Proceedings of the International Conference on Industrial Engineering and Operations Management
Pilsen, Czech Republic, July 23-26, 2019

Statistical and Kinetic Modeling for Investigating Acetyl Salicylic Acid Stability

Najwa Alwazni[1], Chandra Mouli R Madhuranthakam[2], Asmaa Awad[1], Ibrahim Mustafa[1,3], Mohamed Binshams[4], Ali Elkamel[1,5]

1- Chemical Engineering Department,
University of Waterloo,
Waterloo, Ontario, Canada

2 - Chemical Engineering Department,
Abu Dhabi University
Abu Dhabi, United Arab Emirates P.O.Box. 59911
chandra.mouli@adu.ac.ae

3 - Helwan University, Cairo, Egypt

4 - Bahrain University, Bahrain

5- Chemical Engineering Department,
Khalifa University of Science and Technology,
Abu Dhabi, United Arab Emirates, P.O.Box. 2533
ali.elkamel@ku.ac.ae

Abstract

Drug stability is one of the vital subjects in the pharmaceutical industry. All drug products should be kept stable and protected against any chemical, physical, and microbiological degradation to ensure their efficacy and safety until released for public use. Hence, stability is very important to be estimated or predicted. This work involved studying the stability of drug agent using three different mathematical models. These models included both empirical models (linear regression), and mechanistic (kinetic) models. The stability of the drug in the three cases studied was expressed in terms of concentration, hardness, temperature and humidity. The predicted values obtained from the models were compared to the observed values of drug concentrations obtained experimentally and then evaluated by calculating the mean of squared. Among the models used in this work, the mechanistic model was found to be the most accurate and reliable method of stability testing given the fact that it had the smallest calculated errors. Overall, the accuracy of these mathematical models as indicated by the proximity of their stability measurements to the observed values, led to the assumption that such models can be reliable and time-saving alternatives to the analytical techniques used in practice.

Keywords

Regression models, Stability, Data analysis, Mathematical modeling, Neural networks.

1. Introduction

Drug stability is one of the most important aspects of any drug product in pharmaceutical industry. The drug stability plays an important role in its efficacy and pharmacological benefit to the consumers as well as the time after which it would lose its desired quality or even change into a toxic agent [1]. For every drug in the market, the Food and Drug Administration (FDA) requires that a shelf-life must be indicated on the container label. Moreover, the drug companies are required by the FDA to submit data that ensure the drug characteristics such as: strength, potency, and purity, and demonstrate that the average drug characteristics can meet the approved specifications during the claimed shelf life period. Based on the FDA guidelines, the purpose of stability testing is to provide evidence on how the quality of a drug product varies with time under the influence of a variety of environmental factors such as temperature, humidity, and light [2]. Several drug stability testing methods are used in practice to study the rate of drug degradation to estimate its expiration date. Most of these methods are based on analytical laboratory experiments which are usually expensive and time consuming. A newer and a more efficient approach that is currently adopted for measuring the behaviour of the active ingredient through time is the use of statistical analysis. This implies formulation of statistical techniques and mathematical models where specific equations are used to study the drug stability and estimate its shelf life or expiration date [3-5]. The mathematical models that are commonly used nowadays are the empirical models such as linear regression, artificial neural network, and the mechanistic models. It is important to evaluate the applicability and efficiency of statistical modeling in studying drug stability and estimating its degradation with time and to know whether statistical modeling can efficiently substitute the ordinary analytical methods in studying drug stability [6, 7]. In this work, the mentioned three mathematical and statistical models are applied on the drug of Acetyl Salicylic Acid (ASA) to evaluate its stability. The experimental data of ASA were collected and all the three mathematical models (linear regression, artificial neural network, and mechanistic models) were applied to determine the best model to predict its stability. A comparison among these mathematical models was made to conclude the best statistical method should be adopted in practice to achieve the most accurate shelf life [8, 9].

2. Mathematical Modeling and Solution

MATLAB is used to perform the regression analysis using the least squares solution process. In addition, for Artificial neural network (ANN) analysis, back propagation feed forward based on Levenberg-Marquadt training algorithm was employed to train the network. MATLAB® neural network toolbox was used to train and simulate the network model. Log sigmoid was found to give reasonable response. Finally, Microsoft Excel is used for solving the mechanistic model in order to obtain the kinetic parameters in both Arrhenius and the kinetic equations.

3. Results and Discussion

In this case study, the experimental data for the stability of acetyl salicylic acid (ASA) tablets were taken from the article by Snavel et al. [10]. In this experiment, ASA was mixed with a mixed-sugar diluent containing about 8% moisture for 20 minutes and then blended after adding stearic acid for 3 minutes. A part of the blend was compressed into tablets. Both tablets and uncompressed powder blend were packaged into separate glass bottle containers. Stress stability studies were conducted on the compressed tablets (10 tablets per bottle) and the powder blend (3 grams per bottle). The bottles of tablets and powder blend were stored under different temperatures 35, 40, 45, and 50 $^{\circ}$C) for different time intervals then the stability of these drug samples was assessed according to the rate of decomposition. Each time, the content of each bottle rinsed and ground with acetonitril and then filtered and assayed for ASA remaining by using stability indicating simultaneous UV assay. It was concluded by this experiment that aspirin degradation is accelerated with time and with temperature increment. The rate of degradation was found to be following the empirical equation: $y = 100 - Kt^n$; where y is the % ASA remaining, t is the time, and both k and n are constants. The formulation tested proved good stability with only < 1% degradation after 1.75 years at room temperature. Moreover, the moisture content in this formulation had no influence on the decomposition rate without an obvious explanation.

3.1 Experimental Data

Experimental data on ASA stability were collected and all the three mathematical models (linear regression, artificial neural network, and mechanistic models) were conducted on these data to determine the model that best predicts its stability. A comparison among these three mathematical models was made at the end to conclude the best statistical method that should be adopted in practice to achieve the most accurate shelf life. Table 1 shows both independent variables (temperature and time) and dependent variables (percentage of ASA remaining).

Table 1. Effect of Temperature and Time on the %ASA Remaining

Independent Variables	**Dependent Variable**

Temperature (Celsius)	Temperature (Kelvin) X_1	Time (Weeks) X_2	%ASA Remaining Y
50	323	1.19023	98.5603
50	323	4.12644	96.5046
50	323	7.13936	92.0614
50	323	10.1548	86.7537
50	323	13.8791	77.577
45	318	1.25856	99.0543
45	318	7.2619	97.9072
45	318	14.1113	93.7145
45	318	20.188	91.3324
45	318	26.2672	88.0858
45	318	30.1097	86.278
40	313	1.25749	99.4248
40	313	9.14346	98.6499
40	313	15.2181	96.9676
40	313	20.1743	96.0253
40	313	44.1901	90.6133
35	308	1.25749	99.4248
35	308	14.235	99.1485
35	308	20.2368	98.5365
35	308	35.1707	97.274
35	308	45.9181	96.1724

3.2 Linear Regression

As it is known, the linear regression model is concerned with models that are linear in parameters. In this experiment, the values of the independent variables are perfectly known and a linear model was created to estimate the unknown parameters. The coefficient of determination was found to be 84.67% and the standard error was 2.51 as it is shown in the Table 3.

Table 2. Statistical Results of Regression Model

Regression	Statistics
Multiple R	0.920195
R square	0.846759
Adjusted R square	0.808449
Standard Error	2.51667
Observations	21

Table 3. ANOVA Table

Source of variance	DF	SS	MS	F critical	F Significance
Regression	4	559.9597	139.9899	22.10265	2.36E-06
Residual	16	101.338	6.333626		
Total	20	661.2977			

The underlying model:

Table 4 represents the parameters that have significant effects as has been found from ANOVA table.

Table 4. Parameters Estimations

	Parameters Values	Standard error	t observed	P-Value	Lower 95%	Upper 95%
Intercept (β_0)	-8803.22	2618.405	-3.36205	0.003965	-14354	-3252.45
Temp K (β_1)	56.55087	16.53652	3.419757	0.003511	21.49502	91.60672
Time (β_2)	17.26145	3.327954	5.186805	9E-05	10.2065	24.31639
Temp*Time (β_3)	-0.05608	0.010653	-5.26406	7.71E-05	-0.07866	-0.03349
Temp square (β_4)	-0.08977	0.026101	-3.43942	0.003368	-0.14511	-0.03444

$t_{critical}$ =2.05183
General regression model
$\hat{\gamma}=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_1X_2+\beta_4X_2^2+\varepsilon$ where $\varepsilon \sim N(0,\sigma^2)$
Fitted linear regression model
$\hat{\gamma}= -8803.22 + 56.55087\ X_1 + 17.26145\ X_2 - 0.05608\ X_1X_2 - 0.08977\ X_2^2 + \varepsilon$
Model evaluation:
To test the validity of the underlying assumption (error ~N(0,σ^2)) as a diagnostic check, we are going to look at the residuals and normal probability plots from the fitted model. Also determining the coefficient of determination to measure how well the linear regression model represents the data. Using the residual plots is to investigate whether the errors are scattered, independent and normally distributed. Figure 1 shows that the residual has a random pattern. Using normal probability plots to investigate whether the errors are normally distributed. It is shown from Figure 2 that the normal probability plot indicates a straight line. So, it is concluded that the observed sample is normally distributed. Using the coefficient of determination as a diagnostic check is to measure whether the regression line represents the data. From the results, it is shown that R^2 = 0.8467, which means that 84.67% of the total variation in observed sample can be explained by the linear relationship between the regressors and observed response (as described by the regression model). The other 15.3241% of the total variation in the observed sample are unexplained.

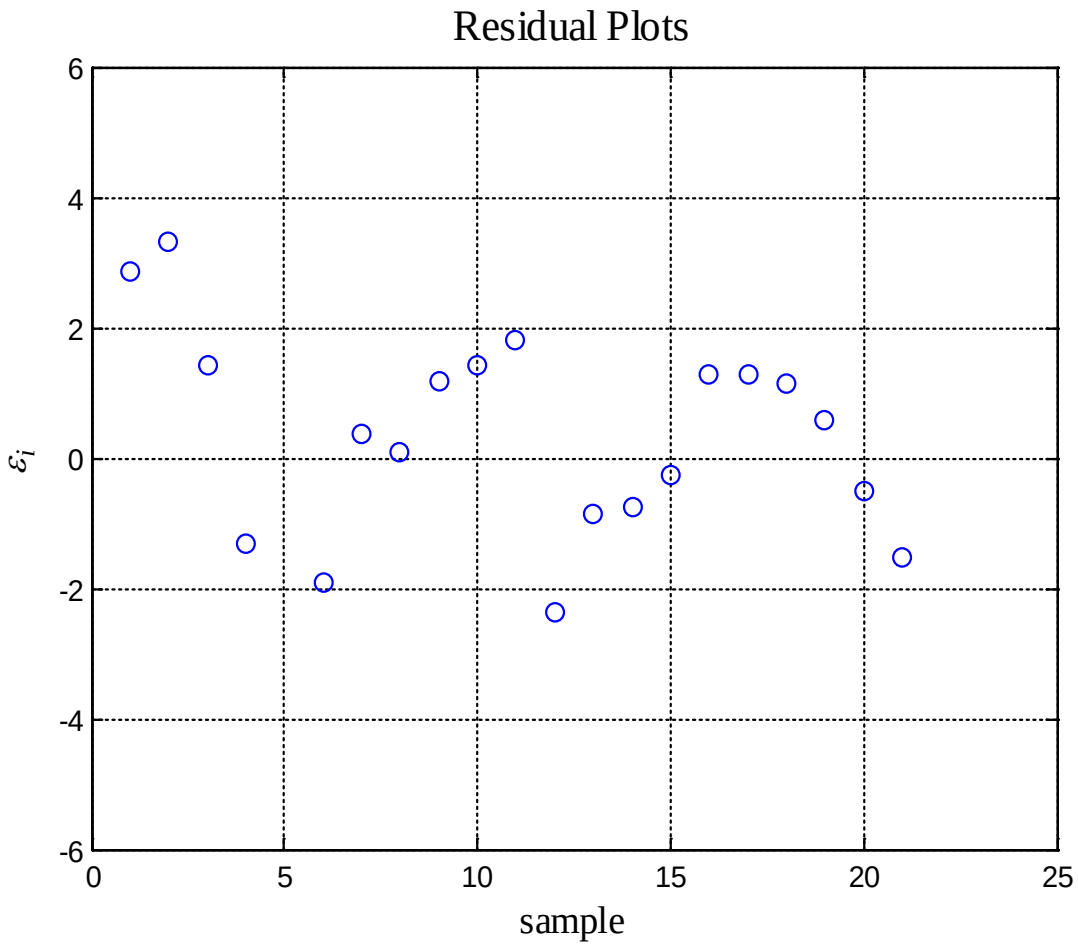

Figure 1. Aspirin Residual Plots

3.3 Artificial Neural Network (ANN)

Back propagation feed forward based on Levenberg-Marquadt training algorithm was employed to train the network. MATLAB® neural network toolbox was used to train and simulate the network model. Log sigmoid was found to give reasonable response. The data was randomly divided into a 15-training set and a 6 testing set as

shown in Tables [6] and [7]. About 10 – 15 % of the data was selected randomly to be the testing data set. Then before training, the data was normalized by using the following formula $\hat{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$ So that, the inputs and outputs are belong to the interval [0, 1].

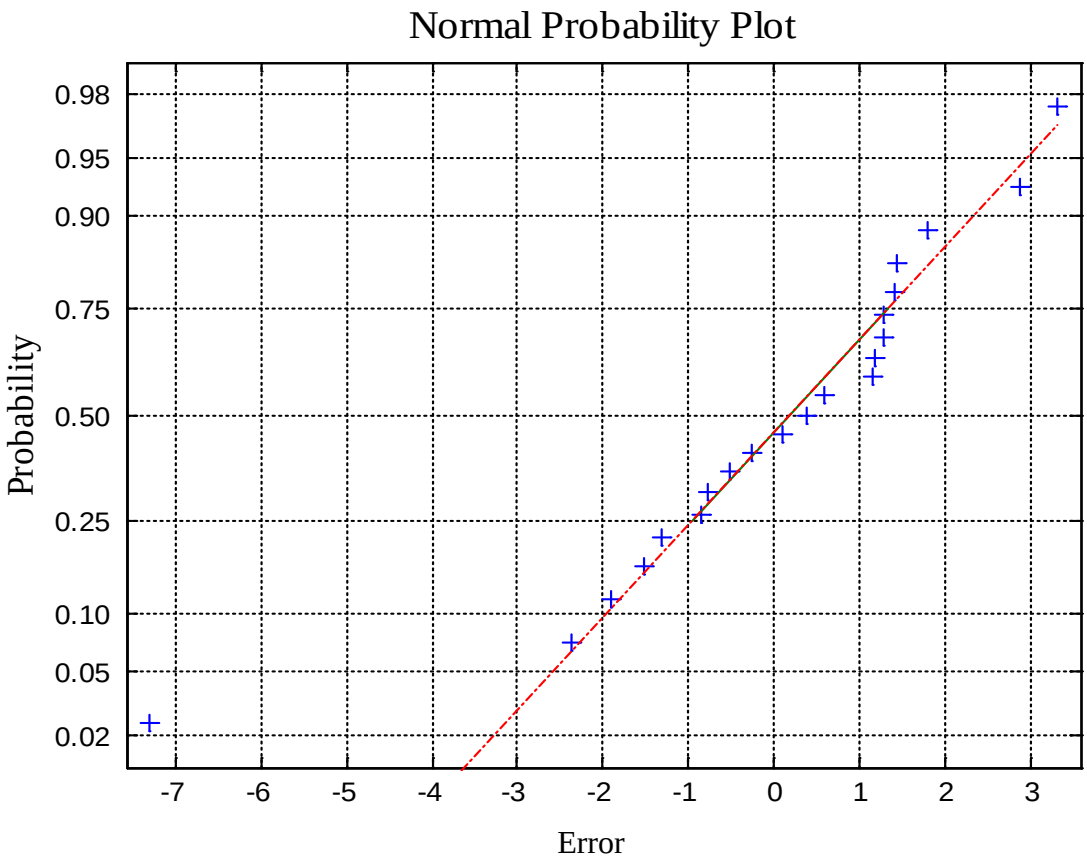

Figure 2 Aspirin Normal Probability Plot

Table 5. Training Data Set of Aspirin

Training Data Set		
Temperature(K)	**Time (Weeks)**	**ASA remaining %**
323	7.1394	92.0614
318	20.1880	91.3324
318	1.2586	99.0543
318	7.2619	97.9072
313	20.1743	96.0253
313	9.1435	98.6499
318	30.1097	86.2780
323	13.8791	77.5770
308	35.1707	97.2740
308	20.2368	98.5365
318	14.1113	93.7145
313	44.1901	90.6133
308	14.2350	99.1485
308	45.9181	96.1724
308	1.2575	99.4248

Table 6. Testing Data Set of Aspirin

Testing Data Set		
Temperature(K)	**Time (weeks)**	**ASA remaining %**
323	1.1902	98.5603
323	4.1264	96.5046
323	10.1548	86.7537
318	26.2672	88.0858
313	1.2575	99.4248
313	15.2181	96.9676

The result was acceptable by using 300 epochs of training, with the goal 10^{-4} and 10 neurons and one hidden layer. Tan sigmoid functions were applied in this set of data. The coefficient of determination (R^2) was 72.016% for training data set and 98.756% for the testing data set as shown in Tables [8] and [9], respectively. Since R^2 approaches 1, it indicates an excellent predictive ability of the neural network model.

Table 7. Statistical Properties for the Training Data Set

Average error	1.94692
Maximum error	10.87060
Minimum error	0.00541
R^2	0.72016

Table 8. Statistical Properties for the Testing Data Set

Average error	3.88989
Maximum error	10.87060
Minimum error	0.00541
R^2	0.94009
SSE	9.1008
MSE	1.5168

As it is shown in Figure 3,the parity line of training data passes through almost all the observed and predicted values on the scatter plot. This indicates high accuracy of the predicted values with close proximity to the observed data. Since there was no enough data to be used both for training and testing, it was not possible to clearly show how close the predicted values are to the observed. This can be seen in the Figure 4, where the straight line does not pass through all the points of testing data. However, the points fall very close to the parity line.

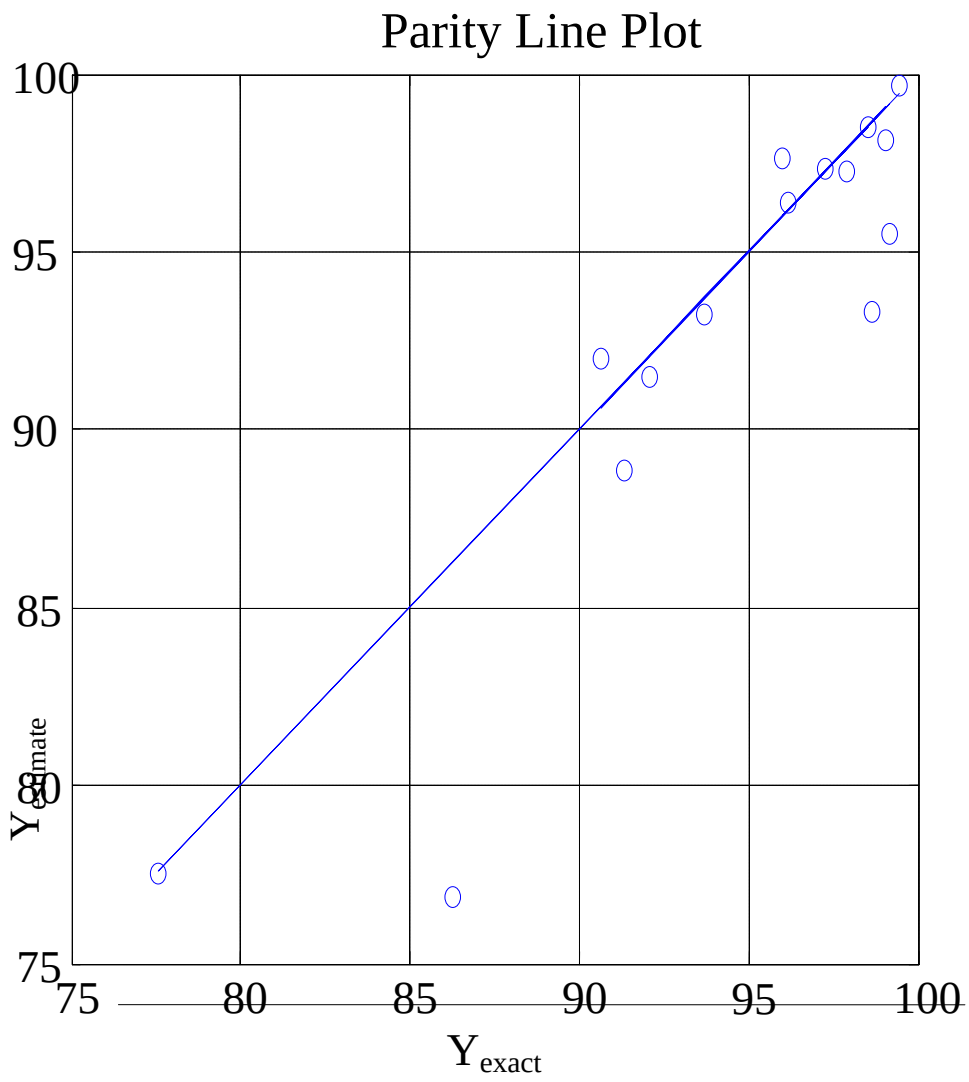

Figure 3 Parity Line Plot for Aspirin Training Datasets

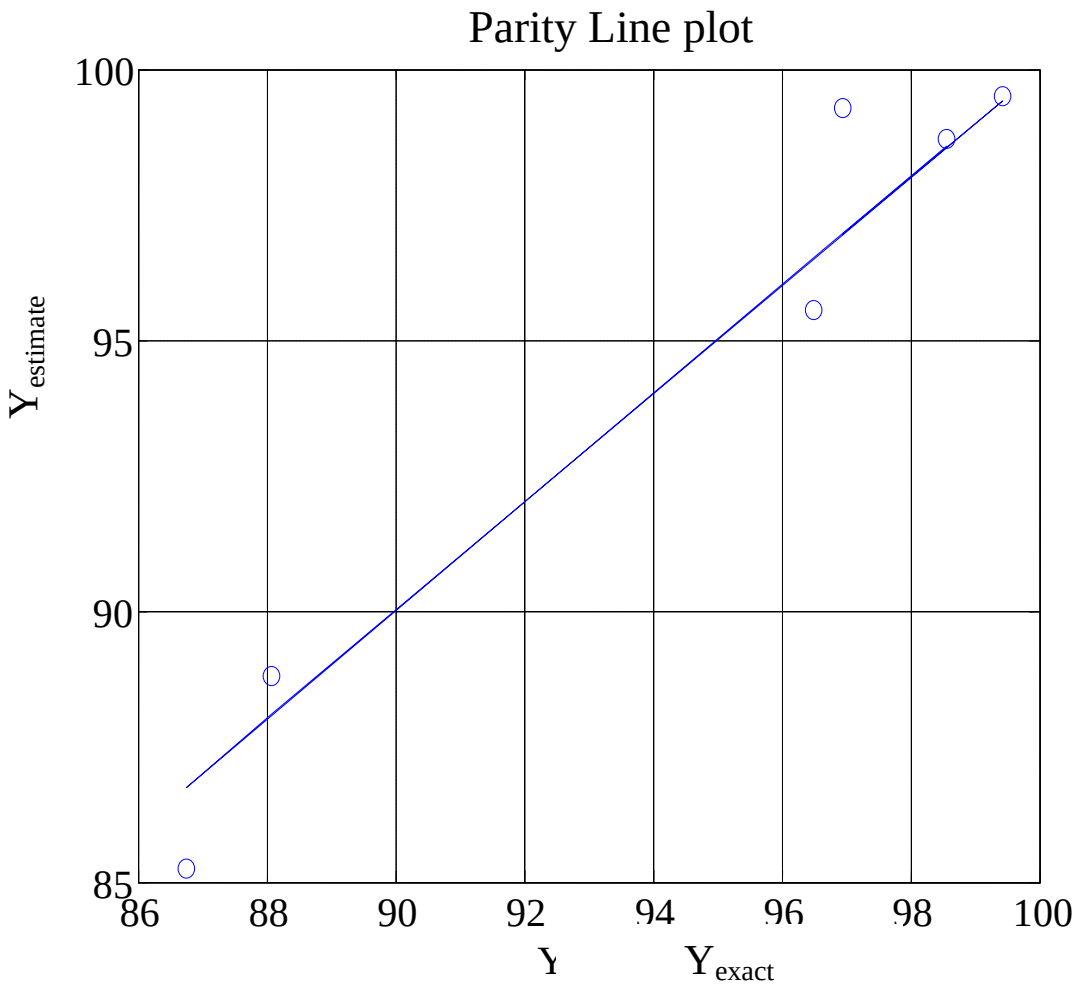

Figure 4. Parity Line Plot for Aspirin Testing Datasets

Figure 4 shows the distribution of the errors. Most of the errors lie in the range between 0-1%.

3.4 Mechanistic Model

Microsoft Office Excel© version 2007 was used for mechanistic model calculations. The determination of degradation kinetic orders as well as the reaction rate constants was attempted. A mechanistic model (also called kinetic model [11]) for Aspirin was proposed as shown in the Equation 1 and both of the rate constant and the rate order were determined at different temperatures using a linear regression method.

$-r_a = k_A C_A^{\alpha}$ [1]

Equation [1] is linearized as:

$\ln(-r_a) = \ln(k_A) + \alpha \ln C_A$ [2]

The results are summarized in the following Table [9].

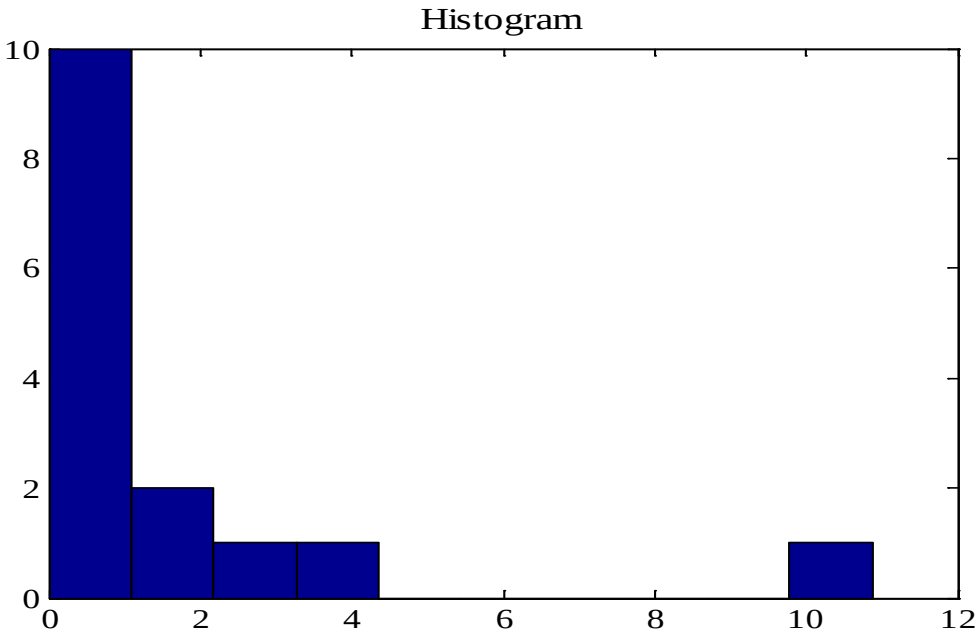

Figure 5. Aspirin Histogram

Table [9]: The Reaction Rate Constant (k) for Aspirin Degradation at Various Temperatures

Temperature(K)	**Reaction rate constant k_A (Week^{-1})**	**The reaction order (α)**
323	0.197484	0.9784
318	0.053547	0.9428
313	0.027645	0.9681
308	0.003763	0.92

The hydrolysis process of aspirin was found to follow first order kinetics [11, 12]. The mean of squared error was found to be $2.92X10^{-7}$ and the coefficient of determination was 96.48%. The natural logarithm of observed rate constants over the temperature range of 40-63°C were plotted versus reciprocal of temperature according to the Arrhenius equation to calculate the frequency factor and the activation energy. The activation energy was calculated from the slope of the straight line obtained by linear regression method and found to be approximately 49 Kcal/mole and the frequency factor was approximately 75 week^{-1}.

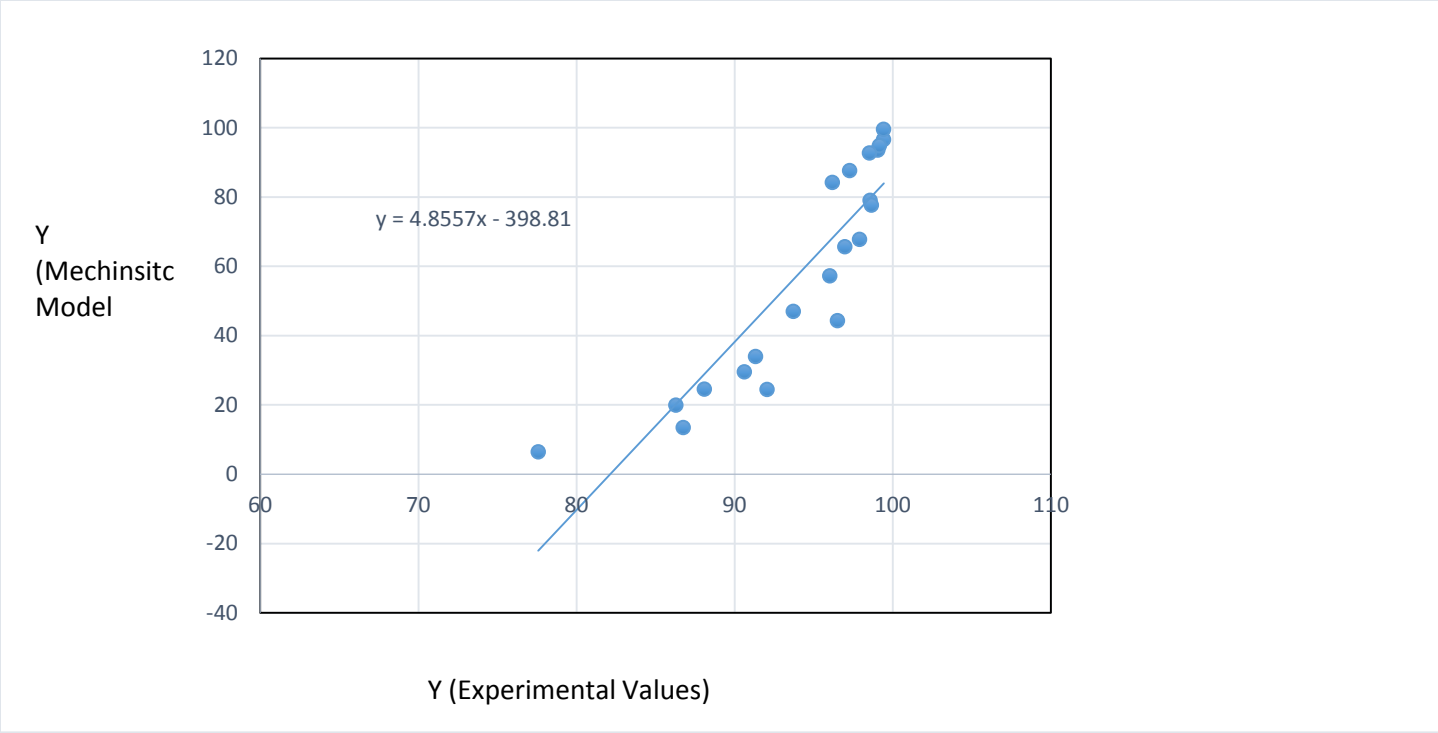

Figure 6. Mechanistic model vs. Experimental values

Figure 6 shows the stability values produced from the mechanistic model vs the experimental real values. As shown in Figure 6, most of the values do exist on the fitted line. It shows that the mechanistic model could express the real values very well.

6. Conclusion

Stability testing requires vigorous work and lengthy studies. Analytical techniques have been the commonly used methods in drug stability testing. These techniques include high performance liquid chromatography (HLPC) or thin layer chromatography (HPTLC), gas chromatography (GC) and electrophoresis. Recently, there has been a new trend towards the use of statistical modelling techniques. The modelling techniques commonly used nowadays are the empirical models such as linear regression, artificial neural network, and the mechanistic models. Such techniques are thought to be reliable and economic alternatives with the advantage of time saving. To assess the reliability and applicability of statistical modelling in studying drug stability and estimating its degradation with time and to know whether statistical modeling can efficiently substitute the ordinary analytical methods, several models were applied in this research on ASA which was chosen from the pharmaceutical literature. Three different models (linear regression, artificial neural network and mechanistic models) were implemented to predict the drug concentration over time. These predicted values were compared to the observed values and then evaluated by calculating the mean of square errors (MSE) to determine how close those values were to the exact values resulted experimentally. Moreover, the coefficient of determination (R^2) was calculated to further evaluate the model accuracy as well. This work focussed on the stability of acetyl salicylic acid (ASA) tablets. The concentration of ASA at variable temperatures (35, 40, 45, 50° C) was determined experimentally at different time intervals using HPLC. Firstly, a linear regression model was proposed, and the unknown parameters were estimated and found to be significant as concluded from ANOVA table. The errors were found to be normally distributed with a mean equal to zero and the residuals residual had a random pattern as seen in normal probability and the residual plots, respectively. The MSE was 6.33, and R^2 was 84.67%. This meant that the regression model represents the data very well. Secondly, the artificial neural network was applied, and it was able to capture the relationships between the inputs and the output in the testing datasets despite the limited number of the experimental data (21 data points). The MSE was 1.5168 whereas the R^2 was 94% indicating a high accuracy for artificial neural network model as well. The third model applied was the mechanistic model. This model was linearized to determine the decomposition rate constants (k_{App}) at various temperatures and the order of the reaction (α). The hydrolysis process was found to follow first order kinetic dependent only on the concentration of one reactant; i.e. aspirin. The apparent activation

energy of aspirin hydrolysis of 49 Kcal/mole was calculated by using the Arrhenius equation method. MSE was found to be $2.92833x10^{-7}$ and the R^2 was 96.48% which indicates that the predicted values were very close to the observed values which in turn indicated that the mechanistic model is reliable. Based on the results the regression model is the best model predicting the stability of ASA.

References

[1] FDA (1987). Food and Drug Administration Guidelines for Submitting Documentation for Stability Studies on Human Drugs and Biologics. Rockville, MD.

[2] Yoshioka, S., and Stella, V.J., (2000). Stability of Drugs and Dosage Forms, Kluwer Academic/Plenum Publishers. New York, USA.

[3] Lu, X., (2006). Simultaneous Confidence Bounds with Applications to Drug Stability Studies. P.H.D, Proquest Information and Learning Company, USA.

[4] Carstensen, J. T., (1995). Drug Stability: Principles and Practices, Marcel Dekker, Inc., New York, USA.

[5] Chow, S.C., and Shaw, J., (1991). Estimating Shelf Life with Random Batches, Biometrics 47:1071-1079.

[6] Chow, S.C., and Liu, J. P., (1995). Statistical Design and analysis in Pharmaceutical Science: Validation, Process Control, and Stability. Marcel Dekker, Inc., New York, USA.

[7] Pong, A., (2001). Comparing Designs for Stability Studies and Shelf life Estimation for Drug Product with Multiple components. PhD Thesis, Temple University.

[8] Dietrich, K.K., (1983). Shelf Life Estimation from Non-accelerated Data. PhD. Thesis, University of Cincinnati, College of Medicine, USA.

[9] Yang, P., and Macdonald, F., (2004). Solution Stability of Factor Xa Inhibitors as a function of pH, Drug Development and Industrial Pharmacy, 30(9):967-973.

[10] Snavely, M.J., Price, J.C., and Jun. J.W., (2008). The Stability of Aspirin in a Moisture Containing Direct Compression Tablet Formulation, Drug Development and Industrial Pharmacy, 19(6):729-738

[11] Oliva, A. et al (2006). Data Analysis of Kinetic Modelling Used in Drug Stability Studies: Isothermal versus Non Isothermal Assays., Pharmaceutical Research, 23(11):2595-2601.

[12] Augburger, L.L., (2008). Pharmaceutical Dosage Forms: Tablets, Rational Design and Formulation. 3rd Edn, Marcel Dekker, Inc., New York, USA.

Najwa Alwazni and **Asmaa Awad** are graduate students in chemical engineering department at the University of Waterloo, Waterloo, Canada.

Ibrahim Mustafa and **Mohamed Binshams** are visiting scholars in chemical engineering department at the University of Waterloo, Waterloo, Canada.

Chandra Mouli R. Madhuranthakam is an assistant professor in the department of chemical engineering at the Abu Dhabi University, Abu Dhabi, United Arab Emirates. He earned B.Tech in chemical engineering from Sri Venkateswara University, India Master's in chemical engineering at the Indian Institute of Science, Bangalore, India and PhD in chemical engineering

University of Waterloo, Canada. He is a member of the American Institute of Chemical Engineers and the Institute for Operations Research and Management Sciences (INFORMS).

Ali Elkamel is a professor in the department of chemical engineering at University of Waterloo, Waterloo, Canada. He earned his B.S. and M.S. in chemical engineering at Colorado School of Mines and PhD in chemical engineering at Purdue University. Ali Elkamel is also a visiting Professor in the Department of Chemical Engineering at Khalifa University, Abu Dhabi, UAE. He is a member of the Canadian Society for Chemical Engineering, the American Institute of Chemical Engineers, the Canadian Operational Research Society, and the Institute for Operations Research and Management Sciences (INFORMS).