Regular Article

PHYSICAL CHEMISTRY RESEARCH
Published by the
Iranian Chemical Society
www.physchemres.org
info@physchemres.org

*Phys. Chem. Res., Vol. 12, No. 2, 525-547, June 2024*
**DOI: 10.22036/pcr.2023.406044.2371**

# Liquid Membrane-based Extraction of Arsenic: Part 2-Optimization through Statistical and Machine Learning Approach

S. Sarkar and P. Saha*

*Department of Chemical Engineering, Indian Institute of Technology Guwahati, Assam 781039, India*

*(Received 15 July 2023, Accepted 6 November 2023)*

This paper is a continuation of the work presented in Part 1 of this series. The aim of this Part is to find optimal operating condition of the process through two comparative approaches that are statistical approach and machine learning-based model. The data were generated through 3 phase SLM experimentations. Variables such as pH, concentration, temperature were selected through the design of experiments, and % extraction/recovery of arsenic were recorded. Three important aspects of the statistical analysis (descriptive, correlational and inferential statistical analyses) were evaluated in comprehensive manner. Associated tests such as Shapiro-Wilk test and Kolmogorov-Smirnov test were conducted to check for normality. The homogeneity of variances of the dependent variable with respect to the independent variable were checked through the Levene's test. The correlational analysis were studied using Spearmann's test and Pearson's correlational analysis along with standard ANOVA and/or MANOVA. In case of two phase study, Spearmann's correlation analysis was carried out based on the Shapiro-Wilk test and the skewed distribution of the data. In three-phase SLM study, Pearson's correlational analysis was assessed as the data was normally distributed and symmetric. The correlational coefficients in two phase study suggested that the extraction of arsenic had a significant negative correlation with pH of the feed phase and a positive correlation with extractant concentration. However, in case of SLM, the dependent variables showed a linear relation. An increase in the extraction percentage of both the arsenic ions led to an increases in the recovery percentage. Five way ANOVA for two phase systems and multivariate ANOVA for three phase systems generated the tests of between-subjects effects. The receiving phase concentration had a significant and main effect in the multivariate test. Since significant main effects obtained for the variables, post hoc (Tukey HSD) analyses were computed to understand the effect of individual and combined independent variables on the dependent variables. Artificial Neural Network has been adopted in machine learning model along with Genetic Algorithm-based optimization tool to compare their performances with the obtained experimental and statistical data. The data points were divided to train, test and validate the ANN based on maximum extraction% and recovery% with minimum mean squared error (MSE). In the two-phase study, the minimum MSE was achieved with less than 8 neurons in the hidden layer for all the arsenic species. However, in three phase study, the minimum MSE was achieved with 4, 5, 27, 14, and 21 neurons in the hidden layer for As(III), As(V), As(III):As(V)::1:1, As(III):As(V)::1:2, and As(III):As(V)::2:1, respectively. The machine learned results on the data of As(III)and As(V) were not up to the mark in the SLM study. There was considerable process/model mismatch. Thus, optimization was not successful. On the other hand, the machine-learned results on the SLM data of As(III):As(V)::1:1, As(III):As(V)::1:2, and As(III):As(V)::2:1 were better than the statistical model. The predicted values of extraction and recovery were very close to the experimental values with less than 1.5% error in most cases.

**Keywords:** Arsenic, Statistical modelling, Machine learning, Extraction, Liquid membrane

## INTRODUCTION

Arsenic is widely available in groundwater in India, especially in Indo-Ganga basin and the Ganga-Brahmaputra delta region [1,2]. It is extremely hazardous to human health [3], and hence, mitigation of arsenic from water is utmost necessary. Arsenic mostly remains in inorganic forms as oxyanions of trivalent arsenite As(III) or pentavalent arsenate As(V) in the groundwater [4]. Extraction-based removal of arsenic from water has been advocated by various researchers

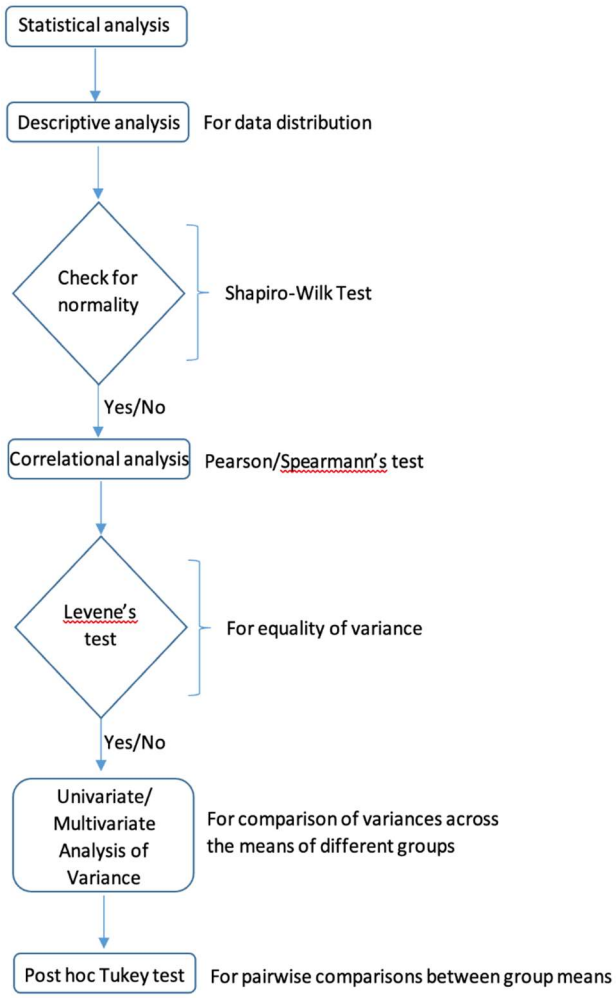
*Corresponding author. E-mail: p.saha@iitg.ac.in

[5,6]. Flat sheet supported liquid membrane (SLM) has been used to study the transfer of As(V) ions [7] and in separation of arsenate and arsenite from aqueous media [8].

This work aims at understanding the modelling of the extraction-based arsenic removal process through both the two-phase and the three-phase experimentations. The entire work was divided into two parts series. In the Part 1, complete mathematical modelling was established, and reaction mechanism of the transport of arsenic from one aqueous phase to the other via thin liquid membrane were explored. It was observed that sesame oil was the most effective environmentally benign diluent for arsenic extraction, while Aliquat 336 was the preferred extractant due to its ability to react with both dissociated and undissociated ions of arsenic. The extractant concentration of 10% (v/v) was found to be the optimum parameter for the extraction of individual and combined arsenic ions in two phase extraction study. The optimum extraction for As(III) was obtained at pH of 6.8, temperature of 37 ºC, stirring speed of 176 rpm, and with 9 h duration of two phase extraction study. In case of As(V), the optimum conditions included pH of 6.7, temperature of 55 ºC, stirring speed of 170 rpm, and 12 h duration. The optimum conditions for the extraction of combined arsenic species was in the mentioned range. It was revealed that the formation of Arsenic-Aliquat 336 complex in the organic phase followed the stoichiometric ratio of 1:1, whereas the extraction of As(V) into the organic phase was more favourable in comparison to As(III) and combined arsenic species. In three phase study, the extraction and/or recovery was optimum at 30:1 ratio of iron and arsenic, pH between 5 and 7, and 40% of the pseudo binary mixture of Aliquat 336, and sesame oil was the optimum composition of the liquid membrane for arsenic extraction and recovery. Mass transfer resistances were more during transport of As(V) than that of As(III). It was revealed in an earlier work of our team [5] that two-phase equilibrium was mainly dependant on 5 parameters including pH of the feed phase, extractant concentration (vol/vol), duration of experiment (hours), temperature (°C), and stirring speed (rpm). The three-phase SLM was mainly dependant on 3 parameters including concentration of receiving phase, pH of the receiving phase, and either stirring speed (rpm) or extractant concentration (%), depending on whether the arsenic ions were present in single species, *i.e.* As(III) or As(V), or combination of both in whatever ratios. Design of experiment was performed on these parameters within a pre-defined variation limits (vide Table ST1). Experimentations were conducted and their corresponding extraction and recovery results were recorded (vide Table ST2, Table ST3, and Table ST4). The crucial process parameters were identified. Mass transfer coefficients, flux, phase resistances, and permeability were computed via mathematical model. In summary, the extraction and recovery of arsenic and their complete mathematical modelling aspect was understood.

In this part (part 2), modelling of arsenic transport process was explored through statistical and machine learning approach. The experimental results, shown in Table ST2, Table ST3, and Table ST4, were useful for the above purpose as well. A proper statistical analysis would be complete if three important aspects of the analysis, including descriptive, correlational and inferential statistical analyses, are performed in comprehensive manner. Descriptive statistical analysis would yield the mean, standard deviation, skewness, and kurtosis for the independent variables with respect to the dependent variables. A normality test can determine if a data set is well-modelled by a normal distribution, and how likely it is for a random independent variable underlying the data set to be normally distributed. Kolmogorov-Smirnov [9,10] and Shapiro-Wilk [11,12] are two well-known models with which the normality tests can be performed, and on the basis of the normality test, Pearson correlation [13] analysis (α) can be executed. The Pearson correlation coefficients ranges from -1 to 1. An absolute value of exactly 1 implies that a linear equation describes the relationship between the independent and dependent variables perfectly [14]. The positive sign implies that dependent variable increases with an increase in the independent variables, while the negative value implies the opposite effect. A value of 0 implies that there is no linear dependency between the variables. Five-way Analysis of Variance (ANOVA) is needed for two-phase experimental results, whereas multivariate ANOVA is needed for three phase experimental results, as it has two dependent variables (extraction % and recovery %). Various trace statistics such as Pillai's trace [15], Wilk's Lambda [16], Hotelling's trace [17] and Roy's largest root [18] are needed to be computed. In summary, the multivariate analysis predicts the individual or combined effect of

independent variables on the dependent variables. The following flowchart demonstrates the order in which the statistical analysis was performed for this work.




On the other hand, machine learning [19] is a technique of data analysis that automates analytical model building. It is a branch of artificial intelligence [20] that states that model can self-learn from data, with minimal human intervention. In case of data analysis, it is possible to employ machine learning techniques to the experimental results, as shown in Table ST2, Table ST3, and Table ST4, and obtain a black-box model that maps independent variables to the dependent variable(s). The significance of such model is based on the fact that the machine learned model is an ever-evolving model which can take new set of experimental results at a future time and capture the alterations in the process behaviour, if any. Artificial Neural Network (ANN) [21] is a well-established tool for such purpose.

This work incorporates the application of statistics and machine learning-based approaches on liquid membrane-based separation of arsenic from water for the first time. The statistical and machine learning model helps in predicting the optimum process condition that ensures the highest yield in terms of optimum extraction % and/or recovery%. The experimentally observed optimum extraction and/or recovery efficiency of arsenic was compared with that found through statistical model and/or machine learning model. The specific measurable objectives for defining optimal conditions include:

1. Achieving a maximum extraction efficiency of arsenic within realistic experimental levels/boundaries.
2. Attaining a high recovery efficiency of arsenic that aligns with actual groundwater conditions.
3. Ensuring compliance with established standards and regulations for groundwater quality during the extraction and recovery process.

This provides a complete framework with experimental, mathematical, statistical, and machine learning-based approaches to work with similar type of metal contaminant in water, and obtain an optimal operating condition for the target metal contaminant. Genetic Algorithm-based optimization tool [22] is useful for this purpose. The MATLAB®(version 2022b) software was used for simulation on a Mac mini (M1, 2020) running on macOS Monterey Version 12.5.1.

## THEORETICAL BACKGROUND

### Statistical Modelling

**Analysis using SPSS.** Statistical analysis is significant to investigate trends, patterns, and find relationships within the data. The analysis can be divided into three parts; first is the descriptive analysis that describes and characterizes the data set, second is the correlational analysis to predict the pattern of the variables, and third (inferential) is the cause and effect or group comparison analysis to find the relation between the variables in the quantitative data. Inferential statistical analysis can be further subdivided into correlational analysis

and analysing differences between the groups. From the descriptive analysis, the mean, median, variance, standard deviation, range, interquartile range, skewness and kurtosis were obtained for each levels of the independent variables with respect to the dependent variables. Then the test for normality was done using two models, Kolmogorov-Smirnov and Shapiro-Wilk. On the basis of the normality test, either Spearmann's test or Pearson's correlation analysis was performed to find the relationship between the independent and dependent variables. In the two-phase equilibrium study, there were five independent variables and one dependent variable for all the single and mixed arsenic species. The independent variables were further subdivided into three levels. Five-way Analysis of Variance (ANOVA) test was carried out to find the differences between independent variables with respect to one dependent variable (% extraction) in the two-phase extraction equilibrium study, whereas multivariate Analysis of Variance (MANOVA) test was done for analysing the differences between groups, as there were three independent variables and two dependent variables (%extraction and %recovery) in three phase SLM study. Pillai's trace, Wilk's Lambda, Hotelling's trace and Roy's largest root are positive-valued statistics ranging from 0 to 1. Increasing values of Pillai's trace indicates the effects that contribute more to the model, while the decreasing values of Wilk's Lambda indicate the effects that contribute more to the model. Hotelling's trace is the sum of the eigenvalues of the test matrix for which increasing values indicate the effects that contribute more to the model. Hotelling's trace is always larger than Pillai's trace, but when the eigenvalues of the test matrix are small, these two statistics are nearly equal. This predicts that the effect does not contribute much to the model. Roy's largest root is the largest eigenvalue of the test matrix. Similar to Hotelling's trace, increasing values indicate the effects that contribute more to the model. Roy's largest root is always less than or equal to Hotelling's trace. When these two statistics are equal, the effect is predominantly associated with just one of the dependent variables. This is due to a strong correlation between the dependent variables, or because the effect does not contribute much to the model. The multivariate analysis predicts the individual or combined significant effect of independent variables on the dependent variables. If there is any significant main effect, then the tests of between-subjects effects is investigated. This indicates the impact of individual or combined independent variable on each dependent variable. If there is any significant main effect, then the post hoc test analysis is carried out to further explain the effect.

## Machine Learning Approach

Machine learning approach employs a set of computational tools that can recognize the pattern of a process using a particular dataset. Artificial Neural Network (ANN) is one such tool that can map a relationship between inputs and outputs of a process and then use it to anticipate the process behaviour [23]. Furthermore, different techniques are available for nonlinear optimization, such as Nelder-Mead simplex Method, Trust-Region method, interior search Method, and genetic algorithm method [24]. However, ANN is not necessarily a differentiable function and it can predict the value of the objective function in a constrained region. The genetic algorithm fulfils both the requirements as it is a non-derivative constrained optimization technique, and thus was preferred in this study.

**ANN based modelling.** ANN comprises three layers that are input, hidden, and output layers. These layers are interconnected to each other by functions $\Psi(\cdot)$ which mimic neurons as shown in Fig. 1. Each neuron consists of an activation function [23] followed by a summing junction. The inputs $u_i$, assigned weight $\omega_i$, and bias $b_i$ are combined into an argument which is fed into the activation function. There are various options of activation functions, and the best fitting function depends upon the type of dataset which needs to be detected through trials. These neurons act on the data and extract useful information from the dataset [23,25]. A mean squared error (MSE) function yields the error between the experimental and predicted values and helps to predict the accuracy of the model,

$$MSE = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(y_{e_i^j} - y_{p_i^j}\right)}{mn} \tag{3}$$

Here, $m$ is the number of input datasets, $n$ is the number of output nodes, $y$ is the value of the output with subscripts $e$ and $p$ referring to experimental and predicted, respectively. The number of nodes in the hidden layer is optimized to minimize the MSE.

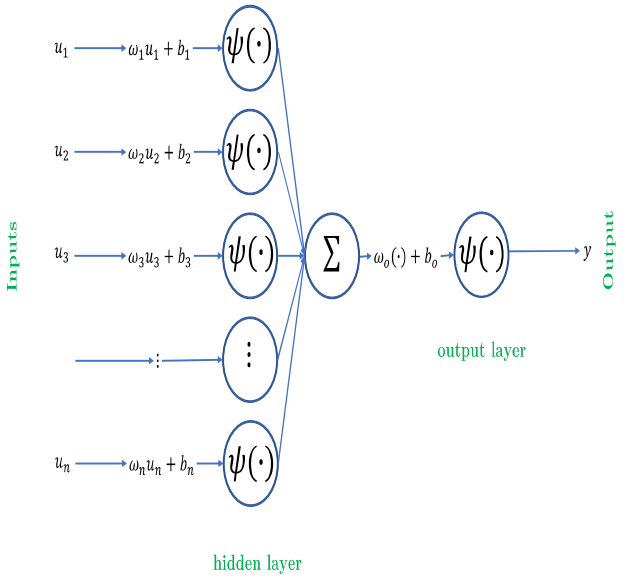



**Fig. 1.** Schematic of ANN structure.

**Optimization through genetic algorithm.** A genetic algorithm (GA) is based on the natural selection theory of biological evaluation [25]. In GA, a problem is initiated with a given population group, which is evaluated based on the fitness function towards the optimum solution. GA-based optimization involves several steps such as selection (selecting the number of individuals based on fitness function from population to breed the new population), crossover (generating the new data point by combining the information of two data points from selected data), and mutation (random change in the information of a new data point similar to the mutation in the biological world). As simulation proceeds, the GA converges in the more suitable population (set of data points) and eventually leads to the optimum solution based on the fitness functions.

## RESULTS AND DISCUSSION

### Two Phase Extraction Equilibrium Studies

It should be noted that the extraction in two-phase equilibrium depends mainly on 5 factors:

A: pH of the feed phase
B: %Extractant concentration (vol/vol)
C: Duration (h)
D: Temperature (°C)
E: Stirring speed (rpm)

The experiments were carried out as per the design of experiment, where all the 5 factors were varied within a pre-designed range (vide Table ST1) in the same manner for all the 5 cases including As(III), As(V), As(III):As(V)::1:1, As(III):As(V)::1:2 and As(III):As(V)::2:1. The corresponding %extraction was recorded and shown in Table ST2.

**Statistical analysis.** The equations of the quadratic models in terms of coded factors for extraction of As(III), As(V), As(III):As(V)::1:1, As(III):As(V)::1:2 and As(III):As(V)::2:1 are given in the following equations:

$$Y_{As(III)} = 71.02-4.68A + 6.19B + 3.25C + 1.88D + 2.34E + 1.79AB + 1.63AC + 0.3497AD + 0.6303AE - 1.87BC - 0.2241BD -0.3797BE + 0.0659CD + 0.1078CE + 0.1741DE-12.49A^2 + 5.64B^2-1.17C^2-4.52D^2 -7.03E^2 \quad (2)$$

$$Y_{As(V)} = 74.76 - 6.65A + 6.35B + 3.24C + 1.71D + 0.8529E - 0.6563AB - 0.0937AC + 0.0312AD - 0.0937AE - 0.4687BC - 0.4687BD - 0.0937BE - 0.1562CD + 0.0938CE - 0.0313DE - 23.15A2 + 1.15B^2 - 0.1474C^2 - 0.1474D^2 - 0.6474E^2 \quad (3)$$

$$Y_{As(III):As(V)::1:1} = 80.51 - 8.46A + 5.85B + 3.83C + 1.22D + 1.75E + 2.75AB + 1.76AC + 0.3431AD + 0.9788AE - 1.24BC - 0.1206BD - 1.03BE - 0.2506CD - 0.1912CE - 0.0225DE - 15.71A^2 + 0.2463B^2 - 3.94C^2 + 1.49D^2- 3.4E^2 \quad (4)$$

$$Y_{As(III):As(V)::1:2} = 81.87 - 6.67A + 4.44B + 2.81C + 1.02D + 0.7679E + 0.4422AB + 0.4022AC + 0.1322AD + 0.3616AE - 0.4641BC - 0.4078BD - 0.2347BE - 0.1141CD + 0.0241CE - \quad (5)$$

$$Y_{As(III):As(V)::2:1} = 0.2584DE - 16.42A^2 - 0.444B^2 - 0.469C^2 - 0.034D^2 - 1.22E^2 \quad (6)$$
$$78.52 - 7.03A + 5.23B + 3.38C + 0.8882D + 0.625E + 1.62AB + 1.47AC - 0.2103AD + 0.5891AE - 0.8397BC - 0.4772BD + 0.3059BE - 0.8747CD + 0.2409CE - 0.0853DE - 13.35A^2 + 0.5106B^2 - 2.18C^2 - 0.7444D^2 - 1.58E^2$$

These are used to make predictions about the response for given levels of each factor. The analysis of variance is shown in Table 1 and Tables ST5-ST8. Table 2 presents the optimum conditions obtained for maximum extraction of arsenic ions. The "Stats" and "ML" columns of Table 2 refer to "Statistical" and "Machine Learning", respectively. This section discusses on the results of statistical modelling, while the simulation results of machine learning model are discussed in section 3.1.2. The predicted maximum extraction of As(III) and As(V) were 82.37% and 84.15%, as opposed to 84% and 86% respectively obtained through experimentations. On the other hand, the predicted maximum extraction of combined compounds such as As(III):As(V)::1:1, As(III):As(V)::1:2 and As(III):As(V)::2:1 were 87.03%, 87.59% and 85.1%, as opposed to 85.5%, 86% and 84.5% respectively obtained through experimentations, as shown in Table 2. A significant

**Table 1.** Analysis of Variance for Two-phase Extraction of As(III)

| Source | Sum of squares | df | Mean square | F-value | p-value |
|---|---|---|---|---|---|
| Model | 6958.94 | 20 | 347.95 | 11.55 | <0.0001 |
| A | 745.99 | 1 | 745.99 | 24.76 | <0.0001 |
| B | 1303.74 | 1 | 1303.74 | 43.27 | <0.0001 |
| C | 359.39 | 1 | 359.39 | 11.93 | 0.0017 |
| D | 119.64 | 1 | 119.64 | 3.97 | 0.0558 |
| E | 186.5 | 1 | 186.5 | 6.19 | 0.0188 |
| AB | 102.21 | 1 | 102.21 | 3.39 | 0.0758 |
| AC | 85.25 | 1 | 85.25 | 2.83 | 0.1033 |
| AD | 3.91 | 1 | 3.91 | 0.13 | 0.7212 |
| AE | 12.71 | 1 | 12.71 | 0.42 | 0.5211 |
| BC | 111.49 | 1 | 111.49 | 3.7 | 0.0643 |
| BD | 1.61 | 1 | 1.61 | 0.053 | 0.819 |
| BE | 4.61 | 1 | 4.61 | 0.15 | 0.6984 |
| CD | 0.14 | 1 | 0.14 | $4.617 \times 10^{-3}$ | 0.9463 |
| CE | 0.37 | 1 | 0.37 | 0.012 | 0.9123 |
| DE | 0.97 | 1 | 0.97 | 0.032 | 0.8589 |
| $A^2$ | 385.96 | 1 | 385.96 | 12.81 | 0.0012 |
| $B^2$ | 78.63 | 1 | 78.63 | 2.61 | 0.1171 |
| $C^2$ | 3.4 | 1 | 3.4 | 0.11 | 0.7395 |
| $D^2$ | 50.57 | 1 | 50.57 | 1.68 | 0.2054 |
| $E^2$ | 122.12 | 1 | 122.12 | 4.05 | 0.0535 |
| Residual | 873.8 | 29 | 30.13 | | |
| Lack of fit | 734.28 | 22 | 33.38 | 1.67 | 0.2485 |
| Pure error | 139.52 | 7 | 19.93 | | |
| Total | 7832.74 | 49 | | | |

model with F-value of 11.55 (vide Table 1) was obtained for As(III). The feed phase pH and extractant concentration were found to be the most significant parameters as the p-value was <0.0001 in both the cases (vide Table 1). The F-value for lack of fit was 1.67, which was not significant; and there was a 24.85% chance that a large value could occur due to noise. The coefficient of determination ($R^2$) was found to be 0.8884 with adjusted $R^2$ value of 0.8115, and the predicted $R^2$ value was 0.6368. The predicted $R^2$ value is in agreement to the adjusted $R^2$ value. Furthermore, a ratio of 15.818 was obtained for adequate precision indicating a desirable signal to noise ratio. The quadratic model obtained for As(V) was significant with F-value equal to 134.26 (vide Table ST5). In this case, the pH of the feed phase, extractant concentration, duration and temperature were the significant parameters as reported in Table ST5. A non-significant lack of fit value of2.38 was obtained with 12.06% chances of occurrence of this value due to noise. A high value of the coefficient of determination ($R^2$) (0.9893) indicated that this model was significant. The predicted $R^2$ (0.9655) was in close agreement with the adjusted $R^2$ value (0.9819). Additionally, the adequate precision value of 39.394 indicated an adequate signal with a very high signal to noise ratio.

The significant F-values of the models for As(III)-As(V) combined in different ratios were 40.77 (for As(III):As(V)::1:1), 58.44 (for As(III):As(V)::1:2) and 53.75 (for As(III):As(V)::2:1), with a non-significant lack of fit values that are shown in Table ST6, Table ST7 and Table ST8, respectively. The pH of the feed phase, extractant concentration and duration were the significant parameters for all the three combined ratios of As(III)-As(V). The high values of $R^2$ stipulated the models to be significant. The reasonable agreement between adjusted $R^2$ and predicted $R^2$ values with a difference less than 0.2 and high adequate precision values further indicated the significance of the models. Figure 2 and Fig. SF1 show the minimum, maximum, lower quartile, median and upper quartile with outliers obtained from the descriptive statistical analysis. The box plots shown in Fig. 2 and Figs. SF6-SF7 help to visualize the interquartile range, in order to understand the distributions of the different levels of independent variables with respect to the dependent variable. The length of the box indicates the variation of the data. In case of As(III), the median was close to the lower quartile or upper quartile, indicating a non-normal skewed distribution as shown in Fig. 2. Figs. SF1a-SF1c for As(V) shows the spread of the data set that is similar to As(III), with mild and strong outliers lying beyond the minimum and maximum values. For As(III)-As(V)combinations, as given in Figs. SF1d-SF1l, the distribution appears to be either left-skewed or right skewed.

**Table 2.** Optimization and Error Analysis for Two-phase Extraction of Arsenic Ions

| Model parameters | As(III) | | As(V) | | As(III):As(V)::1:1 | | As(III):As(V)::1:2 | | As(III):As(V)::2:1 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Stats | ML | Stats | ML | Stats | ML | Stats | ML | Stats | ML |
| $R^2$ | 0.8884 | - | 0.9893 | - | 0.966 | - | 0.976 | - | 0.974 | - |
| Adjusted $R^2$ | 0.8115 | - | 0.9819 | - | 0.942 | - | 0.959 | - | 0.956 | - |
| Predicted $R^2$ | 0.6368 | - | 0.9655 | - | 0.882 | - | 0.916 | - | 0.909 | - |
| Adequate precision | 15.818 | - | 39.394 | - | 26.14 | - | 27.63 | - | 28.94 | - |
| pH of feed phase | 6.81 | 6.91 | 6.65 | 7.6262 | 6.59 | 7.1797 | 6.65 | 7.5039 | 6.26 | 6.4758 |
| Extractant concentration % (vol/vol) | 10 | 10 | 10 | 7.784 | 10 | 10 | 10 | 8.717 | 10 | 10 |
| Duration (h) | 9 | 12 | 12 | 12 | 7 | 7.8729 | 12 | 8.7519 | 9 | 5.4808 |
| Temperature (°C) | 37 | 48.94 | 55 | 37.87 | 47 | 35.334 | 35 | 41.892 | 45 | 55 |
| Stirring speed (rpm) | 176 | 161.43 | 170 | 154.7 | 196 | 202.95 | 173 | 245.39 | 143 | 158.171 |
| Predicted extraction (%) | 82.37 | 87.82 | 84.15 | 83.13 | 87.03 | 85.56 | 87.59 | 86.39 | 85.1 | 85.94 |
| Obtained extraction (%) | 84 | | 86 | | 85.5 | | 86 | | 84.5 | |
| Error (%) | 1.94 | 4.54 | 2.15 | 3.33 | 1.79 | 0.07 | 1.85 | 0.45 | 0.71 | 1.7 |

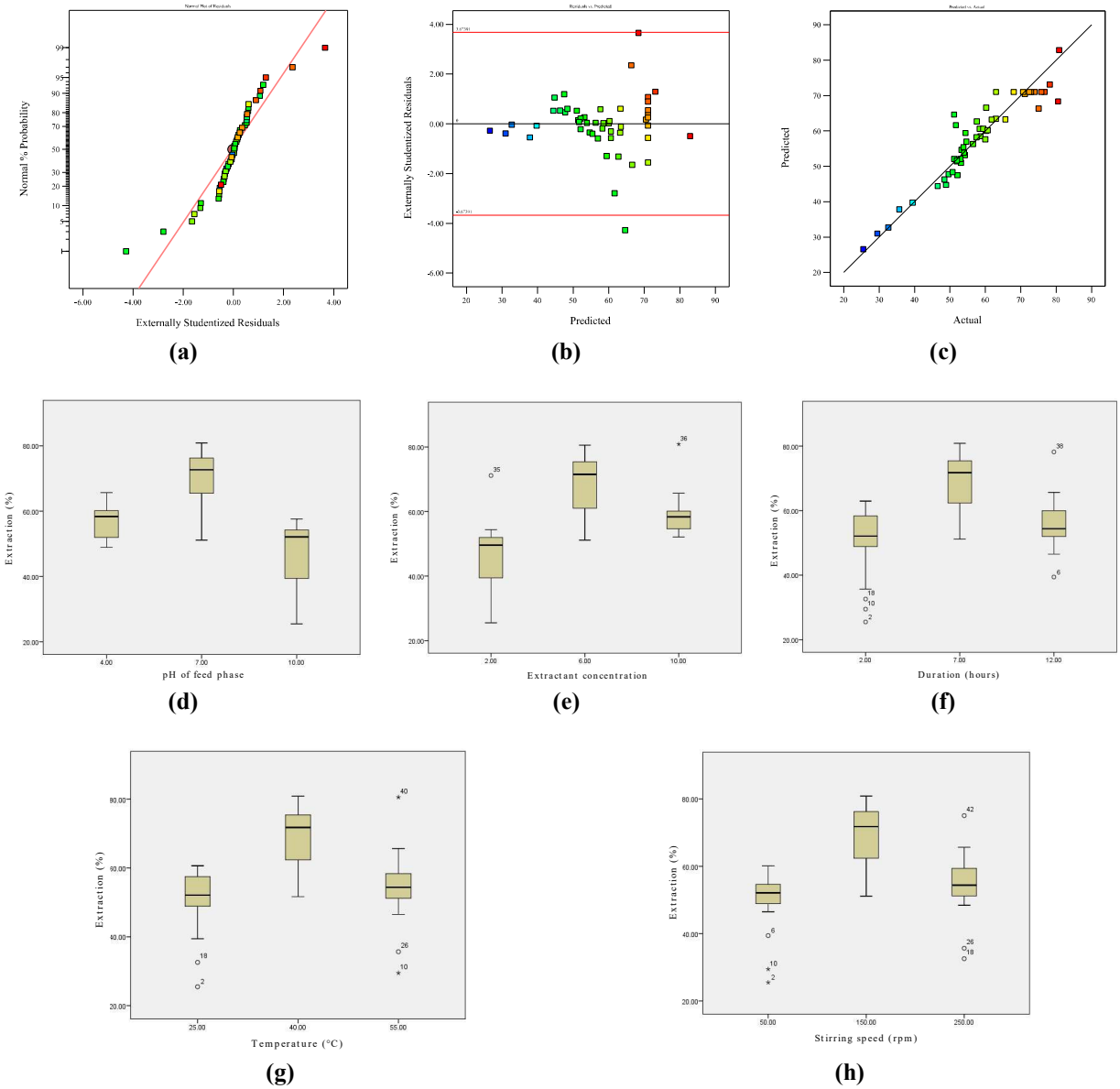



**Fig. 2.** Statistical analysis of the quadratic model predicted and box plots for extraction of As(III).

Shapiro-Wilk test is preferred over the Kolmogorov-Smirnov test for assessing the normality of the data, as it is more appropriate for small sample size (& 50). Table 3 shows the normality test results for the extraction of arsenic ions using Shapiro-Wilk model. Based on the Shapiro-Wilk test and the skewed distribution of the data, Spearmann correlation analysis was carried out to evaluate the relationship between the dependent variable (extraction) and each independent variable, as shown in Table 4 and Table ST9. As it is evident from the data, the extraction of arsenic had a significant correlation with pH of the feed phase and the extractant concentration. The correlation coefficient were consistently negative for the case of pH of feed phase, while it was consistently positive for extractant concentration. A negative value suggests that a monotonic relationship exists when the extraction efficiency increases with ta decrease in

**Table 3.** Normality Test for the Extraction of Arsenic Ions Using Shapiro-Wilk Model

| Independent variables | Range | df | As(III) | | As(V) | | As(III):As(V)::1:1 | | As(III):As(V)::1:2 | | As(III):As(V)::2:1 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Statistic | Significance | Statistic | Significance | Statistic | Significance | Statistic | Significance | Statistic | Significance |
| | 4 | 17 | 0.94 | 0.348 | 0.956 | 0.556 | 0.928 | 0.201 | 0.874 | 0.025 | 0.956 | 0.56 |
| A | 7 | 16 | 0.88 | 0.039 | 0.959 | 0.64 | 0.875 | 0.032 | 0.918 | 0.159 | 0.977 | 0.934 |
| | 10 | 17 | 0.84 | 0.007 | 0.947 | 0.406 | 0.842 | 0.008 | 0.838 | 0.007 | 0.912 | 0.11 |
| | 2 | 17 | 0.913 | 0.113 | 0.98 | 0.96 | 0.896 | 0.057 | 0.949 | 0.445 | 0.92 | 0.147 |
| B | 6 | 16 | 0.907 | 0.103 | 0.701 | 0.001 | 0.825 | 0.006 | 0.726 | 0.001 | 0.812 | 0.004 |
| | 10 | 17 | 0.781 | 0.001 | 0.909 | 0.096 | 0.942 | 0.337 | 0.89 | 0.046 | 0.933 | 0.245 |
| | 2 | 17 | 0.861 | 0.016 | 0.957 | 0.571 | 0.834 | 0.006 | 0.945 | 0.376 | 0.884 | 0.036 |
| C | 7 | 16 | 0.894 | 0.064 | 0.772 | 0.001 | 0.835 | 0.008 | 0.753 | 0.001 | 0.85 | 0.013 |
| | 12 | 17 | 0.929 | 0.208 | 0.923 | 0.166 | 0.977 | 0.93 | 0.937 | 0.284 | 0.946 | 0.392 |
| | 25 | 17 | 0.839 | 0.007 | 0.972 | 0.851 | 0.91 | 0.1 | 0.96 | 0.627 | 0.903 | 0.077 |
| D | 40 | 16 | 0.922 | 0.18 | 0.802 | 0.003 | 0.857 | 0.017 | 0.735 | 0.001 | 0.868 | 0.025 |
| | 55 | 17 | 0.934 | 0.256 | 0.956 | 0.564 | 0.945 | 0.382 | 0.978 | 0.934 | 0.975 | 0.905 |
| | 50 | 17 | 0.827 | 0.005 | 0.984 | 0.986 | 0.859 | 0.015 | 0.958 | 0.593 | 0.914 | 0.119 |
| E | 150 | 16 | 0.929 | 0.233 | 0.806 | 0.003 | 0.831 | 0.007 | 0.743 | 0.001 | 0.875 | 0.033 |
| | 250 | 17 | 0.939 | 0.311 | 0.978 | 0.931 | 0.95 | 0.46 | 0.978 | 0.933 | 0.967 | 0.763 |

the feed phase pH; the reverse is true in case of extractant concentration. The Levene's test of equality of error variances for As(III), As(V) and combinations of As(III)-As(V) was chosen to check the homogeneity of variances of the dependent variable with respect to the independent variable (vide Table 5). The p-value for As(III)(0.998), As(V)(0.913), As(III):As(V)::1:1 (1), As(III):As(V)::1:2 (0.806) and As(III):As(V)::2:1 (0.993) were more than the significance level, showing the null hypothesis was true. This implied that the error variance of the dependent variable was equal across the groups of independent variable.

In the univariate analysis of variance, the significance column in Table 6 and Tables ST10-ST13 refers to the p-values for each tested difference of means pertaining to the independent and dependent variables. In this analysis, the p-values pertaining to the independent variables lying below the standard $\alpha$ of 0.05 indicates that their means differ significantly. In case of As(III), the means of all the five independent variables differ significantly. For As(V), the means of the combined factor of feed phase pH-extractant concentration along with the five independent variables differ significantly.

Similarly, for different ratios of As(III)- As(V) species, pH of the feed phase, extractant concentration, time and the combination of extractant concentration-time are the common factors with means that are varying significantly. Figure 3 and Fig. SF2-SF5 represent the variation of %extraction with various combinations of factors for As(III), As(V), As(III):As(V)::1:1, As(III):As(V)::1:2, and As(III):As(V)::2:1 using statistical model. The observations noted in Sec.4.3.1 of Part 1 of this series were found to be corroborated through these plots.

**Machine learned analysis.** Fifty data points from Table ST2 were used to train, validate and test the ANN. Initially, simulations were performed to optimize the number of neurons in the hidden layer. Figure 4a shows the MSE with the different number of neurons in the hidden layer. The minimum MSE was achieved with 6 neurons in the hidden layer. Thus, the same number of neurons was selected for further simulations. Figure 4b compares the model predictions with all data points with the experimental results. The regression curve showed that model predictions were sufficiently fitting with the model predictions on all data points, confirming the authenticity of the training.

**Table 4.** Spearman's Correlational Analysis for the Extraction of Single Arsenic Ions

| Type | Variables | Parameters | A | B | C | D | E | %Ex |
|---|---|---|---|---|---|---|---|---|
| As(III) | A | Correlation coefficient | 1 | 0 | 0 | 0 | 0 | -0.284 |
| | | Significance (two-tailed) | - | 1 | 1 | 1 | 1 | 0.046 |
| | B | Correlation coefficient | 0 | 1 | 0 | 0 | 0 | 0.477 |
| | | Significance (two-tailed) | 1 | - | 1 | 1 | 1 | 0 |
| | C | Correlation coefficient | 0 | 0 | 1 | 0 | 0 | 0.185 |
| | | Significance (two-tailed) | 1 | 1 | - | 1 | 1 | 0.199 |
| | D | Correlation coefficient | 0 | 0 | 0 | 1 | 0 | 0.139 |
| | | Significance (two-tailed) | 1 | 1 | 1 | - | 1 | 0.334 |
| | E | Correlation coefficient | 0 | 0 | 0 | 0 | 1 | 0.16 |
| | | Significance (two-tailed) | 1 | 1 | 1 | 1 | - | 0.268 |
| | %Ex | Correlation coefficient | -0.284 | 0.477 | 0.185 | 0.139 | 0.16 | 1 |
| | | Significance (two-tailed) | 0.046 | 0 | 0.199 | 0.334 | 0.268 | - |
| As(V) | A | Correlation coefficient | 1 | 0 | 0 | 0 | 0 | -0.39 |
| | | Significance (two-tailed) | - | 1 | 1 | 1 | 1 | 0.005 |
| | B | Correlation coefficient | 0 | 1 | 0 | 0 | 0 | 0.366 |
| | | Significance (two-tailed) | 1 | - | 1 | 1 | 1 | 0.009 |
| | C | Correlation coefficient | 0 | 0 | 1 | 0 | 0 | 0.199 |
| | | Significance (two-tailed) | 1 | 1 | - | 1 | 1 | 0.165 |
| | D | Correlation coefficient | 0 | 0 | 0 | 1 | 0 | 0.105 |
| | | Significance (two-tailed) | 1 | 1 | 1 | - | 1 | 0.468 |
| | E | Correlation coefficient | 0 | 0 | 0 | 0 | 1 | 0.06 |
| | | Significance (two-tailed) | 1 | 1 | 1 | 1 | - | 0.681 |
| | %Ex | Correlation coefficient | -0.39 | 0.366 | 0.199 | 0.105 | 0.06 | 1 |
| | | Significance (two-tailed) | 0.005 | 0.009 | 0.165 | 0.468 | 0.681 | - |

**Table 5.** Levene's Test of Equality of Error Variances in Two Phase

| Type of arsenic species | F | df1 | df2 | Sig. |
|---|---|---|---|---|
| As(III) | 0.251 | 42 | 7 | 0.998 |
| As(V) | 0.515 | 42 | 7 | 0.913 |
| As(III):As(V)::1:1 | 0.207 | 42 | 7 | 1 |
| As(III):As(V)::1:2 | 0.667 | 42 | 7 | 0.806 |
| As(III):As(V)::2:1 | 0.301 | 42 | 7 | 0.993 |

**Table 6.** Univariate Analysis of Variance for Extraction of As(III)

| Source | Type III Sum of Squares | df | Mean square | F | Significance |
|---|---|---|---|---|---|
| Corrected model | 7693.223 | 42 | 183.172 | 9.19 | 0.003 |
| Intercept | 65896.978 | 1 | 65896.978 | 3306.237 | 0 |
| A | 797.301 | 2 | 398.651 | 20.001 | 0.001 |
| B | 878.036 | 2 | 439.018 | 22.027 | 0.001 |
| C | 427.371 | 2 | 213.685 | 10.721 | 0.007 |
| D | 430.544 | 2 | 215.272 | 10.801 | 0.007 |
| E | 467.984 | 2 | 233.992 | 11.74 | 0.006 |
| AB | 102.209 | 1 | 102.209 | 5.128 | 0.058 |
| AC | 85.249 | 1 | 85.249 | 4.277 | 0.077 |
| AD | 3.913 | 1 | 3.913 | 0.196 | 0.671 |
| AE | 12.713 | 1 | 12.713 | 0.638 | 0.451 |
| BC | 111.49 | 1 | 111.49 | 5.594 | 0.05 |
| BD | 1.607 | 1 | 1.607 | 0.081 | 0.785 |
| BE | 4.613 | 1 | 4.613 | 0.231 | 0.645 |
| CD | 0.139 | 1 | 0.139 | 0.007 | 0.936 |
| CE | 0.372 | 1 | 0.372 | 0.019 | 0.895 |
| DE | 0.97 | 1 | 0.97 | 0.049 | 0.832 |
| ABC | 71.85 | 1 | 71.85 | 3.605 | 0.099 |
| ABD | 10.204 | 1 | 10.204 | 0.512 | 0.497 |
| ABE | 15.694 | 1 | 15.694 | 0.787 | 0.404 |
| ACD | 0.279 | 1 | 0.279 | 0.014 | 0.909 |
| ACE | 0.25 | 1 | 0.25 | 0.013 | 0.914 |
| ADE | 3.007 | 1 | 3.007 | 0.151 | 0.709 |
| BCD | 1.292 | 1 | 1.292 | 0.065 | 0.806 |
| BCE | 0.144 | 1 | 0.144 | 0.007 | 0.935 |
| BDE | 8.395 | 1 | 8.395 | 0.421 | 0.537 |
| CDE | 0.76 | 1 | 0.76 | 0.038 | 0.851 |
| ABCD | 2.36 | 1 | 2.36 | 0.118 | 0.741 |
| ABCE | 0.008 | 1 | 0.008 | 0 | 0.984 |
| ABDE | 0.272 | 1 | 0.272 | 0.014 | 0.91 |
| ACDE | 1.151 | 1 | 1.151 | 0.058 | 0.817 |
| BCDE | 0.059 | 1 | 0.059 | 0.003 | 0.958 |
| ABCDE | 0.102 | 1 | 0.102 | 0.005 | 0.945 |
| Error | 139.518 | 7 | 19.931 | | |
| Total | 174372.259 | 50 | | | |
| Corrected total | 7832.741 | 49 | | | |

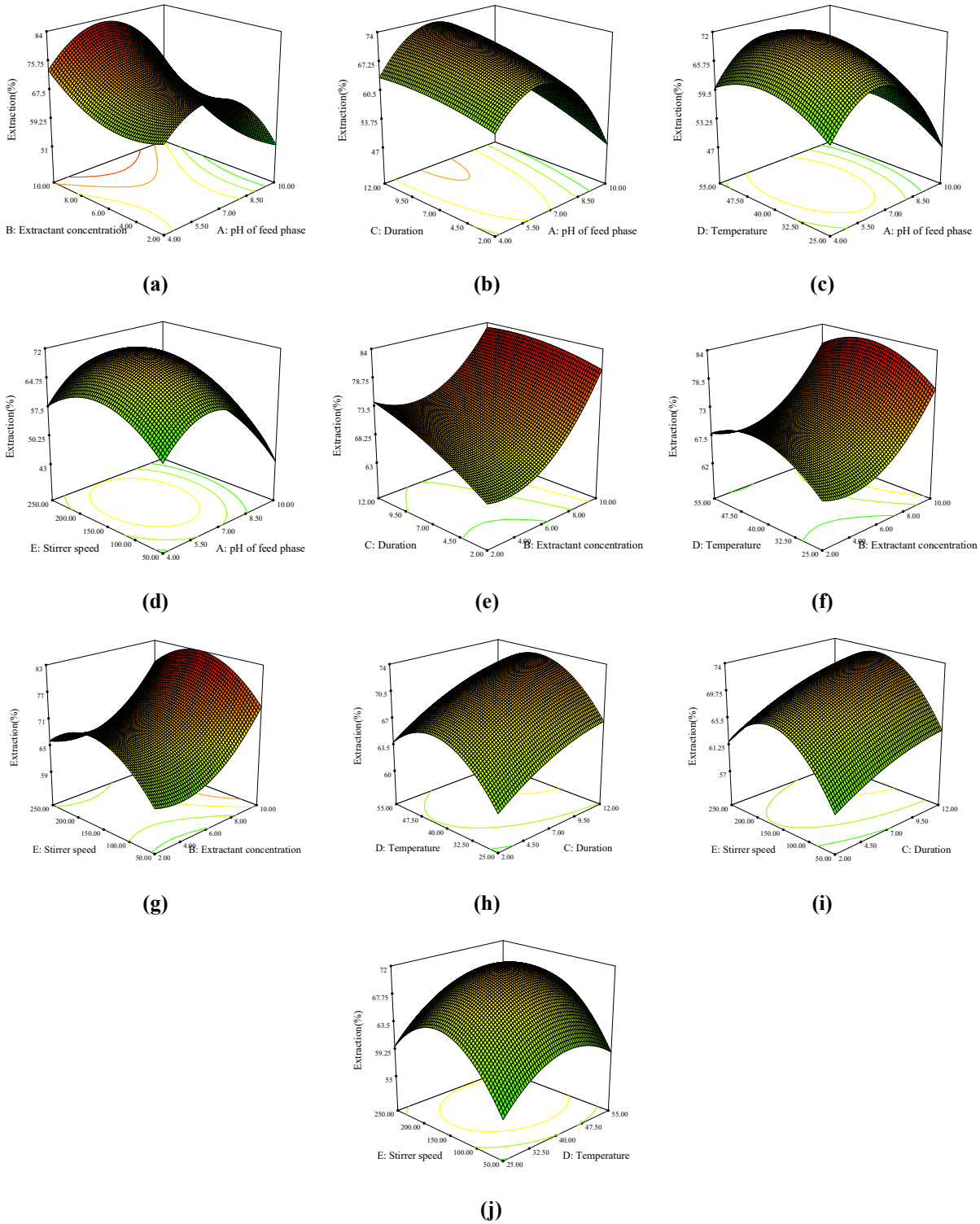



**Fig. 3.** Variation of %extraction in two phase with various combinations of factors for As(III) using statistical model.

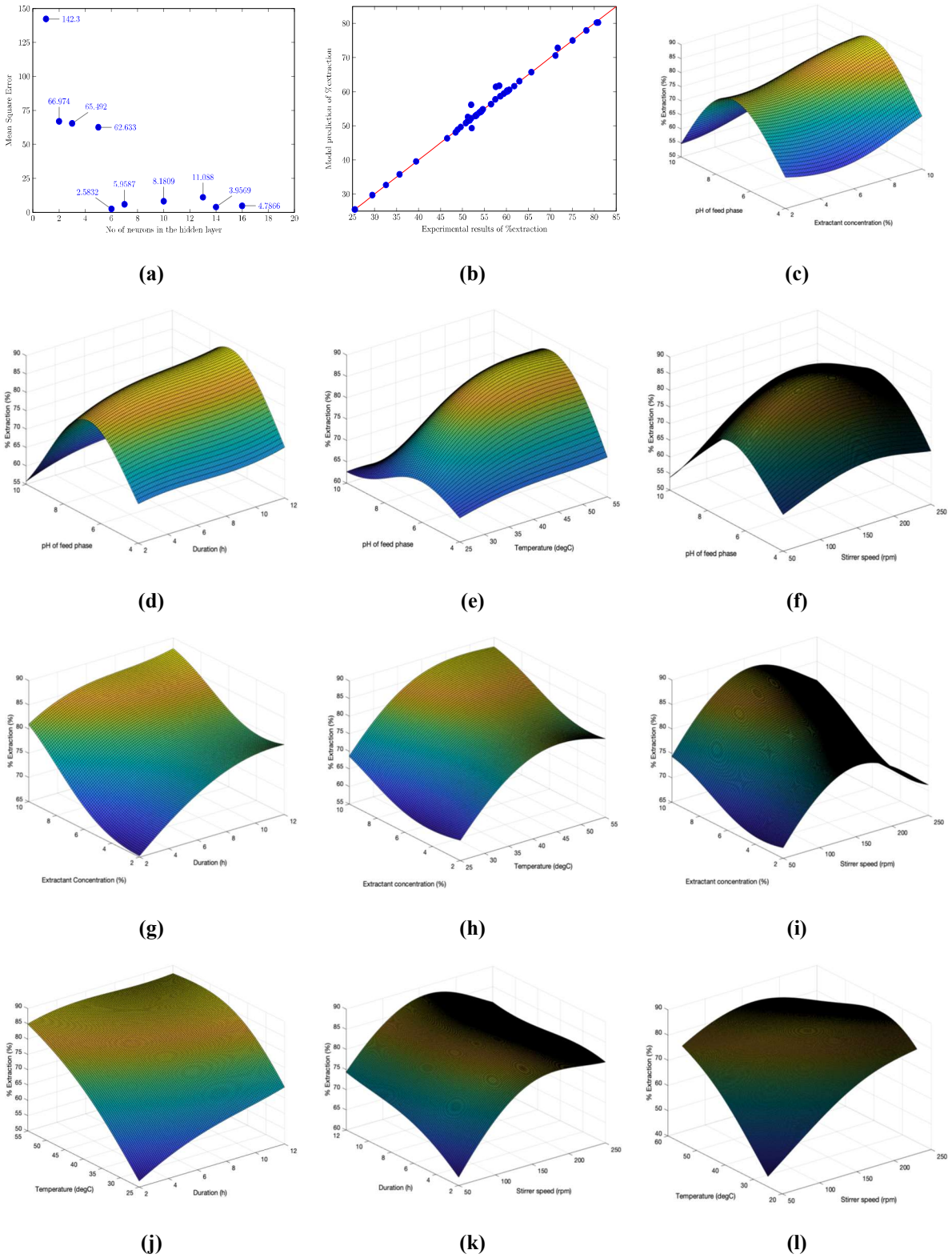



**Fig. 4.** Various plots of two phase extraction for As(III) using machine learning.

The trained ANN was employed to investigate how different operational conditions and membrane parameters impact the transport of arsenic ions. Global optimization of the ANN model was also done to detect the best operating condition to achieve maximum %extraction. The results of the said best operating condition are revealed later. In these simulations, the effects of variations of two input parameters were analysed simultaneously through surface graphs, while values of other input parameters were kept constant at the best condition. The range of input parameters was kept the same, as reported in Table ST1.

Figure 4c and Fig. 4l present variation of %extraction with various combinations of factors. It was revealed from Figs. 4c-4f that neither high nor low pH was suitable for good extraction. At pH 6.7-6.8, the extraction of the arsenic ions into the organic phase was highly favoured. The interaction between aliquat and arsenic ions was maximum by forming arsenic-aliquat complexes; favouring the extraction of arsenic ions from the aqueous phase into the organic phase. The presence of $H^+$ ions at lower pH range and $OH^-$ ions at higher pH range possibly interfered with the complex formation of arsenic-aliquat, leading to lower extraction of arsenic ions at extreme pH conditions [6]. The same results were obtained for stirring speed. The variation in stirring speed was observed over a range of 50 to 250 rpm and the optimum stirring speed for both the arsenic species was 170 rpm. The stirring speed varied from 143-196 rpm for different ratios of As(III) and As(V). The minimum thickness of the diffusion layer was obtained at the optimum stirring rate [7]. At lower stirring speed, the accumulation of arsenic-aliquat complex formed a boundary layer at the interphases, leading to lower extraction efficiencies. At higher stirring speed, the extraction decreased, owing to turbulences that lead to less interfacial contact time for the complex formation [26]. Thus, optimum stirring speed provided effective time for interaction. However, the %extraction increased with an increase in extractant concentration, duration and temperature. It is further clear from Fig. 4g, Fig. 4h and Fig. 4i that the higher concentration of extractant always favours the %extraction. Aliquat 336 is a highly viscous cationic surfactant forming structural micelle aggregates owing to its long hydrocarbon chain. The bulk amount of organic phase utilized in this two-phase study further increases the viscosity. Maximum extractant concentration of 10% is optimum for arsenic extraction. Beyond this, the viscosity of the liquid membrane increased, which in turn impacted the diffusion of the arsenic ions through the organic phase. Figure 4j and Fig. 4k show the relationship of duration with temperature and stirring speed, respectively. The surf figures reiterate the earlier revealed facts too. Figure 4l showed that a combination of high temperature and high stirring speed and/or a combination of low temperature and low stirring speed was not good for effective %extraction. The optimum duration for As(III) was 9 h to obtain an extraction of 84%. This is while the optimum duration for As(V) was 12 h for an extraction of 86%. This difference of three hours could be attributed to the presence of uncharged and charged ions in case of As(III) and As(V), respectively. We examined temperature fluctuations across a spectrum spanning from 25 °C to 55 °C. The optimum temperature for As(III) was 37 ℃ for an extraction of 84% and 55 ℃ for 86% extraction of As(V). This could possibly be due to the presence of $H_3AsO_3$ in case of As(III) that interacted by disrupting the strong ionic interactions between aliquat 336 and sesame oil at 37℃. Because $H_2AsO_4^-$ is an anion, it requires higher temperature of 55℃ to interact with the ionic components of the extractant and diluent. Higher temperature enhances the free energy, leading to a higher diffusion through less viscous liquid membrane. Moreover the arsenic extraction reactions are endothermic in nature within this temperature range [27]. Thus, intermediate temperature range lying between 35 ℃-47 ℃ was found to be optimum for the combined arsenic species, indicating synergistic interactions between the arsenic species.

Moreover, it was explained in Sec.3.1.1 that the transport of the arsenic ions across the membrane strongly depends on the operating parameters shown in Table ST1. Figure 4 also indicated that these parameters had coupled effects on the transport as the optimum value of a parameter, while maximum transport was found to vary with other parameters. Thus, global optimization is required to determine the optimum parameters where maximum transport can be achieved. In this study, a genetic algorithm was used for constrained global optimization. MATLAB Genetic Algorithm solver “ga” was linked with the trained ANN to determine the optimum values of the five operating parameters stated in Table ST1. In the genetic algorithm, the upper and lower limits of the different parameters were kept

the same as given in the experimental study, shown in Table ST1. Optimization was performed to maximize the %extraction of arsenic. After 124 generations and 5881 function counts, the GA reached the best value of %extraction of 87.82% at the optimum operating condition given in Table 2. The machine learning model predicted higher extraction percentage than the statistical model, however it used higher duration, higher temperature and a lower stirring speed.

In a similar manner, the datasets of As(V) and the three combinations of As(III)-As(V) were passed through machine learning algorithm. The results are given in Fig. SF8-SF11, and their efficiencies are reported in Table 2. In all cases, the minimum MSE was achieved with less than 8 neurons in the hidden layer. The variation of %extraction against the individual factors were also similar to that observed in case of As(III), though at a different degree of variability. It is interesting to note that the %extraction of As(III) was favoured by lowering the temperature, whereas the %extraction of As(V) was favoured at higher temperature. For combined salt cases, the trend was favoured as per the salt, which was at higher percentage in the combined species. Temperature had practically no effect on the %extraction when As(III) and As(V) were at equal proportion, i.e. As(III):As(V)::1:1. The reason behind this phenomenon lies with the fact that As(III) is present as an uncharged species in the particular pH range. Whereas, As(V) is present in its anionic form that forms inter- and/or intramolecular bonds, for which higher temperature is required for maximum extraction of a As(V). The %extraction of arsenic, predicted through machine learning algorithm, is comparable to that obtained through statistical model (vide Table 2). In case of combined species, the error% was mostly less in case of machine learned model.

## Three Phase SLM Studies

It should be noted that the extraction and recovery in three-phase SLM depends mainly on three parameters:

F: Concentration of the receiving phase
G: pH of the receiving phase
H1: Stirring speed (rpm)
H2: Extractant concentration (%)

Aqueous solution of ferric chloride was used as the receiving phase. The experiments were carried out as per the design of experiment where all the three factors were varied within a pre-designed range (vide Table ST1) in the same manner for all the 5 cases that are As(III), As(V), As(III):As(V)::1:1, As(III):As(V)::1:2, and As(III):As(V)::2:1. The corresponding %extraction and %recovery were recorded. This can increase the efficiency by minimizing the number of experiments for optimization. The results are reported in Table ST3 and Table ST4.

**Statistical analysis.** A quadratic model for prediction of extraction and recovery of arsenic species are given in the following equations:

$$Z^{E}_{As(III)} = 59.3 + 0.96A + 4.03B - 1.53C - 0.22AB - 0.24AC - 0.16BC - 3.62A^2 - 12.44B^2 - 4.45C^2 \quad (7)$$

$$Z^{R}_{As(III)} = 48.02 + 1.59A + 3.30B\ 1.84C + 0.99AB - 0.19AC + 0.62BC - 2.16A^2 - 10.79B^2 - 6.61C^2 \quad (8)$$

$$Z^{E}_{As(V)} = 0.75 + 4.5A - 0.93B - 1.68C - 0.9AB + 0.035AC + 0.24BC - 13.86A^2 - 7.38B2 - 2.51C^2 \quad (9)$$

$$Z^{R}_{As(V)} = 47.92 + 4.27A - 0.2B - 2.34C + 0.75AB + 0.5AC + 1.5BC - 15.39A^2 - 4.05B^2 - 5.20C^2 \quad (10)$$

Table ST14 presents ANOVA results of As(III) in three-phase SLM. The F-value of 23.14 indicate that the model for the As(III) was significant and there was only a 0.01% chance that this large value could occur due to noise. In the extraction and recovery of As(III) ions, pH of the receiving phase played a significant role as indicated by the model given in Table ST14. The F-value of 0.84 for lack of fit indicated that the value was not significant relative to the pure error and there was a 57.36% chance that a this large value could occur due to noise; this shows that this model can fit well. The predicted $R^2$ of 0.87 is in reasonable agreement with the adjusted $R^2$ of 0.91 (vide Table 7). The adequate precision measured the signal to noise ratio and the value reported in Table 7 is 12.802 (>4 desirable), indicating an adequate signal. Fig. SF12 shows the effect of interaction of parameters on the extraction and recovery of As(III). Furthermore, the F-value of 15.41 for recovery of As(III) ions

showed that the model was significant with 0.01% chance that this large value could occur due to noise as given in Table ST14. The lack of fit value was 0.30 with 89.67% chance that this large value could not occur due to noise, indicating that this model can be fitted well. In addition, the $R^2$ value was 0.93 with the predicted $R^2$ value in reasonable agreement with the adjusted $R^2$ value, and the adequate precision value was 10.53, as shown in Table 7.

In Fig. SF13, the effect of interaction of parameters on the extraction of As(V) was observed. In case of As(V), the F-value of 38.85 (vide Table ST15) with a 0.01% chance implies that the model was significant and that this value couldn't occur due to noise. In this case, the concentration and pH of the receiving phase are suggested to be significant model terms for extraction and recovery of As(V) ions. The lack of fit value was 2.12 that was insignificant relative to the pure error, and there was a 21.51% chance that this value could occur due to noise, as shown in Table ST15. Moreover, the $R^2$ value was 0.97 (vide Table 7) with the predicted $R^2$ value in reasonable agreement with the adjusted $R^2$ value, and the adequate precision was 16.14, as given in Table 7. On the other hand, for the recovery of As(V), the F-value of 41.99 with 0.01% chance implied that the model was significant. The lack of fit value was 1.61 with 30.77% chance, indicating that the value was not significant relative to the pure error and this value couldn't occur due to noise as shown in Table ST15. The $R^2$ value was 0.97 with the predicted $R^2$ value in reasonable agreement with the adjusted $R^2$ value, and the adequate precision value was 17.3, as shown in Table 7. A normal distribution of the residuals were observed from the plots for the extraction and recovery of As(III) and As(V), as given in Fig. SF17.

The minimum, maximum, lower quartile, median and upper quartile with outliers are given in the box plots for descriptive statistical analysis (vide Fig. SF12 and Fig. SF13). As the mean was close to the median, the distribution of the data was symmetric with skewness & 0.5. The high variance and standard deviation indicates that the points of the data set were spread out and not close to the mean. This is while in certain data sets, median was more towards the first (lower) quartile or third (upper) quartile, implying an asymmetrical frequency distribution.

The box plots for concentration of receiving phase helped to visualize the differences in distribution between 1-3 ppm concentration range for extraction and recovery of arsenic species, shown in Fig. SF12 and Fig. SF13. The median was higher for 2 ppm in both the cases. The length of the box indicated the variation of the data, showing positive

**Table 7.** Optimization and Error Analysis for Three-phase SLM on Individual Arsenic Ions

| Model parameters | As(III) | | | | As(V) | | | |
|---|---|---|---|---|---|---|---|---|
| | Statistics | | ML | | Statistics | | ML | |
| | Extraction | Recovery | Extraction | Recovery | Extraction | Recovery | Extraction | Recovery |
| Standard deviation | 2.97 | 3.46 | - | - | 2.64 | 2.64 | - | - |
| Mean | 49.05 | 38.23 | - | - | 48.88 | 35.6 | - | - |
| R2 | 0.95 | 0.93 | - | ... | 0.97 | 0.97 | - | - |
| Adjusted R2 | 0.91 | 0.87 | - | ... | 0.95 | 0.95 | - | - |
| Predicted R2 | 0.87 | 0.85 | - | ... | 0.9 | 0.89 | - | - |
| Adequate precision | 12.8 | 10.53 | - | - | 16.14 | 17.3 | - | - |
| Concentration of receiving phase (ppm) | 2 | | 4 | | 2 | | 2.13 | |
| pH of receiving phase | 5 | | 5.18 | | 5 | | 7 | |
| Stirrer speed (rpm) | 250 | | 200 | | 250 | | 250.72 | |
| Extractant concentration % (v/v) | 10 | | 10 | | 30 | | 30 | |
| Predicted (%) | 61 | 49 | 59.1 | 48.5 | 67 | 48 | 68.5 | 42 |
| Observed (%) | 60 | 48 | 60 | 48 | 66.7 | 47 | 66.7 | 47 |
| Error (%) | 1.64 | 2.04 | 1.52 | 1.03 | 0.45 | 2.1 | 2.63 | 11.9 |

skewness. This further justifies that 2 ppm led to the maximum extraction and recovery for both the arsenic ions. Similarly, the data for the other two independent variables were fairly distributed with pH 5, resulting in a maximum extraction and recovery at 250 rpm. This is in agreement with the optimum result obtained from the design of experiment (vide Table 7).

Shapiro-Wilk test was preferred over the Kolmogorov-Smirnov test for assessing the normality of the data as it is more appropriate for small sample size (<50). The significance values of the Shapiro-Wilk test (vide Table ST19) were more than 0.05, indicating that the data was normally distributed. As the data set was normal, Pearson's correlational analysis was carried out to identify the extent to which two variables were linearly related to each other. From the given Table ST21, it is evident that the dependent variables were linearly related to each other. The increase in extraction percentage of both the arsenic ions in turn increased the recovery percentage or vice versa at $\alpha = 0.01$ for 2-tailed significance analysis. In case of As(V), there was a significant correlation between the concentration of the receiving phase and the extraction percentage at $\alpha = 0.05$ for 2-tailed significance analysis. The multivariate test tabulated the four tests of significance for each model effect as given in Table ST23 and Table ST24. Pillai's trace was more robust than the other statistical tests based on model assumptions [28]. On the basis of the Pillai's trace test shown in Table ST23, the significance values for all effects were greater than 0.1 except for concentration of receiving phase (p-value = 0.085) for As(III). This indicated that the receiving phase concentration had a significant main effect in the multivariate test. The tests of between-subjects effects was carried out, as reported in Table ST25. The combined independent variables, concentration and pH of the receiving phase had a significant effect on the extraction percentage of As(III). For As(V), all the three independent variables and the combined effect of stirring speed and pH of receiving phase were found to have a significant impact on the dependent variables, that was obtained from the multivariate test reported in Table ST24. On this basis, the tests of between-subjects effects showed significant impact of the three individual independent variables on both the dependent variables for As(V), as given in Table ST26.

Since there were significant main effects obtained for the variables, post hoc (Tukey HSD) analyses were calculated to explain the effect. The significant levels of the independent variables were obtained from this analysis, as given in Table ST30 and Table ST31.

A quadratic model for prediction for extraction and recovery of mixed arsenic species are given in the following equations:

$$Z^{E}_{\mathrm{As(III):As(V)::1:1}} = 53.81 + 10.28A + 1.98B + 5.39C - 0.63AB - 1.28AC - 0.58BC - 3.43A^2 + 0.69B^2 - 0.66C^2 \quad (11)$$

$$Z^{R}_{\mathrm{As(III):As(V)::1:1}} = 33.65 + 9.95A + 1.42B + 6.07C - 0.3AB - 1.25AC + 0.19BC - 1.85A^2 - 0.82B^2 - 0.3C^2 \quad (12)$$

$$Z^{E}_{\mathrm{As(III):As(V)::1:2}} = 55.8 + 10.21A + 2.07B + 5.23C - 0.5AB - 1.4AC - 0.45BC - 3.5A^2 + 0.61B^2 - 0.6C^2 \quad (13)$$

$$Z^{R}_{\mathrm{As(III):As(V)::1:2}} = 35.68 + 9.97A + 1.41B + 6.04C - 0.28AB - 1.2AC + 0.23BC - 1.72A^2 - 0.82B^2 - 0.47C^2 \quad (14)$$

$$Z^{E}_{\mathrm{As(III):As(V)::2:1}} = 48.91 + 10.24A + 2B + 5.36C - 0.64AB - 1.24AC - 0.61BC - 3.23A^2 + 0.57B^2 - 0.83C^2 \quad (15)$$

$$Z^{R}_{\mathrm{As(III):As(V)::2:1}} = 28.63 + 10.02A + 1.35B + 6.01C - 0.2AB - 1.15AC + 0.13BC - 1.87A^2 - 0.82B^2 - 0.32C^2$$

The statistical analysis of the extraction and recovery of mixed arsenic species can include similar statistical tools such as ANOVA, normality tests, and Pearson's analysis; their explanations can also be similar in nature. Thus, only salient features are discussed here, in order to avoid repetitive statements. The following tables and figures show various statistical analysis data for mixed arsenic species As(III):As(V)::1:1, As(III):As(V)::1:2 and As(III):As(V):: 2:1.

| | |
|---|---|
| The ANOVA analysis | Table ST16, Table ST17, and Table ST18 |
| Normality test using Shapiro-Wilk model | Table ST20 |
| Pearson's correlational analysis | Table ST22 |
| Tests of between-subjects effects | Table ST27, Table ST28, and Table ST29 |
| Post hoc (Tukey HSD) analyses | Table ST32, Table ST33, and Table ST34 |
| Levene's test of homogeneity of error variances | Table ST35 |
| Normal plot of residuals | Fig. SF17 |
| Effect of interaction of parameters and box plot for statistical analysis | Fig. SF14, Fig. SF15, and Fig. SF16 |

The concentration of the receiving phase solution and extractant concentration were the significant terms of the model for both extraction and recovery of all the combined species of arsenic, as observed in Table ST16, Table ST17, and Table ST18. The $R^2$ values and the predicted $R^2$ values were in reasonable agreement with the adjusted $R^2$ value and the adequate precisions were between 35-40, as given in Table 8. The normal plot of residuals indicated the normal distribution of the data, as shown in Fig. SF17. The distribution of the data was symmetric with skewness ≤ 0.5, while in certain data sets, median was more towards the first (lower) quartile or third (upper) quartile, implying an asymmetrical frequency distribution. Box plots, Fig. SF14, Fig. SF15, and Fig. SF16, indicated that concentration of receiving phase at 2 ppm and 3 ppm yielded maximum extraction and recovery for all the arsenic ions. The data for the other two independent variables were fairly distributed with pH 5 and pH 7, yielding maximum extraction and recovery at 250 rpm. Shapiro-Wilk test, shown in Table ST20, indicated that the data is normally distributed. From the Pearson's correlational analysis (vide Table ST22), it is evident that the dependent variables were linearly related to each other. As this indicates that the receiving phase concentration had a significant main effect in the multivariate test, the tests of between-subjects effects was carried out, as reported in Table ST27, Table ST28, and Table ST29. Since there were significant main effects obtained for the variables, post hoc (Tukey HSD) analyses were calculated to explain the effect. The significant levels of the independent variables were obtained from this analysis, as given in Table ST32, Table ST33, and Table ST34.

**Machine learned analysis.** Fifteen data sets, both from Table ST3 and Table ST4 were used to train, validate and test the ANN. The entire operations were similar to that explained in Sec.3.1.2. Thus, only salient features are discussed here, in order to avoid the repeating statements. The following tables and figures show various machine learned analysis for mixed arsenic species As(III):As(V)::1:1, As(III):As(V)::1:2 and As(III):As(V)::2:1.

| | |
|---|---|
| Mean Squared Error (MSE) with different numbers of neurons | Fig. SF18a, Fig. SF19a, Fig. SF20a, Fig. SF21a, and Fig. 5a |
| Comparative plots of experimental v/s model prediction of %extraction/ %recovery | Fig. SF18b, Fig. SF19b, Fig. SF20b,Fig. SF21b, and Fig. 5b |
| Variation of %extraction/%recovery in three phase SLM with various combinations of factors | Fig. SF18, Fig. SF19, Fig. SF20, Fig. SF21, and Fig. 5 |

Initially, simulations were performed to optimize the number of neurons in the hidden layer. The minimum MSE was achieved with 4, 5, 27, 14, and 21 neurons in the hidden layer for As(III), As(V), As(III):As(V)::1:1, As(III):As(V):: 1:2 and As(III):As(V)::2:1. respectively. Global optimization of the ANN model yielded the best operating condition to achieve maximum extraction and recovery, shown in Table 8. The range of input parameters were maintained the same as reported in Table ST1.

The machine learning results on the data of As(III) and As(V) were not up to the mark. There were considerable process/model mismatch. It could be due to the less number of neurons that needed to arrive at the MSE. And the value of MSE were high for As(III) and As(V) in comparison to that of mixed arsenic species. Thus, the optimization was not successful too. On the other hand, the machine learning results on the data of As(III):As(V)::1:1, As(III):As(V)::1: 2 and As(III):As(V)::2:1 were better than even the statistical model. The optimization results were almost perfect.

**Table 8.** Optimization and Error Analysis for Three-phase SLM on Combined Arsenic Ions

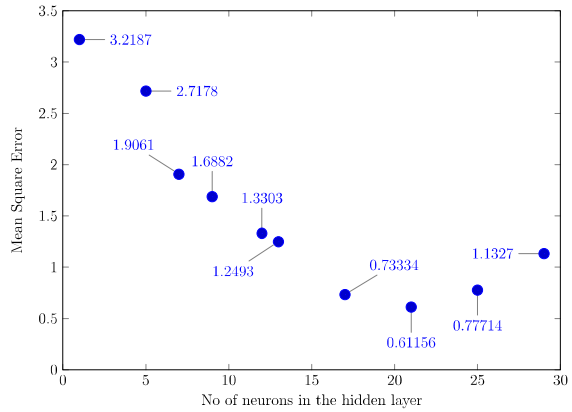
| Model parameters | As(III):As(V)::1:1 | | | | As(III):As(V)::1:2 | | | | As(III):As(V)::2:1 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Statistics | | ML | | Statistics | | ML | | Statistics | | ML | |
| | Ext | Rec | Ext | Rec | Ext | Rec | Ext | Rec | Ext | Rec | Ext | Rec |
| Standard deviation | 1.26 | 1.24 | - | - | 1.43 | 1.24 | - | - | 1.23 | 1.25 | - | - |
| Mean | 52.11 | 32.16 | - | - | 54.1 | 34.17 | - | - | 47.17 | 27.12 | - | - |
| $R^2$ | 0.989 | 0.989 | - | - | 0.986 | 0.989 | - | - | 0.989 | 0.989 | - | - |
| Adjusted $R^2$ | 0.98 | 0.98 | - | - | 0.97 | 0.98 | - | - | 0.98 | 0.98 | - | - |
| Predicted $R^2$ | 0.94 | 0.94 | - | - | 0.92 | 0.94 | - | - | 0.94 | 0.94 | - | - |
| Adequate precision | 39.6 | 39.9 | - | - | 34.7 | 39.6 | - | - | 40.5 | 39.4 | - | - |
| Concentration of receiving phase (ppm) | 3 | | 3 | | 3 | | 3 | | 3 | | 3 | |
| pH of receiving phase | 7 | | 7 | | 7 | | 7 | | 7 | | 7 | |
| Stirrer speed (rpm) | 250 | | 250 | | 250 | | 250 | | 250 | | 250 | |
| Extractant concentration % (v/v) | 40 | | 38.1 | | 40 | | 36 | | 40 | | 40 | |
| Predicted (%) | 66 | 47 | 65.7 | 46.6 | 67.5 | 49 | 67.7 | 48.7 | 60.5 | 42 | 61.2 | 41.6 |
| Observed (%) | 65 | 46.5 | 65 | 46.5 | 66.5 | 48.5 | 66.5 | 48.5 | 59 | 41 | 59 | 41 |
| Error (%) | 1.5 | 1.1 | 1.07 | 0.21 | 1.48 | 1.02 | 1.77 | 0.41 | 2.47 | 2.4 | 3.6 | 1.44 |

Interestingly, concentration and pH of receiving phase were the same for all three cases at 3 and 7, respectively. The differences in extractant concentration were very minimum. The predicted values of extraction and recovery were very close to the experimental values with less than 1.5% error in most cases.

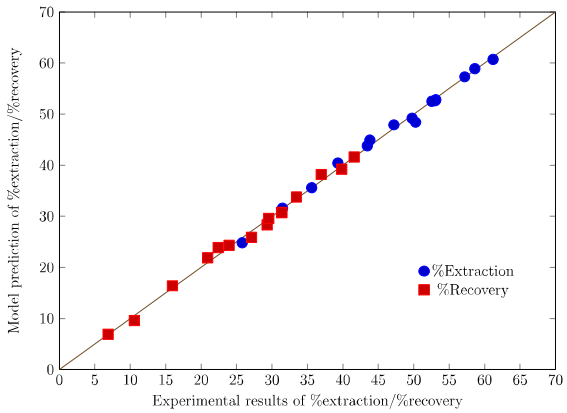
## CONCLUSION

A quadratic significant model with non-significant lack of fit was obtained for each of the arsenic species in both two-phase and three phase studies through the face-centered response surface methodology of central composite design of experiments. In two phase, the predicted maximum extraction of As(III) and As(V) were 82.37% and 84.15%, as opposed to 84% and 86% respectively obtained through experimentations. On the other hand, the predicted maximum extraction of combined As(III):As(V)::1:1, As(III):As(V)::1:2 and As(III):As(V)::2:1 were 87.03%, 87.59% and 85.1%, as opposed to 85.5%, 86% and 84.5% respectively obtained through experimentations in the two-phase study. The feed phase pH and extractant concentration were found to be the common significant parameters in most of the cases. In three-phase study, the optimum conditions that led to the maximum extraction and recovery for both the arsenic ions were found to be 2 ppm concentration of receiving phase and pH 5 at 250 rpm. This is while the optimum conditions for the combined arsenic ions are 3 ppm concentration of receiving phase and pH 7 at 250 rpm. Here, the optimum extractant concentration was different; 10% for As(III), 30% for As(V) and 40% for all the three different combination of arsenic species. The minimum, maximum, lower quartile, median and upper quartile with outliers obtained from the descriptive statistical analysis with the interquartile range helped to understand the distributions of the different levels of independent variables with respect to the dependent variable. This was followed by the normality test in which the Shapiro-Wilk test was preferred over the Kolmogorov- Smirnov test due to the sample size.

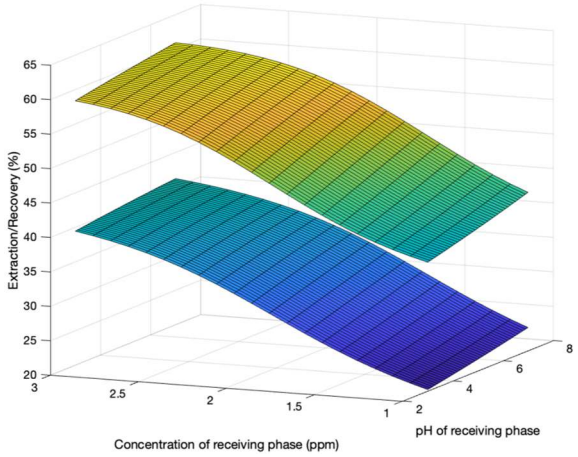
In case of two-phase study, Spearmann's correlation analysis was carried out based on the Shapiro-Wilk test and the skewed distribution of the data. For the three-phase SLM study, Pearson's correlational analysis was assessed, as the data was normally distributed and symmetric. The correlational coefficients in two phase study suggested that the extraction of arsenic had a significant negative correlation with pH of the feed phase and a positive correlation with the extractant concentration. However, in case of SLM, the



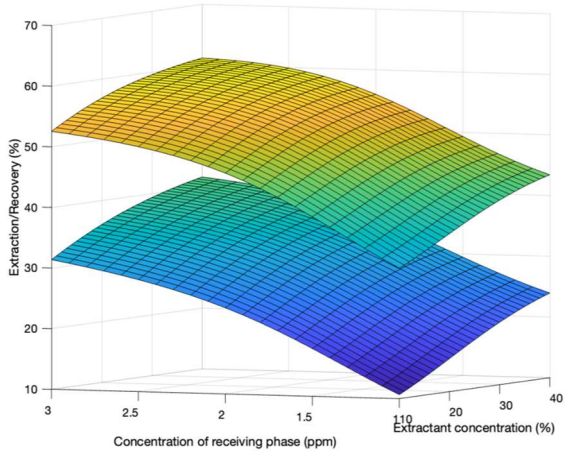

**(a)**



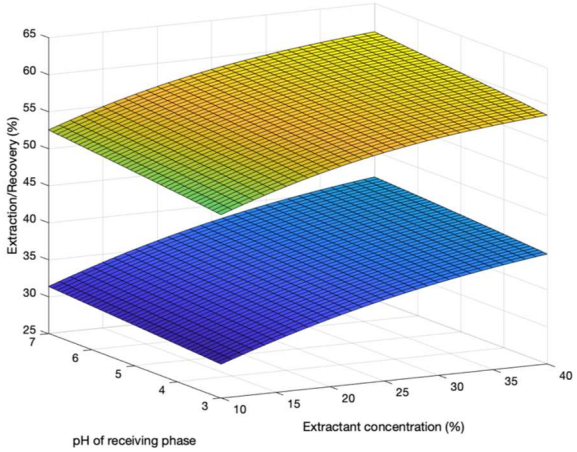

**(b)**




**(c)**




**(d)**




**(e)**

**Fig. 5.** Various plots of three phase SLM for As(III):As(V)::2:1 using machine learning.

dependent variables showed a linear relation. An increase in extraction percentage of both the arsenic ions in turn caused an increase in the recovery percentage. Five way ANOVA for the two-phase systems and multivariate ANOVA for the three-phase systems tabulated the four tests of significance for each model effect, and generated the tests of between-subjects effects. Pillai's trace was more robust than the other statistical tests based on the model assumptions. This indicated that the receiving phase concentration had a significant main effect in the multivariate test. Since there were significant main effects obtained for the variables, post hoc (Tukey HSD) analyses were calculated to explain the effect. All the data points from the two-phase study and SLM experimentations were used to train, validate and test the ANN. Initially, simulations were performed to optimize the number of neurons in the hidden layer. Global optimization of the ANN model yielded the best operating condition to achieve maximum extraction and recovery. The %extraction of arsenic in two-phase study predicted through machine learning algorithm, which was comparable to that obtained through statistical model. In case of combined species, the error% was mostly less in the case of machine learning model. In two-phase study, the minimum MSE was achieved with less than 8 neurons in the hidden layer for all the arsenic species. However, in three-phase study, the minimum MSE was achieved with 4, 5, 27, 14, and 21 neurons in the hidden layer for As(III), As(V), As(III):As(V)::1:1, As(III):As(V)::1:2 and As(III):As(V)::2:1, respectively. The machine learning results on the data of As(III) and As(V) were not up to the mark in the SLM study. There was considerable process/model mismatch. Thus, the optimization was not successful too. On the other hand, the machine learning results on the SLM data of As(III):As(V)::1:1, As(III):As(V)::1:2 and As(III):As(V)::2:1 were better than the statistical model. The predicted values of extraction and recovery were very close to the experimental values with less than 1.5% error in most cases.

## ACKNOWLEDGEMENTS

We thankfully acknowledge the financial support provided by the Department of Science and Technology (India) through sanction order no. DST/TM/WTI/2K13/125(G) for carrying out this research work.

## Supporting Information

The supporting information contains Tables and Figures whose numbers begin with the letter "S" and are available online and also through email communication to the author for correspondence.

## Glossary

| | |
|---|---|
| ANN: | Artificial neural network |
| A: | pH of the feed phase (factors/variables in two-phase studies) |
| B: | %Extractant concentration (vol/vol) (factors/variables in two-phase studies) |
| C: | Duration (hours) (factors/variables in two-phase studies) |
| D: | Temperature (°C) (factors/variables in two-phase studies) |
| E: | Stirring speed (rpm) (factors/variables |
| F: | in two-phase studies) |
| G: | Concentration of the receiving phase |
| H1: | (factors/variables in three-phase studies) |
| H2: | pH of the receiving phase |
| GA: | (factors/variables in three-phase studies) |
| MSE: | Stirring speed (rpm) (factors/variables |
| ML: | in three-phase studies) |
| Stats: | Extractant concentration (%) |
| SLM: | (factors/variables in three-phase studies) |
| | Genetic algorithm |
| $Y_{As(III)}$: | Mean squared error |
| | Machine learning |
| $Y_{As(V)}$: | Statistical |
| | Supported liquid membrane |
| $Y_{As(III):As(V)::1:1}$: | |
| $Y_{As(III):As(V)::1:2}$: | Extraction efficiency predicted from the quadratic model for As(III) in two-phase study |
| $Y_{As(III):As(V)::2:1}$: | Extraction efficiency predicted from the quadratic model for As(V) in two-phase study |
| $Z^{E}_{As(III)}$: | |
| $Z^{E}_{As(V)}$: | Extraction efficiency predicted from the |
| $Z^{E}_{\mathrm{As(III):As(V)::1:1}}$: | quadratic model for As(III):As(V)::1:1 in two-phase study |

$Z^{E}_{\mathrm{As(III):As(V)::1:2}}$: Extraction efficiency predicted from the quadratic model for As(III):As(V)::1:2 in two-phase study

$Z^{E}_{\mathrm{As(III):As(V)::2:1}}$: Extraction efficiency predicted from the quadratic model for As(III):As(V)::2:1 in two-phase study

$Z^{R}_{As(III)}$: Extraction efficiency predicted from the quadratic model for As(III) in SLM

$Z^{R}_{As(V)}$: Extraction efficiency predicted from the quadratic model for As(V) in SLM

$Z^{R}_{\mathrm{As(III):As(V)::1:1}}$: Extraction efficiency predicted from the quadratic model for As(III):As(V)::1:1 in SLM

$Z^{R}_{\mathrm{As(III):As(V)::1:2}}$: Extraction efficiency predicted from the quadratic model for As(III):As(V)::1:2 in SLM

$Z^{R}_{\mathrm{As(III):As(V)::2:1}}$: Extraction efficiency predicted from the quadratic model for As(III):As(V)::2:1 in SLM

Recovery efficiency predicted from the quadratic model for As(III) in SLM

Recovery efficiency predicted from the quadratic model for As(V) in SLM

Recovery efficiency predicted from the quadratic model for As(III):As(V)::1:1 in SLM

Recovery efficiency predicted from the quadratic model for As(III):As(V)::1:2 in SLM

Recovery efficiency predicted from the quadratic model for As(III):As(V)::2:1 in SLM

## REFERENCES


[1] Hughes, M. F.; Thomas, D. J.; Kenyon, E. M., *Toxicology and Epidemiology of Arsenic and its Compounds*. Arsenic: Environmental Chemistry, Health Threats and Waste Treatment, Henke J. (Ed), John Wiley & Sons **2009**, Chapter 4, 237-275, DOI: 10.1002/9780470741122.

[2] Ahuja, S., *Solutions for Arsenic Contamination of Groundwater*. In Arsenic Contamination of Groundwater, John Wiley & Sons **2008**, Chapter 15, 367-376, DOI: 10.1002/9780470371046.ch15.

[3] Shaji, E.; Santosh, M.; Sarath, K. V.; Prakash, P.; Deepchand, V.; Divya, B. V., Arsenic contamination of groundwater: A global synopsis with focus on the Indian Peninsula. *Geosci Frontiers* **2021**, *12* (3), 101079, DOI: 10.1016/j.gsf.2020.08.015.

[4] Smedley, P. L.; Kinniburgh, D. G., A review of the source, behaviour and distribution of arsenic in natural waters. *Appl. Geochem.* **2002**, *17* (5), 517-568, DOI: 10.1016/S0883-2927(02)00018-5.

[5] Sarkar, S.; Hazra, S.; Chakraborty, K.; Nayak, A.; Saha, P., Hybrid technique for removal of arsenic from drinking water. *Chem. Engg. Tech.* **2023**, *46* (2),242-255, DOI: 10.1002/ceat.202100603.

[6] Assis, R. C.; de Araujo Faria, B. A.; Caldeira, C. L.; Mageste, A. B.; de Lemos, L. R.; Rodrigues, G. D., Extraction of arsenic(III) in aqueous two-phase systems: A new methodology for determination and speciation analysis of inorganic arsenic. *MicroChem. J.* **2019**, *147*, 429-436, DOI: 10.1016/j.microc.2019.03.058.

[7] Perez, M. E.; Reyes-Aguilera, J. A.; Saucedo, T. I.; Gonzalez, M. P.; Navarro, R.; Avila-Rodriguez, M., Study of As(V) transfer through a supported liquid membrane impregnated with trioctylphosphine oxide (Cyanex 921). *J. Mem. Sci.* **2007**, *302* (1), 119-126, DOI: 10.1016/j.memsci.2007.06.037.

[8] Güell, R.; Fontàs, C.; Anticó, E.; Salvadó, V.; Crespo, J. G.; Velizarov, S., Transport and separation of arsenate and arsenite from aqueous media by supported liquid and anion-exchange membranes. *Sep. and Purif. Tech.* **2011**, *80* (3), 428-434, DOI: 10.1016/j.seppur.2011.05.015.

[9] Li, J.; Wang, J.; Yang, L., Kolmogorov-Smirnov simultaneous confidence bands for time series distribution function. *Comput. Stat.* **2022**, *37* (3), 1015-1039, DOI: 10.1007/s00180-021-01149-5.

[10] Susdarwono, E. T., Chemistry Learning Through CIRC on DNA Base Materials: Hypothesis testing the Kolmogorov-Smirnov method on Male and Female Population Samples. *Int. J. Edu. and Teaching Zone* **2022**, *1* (2), 78-86, DOI: 10.57092/ijetz.v1i2.32.

[11] González-Estrada, E.; Villaseñor, J. A.; Acosta-Pech, R., Shapiro-Wilk test for multivariate skew-normality.

*Comput. Stat.* **2022**, *37*, 1-17, DOI: 10.1007/s00180-021-01188-y.

[12] Meijer, L. L.; Hasenack, B.; Kamps, J. C. C.; Mahon, A.; Titone, G.; Dijkerman, H. C.; Keizer, A., Affective touch perception and longing for touch during the COVID-19 pandemic. *Sci. Rep.* **2022**, *12* (1), 1-9, DOI: 10.1038/s41598-022-07213-4.

[13] Chen, K.; Chen, L.; Xiao, J.; Li, J.; Hu, Y.; Wen, K., Speckle reduction in digital holography with non-local means filter based on the Pearson correlation coefficient and Butterworth filter. *Optics Let.* **2022**, *47* (2), 397-400, DOI: 10.1364/OL.444769.

[14] Andrianou, X. D.; van der Lek, C.; Charisiadis, P.; Ioannou, S.; Fotopoulou, K. N.; Papapanagiotou, Z.; Botsaris, G.; Beumer, C.; Makris, K. C., Application of the urban exposome framework using drinking water and quality of life indicators: a proof-of-concept study in Limassol, Cyprus. *Peer J.* **2019**, 7, e6851, DOI: 10.7717/peerj.6851.

[15] Zhang, Q.; Hu, J.; Bai, Z., Modified Pillai's trace statistics for two high-dimensional sample covariance matrices. *J. Stat. Plann. Infer.* **2020**, *207*, 255-275, DOI: 10.1016/j.jspi.2020.01.002.

[16] El Ouardighi, A.; El Akadi, A.; Aboutajdine, D., Feature selection on supervised classification using Wilks lambda statistic. *IEEE International Symposium on Computational Intelligence and Intelligent Informatics*, **2007**, 51-55, DOI: 10.1109/ISCIII.2007.367361.

[17] Akbari, V.; Anfinsen, S. N.; Doulgeris, A. P.; Eltoft, T.; Moser, G.; Serpico, S. B., Polarimetric SAR change detection with the complex Hotelling--Lawley trace statistic. *IEEE Trans. on Geosci. Rem. Sen.* **2016**, *54* (7), 3953-3966, DOI: 10.1109/TGRS.2016.2532320.

[18] Davis, A. W., On the effects of moderate multivariate nonnormality on Roy's largest root test. *J. Am. Stat. Assoc.* **1982**, *77* (380), 896-900, DOI: 10.1080/01621459.1982.10477904.

[19] Zhou, Z. H., *Machine learning*. Springer Nature: **2021**, DOI: 10.1007/978-981-15-1967-3.

[20] Hunt, E. B., *Artificial intelligence*. Academic Press: **2014**, DOI: 10.1016/C2013-0-10882-3.

[21] Yegnanarayana, B., Artificial neural networks for pattern recognition. *Sadhana* **1994**, *19*, 189-238, DOI: 10.1007/BF02811896.

[22] Min, D.; Song, Z.; Chen, H.; Wang, T.; Zhang, T., Genetic algorithm optimized neural network based fuel cell hybrid electric vehicle energy management strategy under start-stop condition. *Appl. Energy* **2022**, *306*, 118036, DOI: 10.1016/j.apenergy.2021.118036.

[23] Khan, S., Ethem Alpaydin. Introduction to Machine Learning (Adaptive Computation and Machine Learning Series. *Nat. Lang. Eng.* **2008**, *14* (1), 133-137, DOI: 10.1017/S1351324906004438.

[24] Sharma, R.; Jain, M., Variance based sensitivity analysis and statistical optimization of design and operating parameters of spiral wound pervaporation modules for thiophene removal from FCC gasoline. *Comp. Chem. Eng.* **2020**, *141*, 106987, DOI: 10.1016/j.compchemeng.2020.106987.

[25] Hammerstrom, D., Working with neural networks. *IEEE Spec.* **1993**, *30* (7), 46-53, DOI: 10.1109/6.222230.

[26] Yang, X.; Duan, H.; Shi, D.; Yang, R.; Wang, S.; Guo, H., Facilitated transport of phenol through supported liquid membrane containing bis (2-ethylhexyl) sulfoxide (BESO) as the carrier. *Chem. Eng. Proc.: Proc. Inten.* **2015**, *93*, 79-86, DOI: 10.1016/j.cep.2015.05.003.

[27] Ali, H.; Ahmed, S.; Hsini, A.; Kizito, S.; Naciri, Y.; Djellabi, R.; Abid, M.; Raza, W.; Hassan, N.; Rehman, M. S. U.; Khan, A. J., Efficiency of a novel nitrogen-doped $Fe_3O_4$ impregnated biochar (N/$Fe_3O_4$@ BC) for arsenic (III and V) removal from aqueous solution: Insight into mechanistic understanding and reusability potential. *Arabian J. Chem.*, **2022**, *15* (11), 104209, DOI: 10.1016/j.arabjc.2022.104209.

[28] Olson, C. L., Comparative robustness of six tests in multivariate analysis of variance. *J. Am. Stat. Assoc.* **1974**, *69* (348), 894-908, DOI: 10.1080/01621459.1974.10480224.