Application of response surface methodology (RSM) in the treatment of final effluent from the sugar industry using Chitosan

R. L. L. Pambi[1] & P. Musonge[2]
[1]*Department of Physics, Durban University of Technology, South Africa*
[2]*Department of Chemical Engineering,*
Institute for Water and Wastewater Technology,
Durban University of Technology, South Africa

Abstract

The sugar industry contributes significantly to the economic growth of South Africa by creating jobs in the agricultural and industrial sectors. However, this industry discharges large amounts of effluent containing a high level of suspended and dissolved solids, which impart colour to the wastewater stream and add treatment cost. Chitosan, a natural polymer, has been used in the coagulation of impurities from the sugar refinery using the one-factor-at-a-time (OFAT) method. The results indicated the removal of total suspended solids (TSS) and colour of 87% and 76% respectively at the pH of 9. Response surface methodology (RSM) was used to maximize the efficiency of this coagulant according to the Box-Behnken design (BBD). The use of RSM was found to have several advantages in comparison to the OFAT, such as the identification of interaction, the use of statistical analysis that produce model equations for optimization and prediction of the behavior of a particular system. Furthermore, at the pH of 9, the BBD yielded TSS and colour removals of 99% and 90% respectively. This should be a motivation for an industrial researcher to deviate from the traditional OFAT especially in process optimization studies.
Keywords: Chitosan, coagulation, response surface methodology, sugar effluent.

1 Introduction

Design of experiments (DOE) can be defined as the systematic planning of information gathering with the use of experimental and statistical methods to identify the optimum factors and levels for a particular problem either of industrial scale or for research purposes. The most commonly used DOE is the one-factor-at-a-time (OFAT) design which vary only one variable at a time while keeping others constant. Pambi and Musonge [1] investigated the removal of impurities from the two effluent streams discharged by a local sugar refinery using the OFAT experimental design. Chi and Cheng [2] also applied the OFAT method in their study of the coagulation of milk processing plant wastewater using chitosan. The main advantages of the OFAT method is the fact that it allows a rapid identification of the influence of the factors and the experimental outcomes can be readily understood [3, 4]. However, the use of OFAT is being discouraged due to the following reasons [5–8]:

- The OFAT design does not show interactions between the variables.
- The prediction of the response in the factor space is poor.
- The lack of randomization can lead to biased conclusions.

One way of overcoming these shortcomings is the use of response surface methodology (RSM). The RSM uses a set of mathematical and statistical procedures to describe the relationship between a set of data. This relationship is described by a polynomial equation that relates to the experimental data, in order to concurrently optimize the responses [8, 9]. The most common types of DOE used for the RSM analysis are the three-level full factorial design central composite design (CCD), the Doehlert design (DD) and the Box-Behnken design (BBD) [9, 10].

The CCD, DD and BBD are two-level factorial designs (2^n) that have been modified by the addition of centre-points that generate three-level designs that are suitable for higher degree polynomial in RSM. Without the centre-points, the 2^n design generates a linear function of the response (Y), as shown in Eqn (1), which does not accommodate second degree polynomial equations due to the fact that it does not represent curvatures.

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \varepsilon \qquad (1)$$

where β_0 represents the constant term and β_i represent the coefficient of the linear parameters respectively. X_i represents the factor and ε is the residual from the treatments.

The critical points in the RSM plot are obtained from the following quadratic polynomial equation [9]:

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \sum_{i=1}^{n} \beta_{ii} X_{ii}^2 + \sum_{i<j}^{n} \beta_{ij} X_i X_j + \varepsilon \qquad (2)$$

X_i and X_j represent the factors, β_{ii} represents the coefficient of the quadratic parameter, and β_{ij} represents the coefficient for the interaction parameters.

Interaction occurs when the effect of one factor is dependent on the level of another. This a special feature DOE has in comparison to the OFAT design. A factor with a small individual effect can contribute greatly to the response by interacting with another. The contribution of each terms in the regression and the significance of the model equations can be obtained through a fit test known as analysis of variance (ANOVA) which provides a comparison between the variations caused by the experimental runs and the variations caused by the measurement errors. This is done by analyzing the magnitude of the sum of square (SS), mean square (MS), the Fischer test (F-test) and the lack-of-fit test (LOF).

The sum of square (SS) is a measure of variability, and can be used to estimate the variance of the mean value of a statistical analysis when scaled for the degrees of freedom (df).

The F-test of a model evaluates the significance of the model by calculating the ratio of the mean square of regression to the mean square of residuals.

$$MS = SS/df \qquad (3)$$

A small F-value for the model is not desired since it indicates that the variance is caused by random unexplained disturbances referred to as noise. The p-value $(p > F)$ provides an indication of the significance of a model in relation with the F-value. It can be defined as the probability that a variable did not affect the response for a given F-value. If the $p > F$ for the model is less than 0.05 a model is said to be significant, meaning that there is 5% chance that the F-value is due to noise. If the $p > F$ is above 0.1, the model is insignificant [6, 11].

The lack-of-fit test (LOF) determines the inability of a model to fit experimental data that are not represented in the experimental domain. This is commonly done by calculating its F-value. A small F-value for the LOF is desired, since the experimenter wants the model to fit. If the $p > F$ is greater than 0.05 the LOF for the model is insignificant and the model is able to fit any data that are not specified in the experimental domain [11]. A good LOF does not guarantee the adequacy of a model, the coefficient of determination (R^2) must be considered given the fact that it measures the overall performance of a model and its value should be close to 1 [10].

The main aim of this study was to optimize the removal of total suspended solids (TSS) and colour from final effluent (FE) discharged by the sugar refinery the using RSM. A previous study using OFAT demonstrated that the treatment of FE with the chitosan coagulant (CCo) yielded good results for TSS and colour [1]. This study uses the BBD to optimize the performance of CCo for the sugar effluent.

2   Experimental methods

2.1 Jar-tests

The coagulant was prepared according to the method found in literature [1, 12–14]. The jar-tests were done using a non-programmable Voss flocculator

(6 paddles). The coagulant was mixed with the FE for 3 minutes at high speed (100 rpm). This is referred to as flash mixing. In order to promote flocs growth, the mixing speed was decreased to 40 rpm for 15 minutes. The mixture was left to settle in order for the flocs to precipitate.

2.2 RSM

The Box-Behnken design (BBD) was used to generate the experimental matrix and the model equations for the optimization of the performance of chitosan for the FE.

In factorial designs and RSM, the variables are codified to normalize the variables before regression analysis and eliminate the effect of different units and ranges in the experimental domain and allows parameters of different magnitude to be investigated more evenly in a range between -1 and +1. The equation used for coding is

$$X = \frac{x - (\frac{x'' + x'}{2})}{(\frac{x'' - x'}{2})} \qquad (4)$$

where X is the coded factor/variable, x is the natural variable, x'' and x' represent the maximum and the minimum of the values of the natural variable respectively [10].

Three manipulated variables were varied on three levels: a high level, represented as (+1), a low level represented as (-1) a middle point (0) as shown in Table 1.

Table 1: Manipulated variable and BBD levels.

Variables	-1	0	+1
X_1: Coagulant loading (ml)	10	20	30
X_2: pH	4	7	10
X_3: Settling time (minutes)	30	60	90

The response variables represented as Y_1 and Y_2 are the percentage removal of TSS and colour (represented as   in Eq.2) respectively using CCo calculated form the expression below:

$$Y_n = \left(\frac{y_0 - y_i}{y_0}\right) \times 100 \qquad (5)$$

y_o and y_i respectively, represent the initial values of the response variable (TSS and colour) and their values after the coagulation process.

3   Results and discussion

The experiments were conducted in a random order instead of the standard order to avoid biases during the trials, and the analyses of the data were done at standard BBD run order as seen in Table 2 using the Design expert 9.0 software.

Table 2:   Experimental matrix.

Standard runs	Randomized runs	X_1	X_2	X_3	Y_1 (TSS)	Y_2 (Colour)
1	6	-1	-1	0	88	72
2	5	+1	-1	0	80	72
3	10	-1	+1	0	89	80
4	11	+1	+1	0	91	82
5	1	-1	0	-1	87	69
6	3	+1	0	-1	79	71
7	9	-1	0	+1	96	87
8	13	+1	0	+1	89	80
9	2	0	-1	-1	80	69
10	4	0	+1	-1	88	76
11	8	0	-1	+1	92	72
12	17	0	+1	+1	99	90
13	15	0	0	0	93	84
14	16	0	0	0	92	85
15	12	0	0	0	93	87
16	7	0	0	0	92	83
17	14	0	0	0	93	86

3.1 Analysis for TSS reduction

The results of the analysis of variance (ANOVA) and lack-of-fit (LOF) tests as generated from the experimental data are shown in Table 3. The quadratic regression has a large F-value and the p>F value is less than 0.05, meaning that the model is significant. The term X_3 has the highest F-value and the lowest p>F value, meaning that the settling time has the largest effect on the TSS removal, followed by the pH and the coagulant concentration. It can also be seen that there is 6.35% chance that the LOF is due to noise which is not a very significant value, as the LOF values greater than 0.05 (5%) are not significant, showing that the model can fit any data.

The model equation for the removal of TSS is expressed as follows:

$$Y_1 = 92.6\text{-}2.81X_1+3.39X_2+5.4X_3+2.16X_1X_2$$
$$+0.15X_1X_3\text{-}0.05X_2X_3\text{-}3.77X_1{}^2\text{-}1.62X_2{}^2\text{-}0.98X_3{}^2 \qquad (6)$$

The pareto (Fig. 1) was plotted using coefficients in eqn (6). It can be seen that the removal of TSS is mostly affected by the settling time (X_3). The fact that the coefficient for X_3 in eqn (6) carries a positive signs shows that the TSS removal increases when the sample is left to settle for longer.

It can also be noted that X_1X_2 is the most significant interaction term. This indicates that the removal of TSS is positively influenced by the combination of the variation in both pH and coagulant dosage.

Table 3:   ANOVA for TSS removal.

Variation	SS	df	MS	F-value	p-value p>F
Regression	487.90	9	54.21	31.70	< 0.0001
X_1	63.06	1	63.06	36.87	0.0005
X_2	92.07	1	92.07	53.84	0.0002
X_3	233.28	1	233.28	136.41	< 0.0001
X_1X_2	18.75	1	18.75	10.96	0.0129
X_1X_3	0.090	1	0.090	0.053	0.8251
X_2X_3	1E-02	1	1,E-02	6E-003	0.9412
X_1^2	59.76	1	59.76	34.95	0.0006
X_2^2	11.02	1	11.02	6.44	0.0388
X_3^2	4.06	1	4.06	2.38	0.1671
Residual	11.97	7	1.71		
LOF	9.69	3	3.23	5.67	0.0635
Pure Error	2.28	4	0.57		
Total	499.87	16			

$R^2 = 0.976$

From eqn (6) and Fig. 1, it can be seen that increasing the coagulant dosage can have a negative effect on the TSS removal. The term X_1X_3 in this equation means that the TSS removal is slightly boosted by the interaction between the chitosan dosage and the settling time.
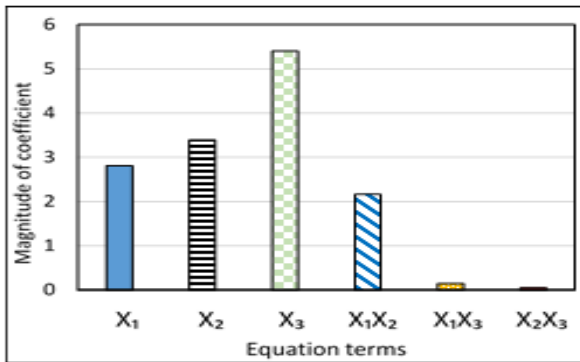

Figure 1:   Pareto chart for TSS removal.

It can also be noticed that the TSS removal is negatively affected by the interaction between the pH and the settling time, even though these two factors individually increase the TSS removal, the negative sign of the term X_2X_3 means that the chitosan precipitate as settling time is prolonged, depending on the pH of the sample. However, in this equation, the effects of the interaction terms X_1X_3 and X_2X_3 are minimal.

The quadratic terms indicate the presence curvatures. The negative signs in eqn (6) reveal that the quadratic curves for X_1^2, X_2^2 and X_3^2 are concave. This means that removal of TSS increases with an increase in coagulant dosage, pH and settling time up to a maximum value beyond which the efficiency decreases with further increase of the three variables.

The magnitude of the coefficients of the quadratic terms indicates the steepness of the curvature. Thus it can be seen from Table 3 and eqn (6) that X_1^2 is steeper than X_2^2 and X_3^2. The value of the correlation coefficient of R^2 of 0.976 shows that only 2.4% of the total variation could not be explained by the empirical model.

The RSM plot varies two factors while keeping the other constant. Fig 2(a), 2(b) and 2(c) were plotted by keeping the settling time constant at 60 minutes, at pH 7 and coagulant dosage of 20 ml respectively.

From Fig. 2(a), it can be seen that the optimum chitosan dosage was 20 ml which corresponds to 7.41 mg/l, beyond which the efficiency decreased. The optimum pH range was found to be between 8.5 and 10. The removal of TSS increases with the increase in settling time (Fig. 2(b)), as mentioned earlier in Fig. 1.

The plots in Fig. 2(c) confirm the pH ranges and stresses on the significance of the settling time. The maximum TSS removal efficiency was achieved at 90 minutes.
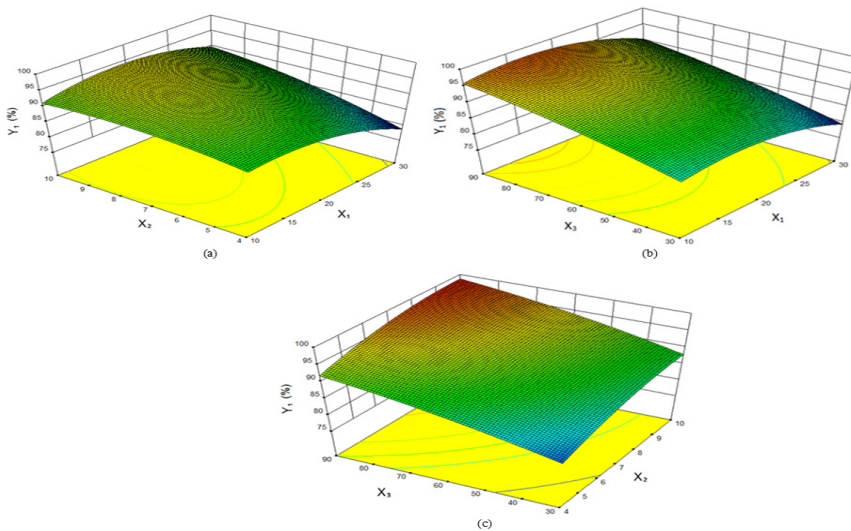

Figure 2:   RSM plots for TSS removal (a) coagulant loading and pH (b) coagulant loading and time (c) time and pH.

3.2 Analysis for colour reduction

The ANOVA (Table 4) indicates that the regression is significant due to the F-value of 11.26 and a p>F value of 0.0021. There is 19.45 % chance that the LOF is due to noise which is insignificant; meaning that the model can fit any data.

The regression is expressed as follows:

$$Y_2 = 84.8 - 1.36X_1 + 5.26X_2 + 4.46X_3 + 0.58X_1X_2 - 0.15X_1X_3$$
$$+ 2.88X_2X_3 - 3.25X_1{}^2 - 5.38X_2{}^2 - 2.8X_3{}^2 \qquad (7)$$

It can be noticed from Fig. 4 that X_2 is the most significant linear term, followed by X_3. Whereas the p-value of X_1 indicates that it is an insignificant term in the model equation. This means that the removal of colour using chitosan is affected more by pH of the sample and the settling time than it is affected by the chitosan dosage.

The positive signs of the linear coefficients indicate that the colour removal increases with an increase in pH, settling time and coagulant dosage. The p-value of X_2X_3 (Table 4 and Fig. 3) shows that it is the most significant interaction term.

Therefore, the colour removal will increase with the simultaneous increase of the pH and the settling time. The colour removal is also increased to a smaller extent by the interaction between the coagulant dosage and the pH of the sample.

The negative sign of the term X_1X_3 show that the colour removal efficiency decreases with an increase in CCo dosage and settling time. The R^2 of 0.935 shows that only 6.5% of the total variation could not be explained by the model.

Table 4:   ANOVA for colour removal.

| Variation	SS	df	MS	F-value	p-value p>F
Regression	642.13	9	71.35	11.26	0.0021
X_1	14.85	1	14.85	2.34	0.1697
X_2	218.40	1	218.40	34.46	0.0006
X_3	159.31	1	159.31	25.14	0.0015
X_1X_2	1.32	1	1.32	0.21	0.6616
X_1X_3	0.090	1	0.090	0.014	0.9085
X_2X_3	33.06	1	33.06	5.22	0.0563
$X_1{}^2$	44.47	1	44.47	7.02	0.0330
$X_2{}^2$	117.16	1	117.16	18.49	0.0036
$X_3{}^2$	33.01	1	33.01	5.21	0.0564
Residual	44.36	7	6.34		
LOF	29.10	3	9.70	2.54	0.1945
Pure Error	15.26	4	3.81		
Total	686.50	16			
$R^2 = 0.935$					
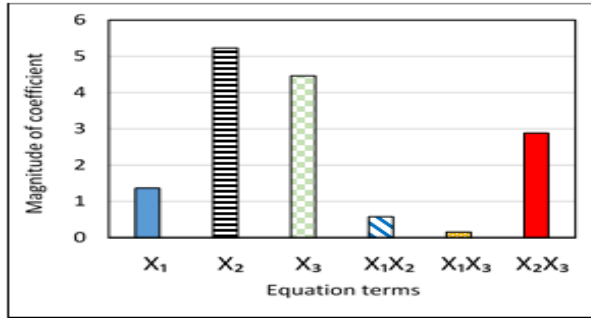
Figure 3:   Pareto chart for colour removal.

Fig 4(a), 4(b) and 4(c) were plotted by keeping the settling time constant at 60 minutes, pH at 7 and coagulant dosage at 20 ml respectively. From Fig. 4, it can be noticed that the optimum colour removal was achieved at a pH of 10, CCo loading of 20 ml (7.41 mg/l) and a settling time of 90 minutes.
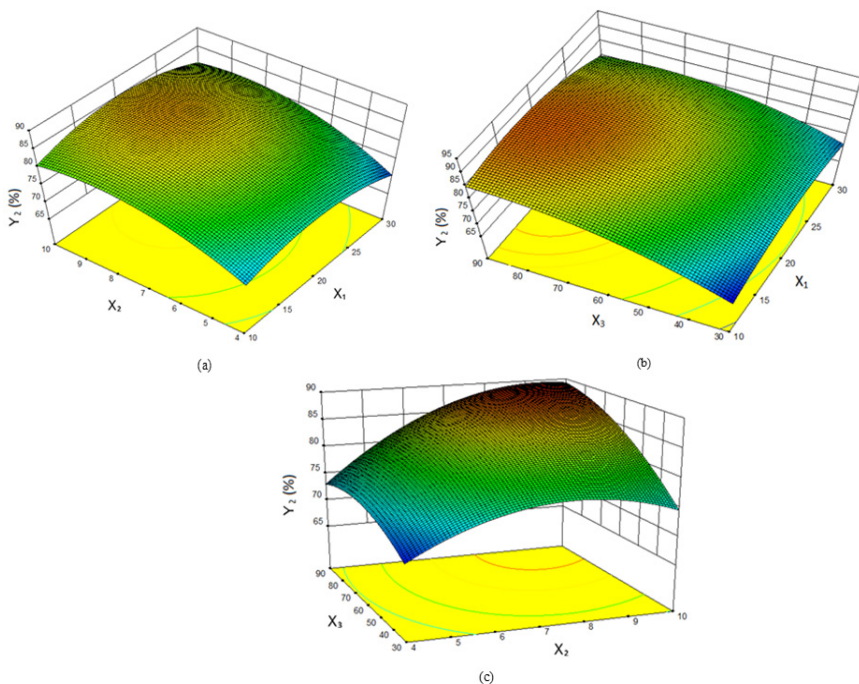

Figure 4:   RSM plots for colour removal (a) coagulant loading and pH (b) coagulant loading and time (c) time and pH.

Using the RSM for industrial experiments had the following advantages:
(1)   The optimum conditions could be spotted from the experimental matrix without requiring special statistical training: The experimenter can see the

best combination of variables from the matrix (Table 1). And the statistical analyses generated model equations that can predict the variation in the system without laboratory experiments.

(2) The RSM allowed the experimenter to analyze all factors and responses simultaneously. It also allowed the identification of interacting variables.

In previous studies done for this particular effluent using the OFAT method [1] the optimum conditions were obtained under acidic conditions. This is undesirable due to the possible damage to the pipelines and the fact that it does not comply with government regulations. However, when all the parameters were studied simultaneously, it was found that good results could be obtained under basic conditions.

4   Conclusions

In this study the Box-Behnken method was used to design the order of the experimental runs and to statistically analyze the experimental data. The settling time was found to have a great impact on all the responses investigated. The performance of the chitosan increase until a maximum point beyond which the efficiency decreased. The TSS removal (99%) and the colour removal (90%) obtained using the BBD are better than the results obtained in the treatment of the FE using OFAT (87% TSS removal and 76% colour reduction). The ANOVA table and the pareto chart provide the identification of important variables. The use of RSM generate 3D plots that allow the experimenter to visualize the behaviour of the system being investigated.

The application of RSM in industrial research was found to be more advantageous than the traditional OFAT due to the use of experimental matrices such as the BBD, the use of statistical analysis that generate model equations that can be used to predict and optimize the output of a particular process.

References

[1]   Pambi, R. L. L. & Musonge, P., The efficiency of chitosan as a coagulant in the treatment of the effluents from the sugar industry, *Journal of polymers and materials, 32*(*1*), pp. 59-65, 2015.

[2]   Chi, F. H. & Cheng, W. P., Use of Chitosan as Coagulant to Treat Wastewater from Milk Processing Plant, *Journal of polymer and environment, 13*, pp. 411-417, 2006.

[3]   McDaniel, W. R. & Ankenman, B. E., Comparing experimental design strategies for quality improvement with minimal changes to factor levels, *Quality and Reliability Engineering International, 16*(*355-362*), 2000.

[4]   Qu, X. & Wu, C. F. J., One-factor-at-a-time designs of resolution V, *Journal of statistical planning and inference, 131*(*2*), pp. 407-416, 2005.

[5]   Czitrom, V., One-factor-at-a-time versus designed experiments, *The American statistician, 53*(*2*), pp. 126-131, 1999.

[6]   Montgomery, D. C., *Design and nalysis of experiments*, 6th ed. Hoboken: NJ: John Wiley & Sons, Inc., 2005.

[7]   Wahid, Z. & Nadir, N., Improvement of one factor at a time through design of experiments, *World Applied Sciences Journal, 21*(*Mathematical Applications in Engineering*), pp. 56-61, 2013.

[8]   Akintunde, A. M., Ajala, S. O., & Betiku, E., Optimization of Bauhinia monandra seed oil extraction via artificial neural network and response surface methodology: A potential biofuel candidate, *Industrial Crops and Products, 67*(*0*), pp. 387-394, 2015.

[9]   Bezerra, M. A., Santelli, R. E., Oliveira, E. P., Villar, L. S., & Escaleira, L. A., Response surface methodology (RSM) as a tool for optimization in analytical chemistry, *Talanta, 76*(*5*), pp. 965-977, 2008.

[10] Baş, D. & Boyaci, I. H., Modeling and optimization : Usability of response surface methodology, *Journal of Food Engineering, 78*(*3*), pp. 836-845, 2007.

[11] Trinh, T. K. & Kang, L. S., Response surface methodological approach to optimize the coagulation-flocculation process in drinking water treatment, *Chemical Engineering Research and Design, 89*(*7*), pp. 1126-1135, 2011.

[12] Pambi, R. L. L. & Musonge, P., Influence of effluent type on the performance of Chitosan as a coagulant, *International journal of futuristic trends in engineering and technology, 2*(*1*), pp. 1-6, 2014.

[13] Freese, S. D., Trollip, D. L., & Nozaic, D. J. (2003, K5/1184). *Manual for testing of water and wastewater treatment chemicals.* Available: http://www.wrc.org.za/knowledge%20hub%20documents/.../1184%20web.pdf

[14] Bina, B., Mehdinejad, M. H., Nikaeen, M., & Movahedian Attar, H., Effectiveness of chitosan as natural coagulant aid in treating turbid waters, *Iranian journal of environmental health science & engineering editorial, 6*(*4*), pp. 247-252, 2009.