Sawa Medical Journal

Vol.0 No.1 2025

Sawa Medical Journal

Reprint from

A new and accurate equation for constructing a calibration curve and handling Clinical, pharmaceutical, and all analysis data

Dheyaa Yahaia Alhameedi

Department of Anesthesiology, College of Health and Medical Techniques, Sawa University, Almuthana, Iraq.

Abstract

A simple and accurate regression equation was established to construct a regression curve that agreed with analytical requirements. This method uses new equations for the slope and y-intercept based on a new criterion, which is the sum of the absolute relative errors of the x-values (SARE). This paper derived new parameters SARE and AARE to determine the best linear relationship between the variables. The idea in this paper is that the best line fit for the points is that it has the smallest sum of the absolute relative error of the value of x. The SARE and AARE parameters can be used to find the best line and best dynamic range or working range to use in the experiments.

Key words: AARE, DH Equation, Regression equation, SARE

Introduction

Regression equation

The calibration curve is necessary in most measurements of the experiments. It is a set of operations that detects the relationship between the instrument reading of the experiments (e.g., the response of an instrument) and the accepted values of the standard (e.g., the amount present of analyte). Many analytical methods require calibration curves. This typically involves using a set of standards containing a known amount of the analyte of interest, reading the instrument's response to each standard, and finding the relationship between the instrument's response and the amount of the analyte (calibration curve or regression curve). This relationship is then used to convert the measurements made on the test samples into estimates of the amount of analyte present, for example is the labbert- beer plot of absorbance against concentration. [1,2]

There is more than one line can pass through a set of points. As example four data points in table (1). Several lines can be fitting to the four data points. Figures (1) and (2) show two lines possibilities.

Table 1.

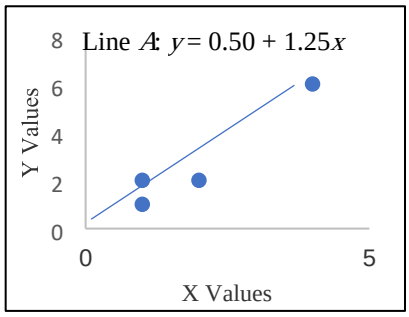

Fig 1. Line A

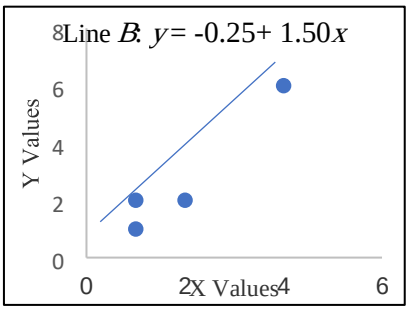

Fig 2. Line B

To determine how a line fits to the data, the errors (e) must be computed using the same line to quantitatively measure the y values of the data points. The best line is that has the smaller sum of squared errors ($\sum$ e2), [3,4] Figure 2 shows line B has the better data. The least-squares criterion is the line that achieves the smallest sum of squared errors, and it is the best line fitting to the points. Table (2) shows the result.

Table 2. Sum of squared errors to two lines.

Linei *A*: *y* = 0.50 + 1.25*x*					Linei *B*: *y* = -0.25+ 1.50*x*		
y^-	*e*	e^2	X	Y	y^-	*E*	e^2
1.75	-0.75	0.5625	1	1	1.25i	-0.25	0.0625
1.75	0.25	0.0625	1	2	1.25	0.75	0.5625
3	-1	1.0000	2	2	2.75	-0.75	0.5625
5.5	0.5	0.2500	4	6	5.75	0.25	0.0625
	Sum	1.8750				Sum	1.2500

Where y- is computed by the regression equation of the line at each X value and e (error) is the difference between y and y^- ($y - y^-$)

There are equations for determining the line with the smallest value of the sum of squared errors ($\sum e^2$), where $y = b_1x + b_0$, b_1 is the slope, and b_0 is the y-intercept.

For, the number of data points (n), b_0, and b_1, can be computed using the following equation: [5]

$$b_1 = \frac{s_{xy}}{s_{xx}}$$

$$0= \frac{1}{n}(\sum y_{i-b_1} \sum x_i) \qquad b$$

Where, s_{xx} ,s_{xy} called computing formula

$$S_{xx} = \sum x_i^2 - \left(\sum x_i\right)^2 /n$$

$$S_{xy} = \sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)/n$$

b_1 and b_o are the slope and y-intercept of the regression equation for the best line with least squared errors. For an example, the regression equation was computed for the experiment established for spectrophotometric determination of chloramphenicol. [6] Tables (3) shows the results.

Table 3. b_0 and b_1 for the method

	Y	Xy	^{2}	S_{xx}	S_{xy}	b_0	b_1	X
					x			
	0.1	0.032	0.0032	0.01	180.8092	10.1004	0.0559	0.0228
		1	0.083	0.083	1			
		2	0.131	0.262	4			
		3	0.188	0.564	9			
		4	0.244	0.976	16			
		5	0.299	1.495	25			
		6	0.354	2.124	36			
		7	0.409	2.863	49			
		8	0.464	3.712	64			
		9	0.542	4.878	81			
		10	0.598	5.98	100			
		11	0.63	6.93	121			
		12	0.685	8.22	144 m	78.1	4.659	38.0902 650.01

The regression equation is y = 0.0559x + 0.0228

To plot the line, two independent variables (x) were chosen, and the dependent variable (y) was computed from the equation (y^-). Table (4) and figure (3) show the results.

Table 4. Two points

X	y-
0.1	0.02839
12	0.6936

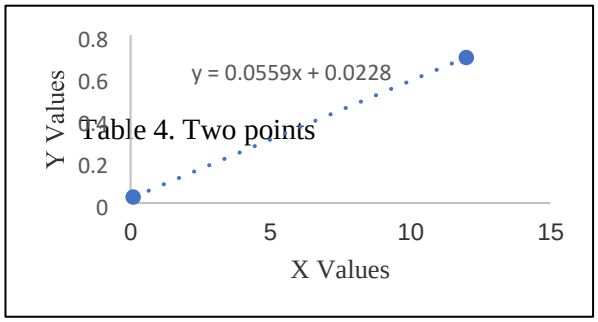

Fig 3. Regression line of method

The line that best fits the data points according to the least-squares criterion (regression curve) is the line that has the smallest sum of square errors. Table (5) shows the sum of the square errors for the experiment.

Table 5. Sum of square errors for method .

Y_i	y_i	E	e^2	$\sum e^2$
	0.03016	0.00184	$3.3856*10^{-06}$	0.000 615
	0.0802	0.0028	$7.84*10^{-06}$	
	0.1358	-0.0048	0.00002304	
	0.1914	-0.0034	$1.156*10^{-05}$	
	0.247	-0.003	$9*10^{-06}$	
	0.3026	-0.0036	$1.296*10^{-05}$	
0.03	0.3582	-0.0042	$1.764*10^{-05}$	
0.032	0.4138	-0.0048	$2.304*10^{-05}$	
0.083	0.4694	-0.0054	$2.916*10^{-05}$	
0.131	0.525	0.017	0.000289	
0.188	0.5806	0.0174	0.00030276	
0.244	0.6362	-0.0062	$3.844*10^{-05}$	
0.299	0.6918	-0.0068	$4.624*10^{-05}$	
0.354				
0.409				

0.464

0.542

0.598

0.63

0.685

The sum of the squared errors was (0.000615), and no other line can give a smaller value than (0.000615).

1.2 Coefficient of Determination

The coefficient of determination which is symbolized by the (r^2), is an important quantity that measures the fraction of the observed variation in y that is explained by the linear relationship. The value of r^2 between 0

t0 1. [7] The equation below explains how r^2 calculated. [8]

$$r^2 = \frac{SSR}{SST}$$

Where, *SSR* is **Regression sum of squares,** which explains by:

$\hat{y}_i$, is the computed y by the

equation. $\bar{y}$, is the mean of

y observed

$$SSR = \sum(y_i - \bar{y})^2$$

SST is total sum of squares**,** which explains by:

$$SST = \sum(y_i - \bar{y})^2$$

Analysts decide that the best range for the regression curve is one that gives an r^2 value close to 1. Table (6) shows using equations above for computing r^2 to chloramphenicol experiment.

(6)

Table 6. R^2 for method.

X	Y	$\hat{y}$	$y-\bar{y}$	$(y-\bar{y})^2$	$(\hat{y}-\bar{y})$	$(\hat{y}-\bar{y})^2$
	0.032	0.030	-0.285	0.081	-0.2869	0.0823
1	0.083	0.080	-0.234	0.054	-0.2368	0.0561
2	0.131	0.135	-0.186	0.034	-0.1812	0.0328
3	0.188	0.191	-0.129	0.016	-0.1255	0.0157
4	0.244	0.247	-0.073	0.005	-0.0699	0.0048
5	0.299	0.302	-0.018	0.0003	-0.0142	0.0002
6	0.354	0.358	0.036	0.0013	0.0413	0.0017
7	0.409	0.414	0.091	0.008	0.0970	0.0094
8	0.464	0.469	0.146	0.0215	0.1526	0.0233
9	0.542	0.525	0.224	0.0505	0.2083	0.0433
10	0.598	0.581	0.280	0.0788	0.2639	0.069
11	0.63	0.636	0.312	0.0978	0.3195	0.1021
12	0.685	0.692	0.367	0.1353	0.3752	0.1408
$\bar{y} = 0.35838461$	$\sum(y-\bar{y})^2 = 0.565027077$			$\sum(\hat{y}-\bar{y})^2 = 0.564994492$		**r^2= 0.9999**

0.1

Problem and suggestion

1.1. The problem

In the least squares criterion, there is a problem when points of concentration of a substance (x-value) are plotted against the physical quantity associated with the concentrations (y-value), if the same regression equation is used to calculate the unknown concentration of the same substance, there is an error in the concentration recovery percentage. For example, if the concentration of chloramphenicol is taken as 1 ppm (xvalue) and its corresponding absorbance is 0.083 (The observed y-value) as in the table above and if the concentration is assumed to be unknown and the regression equation is used to calculate it, the result will be 1.08 ppm with a relative error of 7.80% and a recovery percentage of 107.80. because of the error, or called (residuals). Residuals, which are represented by (e), are the differences between the observed reading and measured values from the equation of the response variable. [9]

$$Residual = e_i = y_i - \hat{y}_i$$

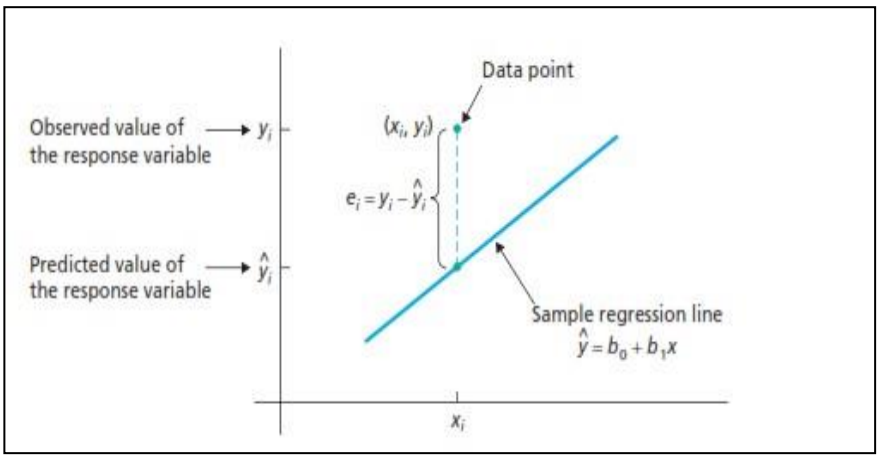

Fig 4 The residuals

In fact, this error is inevitable because there is no exact linear relationship between two variables therefore, the best line passing between points is the one with the smallest sum of squared error according to the least squares criterion. (10)

Does this equation provide the best application for determining the concentration (x-value)? And is the effect of the residuals on a small concentration the same as that on a large concentration? To answer this question, let us return to Table 3 and assume that the absorbance of the solution with a concentration of 1 ppm changes from 0.083 to 0.09, with a difference equal to 0.007, and the absorbance of the solution with a concentration of 10 ppm changes from 0.598 to 0.605 with a difference equal to 0.007, the line will change to give new equations. If the equations are used to calculate the concentration and the relative error in two cases, the result is the calculated concentration for 1 ppm will change from 1.08 to 1.18 ppm and the relative error from 7.8% to 18.06% and for 10 ppm will change from 10.3 to 10.41 ppm and relative error from 2.97% to 4.06%. The results are shown in Table 7.

Table 7. The relative error after occur changing in absorbance.

Before the change y = 0.0559x + 0.0228				After the change y = 0.0558x + 0.0241			
Conc.(X)	Abs(y)	Conc. found	E%	Abs(y)	Conc. found	E%	Rec.%
1	0.083	1.08	7.8	0.09	1.18	18.06	118.06
10	0.589	10.3	2.97	0.605	10.41	4.06	104.06

As shown in the above results, the effect of the error on the smaller value of x is greater than that for a large x value.

In fact, the difference in the effect of the error on the smallest value of concentration or the value of X is due to the idea of the least-squares criterion focuses on minimizing the absolute error of the computed y-value. Since the effect of two identical absolute errors in the y-value on the computed x-value (when the

y-value is used to calculate x) varies with the magnitude of the x-value, using the least square criterion does not give an accurate result, especially at small x-values.

Analysts focus on minimizing the relative error in calculating the concentration or value of x, especially for smaller values of x. [3] Therefore, this study focuses on finding a new equation to calculate the value of x with the smallest relative error.

Analysts use standard materials to measure the value of a material property associated with several concentrations, such as absorbance, and then plot a regression curve. Analysts seek to know the working range or linear dynamic range, which is the set of points that are related to each other by a linear relationship, in other words, the concentration range over which the analyte can be determined using a calibration curve. [11] Some higher or lower points deviate from that relationship for a variety of chemical or physical reasons. Analysts use the r^2 criterion to find the working range by determining how close the r^2-value is to 1. But there is a problem with this method because there is no rule that decides whether or not to exclude the point, by comparison with the amount of change in the value of r^2. For example, if a measurement point is added to the working range and it is found that the r^2 -value has changed from 0.9997 to 0.9996, is this difference (0.0001) enough for us to decide to exclude the point? In fact, there is no clear rule for this. Therefore, this study developed an accurate criterion to know the appropriate working range to be used and the possibility of adding or excluding a measurement point.

2.2 The suggested solution

This study suggests a new method for computing the regression equation in an easy manner and provides the best result for the smallest x values. The symbol of this equation is suggested to be (DH Equation).

The known regression equation is based on the sum of the square error to the y value, whereas the suggested equation based on the new criterion is the sum of the absolute relative error to the x value.

The absolute relative error is used instead of the squared error, because the amount of error contribution to the sum of absolute errors is different from the sum of squared errors. For example, an error of 1.1 becomes greater when squared (1.21), while an error of 0.11 becomes smaller when squared (0.0121).

Therefore, the use of the absolute value is appropriate to make the contribution of the relative error to the sum of the relative errors proportional to the value of the relative error. An equation was derived to calculate the sum of the absolute relative errors (ASRE) and average of absolute relative error (AARE).

$$y - \text{SARE}| = 100 \sum \left| \frac{i \quad b_o}{x} - 1 \right|$$

$$b_1 \; i$$

$$AARE| = \frac{100}{n} \sum \left| \frac{y_i - b_o}{b_1 x_i} - 1 \right|$$

The AARE value must be smaller than five because the relative error for the individual value must be between -5 and +5 [3]. The best line and most suitable for points is the one that has the smallest (SARE). The SARE and AARE for the data in table 3 show in Table 8.

Table 8 The SARE and AARE for the data in table 3

y = 0.0559x + 0.0228	
ASRE	AARE
91.77	7.06

SARE is not the smallest value because this line depends on the smallest sum of the square error not the sum of the absolute relative error. Therefore, a new equation was developed (DH Equation) to find the line with the smallest SARE from the known equation. The slope (b_1) and y-intercept (b_o) to (DH Equation) was defined as

$$b^1 = \frac{\sum y_i - n y_1}{\sum x_i - n x_1}$$

$$b_o = \overline{y_1 \sum x_i - x_1 \sum y_i \over \sum x_i - n x_1}$$

2.3 DH Equation VS Known Equation

Some experimental data were used to compare the DH equation with a known equation with SARE and AARE.

Table 9. The comparison between DH Equation and known equation for the experiment 1[6]

X	0.1	l	2	3	4	5	6	7	8	9	10	11	12
Y	0.032	).083	).131	0.188	0.244	).299	).354	).406	).464	0.542	0.598	0.63	).685

Equation	DH Equation y = 0.0552x + 0.0265	Known equation y = 0.0559x ⊦ 0.0228
SARE	24.95	)1.77
AARE	1,92	7.06

Table 10. The comparison between DH Equation and known equation for the experiment 2 [12]

X	1	2	3	4	5	6	7	8	9
y	0.061	0.111	0.166	0.236	0.299	0.337	0.399	0.462	0.522

Equatio	DH Equation	Known equation
	y = 0.0568x + 0.0042	y = 0.0577x - 0.0006
SARE	21.98	24.45
AARE	2,44	2.72

Table 11. The comparison between DH Equation and known equation for the experiment 3 [13]

X	0.1	0.5	1	2	3	4	5	6
y	0.088	0.183	0.285	0.528	0.757	1.009	1.222	1.446

Equatio	DH Equation	Known equation
	y = 0.2314x + 0.0649	y = 0.2318x + 0.0639
SARE	9.88	14.55
AARE	1,23	1.82

Table 12. The comparison between DH Equation and known equation for the experiment 4 [14]

X	1	2	3	4	5
Y	0.262	0.539	0.762	1.022	1.334

Equation	DH Equation y	Known equation y
	= 0.2685x - 0.0465	= 0.2708x - 0.0534
SARE	5.73	7.68
AARE	1,43	1.92

Table 13. The comparison between DH Equation and known equation for the experiment 5 [15]

	0.133	0.283	0.565	0.828	1.125	1.391
Equation	DH Equation	Known equation		y	=	
3.527x - 0.0081	y = 3.4895x - 0.0003					
SARE		7.17			9.42	
AARE		1,19			1.57	

The above examples show that the SARE and AARE of the DH Equation are smaller than those of the normal equation.

2.4 SARE and AARE VS R^2

SARE and AARE can be useful for determining the best range of data where the best range has an AARE nearest to zero and less than 5, also when SARE plummets more than 5 with exclusion of a point from the beginning or end range. This indicates the possibility of excluding that point because it has a high relative error (more than five).

As example, when two ranges of data 0.1-12 and 1-12 are taken from table 9 and from the results that were shown in the table below, we conclude the possibility to delete (0.1) from the range because SARE and AARE decrease significantly. In addition, the DH equation has shown the ability to provide fewer values of SARE and AARE.

Table 14 Shows the parameters of equations of range 0.1-12 and 1-12

Range from 0.1 to 12				
Equation	Formula	R^2	SARE	AARE
DH Equation	y = 0.0552x + 0.0265		24.95	1.92
Normal equation	y = 0.0559x + 0.0228	0.9986	91.77	7.06
Range from 1 to 12				
Equation	Formula	R^2	SARE	AARE
DH Equation	y = 0.055x + 0.028		24.37	2.03
Normal equation	y = 0.056x + 0.0214	0.9983	26.79	2.23

As shown in Table 14, when the range changes from 0.1-12 to 1-12, the r^2 of the normal equation changes from 0.9986 to 0.9983. This means that point 0.1 should not be exclude and the best range is 0.1-12.

However, this is not correct because when calculating the smallest point value (0.1, suppose the smallest point value is unknown and is calculated from the normal equation where the y-value is 0.032), the result is 0.16 with a relative error of 65%. Thus, this point must be excluded.

This is evident when using SARE and AARE, as SARE was reduced from 91.77 to 26.79 when the range was changed from 0.1-12 to 1-12 with a difference equal to 65, so SARE tells us precisely that the range should be 1-12.

On other hand, when using the DH Equation and changing the range from 0.1-12 to 1-12 was found that SARE only change from 24.95 to 24.37 with a difference of 0.58 only, that means in some cases the DH equation reduced the error of the smallest point and allow the range to be wider, in general DH Equation give the smaller sum of relative error and the SARE and AARE parameter give the best indicate to know the best range.

Another example is when the data in table 11 are taken and two ranges of data 0.1 -6 and 0.05-6 are taken. Table 15 shows the SARE and AARE values in the range of 0.05 6.

Table 15 Shows the parameters of equations of range 0.1 to 6 and 0.05-6

Range from 0.1 to 6				
Equation	Formula	R^2	SARE	AARE
DH Equation	y= 0.2314x+0.0649		9.88	1.23
Known equation	y= 0.2318x+0.0639	0.9997	14.55	1.82
Range from 0.05 to 6				
Equation	Formula	R^2	SARE	AARE
DH Equation	y= 0.2376x+0.0481		87.45	9.72
Known equation	y= 0.2327x+0.0599	0.9997	132.56	14.73

As shown from the above results, r^2 does not change (0.9997) with a difference of 0; thus, the question is whether this point is excluded. The answer is that there is no clear rule, so it may seem that the difference is small or no difference, and that the decision is not to exclude the point.

However, this is not correct because when calculating the smallest point value (0.05, suppose the smallest point value is unknown and is calculated from the normal equation where the y-value is 0.06), the result is 0 with a relative error of approximately -100, this means that it must exclude the point.

The data from SARE and AARE suggested that the range should be 0.1-6 because the SARE rose from 14.55 to 132.56 when they modified the range from 0.1-6 to 0.05-6 with a difference equal to 118. Again, DH Equation gives the smallest SARE and AARE.

Parameters	T	DH Equation	Known Equation
Slop (b_1)	a	$\frac{\sum y_i - n y_1}{\sum x_i - n x_1}$	$\frac{\sum x_i y_i - (\sum x_i)(\sum y_i)/n}{\sum x_i^2 - (\sum x_i)}$
	b		
	l		

e 16 the comparison between DH Equation and Normal Equation.

2

/n

y-intercept (b_o) $y_1 \sum x_i - x_1 \sum y_i \quad {}_{n}{}^{-1}(\sum y_{i-b1} \sum x_i)$

$\sum x_i - n x_1$

R^2 $\sum(\hat{y}_i - \bar{y})^2$

$\sum(y_i - \bar{y})^2$

R $\sum x$

$_i y_i - (\sum x_i)(\sum y_i)/n$

$\sqrt{[\sum x_i^2 - (\sum x_i)^2/n][\sum y_i^2 - (\sum y_i)^2 \quad /n]}$

SARE $y_i - b_o \quad 100 \sum \left| \frac{}{x} - 1 \right.$

|

b_1 i

AARE 100 $y_i - b_o$

$\frac{}{n} \Sigma \ \left| \frac{}{1 \ i} \right. -$

Conclusions

In this paper, a new regression equation (DH equation) was suggested. It was a good equation for more than one reason.

1. It is easy to calculate slop and y-intercept values.

3. It is the best equation for calculating the x-value with the smallest relative error.

It uses the smallest sum absolute relative error (SARE) and the smallest value of average absolute relative error (AARE) in the calculated x value to find out the best fitting line of points that meets the requirements of analytical analysis.

It gives the best way to find out the best working range of data by using SARE and AARE parameter.

References

1. V. Barwick and L. Prichard, ‹Preparation of Calibration Curves ', LGC Limited, USA , 2003.

2. C. M. Peter and E.Z.Richard, ‹Statistical method in analytical chemistry ', 2nd Ed., John Wiley & Sons, Inc., USA, 2000.

3. D. A. Skoog and D. M. Wes, ‹Fundamentals of Analytical Chemistry' ,10th Ed., Cengage Learning, USA, 2022.

4. B. A. Zeev, B. Zvi and R.Yigal, ‹Statistical treatment of analytical data ', CRC Press LLC, USA, 2005.

5. L. R. E. Stephen, J. B. Vicki and J. D. F.Trevor, ‹Practical Statistics for the analytical scientist ', LGC Limited, UK, 2009.

6. A. N. Alshirifi and D.Y. Alhameedi, International Journal of ChemTech Research, 9(5), 712-722 (2016).

7. J. N. Miller and J. C. Miller, ‹Statistics and Chemometrics for Analytical Chemistry', 6th Ed., Pearson Education Limited, England, 2010.

8. N. A. Weiss, ‹introductory statistics', 9th Ed., Addison Wesley /Pearson Learning, Boston, 2011.

9. D. C. Montgomery, and G. C. Runger, ‹Applied Statistics and Probability for Engineers', 3rd Ed., John Wiley & Sons, Inc., USA, 2003.

10. R. Jayaprakash, ‹Advanced Quantitative Techniques', S. B. Nangia, New Delhi, 2004.

11. D. Kealey and P.J. Haines, ‹the instant notes series', BIOS Scientific Publishers Limited, UK, 2002.

12. A. N. Alshirifi and D.Y. Alhameedi, International Journal of ChemTech Research, 9(9), 281-293 (2016).

13. A. N. Alshirifi and D.Y. Alhameedi, International Journal of ChemTech Research, 9(9), 294-308 (2016).

14. D.Y. Alhameedi and A. N. Alshirifi, AIP Publishing, 2398, 030022-1–030022-16 (2022).
15. D.Y. Alhameedi and A. N. Alshirifi, AIP Publishing, 2547, 040002-1–040002-12 (2022).