

Modelling the Combined Effect of Salt Precipitation and Fines Migration on CO₂ Injectivity Changes in Sandstone Formation

M Nabil B Ziaudin Ahamed ¹⁾, M Azfar bin Mohamed ^{1*)}, M Aslam B Md Yusof ¹⁾, Iqmal Irshad ²⁾, Nur Asyraf Bt Md Akhir ¹⁾, Noorzamzarina binti Sulaiman ³⁾

¹⁾ Petroleum Engineering, Universiti Teknologi PETRONAS

²⁾ Data Science, Petroliaam Nasional Berhad (PETRONAS)

³⁾ Geoscience, Universiti Malaysia Kelantan

* corresponding e-mail: azfar.mohamed@utp.edu.my

ABSTRACT

Carbon dioxide, CO₂ emissions have risen precipitously over the last century, wreaking havoc on the atmosphere. Carbon Capture and Sequestration (CCS) techniques are being used to inject as much CO₂ as possible and meet emission reduction targets with the fewest number of wells potential for economic reasons. However, CO₂ injectivity is being reduced in sandstone formations due to significant CO₂-brine-rock interactions in the form of salt precipitation and fines migration. The purpose of this project is to develop a regression model using linear regression and neural networks to correlate the combined effect of fines migration and salt precipitation on CO₂ injectivity as a function of injection flow rates, brine salinities, particle sizes, and particle concentrations. Statistical analysis demonstrates that the neural network model has a reliable fit of 0.9882 in R Square and could be used to accurately predict the permeability changes expected during CO₂ injection in sandstones.

Keywords: CO₂ injection; neural network; regression model

I. INTRODUCTION

Carbon Capture and Sequestration (CCS) has developed into a critical component of a portfolio of technologies to reduce emissions of the world's most prevalent greenhouse gas, Carbon Dioxide, CO₂. Untreated CO₂ emissions remain in the atmosphere, wreaking havoc on the climate and weather, commonly referred to as global climate change. By 2015, CO₂ emissions had contributed approximately 0.8 °C to transient global warming (Y. Xu and V. Ramanathan, 2017). CCS is based on the capture and permanent storage of CO₂ in deep underground geologic formations such as depleted oil fields. These formations are buried beneath layers of dense, impermeable rock or cap rock, which effectively prevents the injected CO₂ from escaping to the surface, just as it did for the oil and gas previously contained in the formation over thousands of years.

Adequate well injectivity and storage capacity are required to successfully inject large volumes of CO₂ and achieve emission reduction targets with the fewest possible wells. According to previous observations, sandstone formations exhibit a significant average loss of permeability of up to 35-55 percent following a typical CO₂ injection (I. M. Mohamed, J. He, and H. A. Nasr-El-Din, 2012). Numerous studies are being conducted to optimize the CO₂ injection process, with numerous mathematical models being developed to observe changes in the permeability of porous media (W. D. Carrier III, 2003 & A. Verma and K. Pruess, 1988). The primary factors limiting injectivity are determined to be the combined effect of salt precipitation and fines migration, which results in a severe impairment of the rock formation's permeability (Y. A. Sokama-Neuya et al., 2017).

II. LITERATURE REVIEW

The rate at which CO₂ can be injected into a formation without fracturing it can be expressed in terms of the injectivity index, I, which is defined as the ratio of volumetric injection flow rate, q, to pressure drop

$$I = \frac{q}{\Delta P} \quad (1)$$

This injectivity, however, will be influenced by the physical changes that occur in porous media during CO₂ injection. These changes in fluid injectivity are indicative of an injectivity impairment and can be expressed as Relative Injectivity Change,

$$RIC = \left(\frac{I_i - I_f}{I_i} \right) \quad (2)$$

Porosity and permeability changes can be induced by a variety of different mechanisms, including mechanical stress changes and biomass growth (J. Hommel, E. Coltman, and H. Class, 2018). However, we will focus on the factors contributing to salt precipitation and fines migration in this study.

Salt precipitation is a well-known phenomenon in the oil and gas industry and is widely regarded as a common factor reducing CO₂ injectivity for sequestration (Y. A. Sokama-Neuyam, J. R. Ursin, and P. Boakye, 2019). When CO₂ is injected into sandstone formations, the initially saturated water is removed from the formation via advection and vaporization due to the mass and heat transfer of the injected fluid. This decrease in water saturation causes solids to precipitate out of the brine, posing a significant impediment to fluid flow, particularly in the vicinity of the wellbore. This region of dryness will then extend further away from the injection point, spreading its effect. This would result in a significant reduction in permeability due to the blockage and jamming caused by the precipitated salt. However, the volume of water removed may result in a minor increase in permeability, which should not be overlooked (R. Miri and H. Hellevang, 2016).

Fines are also released as a result of the CO₂ - brine - rock interactions (Y. A. Sokama-Neuyam et al., 2017), which may exacerbate the injectivity impairment during continuous CO₂ injection. Carbonic acid is formed when CO₂ is injected into a porous medium saturated with an aqueous solution (3).



However, this dissolution results in the dislodgment of less reactive minerals such as clay and quartz, which aggregate into microscopic particles and are mobilized along the flowing stream, resulting in plugging of the reservoir rock's narrow pore channels and a reduction in permeability (F. Othman, M. Yu, F. Kamali, and F. Hussain, 2018). Normally, mineral dissolution is preferred, as is the case with matrix acidifying activities, but this subsequent mechanism of precipitation and fines migration during continuous CO₂ injection would reduce permeability.

This interaction of salt precipitation and fines migration results in a significant change in permeability during the CO₂ injection process. As a result, the total amount of CO₂ injected will be significantly reduced, which contradicts the primary objective of meeting specified CO₂ emission reduction targets. The primary factors that have a direct effect on this mechanism are the CO₂ injection parameters, which include the injection flow rate, brine salinities (F. Othman, M. A. Naufaliansyah, and F. Hussain, 2019) particle sizes, and concentrations of particles (M. A. Md Yusof et al, 2020). Thus, the purpose of this article is to develop a CO₂ injectivity regression model that describes the change in the permeability of porous media as a function of these four injection parameters. This model should be capable of predicting the expected RIC during the CO₂ injection process and can be an extremely useful tool in making future core flooding decisions.

III. METHODS

Predictive Modelling

Md Yusof and Arif Ibrahim (M. A. Md Yusof et al, 2020) measured and prepared 45 data points of Relative Injectivity Change (RIC) for this analysis based on permeability changes in two Berea sandstones that are sensitive to injection flow rate, brine salinities, particle sizes, and particle concentrations via a specific design of experiment. Following that, these data sets will be analyzed using linear regression and neural network models.

Linear regression is a method for modelling the relationship between a scalar response and one or more explanatory variables using a linear relationship. In this project, we used a linear regression model, which is one of the data analysis methods included in Microsoft Excel's Analysis ToolPak. Linear regression is a widely used analytical technique for analyzing financial, statistical, and engineering data.

By contrast, neural network models are forecasting techniques that are used in a wide variety of deep learning applications and forecasting methods. It performs admirably in mapping the response variable to its predictions, even when the

relationship is non-linear or complex. The supervised neural network model used in this project was created with the Python library's sci-kit-learn module and is hosted in Jupyter Notebook, an interactive open-source web application for statistical modelling, data visualization, and machine learning exercises.

Both regression models will be used to generate predictions of expected injectivity changes as a result of various injection parameter combinations. The differences between the experimental and predicted datasets will be the primary focus of this project's analysis.

Statistical Analysis

There are four main statistical parameters that are being considered in this study to help in evaluating the accuracy of the predicted relative injectivity change value from both regression models.

The main statistical analysis parameter for this study is the R-squared value. This serves as a good indicator of how well the data fits the model. As shown in Figure 1 below, a low R-squared value indicates that the model is inaccurate and inapplicable to the dataset (left graphic), whereas a high value converging to 1 will indicate that the model fits very well and is an accurate representation of the data (right graphic).

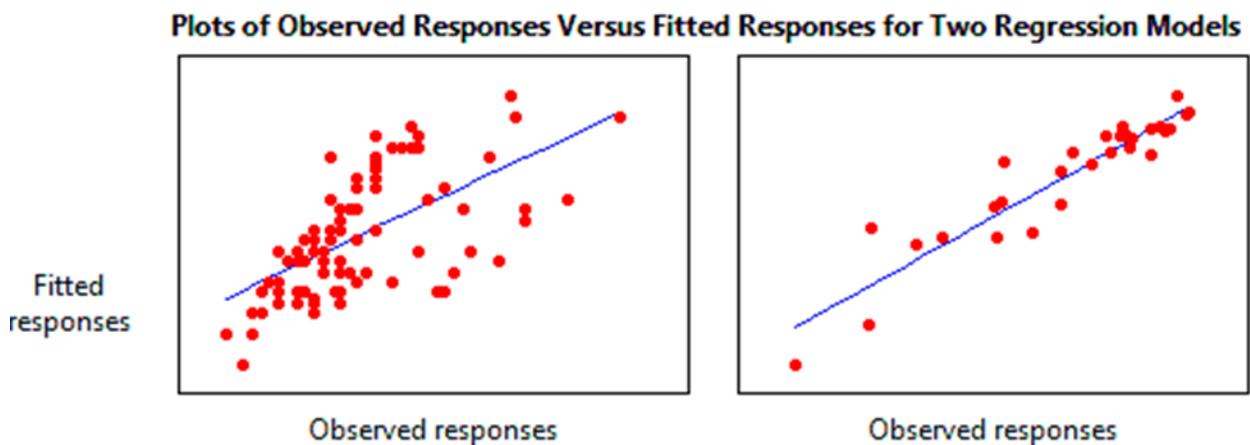


Figure 1. Graphical representation of R-squared analysis

(Retrieved from The Minitab Blog)

Additionally, the accuracy of mathematical correlations will be assessed using the Average Absolute Percentage Error (AAPE) as demonstrated in (2). This represents the difference between the actual value (A_t) and the forecast value numerically (F_t). A low AAPE value indicates that the predicted datasets are more accurate than the original datasets.

$$AAPE = \left| \frac{A_t - F_t}{A_t} \right| * 100\% \quad (4)$$

The Mean Absolute Error (MAE) is the average of the absolute values of the discrepancies between the predicted and observed values across the verification sample. The MAE is a linear score, which means that all individual variations are equally weighted in the aggregate.

On the other hand, the Root Mean Squared Error (RMSE) is a quadratic scoring rule that quantifies the average magnitude of an error. It squares the difference between the predicted and observed values and then averages them over the sample. Because the errors are squared prior to being averaged, the RMSE weights large errors heavily. As a result, the root mean square error is most useful when large errors are particularly undesirable. In general, both the MAE and RMSE indicate inaccuracy, and statistical analysis prefers lower values.

IV. RESULTS AND DISCUSSION

The purpose of this project is to develop a mathematical model that employs both linear regression and neural networks to establish a mathematical relationship between the measured and predicted datasets. This enables extensive forecasting of expected Relative Injectivity Change (RIC), for a variety of injection parameter combinations, assuming significant statistical parameters are obtained in the statistical error analysis.

Linear Regression Model

The data is separated using the train-test-split procedure prior to handing over the dataset in this predictive modelling problem. This is a necessary step in machine learning methods because the dataset is divided into two parts: the "train" dataset, which the model will use to generate an effective mapping of the parameters to the desired outputs, and the "test" dataset, which will be used to test and measure the accuracy and performance of the generated model from the "train" dataset. The percentage split is chosen to ensure that both the training and testing datasets are representative of the total data. Eighty percent of the 45 data points are classified as training data (36 data points), while the remaining twenty percent will be used for evaluation purposes (9 data points). When a model is able to predict the outputs from the varying parameters entered with a small deviation from the actual measured values of the output, it is said to be efficient.

The correlational analysis of the linear regression method's results is summarized in Table 1 below.

Table 1. Performance of the linear regression model

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.959050223
R Square	0.919777329
Adjusted R Square	0.911755062
Standard Error	6.752552223
Observations	45

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	20911.32154	5227.830385	114.6530429	2.36398E-21
Residual	40	1823.878461	45.59696152		
Total	44	22735.2			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	0.392225467	2.319940829	0.169067013	0.866596017	-4.296549849	5.081000782	-4.296549849	5.081000782
Brine Salinity (ppm)	0.000450726	2.93029E-05	15.38159623	2.16144E-18	0.000391503	0.00050995	0.000391503	0.00050995
Flow rate (ml/min)	1.428836649	0.347416211	4.112751807	0.000189047	0.726682295	2.130991003	0.726682295	2.130991003
Jamming ratio	510.8663606	74.09214949	6.895013359	2.62661E-08	361.1205406	660.6121805	361.1205406	660.6121805
Particle Concentrati	8.256967525	6.153118296	1.34191594	0.187191445	-4.178948436	20.69288349	-4.178948436	20.69288349

A respectably high value of 0.91978 is obtained, indicating a significant degree of positive correlation, given that it approaches unity. This does not, however, imply a perfect fit, as the relationship between the four injection parameters is not perfectly linear with the measured RIC. This encourages similar analyses to be conducted using a more sophisticated and sensitive regression model that is also more computationally efficient. Having said that, the linear regression model is still quite capable and can be used as a valid comparison and evaluation criterion in this project.

In Figure 2, the predicted data from the linear regression model is plotted against the measured data. Similarly, the same plot is generated for the training and testing datasets, respectively, from which a R Square value of 0.914 and an AAPE value of 0.198 for the training data and a R Square value of 0.944 and an AAPE value of 0.152 for the testing data are obtained (Figure 3 and 4). This analysis demonstrates that the model is representative of both the training and testing datasets and justifies the original data splitting ratio of 80:20. The significant fitting observed in the test datasets also bolsters the generated linear model's accuracy and validity, as well as its predictive ability.

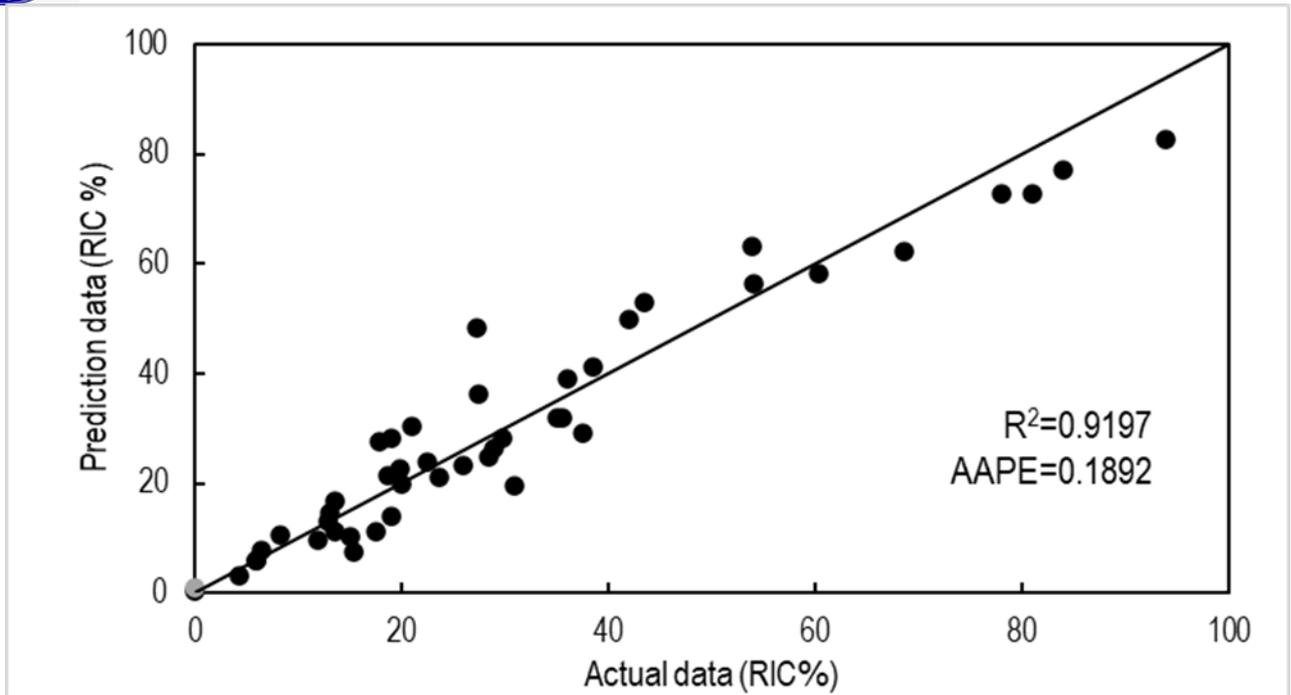


Figure 2. Plot of the predicted data from the linear regression model against the measured data

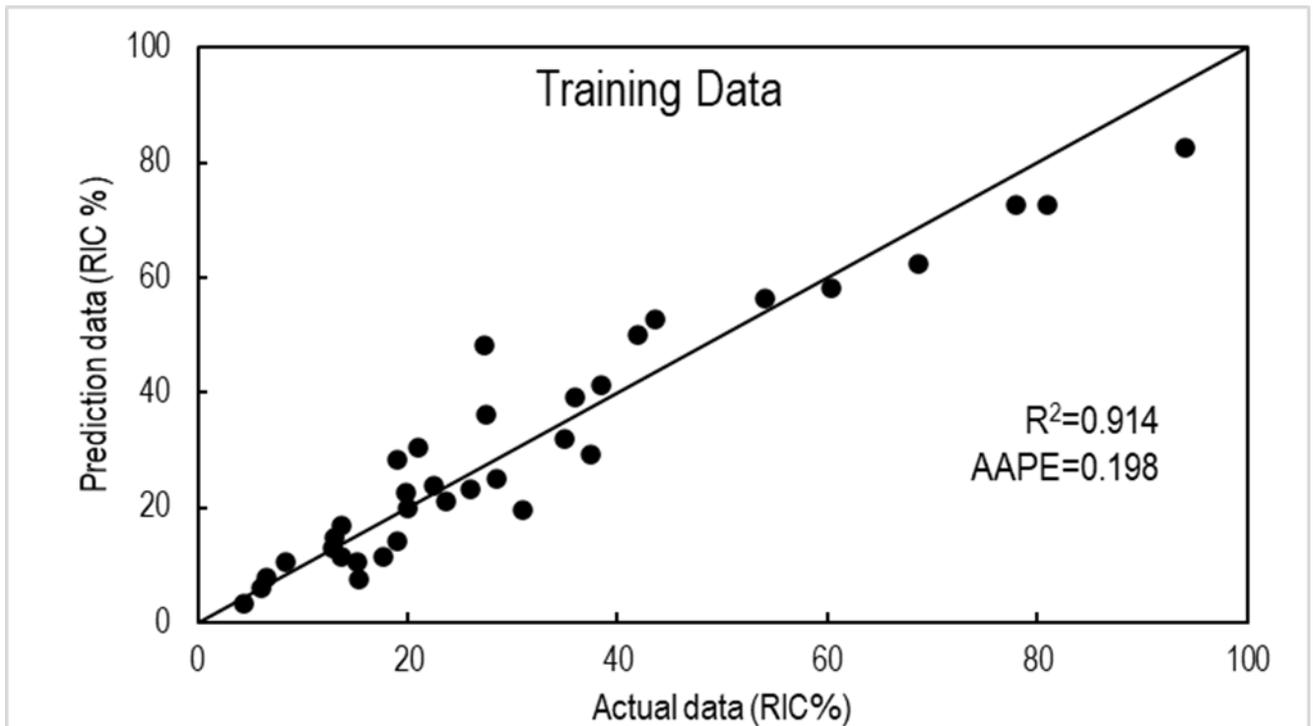


Figure 3. Plot of the predicted data from the linear regression model against the measured data for the training datasets

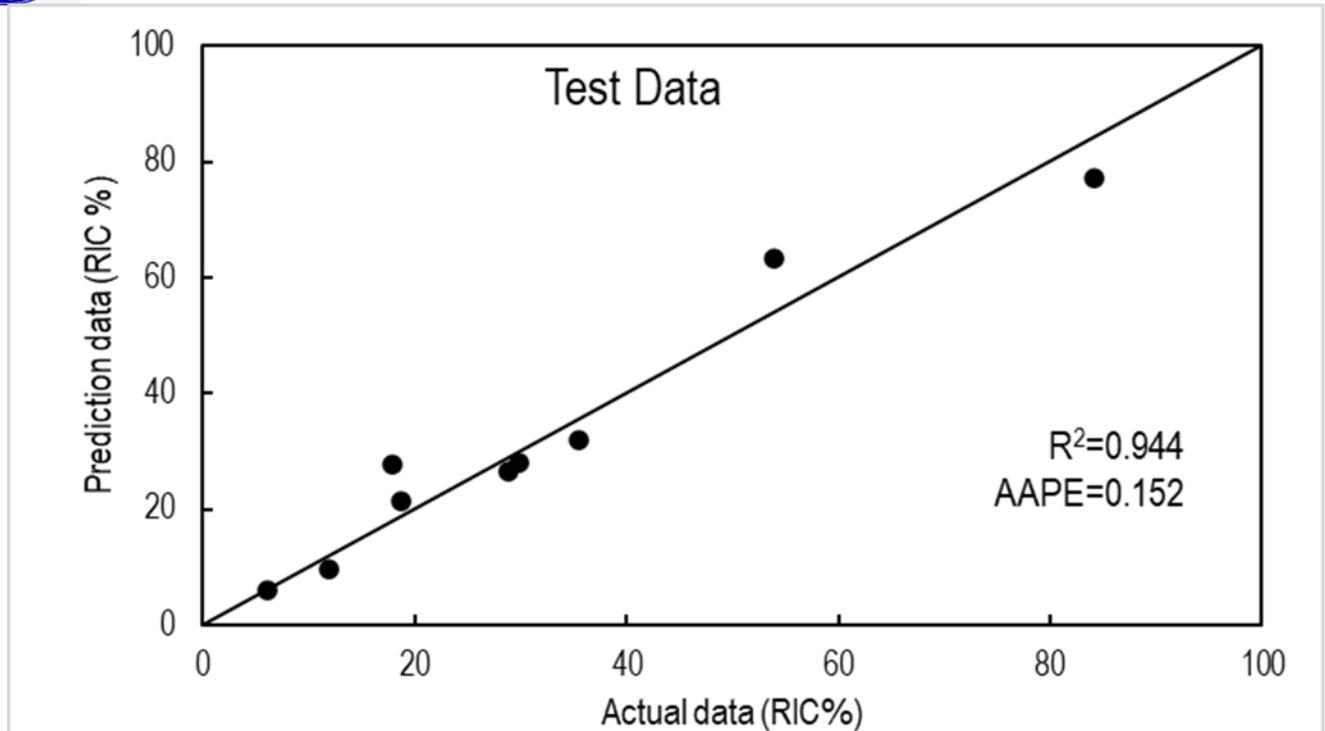


Figure 4. Plot of the predicted data from the linear regression model against the measured data for the testing datasets

Neural Network Regression Model

As with the linear regression model, the neural network receives an 80:20 split of the data points, with 80 percent (36 data points) being used for training and fitting the input parameters and the remaining 20% serving as a validation set against which the model's performance will be evaluated. The overall performance of a neural network model is highly dependent on its computational complexity and is measured in terms of the precision and robustness of the predicted data. Apart from that, when dealing with these machine learning modules, convergence and computational speed are significant concerns. As such, thorough and critical optimization of the model framework is required to produce a high-performing neural network algorithm. To be more precise, the neural network model is based on a flexible set of "hyperparameters" that define how the algorithm adapts to the input data. The size of the hidden layers, the alpha parameter, the momentum parameter, and the initial learning rate are all tuned in this project. The optimization technique is entirely based on which hyperparameter values provide the highest level of accuracy in terms of R Square.

To begin, the number of layers between the input and output nodes has a significant effect on the data being processed because it is necessary for converging and regressing the input variables towards the desired output. Adding additional layers does not always result in improved results, but rather results in unnecessarily longer computational times. Additionally, excessive growth in the number of neurons or layers can create an overfitting problem. As a result, it is necessary to conduct a thorough analysis of the datasets in order to determine the optimal neural network topology.

The plot (Figure 5) compares the R Square values obtained from various hidden layer configurations with an increasing number of neurons and layers from left to right. This plot demonstrates that three hidden layers with a size of 5 to 6 neurons each produce the highest R Square value of approximately 0.988. This result is consistent with the general rule of thumb for selecting the size of the hidden layer neurons, which states that the number of hidden layer neurons should be between 70% and 90% of the sum of the input and output layer neurons (S. Karsoliya, 2021). In this analysis, an input consisting of four injection parameters and one permeability change value (for a total of five neurons) corresponds to the approximated three hidden layers. Considering both the sensitivity analysis and the general rule of thumb, the hidden layer (6, 6, 6) is chosen, which is a three-layered neuron system composed of six neurons per layer.

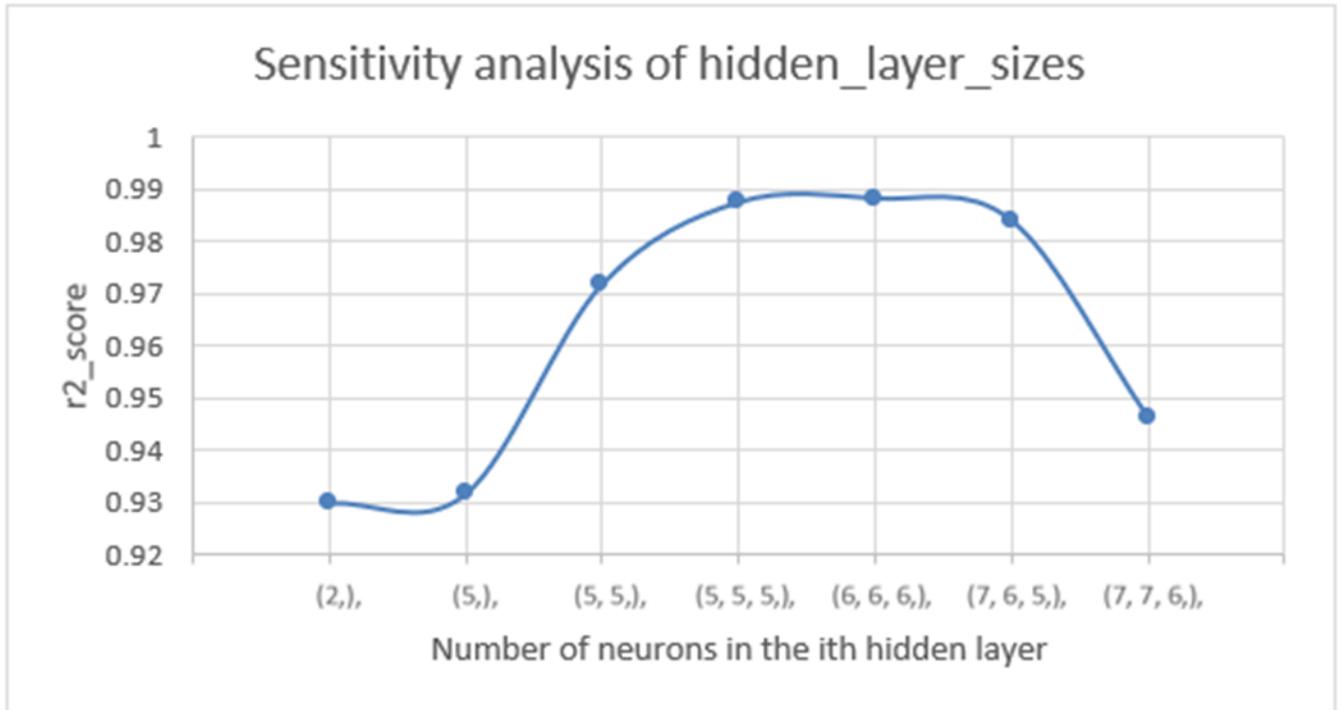


Figure 5. Sensitivity analysis on hidden layer size for the neural network model

The following section of the sensitivity analysis focused on the neural network model's alpha parameter. Alpha can be thought of as the model's learning rate, or the rate at which the model "learns" from previous iterations and applies specific weighting to converge on a value. It is a value between 0 and 1, with a value close to 0 indicating more conservative weight modifications and a value close to 1 indicating more radical weight modifications. The testing for alpha parameters ranging from 1E-2 to 8E-2 is shown in Figure 6 because they have a significant R Square value in comparison to the others in the 0 to 1 range. What is striking about this plot is that the R Square performance has a minimum variance of 0.0001 within these alpha values. However, there is a distinct peak at alpha = 7.75E-2, beyond which a decreasing trend is observed, and thus this value is chosen as the optimal value.

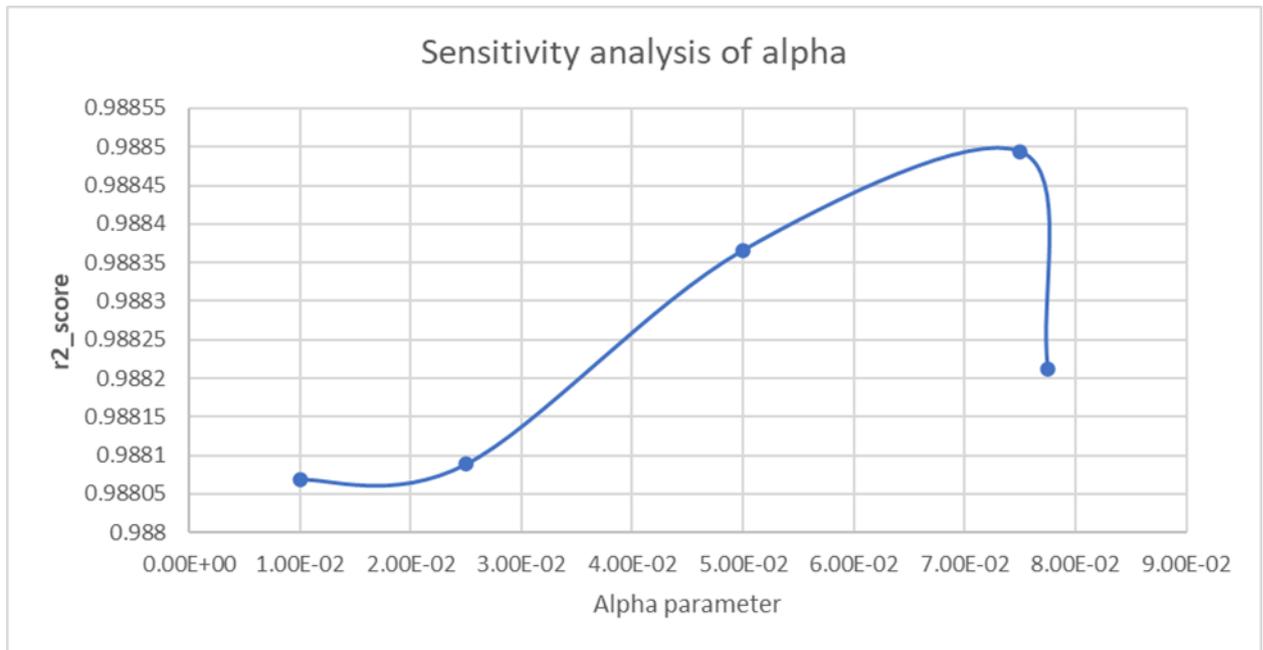


Figure 6. Sensitivity analysis on the alpha parameter for the neural network model

On the other hand, sensitivity analysis of the neural network model's initial learning rate reveals a greater variance within its range. The R Square performance increases significantly from the initial learning rate of 0.0001 to 0.001, before gradually decreasing to 0.003, as illustrated in Figure 7 below. This parameter specifies the initial value at which the model begins considering the weighting of the input parameters throughout the neural network algorithm. The performance of neural network models can be dramatically altered by selecting a small or large initial learning rate value (Y. Li, C. Wei, and T. Ma, 2019). While a low initial learning rate enables faster training and improved test performance initially, a high initial learning rate results in improved generalization shortly after the initial learning rate is annealed. This is dependent on how difficult it is to generalize the input parameters and fit patterns. For this particular research, the neural network model will use the highest performing initial learning rate of 0.001.

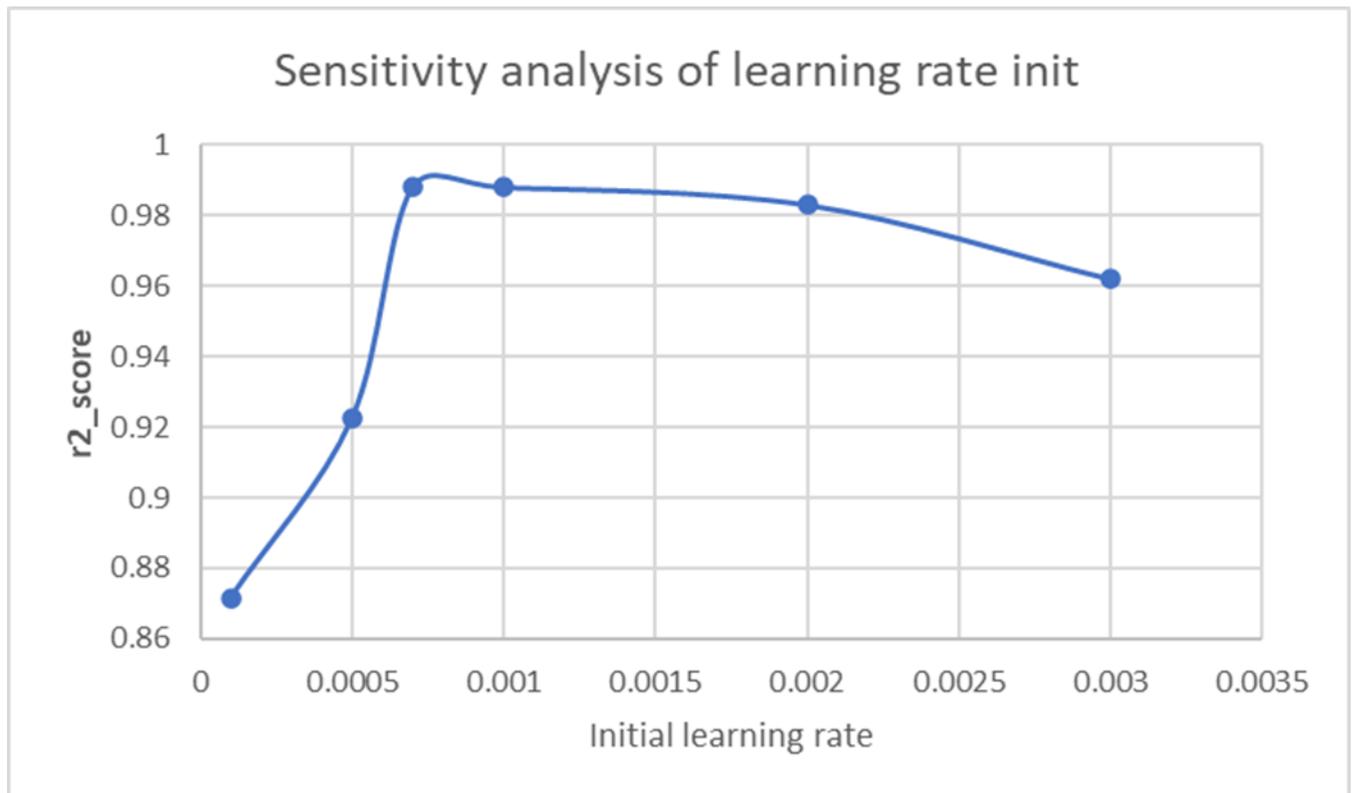


Figure 7. Sensitivity analysis on initial learning rate for the neural network model

Momentum is the final hyperparameter that is analyzed. Momentum determines the extent to which a previous weighting update influences the current weighting update. As illustrated in Figure 8, the momentum parameter, which also has a range of 0 to 1, exhibits a clear trend of increasing fit in the R Square value up to a value of 0.9. This appears to be the optimal velocity at which the neural network model operates, as increasing the velocity further appears to result in a decrease in performance.

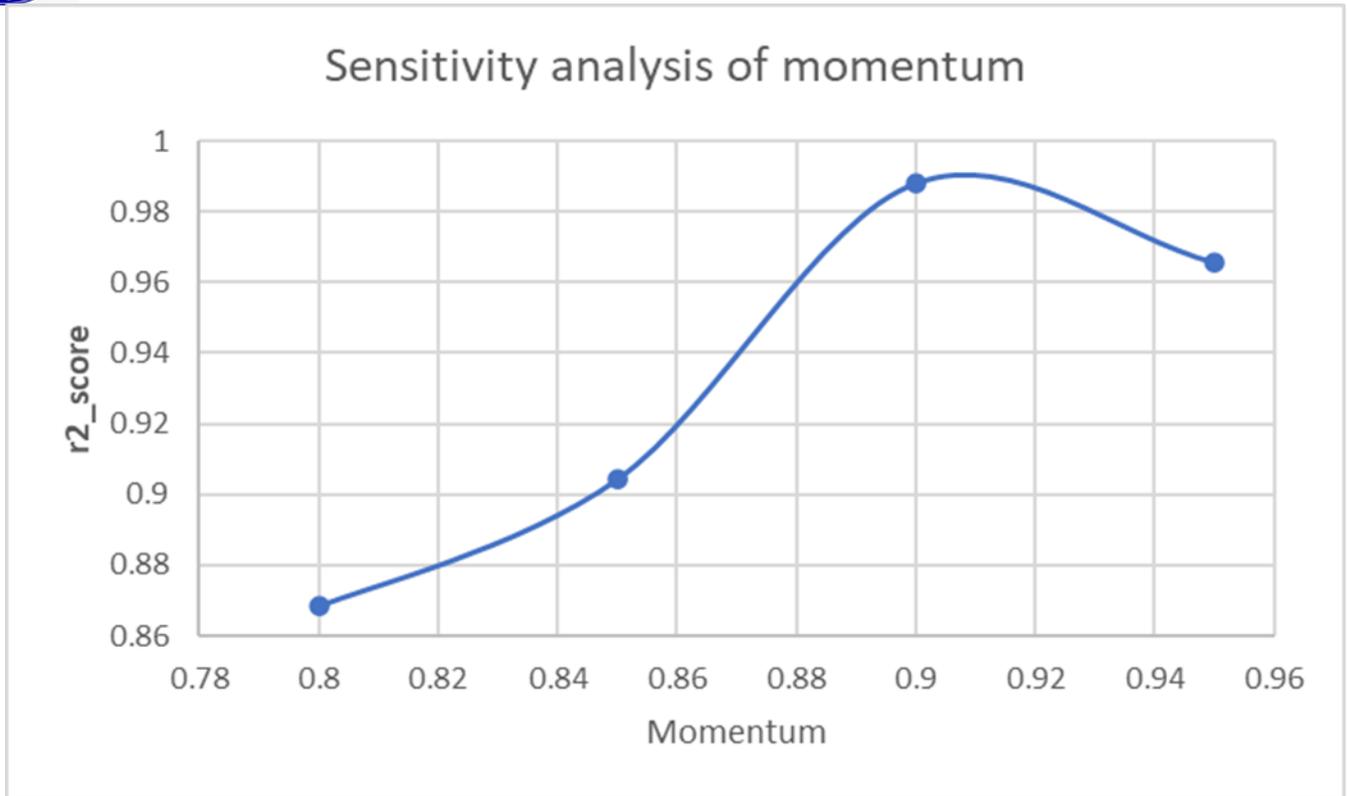


Figure 8. Sensitivity analysis on momentum for the neural network model

This sensitivity analysis of the hidden layer size, alpha parameter, initial learning rate, and momentum parameters provides valuable guidance for optimizing the hyperparameter settings for the neural network model based on the input datasets analyzed. As expected, the R Square value increases significantly when compared to the linear regression model, which achieves a remarkable value of 0.988212. (Table 2). This is an excellent indicator that the model is capable of fitting the data to the actual value with a minimum of variance.

Table 2. Performance of the neural network model

```
In [780]: x_train, x_test, y_train, y_test = split_data(df)
          y_pred, regr, scaler = neural_network_regression(x_train, x_test, y_train, y_test)
          # save_model(regr, scaler, "Exp.pkl")
          regression_evaluation_metrics('Neural Network Regression', 'Experiment Model', y_test, y_pred, x_test)
```

ANN score:0.9882118803730364.
 ANN Training done in 0:00:02.197636.

```
Out[780]:
```

	model	detail	mean_absolute_percentage_error	mean_absolute_error	root_mean_squared_error	r2_score	r2_adjusted
0	Neural Network Regression	Experiment Model	11.658284	2.696197	3.231454	0.988212	0.976424

As with the linear regression model, Figure 9 to Figure 11 below shows a plot of the predicted RIC and the actual RIC. This illustration clearly depicts a consistent and reliable fit between the two datasets with very little deviation from the unit slope line.

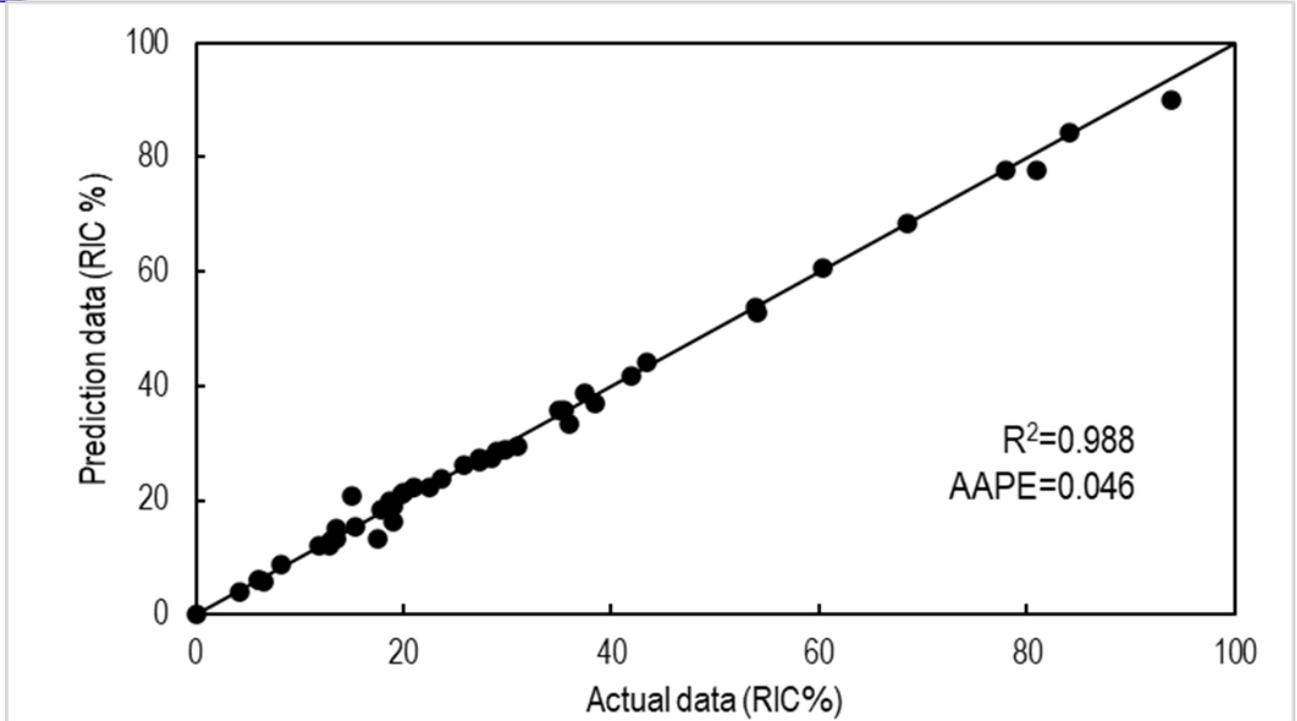


Figure 9. Plot of the predicted data from the neural network model against the measured data

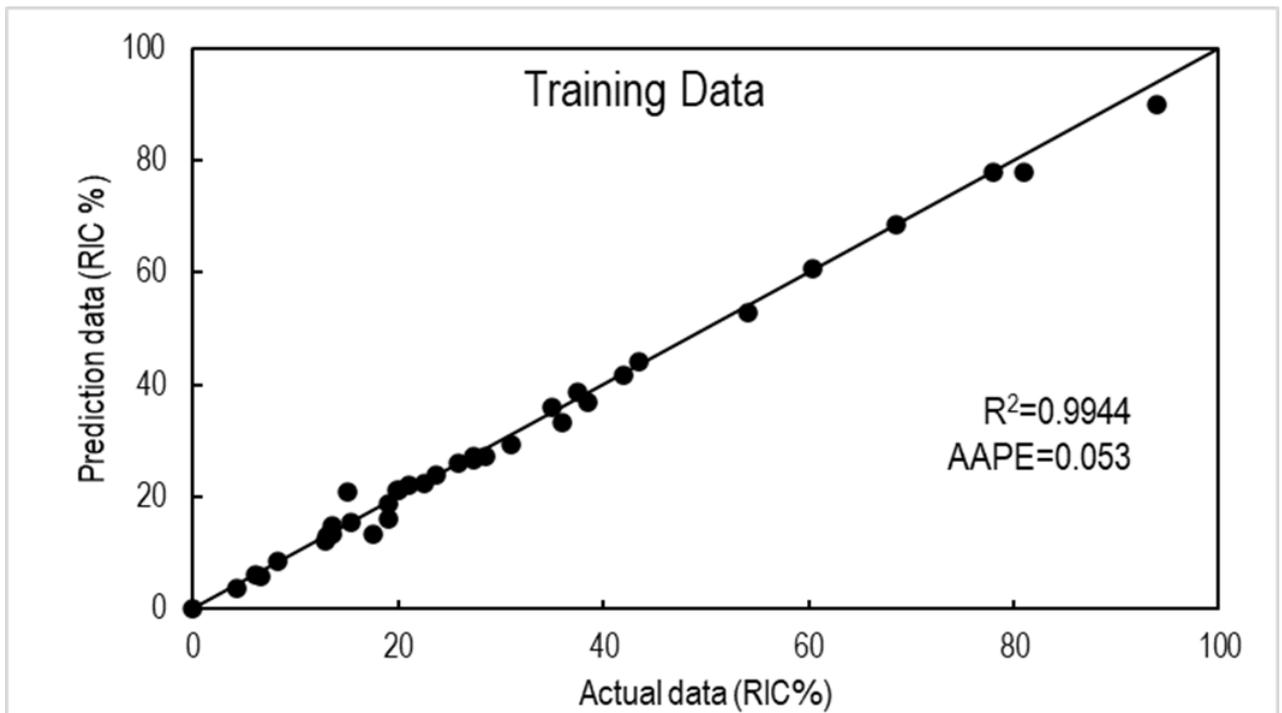


Figure 10. Plot of the predicted data from the neural network model against the measured data for the training datasets

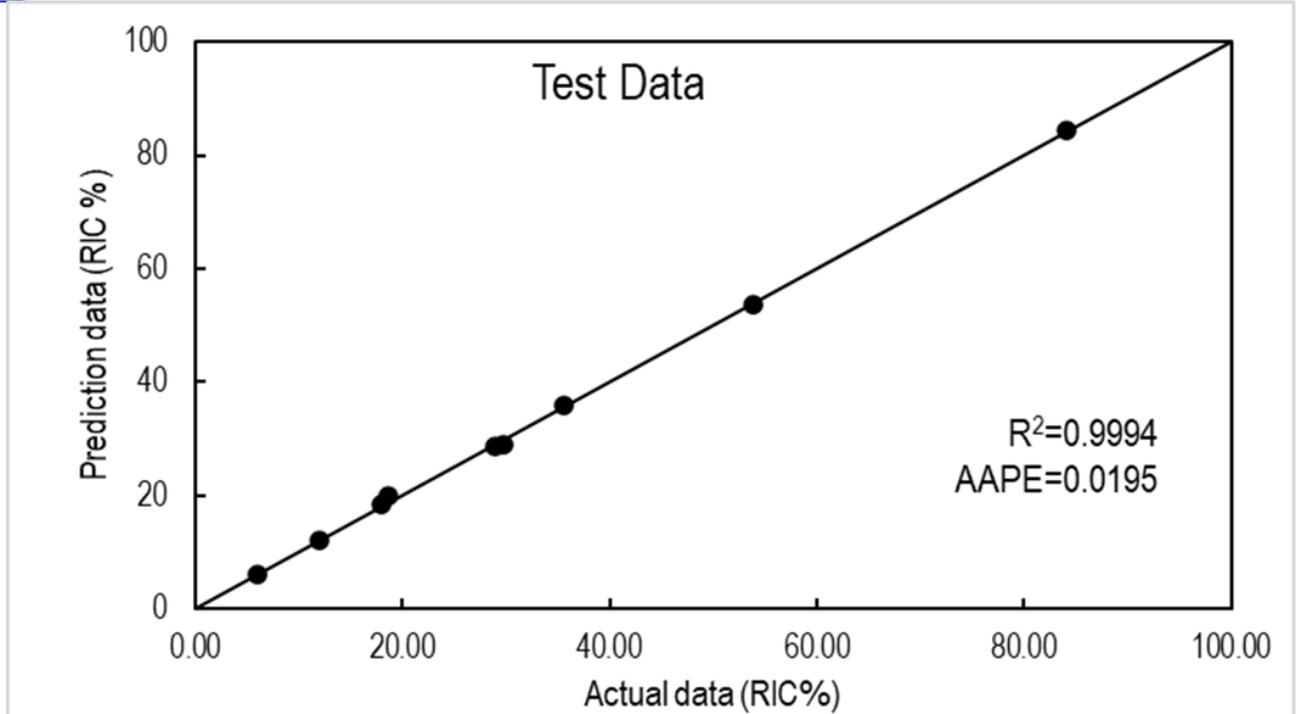


Figure 11. Plot of the predicted data from the neural network model against the measured data for the testing datasets

Comparison between Linear Regression Model and Neural Network Regression Model

The results of this chapter are compared and analyzed using both the linear regression and neural network models. Apart from the R Square and AAPE values, the mean absolute error, MAE, and root mean squared error, RMSE are also included in the statistical analysis to help determine which of the two regression models is more appropriate (Table 3).

Table 3. Comparison of statistical analysis parameters between the linear regression model and the neural network model

	Linear Regression Model	Neural Network Model
R Square	0.9198	0.9882
Average Absolute Percent Error (AAPE)	0.1893	0.0460
Mean absolute error (MAE)	5.8394	2.6962
Root mean squared error (RMSE)	7.1690	3.2315

The first set of statistical analysis parameters, the R Square value, has already indicated that the neural network model is more favorable than the linear regression model, with a value of 0.9882 versus 0.9198. This is a substantial improvement, indicating that the neural network model is more capable of fitting and handling the given dataset. Similarly, the AAPE values are significantly lower in the neural network model, at 0.0460, than in the linear regression model, at 0.1893.

Additionally, additional statistical tests using the MAE and RMSE revealed a trend in favor of the neural network model. The MAE is used to quantify accuracy for continuous variables, whereas the RMSE is a quadratic scoring rule that emphasizes the error's larger variances. Both the MAE and the RMSE are negative-oriented scores that can be used in conjunction to diagnose error variation. In comparison to the linear regression model, the neural network model achieves lower values on both scorings, indicating less error and variance.

This comparison of the two regression models demonstrates that the neural network model is superior to the linear regression model at modelling the four different injection parameters toward the desired permeability changes observed.

V. CONCLUSION

The objective of this project is to demonstrate a predictive modelling approach for estimating the expected Relative Injectivity Change, RIC, in sandstone reservoirs during a typical CO₂ injection process. It is well established that the combined effect of fines migration and salt precipitation on CO₂ injectivity varies with injection flow rate, brine salinity, particle size, and concentration of particles. Each injection parameter is mathematically mapped to the observed change in injectivity using both the linear regression and neural network methods. This provides a visual representation and comprehension of the effect and influence of each injection parameter on the RIC.

Additionally, it is discovered that the regression model developed using the neural network method outperforms the linear regression method in statistical analysis. The high degree of fitness observed is a strong indicator of the developed model's accuracy and reliability. This means that the regression model can be used effectively to forecast the expected permeability change in sandstones as a function of the four manipulated variables. The findings from this project may be used to stimulate future CO₂ injection and to help design the optimal injection parameters that result in the least amount of RIC, thereby assisting in achieving the emission reduction targets associated with these CO₂ sequestration procedures.

REFERENCES

- A. Verma and K. Pruess (1988). Thermohydrological Conditions and Silica Redistribution Near High-Level Nuclear Waste Emplaced in Saturated Geological Formations. *Journal of Geophysical Research: Solid Earth*, vol. 93, no. B2, pp.1159-1173.
- F. Othman, M. Yu, F. Kamali, and F. Hussain (2018). Fines Migration during Supercritical CO₂ Injection in Sandstone. *Journal of Natural Gas Science and Engineering*, vol. 56, pp. 344-357.
- F. Othman, M. A. Naufaliansyah, and F. Hussain (2019). Effect of Water Salinity on Permeability Alteration during CO₂ Sequestration. *Advances in Water Resources*, vol. 127, pp. 237-251.
- I. M. Mohamed, J. He, and H. A. Nasr-El-Din (2012). Carbon Dioxide Sequestration in Sandstone Aquifers: How Does It Affect the Permeability?. in *Carbon Management Technology Conference, 2012: Carbon Management Technology Conference*.
- J. Hommel, E. Coltman, and H. Class (2018). Porosity–Permeability Relations for Evolving Pore Space: A Review with a Focus on (Bio-)geochemically Altered Porous Media. *Transport in Porous Media*, vol. 124, no. 2, pp. 589-629, 2018/09/01. doi: 10.1007/s11242-018-1086-2.
- M. A. Md Yusof et al. (2020). Effects of CO₂/Rock/Formation Brine Parameters on CO₂ Injectivity for Sequestration. *SPE-203843-PA*, vol. Preprint, no. Preprint, p. 14, 2020/10/1/. doi: 10.2118/203843-PA.
- R. Miri and H. Hellevang (2016). Salt Precipitation during CO₂ Storage—A review," *International Journal of Greenhouse Gas Control*, vol. 51, pp. 136-147.
- S. Karsoliya (2012). Approximating Number of Hidden Layer Neurons in Multiple Hidden Layer BPNN Architecture. *International Journal of Engineering Trends and Technology*, vol. 3, no. 6, pp. 714-717.
- W. D. Carrier III (2003). Goodbye, Hazen; Hello, Kozeny-Carman. *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 129, no. 11, pp. 1054-1056.
- Y. Xu and V. Ramanathan (2017). Well below 2 C: Mitigation Strategies for Avoiding Dangerous to Catastrophic Climate Changes. *Proceedings of the National Academy of Sciences*, vol. 114, no. 39, pp. 10315-10323.
- Y. A. Sokama-Neuyam, S. L. Forsetløyken, J.-e. Lien, and J. R. Ursin (2017). The Coupled Effect of Fines Mobilization and Salt Precipitation on CO₂ Injectivity. *Energies*, vol. 10, no. 8, p. 1125.



- Y. A. Sokama-Neuyam, J. R. Ursin, and P. Boakye (2019). Experimental Investigation of the Mechanisms of Salt Precipitation during CO₂ Injection in Sandstone. C—Journal of Carbon Research, vol. 5, no. 1, p. 4.
- Y. A. Sokama-Neuyam, P. U. R. Ginting, B. Timilsina, and J. R. Ursin (2017). The Impact of Fines Mobilization on CO₂ Injectivity: An Experimental Study. International Journal of Greenhouse Gas Control, vol. 65, pp. 195-202.
- Y. Li, C. Wei, and T. Ma (2019). Towards Explaining the Regularization Effect of Initial Large Learning Rate in Training Neural Networks," arXiv preprint arXiv:1907.04595.