Predictive modelling of arsenate (As(V)) adsorption onto surface-engineered magnetite nanoparticles [version 1; peer review: 1 approved]

Nisha Kumari Devaraj, Ameer Al Mubarak Hamzah

Faculty of Engineering, Multimedia University, Cyberjaya, Selangor, 63100, Malaysia

Abstract

Background: Since adsorption is a complex process, numerous models and theories have been devised to gain general understanding of its underlying mechanisms. The interaction between the adsorbates and adsorbents can be identified via modelling of the adsorption data with different adsorption isotherms as well as kinetic models. Many studies are also focused on developing predictive modelling techniques to facilitate accurate prediction of future adsorption trends. Methods: In this study, a predictive model was developed based on a multiple linear regression technique using existing data of As(V) adsorption onto several coated and uncoated magnetite samples. To understand the mechanisms and interactions involved, the data was first modelled using either Temkin or Freundlich linear isotherms. The predicted value is a single data point extension from the training data set. Subsequently, the predicted outcome and the experimental values were compared using multiple error functions to assess the predictive model's performance. Results: In addition, certain values were compared to that obtained from the literature, and the results were found to have low error margins. Conclusion: To further gauge the effectiveness of the proposed model in accurately predicting future adsorption trends, it should be further tested on different adsorbent and adsorbate combinations.

Keywords
Arsenate adsorption, adsorption predictive technique, surface-engineered magnetite nanoparticles, linear regression technique, error function
**Introduction**

Heavy metal ions are toxic and dangerous towards living organisms depending on the concentration and duration of exposure. Among the most toxic heavy metals are chromium (Cr), nickel (Ni), copper (Cu), zinc (Zn), cadmium (Cd), lead (Pb), mercury (Hg) and arsenic (As). The metalloid As itself has been considered as the highest priority pollutant to be controlled in China since 2009 together with four other heavy metal ions. Heavy metals come from two major sources, namely natural and anthropogenic activities. Natural sources include volcanic-related occurrences, soil erosion, eroded rocks and minerals whilst anthropogenic sources involve landfills, sewage, agricultural activities, industrial pollutants, as well as mining. The increased discharge of heavy metals into the environment may be attributed to global industrialisation and increased urbanisation processes. In addition, the metalloid As is a ubiquitous element which exists naturally in the environment, hence resulting in enhanced exposure that could prove to be fatal if not controlled. A recent spate of toxic waste dumping incidences into rivers in Peninsular Malaysia necessitates the development of more effective water remediation methods.

Adsorption is one of the numerous solutions that has gained popularity lately due to its effectiveness and low cost, especially adsorption using nanoparticles (NPs) as adsorbents. Extensive research is being carried out to improve the adsorption capacity of the wide range of NPs available. However, the comprehension of the adsorption mechanism by NPs is still on-going, thus isotherm and kinetic equations are continuously being formulated. Having the ability to predict certain outcomes especially regarding the behaviour of particles involved in the process of adsorption or desorption would be beneficial for the effective design of experiments. An accurate prediction will bring forth a great breakthrough not only for the understanding of the underlying mechanisms but for future experiments as well.

In this project, existing arsenate (As(V)) adsorption data onto uncoated, humic acid-coated and ceria-coated magnetite samples were initially modelled. As for the predictive model, multiple linear regression was implemented to predict the equilibrium adsorption capacity, \( q_e \), of each sample. The predicted results were compared to the experimental \( q_e \) results and the model’s accuracy was thus ascertained.

**Literature review**

In general, the first step involved in predicting an adsorption process is to emulate the process itself. An example is the study by Mandal et al., in which a hybrid material was synthesised to run an actual As adsorption experiment. The optimum parameters such as adsorbent dose, pH and initial concentration were varied for the predictive model’s design.

A predictive model consists of a set of data framework which is correlated to the experimental variables. In the same study by Mandal et al. the evolutionary Genetic Programming (GP) and Least Squares Support Vector Machine (LS-SVM) models were utilised. The optimal physical parameters were inserted into these models for computation. The result was a model that could predict the adsorption process with a certain degree of confidence in emulating the actual adsorption process results.

Another study by Onur et al. implemented Quantitative Structure-Activity Relationship (QSAR) and Linear Solvation Energy Relationship (LSER) techniques to develop a predictive model for aromatic contaminants adsorption onto multi-walled carbon nanotubes. In the study, 29 aromatic adsorption datasets obtained from literature and experiments were used in developing the predictive model. A few variables such as the molecular connectivity indices and solvatochromic descriptors were included for both the QSAR and LSER models, thus increasing the accuracy of predicting future trends. Besides that, Sisi et al. predicted the adsorption of As(V) onto iron-coated sand, which was based on a model known as surface complexation. The input parameters were the experimental maximum adsorption capacities as well as results from an extended X-ray absorption fine structure spectroscopy (EXAFS) as a function of pH and adsorbent dose. The goal was to derive a model that could predict and determine the adsorption performance over a pH range of 5–8. This indicates that the predictive model can be implemented to determine the other factors of adsorption as well. Meanwhile, a study by Fukushi et al. implemented the extended triple layer model (ETLM), which is also a form of the surface complexation model. The model was used to predict As(V) adsorption onto oxides with the aid of in situ spectroscopic evidence as well as molecular theories and calculation. Based on the various studies reviewed, it can be surmised that there are multiple techniques that could be applied in predicting the outcomes of an adsorption process. The understanding of the adsorption process and as well as the availability of data parameters are the deciding factors of the suitable predictive model(s) to be selected and implemented.

**Methods**

The flow of executing the predictive model from providing input to acquiring the output is shown in Figure 1. After determining the best model that fits the experimental data obtained from our previous experiments involving the different types of magnetite samples, that specific model’s equation was used to provide the linearity in determining the equilibrium adsorption capacity, \( q_e \). Aside from the model’s parameters, the other values which acted as input are the raw data (the experimental \( q_e \) that was linearised by the model chosen and the value of physical parameters such as concentration, temperature and the adsorbent sizes.

**Input parameters**

The respective particles’ adsorption data: humic-acid-coated magnetite NPs (modelled with Temkin isotherm), uncoated magnetite and ceria-coated magnetite particles (both modelled with Freundlich) were chosen as inputs to the predictive model. The summary of the choice of the isotherm parameters such as \( C_e \) (equilibrium concentration of adsorption), \( b \) (Temkin constant), \( 1/n \) (intensity of adsorption) as well as \( K_f \) (Freundlich constant) and the physical parameters are as shown in Table 1. Y-axis parameter values, which are forms of \( q_e \)
(dependant on the chosen isotherm model), were predicted and compared to the experimental Y-axis parameter values.

### Multiple regression model
Linear predictive analysis was implemented using the software Python 3.0 (Python Programming Language, RRID: SCR_008394) to code the model’s framework. Below are some key points of the linear regression model:

- Sample size of four including the starting point.
- The first three rows of data were the training data set while the last one was the test set.
- No iterations involved.

### Validation of model performance
The performance of the developed predictive model was validated using multiple error functions: mean squared error (MSE), mean absolute percentage error (MAPE), average absolute relative error (AARE), non-linear chi-squared value ($\chi^2$) and root mean squared error (RMSE) as shown in Equation 1, Equation 2, Equation 3, Equation 4 and Equation 5 respectively. These values were also compared to other models’ results from the literature to benchmark our model’s performance. Besides that, Equation 6 and Equation 7 are the absolute error (AE) and percentage error (PE) calculation formulae.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (q_{\text{calc}} - q_{\text{meas}})^2$$

### Table 1. Summary of isotherm model and physical parameters used.

<table>
<thead>
<tr>
<th>NPs’ surface coating</th>
<th>Isotherm model</th>
<th>X-axis parameter</th>
<th>X-axis value</th>
<th>Isotherm parameters</th>
<th>Physical parameter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humic acid</td>
<td>Temkin</td>
<td>$q_e$</td>
<td>$\ln C_e$, $b$</td>
<td>Particle size, sonication time</td>
<td></td>
</tr>
<tr>
<td>Ceria</td>
<td>Freundlich</td>
<td>$\log q_e$, $\log C_e$, $1/n$, $K_F$</td>
<td>Adsorbent dosage, temperature, sonication time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No coating</td>
<td>Freundlich</td>
<td>$\log q_e$, $\log C_e$, $1/n$, $K_F$</td>
<td>Particle size, sonication time, temperature</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Figure 1. Framework of the predictive analysis.](image-url)
\[
MAPE = 100 \times \frac{1}{N} \sum_{i=1}^{N} \left| \frac{q_{\text{cal}} - q_{\text{meas}}}{q_{\text{meas}}} \right| \\
AARE = 100 \times \frac{1}{N} \sum_{i=1}^{N} \left| \frac{q_{\text{cal}} - q_{\text{meas}}}{q_{\text{meas}}} \right| \\
\chi^2 = \sum_{i=1}^{N} \left( \frac{q_{\text{cal}} - q_{\text{meas}}}{q_{\text{meas}}} \right)^2 \\
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( \frac{q_{\text{cal}} - q_{\text{meas}}}{q_{\text{meas}}} \right)^2} \\
AE = \frac{q_{\text{cal}} - q_{\text{meas}}}{q_{\text{meas}}} \\
PE = \frac{100 \times (q_{\text{cal}} - q_{\text{meas}})}{q_{\text{meas}}}
\]

Results
Table 2, Table 3 and Table 4 depict the predicted results which are based on the actual data sets provided as input. The overall absolute error is particularly low. However, this may be because linearity is the main component for the model framework, where almost zero error is possible so as long as the provided input gives good data fit as well. In short, the data simply stretches out linearly and can predict values for every additional 20 minutes beyond the modelled data’s duration. In addition, the linearity extends indefinitely, which neglects the fact that there is an adsorption equilibrium capacity. Further analysis based on the results from the Table 2–Table 4 is tabulated and summarised in Table 5.

From Table 2 it can be observed that the lowest absolute error of 0.000805 was obtained for the 2238 nm particle, while the largest error, 0.087876 was for the 3203 nm particle. Also observed is that at a constant humic acid concentration, the predicted value increased with increasing particle sizes and surpasses the actual value.

Based on the tabulated results in Table 3, the lowest absolute error was for the 998.3 nm particle under a sonication temperature of 70°C, while the largest was for the 782 nm particle at

Table 2. The predictive value analysis of equilibrium As(V) adsorption by humic-acid coated magnetite.

<table>
<thead>
<tr>
<th>Particle size (nm)</th>
<th>Humic-acid concentration (mmol)</th>
<th>Actual Value ((q_e))</th>
<th>Predicted Value ((q_e))</th>
<th>Absolute Error</th>
<th>Percentage Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2238</td>
<td>0</td>
<td>1.78703</td>
<td>1.787835</td>
<td>0.000805</td>
<td>0.045017</td>
</tr>
<tr>
<td>2931</td>
<td>4.4</td>
<td>1.791259</td>
<td>1.795618</td>
<td>0.004359</td>
<td>0.243351</td>
</tr>
<tr>
<td>2400</td>
<td>6.6</td>
<td>1.811466</td>
<td>1.759903</td>
<td>0.051563</td>
<td>2.846493</td>
</tr>
<tr>
<td>3203</td>
<td>2.2</td>
<td>1.651316</td>
<td>1.56344</td>
<td>0.087876</td>
<td>5.321543</td>
</tr>
<tr>
<td>3450</td>
<td>2.2</td>
<td>1.638816</td>
<td>1.659572</td>
<td>0.020756</td>
<td>1.266545</td>
</tr>
<tr>
<td>3712</td>
<td>2.2</td>
<td>1.678853</td>
<td>1.75345</td>
<td>0.074597</td>
<td>4.443300</td>
</tr>
</tbody>
</table>

Table 3. The predictive value analysis of equilibrium As(V) adsorption by ceria-coated magnetite.

<table>
<thead>
<tr>
<th>Particle size (nm)</th>
<th>Sonication temperature (°C)</th>
<th>Actual Value ((\log q_e))</th>
<th>Predicted Value ((\log q_e))</th>
<th>Absolute Error</th>
<th>Percentage Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>442</td>
<td>30</td>
<td>2.165309</td>
<td>2.166761</td>
<td>0.001452</td>
<td>0.067057</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>2.262581</td>
<td>2.262747</td>
<td>0.000166</td>
<td>0.007337</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>2.294981</td>
<td>2.301577</td>
<td>0.005317</td>
<td>0.227410</td>
</tr>
<tr>
<td>782</td>
<td>30</td>
<td>2.298799</td>
<td>2.388090</td>
<td>0.089291</td>
<td>3.884246</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>2.252853</td>
<td>2.252866</td>
<td>0.000013</td>
<td>0.000577</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>2.281237</td>
<td>2.275920</td>
<td>0.005317</td>
<td>0.223075</td>
</tr>
<tr>
<td>992.4</td>
<td>30</td>
<td>2.265698</td>
<td>2.266304</td>
<td>0.000606</td>
<td>0.026747</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>2.321024</td>
<td>2.322746</td>
<td>0.001722</td>
<td>0.074191</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>2.304195</td>
<td>2.300033</td>
<td>0.004162</td>
<td>0.180627</td>
</tr>
<tr>
<td>998.3</td>
<td>30</td>
<td>2.266283</td>
<td>2.266467</td>
<td>0.000184</td>
<td>0.008119</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>2.331221</td>
<td>2.335541</td>
<td>0.004320</td>
<td>0.185311</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>2.289469</td>
<td>2.289465</td>
<td>0.000004</td>
<td>0.000175</td>
</tr>
</tbody>
</table>
Table 4. Predictive value analysis of equilibrium As(V) adsorption by uncoated magnetite.

<table>
<thead>
<tr>
<th>Particle size (nm)</th>
<th>Sonication temperature (°C)</th>
<th>Actual value ((\log q_e))</th>
<th>Predicted Value ((\log q_e))</th>
<th>Absolute Error</th>
<th>Percentage Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>445</td>
<td>30</td>
<td>2.264800</td>
<td>2.265266</td>
<td>0.000466</td>
<td>0.020576</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>2.331700</td>
<td>2.331611</td>
<td>0.000089</td>
<td>0.003817</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>2.292100</td>
<td>2.291395</td>
<td>0.000705</td>
<td>0.030758</td>
</tr>
<tr>
<td>617</td>
<td>30</td>
<td>2.262200</td>
<td>2.262155</td>
<td>0.000045</td>
<td>0.001989</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>2.315800</td>
<td>2.316210</td>
<td>0.000410</td>
<td>0.017704</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>2.267300</td>
<td>2.264782</td>
<td>0.002518</td>
<td>0.111057</td>
</tr>
<tr>
<td>822</td>
<td>30</td>
<td>2.355700</td>
<td>2.324627</td>
<td>0.031073</td>
<td>1.319056</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>2.262200</td>
<td>2.262363</td>
<td>0.000163</td>
<td>0.007205</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>2.292100</td>
<td>2.292517</td>
<td>0.000417</td>
<td>0.018193</td>
</tr>
<tr>
<td>1259</td>
<td>30</td>
<td>2.291700</td>
<td>2.258091</td>
<td>0.033609</td>
<td>1.465553</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>2.256000</td>
<td>2.273768</td>
<td>0.017768</td>
<td>0.787589</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>2.276100</td>
<td>2.319241</td>
<td>0.043141</td>
<td>1.895391</td>
</tr>
</tbody>
</table>

Table 5. Summary of errors between the actual and predicted values.

<table>
<thead>
<tr>
<th>Adsorbent</th>
<th>Absolute error</th>
<th>Percentage error (%)</th>
<th>Smallest</th>
<th>Highest</th>
<th>Smallest</th>
<th>Highest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humic-acid coated</td>
<td>0.000805</td>
<td>0.087876</td>
<td>0.045017</td>
<td>5.321543</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ceria-Coated</td>
<td>0.000004</td>
<td>0.089291</td>
<td>0.000175</td>
<td>3.884246</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncoated</td>
<td>0.000045</td>
<td>0.043141</td>
<td>0.001989</td>
<td>1.895391</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

30 °C. At 50°C, the absolute errors for the 442 nm and 782 nm particles are lower compared to the other temperatures. On the other hand, the absolute error was the highest at the same temperature for the 998.3 nm particle. In addition, at 70°C, as the particle size increased, the absolute error values decreased. From Table 4, the lowest and the highest values of absolute error are observed for the 617 nm particle at 30°C and 1259 nm particle at 70°C, respectively.

The validated results are shown in Table 6. The value of RMSE obtained from various predictive models in the literature to compare against our predictive model’s accuracy are as shown in Table 7.

Discussion
Compared to other studies\(^5\)\(^-\)\(^13\) shown in Table 7, our predictive modelling has lower error percentages, which may imply a higher prediction accuracy. However, this was only because our model is based solely on linearity, given very few input data sets with no iterations and no complex theories derived. The isotherm model’s linear regression was simply recycled and transformed to a multiple linear regression by adding in a few X-axis parameters (including the chosen isotherms model’s parameters) to predict the \(q_e\) value. Compared to other methodologies employed as outlined in Table 7,\(^5\)\(^-\)\(^13\) their results were simulated and obtained from a more complex framework of codes, not to mention having many data sets as well as multiple repetitions of trial and error. Nonetheless, the framework of the predictive technique developed here could act as a foundation in the formulation of a more complex model with more input datasets to improve its reliability.

To explain further, since the linear isotherm model chosen for each sample provided the best fit to the adsorption data, the assumptions and theories of the model would be applicable for our samples as well. Hence, the existing model’s equation can be used as a basis to predict future adsorption trends under different experimental variables once the preliminary adsorption data has been modelled. This is crucial especially as the linearity of the modelled data is the input to the predictive model.
model analysis. Hence, this would reduce the need to derive or formulate new complex theories to emulate the adsorption process. The simplicity and straightforward nature of the predictive model developed herein makes it attractive for further development and testing with the adsorption data of numerous types of samples under different experimental parameters for a more effective design of experiment.

**Conclusions**

Many predictive modelling techniques were researched to emulate the actual adsorption process. In this project, a simple and straightforward multiple linear regression method was implemented. Satisfactory results were obtained with small error values. Despite lacking in theoretical derivations, choosing an existing model may prove to be sufficient to predict the $q_e$ values as long as sufficient experimental data are available, and the best-fit model is decided beforehand.

**Data availability**

**Underlying data**

Open Science Framework: Underlying data for ‘Predictive modelling of As(V) adsorption onto surface-engineered magnetite nanoparticles’, [https://doi.org/10.17605/OSF.IO/5ZMKR](https://doi.org/10.17605/OSF.IO/5ZMKR)

This project contains the following underlying data:
- Data file 1: CM1.2.csv
- Data file 2: CM234.csv
- Data file 3: ErrorCalc_ceria magnetite.csv
- Data file 4: ErrorCalc_Humic acid.csv
- Data file 5: ErrorCalc_Magnetite.csv
- Data file 6: HA1_6.csv
- Data file 7: M1234.csv
- Data file 8: Magnetite As(V) adsorption data.pdf

Data are available under the terms of the CC0 1.0 Universal.

**Software availability**


Archived source code at time of publication:
License: Apache-2.0 License.
References


Open Peer Review

Claudia Maria Simonescu
Department of Analytical Chemistry and Environmental Engineering, Faculty of Applied Chemistry and Materials Science, Politehnica University of Bucharest, Bucharest, Romania

This interesting manuscript presents new data regarding development of a predictive model based on a multiple linear regression technique using existing data of As(V) adsorption onto several coated and uncoated magnetite samples. Temkin and Freundlich linear isotherms were used to model the experimental results. The predictive model's performance was assessed by comparing of predicted outcome and the experimental values using multiple error functions.

This work has enough content to be published in the journal, is well-argued, well-written and well organized information. The organization of the article is satisfactory. The paper's title is brief and reflects the main theme of the paper.

- The abstract is sufficiently informative. It is completely self-explanatory, briefly presents the topic, states the scope of the experiments, indicates significant data, and points out major findings and conclusions.

- The keywords are suitable so the article can be found in the current registers or indexes.

- The introduction shows how this work builds on previous work on the subject. Introduction clearly states the problem being investigated. The purpose and objectives of the manuscript are adequate and appropriate in view of the subject matter. The authors highlighted the novelty of the paper.

- The authors accurately explain how the model was developed.

- The input parameters, the key points of the linear regression model and multiple error functions determined are specified. There is sufficient information to allow replication of the research.

- In the Results and Discussion part, the authors present and interpret the results of the
modelling performed. This section is well organized.

- The authors used the unit measurements according to the International Standards.

- The figure and tables are numbered sequentially, and they are clearly labeled and positioned close to the relevant text. Titles of figure and tables are brief and informative.

- In the Conclusions part the authors mentioned the major and specific conclusions of their research study. All the conclusions are justified and supported by the results obtained.

- The abbreviations and nomenclature are used according to applicable international standards and rules.

- The length of the manuscript is adequate.

- The authors compared their results with other research findings.

- Significant findings regarding implementation of one simple and straightforward multiple linear regression method to describe As(V) adsorption onto several coated and uncoated magnetite samples are provided by this research study. In my opinion, the work provides an advance towards the current knowledge.

Consequently, I recommend the publication of this manuscript without modifications.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Environmental engineering and protection

I confirm that I have read this submission and believe that I have an appropriate level of
expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com