005024	.000432	.000023	.000001	.000000				
9	.612880	.225159	.052972	.008331	.000691	.000064	.000003	.000000	.000000			
10	.651322	.263901	.070191	.012795	.001635	.000147	.000009	.000000	.000000	.000000		
11	.686169	.302643	.089562	.018535	.002751	.000296	.000023	.000001	.000000	.000000	.000000	
12	.717570	.340998	.110670	.025637	.004329	.000641	.000050	.000003	.000000	.000000	.000000	.000000
13	.745813	.378855	.133683	.034161	.006460	.000920	.000099	.000008	.000000	.000000	.000000	.000000
14	.771232	.415371	.158360	.044133	.009230	.001474	.000181	.000017	.000001	.000000	.000000	.000000
15	.794109	.450957	.184061	.055556	.012720	.002250	.000311	.000034	.000003	.000000	.000000	.000000
16	.814698	.485272	.210751	.068406	.017004	.003297	.000505	.000061	.000005	.000000	.000000	.000000
17	.833228	.518215	.238203	.082641	.022144	.004667	.000784	.000106	.000011	.000001	.000000	.000000
18	.849905	.549716	.266204	.098197	.028194	.006415	.001172	.000173	.000021	.000002	.000000	.000000
19	.864915	.579735	.294555	.114998	.035194	.008593	.001696	.000273	.000036	.000004	.000000	.000000
20	.878423	.608253	.323073	.132953	.043174	.011253	.002386	.000416	.000060	.000007	.000001	.000000
21	.890591	.635270	.351591	.151965	.052152	.014445	.003273	.000613	.000095	.000012	.000001	.000000
22	.901523	.660801	.379959	.171928	.062134	.018216	.004390	.000879	.000147	.000021	.000002	.000000
23	.911371	.684873	.408043	.192731	.073113	.022508	.005773	.001230	.000220	.000033	.000004	.000000
24	.920234	.707523	.435726	.214262	.085075	.027658	.007456	.001694	.000321	.000052	.000007	.000001
25	.928210	.728754	.462306	.236409	.097934	.033400	.009476	.002261	.000459	.000079	.000012	.000001
26	.935388	.748736	.489495	.259058	.111835	.039859	.011869	.002983	.000638	.000117	.000018	.000003
27	.941950	.767401	.515419	.282102	.126557	.047057	.014668	.003871	.000872	.000169	.000028	.000004
28	.947655	.784846	.540517	.305434	.142112	.055007	.017907	.004951	.001172	.000239	.000043	.000007
29	.952699	.801128	.565040	.328952	.158444	.063717	.021617	.006247	.001550	.000333	.000062	.000010
30	.957009	.816305	.588649	.352561	.175495	.073190	.025827	.007784	.002020	.000454	.000089	.000015
31	.961048	.830435	.611414	.376170	.193201	.083421	.030563	.009588	.002596	.000611	.000136	.000023
32	.965663	.843577	.633316	.399694	.211498	.094399	.035849	.011685	.003295	.000809	.000174	.000030
33	.969907	.855785	.654342	.423056	.230318	.106109	.041704	.014102	.004134	.001058	.000238	.000047
34	.972167	.867116	.674487	.446185	.249592	.118530	.048144	.016862	.005131	.001366	.000320	.000066
35	.974068	.877624	.693750	.469015	.269251	.131636	.055183	.019990	.006304	.001742	.000424	.000091
36	.977472	.887358	.712137	.491489	.289227	.145397	.062828	.023509	.007673	.002198	.000556	.000125
37	.979724	.896359	.729659	.513553	.309454	.159780	.071095	.027441	.009256	.002746	.000720	.000168
38	.981752	.904705	.746330	.535164	.329864	.174748	.079955	.031806	.011076	.003397	.000923	.000223
39	.983577	.912410	.762168	.556281	.350394	.190259	.089434	.036521	.013148	.004165	.001172	.000282
40	.985219	.919526	.777192	.576969	.370982	.206273	.099516	.041902	.015495	.005063	.001472	.000381

TABLE II

The probability that M or more points out of N will fall above the 99th percentile of the chart (single event probability = 0.010000)

N	M=1	M=2	M=3	M=4	M=5	M=6	M=7	M=8	M=9	M=10	M=11	M=12
1	.010000											
2	.019900	.000100										
3	.029701	.000298										
4	.039404	.000592	.000001									
5	.049010	.000980	.000010	.000000	.000000							
6	.058520	.001460	.000020	.000000	.000000	.000000						
7	.067935	.002031	.000034	.000000	.000000	.000000	.000000					
8	.077255	.002690	.000054	.000001	.000000	.000000	.000000	.000000				
9	.086483	.003436	.000080	.000001	.000000	.000000	.000000	.000000	.000000			
10	.095618	.004266	.000114	.000002	.000000	.000000	.000000	.000000	.000000	.000000		
11	.104662	.005180	.000155	.000003	.000000	.000000	.000000	.000000	.000000	.000000	.000000	
12	.113615	.006175	.000206	.000005	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000
13	.122479	.007249	.000265	.000007	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000
14	.131254	.008401	.000336	.000009	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000
15	.139942	.009630	.000416	.000012	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000
16	.148542	.010933	.000508	.000017	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000
17	.157057	.012309	.000612	.000021	.000001	.000000	.000000	.000000	.000000	.000000	.000000	.000000
18	.165486	.013756	.000729	.000027	.000001	.000000	.000000	.000000	.000000	.000000	.000000	.000000
19	.173831	.015274	.000858	.000034	.000001	.000000	.000000	.000000	.000000	.000000	.000000	.000000
20	.182093	.016859	.001004	.000043	.000001	.000000	.000000	.000000	.000000	.000000	.000000	.000000
21	.190272	.018512	.001162	.000052	.000002	.000000	.000000	.000000	.000000	.000000	.000000	.000000
22	.198369	.020229	.001336	.000063	.000002	.000000	.000000	.000000	.000000	.000000	.000000	.000000
23	.206386	.022011	.001525	.000076	.000003	.000000	.000000	.000000	.000000	.000000	.000000	.000000
24	.214322	.023854	.001729	.000091	.000004	.000000	.000000	.000000	.000000	.000000	.000000	.000000
25	.222179	.025759	.001951	.000107	.000004	.000000	.000000	.000000	.000000	.000000	.000000	.000000
26	.229957	.027723	.002189	.000125	.000005	.000000	.000000	.000000	.000000	.000000	.000000	.000000
27	.237657	.029746	.002444	.000146	.000007	.000000	.000000	.000000	.000000	.000000	.000000	.000000
28	.245281	.031825	.002717	.000169	.000008	.000000	.000000	.000000	.000000	.000000	.000000	.000000
29	.252828	.033959	.003008	.000194	.000010	.000000	.000000	.000000	.000000	.000000	.000000	.000000
30	.260300	.036148	.003318	.000223	.000012	.000000	.000000	.000000	.000000	.000000	.000000	.000000
31	.267697	.038390	.003646	.000254	.000014	.000001	.000000	.000000	.000000	.000000	.000000	.000000
32	.275020	.040683	.003993	.000287	.000016	.000001	.000000	.000000	.000000	.000000	.000000	.000000
33	.282269	.043026	.004360	.000325	.000019	.000001	.000000	.000000	.000000	.000000	.000000	.000000
34	.289447	.045418	.004747	.000365	.000022	.000001	.000000	.000000	.000000	.000000	.000000	.000000
35	.296552	.047859	.005154	.000409	.000025	.000001	.000000	.000000	.000000	.000000	.000000	.000000
36	.303587	.050346	.005581	.000456	.000029	.000002	.000000	.000000	.000000	.000000	.000000	.000000
37	.310551	.052879	.006028	.000507	.000033	.000002	.000000	.000000	.000000	.000000	.000000	.000000
38	.317445	.055455	.006497	.000563	.000038	.000002	.000000	.000000	.000000	.000000	.000000	.000000
39	.324271	.058075	.006986	.000622	.000043	.000002	.000000	.000000	.000000	.000000	.000000	.000000
40	.331028	.060737	.007497	.000686	.000048	.000003	.000000	.000000	.000000	.000000	.000000	.000000

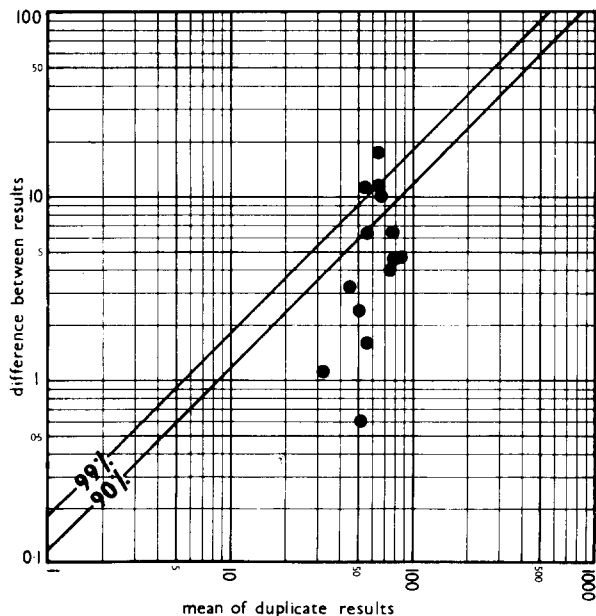


Fig. 3. Precision chart used in Applied Geochemistry Research Group.

trography, colorimetry and atomic absorption respectively. Computations were carried out on the Imperial College Computer Centre CDC 7314/6400 facility.

REFERENCES

- Howarth, R.J., 1977. Automatic generation of randomised sample submittal schemes for laboratory analysis. *Comput. Geosci.*, 3: 327–334.
- Thompson, M., 1978. Interpretation of duplicated data in geochemical analysis. *Comput. Geosci.*, in press.
- Thompson, M. and Howarth, R.J., 1976. Duplicate analysis in geochemical practice (2 parts). *Analyst*, 101: 690–709.