IJAME
Int. J. of Applied Mechanics and Engineering, 2026, vol.31, No.2, pp.155-167
DOI: 10.59441/ijame/214981

PROBABILISTIC PREDICTION OF PERMEABILITY DAMAGE IN WATERFLOODING USING GAUSSIAN PROCESS REGRESSION

R. Saifi[1*], N. Zeraibi[1], M. Gareche[1], M. Nait Amar[2] and C. Benamara[2]
[1]Department of Mining and Petroleum Deposits, Physics Engineering of Hydrocarbons, M'Hamed BOUGARA University, Boumerdes, ALGERIA
[2]Department of Thermodynamic Studies, Laboratory Division, Sonatrach, ALGERIA
E-mail: r.saifi@univ-boumerdes.dz

Waterflooding represents one of the most extensively employed techniques for secondary oil recovery, where water is injected into reservoirs to displace oil toward production wells and enhance hydrocarbon recovery. However, a significant challenge in waterflooding operations is salt precipitation, which results from the chemical incompatibility between the injected water, commonly enriched with divalent cations such as calcium, strontium, and barium, and the formation water, which generally exhibits elevated concentrations of sulfate ions. This chemical interaction leads to the formation of sulfate scales, significantly reducing reservoir permeability and hindering oil recovery efficiency.

This study employed Gaussian Process Regression (GPR), a nonparametric, probabilistic machine learning method, to predict the extent of permeability damage resulting from sulfate scale deposition during waterflooding. A dataset of *431* experimental tests was used, incorporating input variables such as ion concentrations, differential pressure, temperature, pore volume, and initial permeability. The GPR model successfully captured the nonlinear relationships between these inputs and the resulting permeability damage. Both graphical and statistical evaluations demonstrated strong agreement between the model predictions and experimental results, with a high coefficient of determination ($R^2 = 0.99$) and low prediction errors (RMSE = *0.0839* ; MAE = *0.0529*).

The GPR model demonstrated enhanced predictive accuracy compared to alternative machine learning algorithms, including decision trees, support vector machines (SVMs), and artificial neural networks. Furthermore, the probabilistic framework of GPR facilitated the quantification of predictive uncertainty, thereby establishing it as a dependable and robust tool for informed operational decision-making in reservoirs susceptible to scaling.

Key words: waterflooding; gaussian process; scaling deposition; machine learning; damage permeability.

1. Introduction

Flow assurance is critical to oil and gas production, ensuring the efficient and uninterrupted transport of hydrocarbons from the reservoir to surface processing facilities. One of the most persistent challenges in this domain is the formation of inorganic scales, which are solid mineral deposits that precipitate when incompatible water chemistries interact [1]. Commonly encountered mineral scales, such as calcium carbonate $(CaCO_3)$, calcium sulfate $(CaSO_4)$, and barium sulfate $(BaSO_4)$, can precipitate at various stages throughout the hydrocarbon production process [2]. Their precipitation is strongly influenced by fluid composition, temperature, pressure, and flow rate [3]. When left unmanaged, scale deposition can block pore spaces and constrict flow paths in wellbores, production tubing, valves, downhole equipment, and even the reservoir rock. This obstruction reduces permeability, impairs injectivity, and significantly impacts

* To whom correspondence should be addressed

hydrocarbon recovery, often leading to increased operational costs and unplanned downtime [4]. Scale formation results from complex geochemical reactions between formation water, injection water, and reservoir rock, all shaped by the geological and hydrological history of the field [5]. The American Society for Testing and Materials (ASTM) has identified over 100 types of deposits in water systems. The predominant scales encountered in oilfield environments are carbonates and sulfates of divalent alkaline-earth metals, particularly calcium, barium, and strontium. Iron-based deposits (including oxides, carbonates, sulfides, and hydroxides) may also form due to corrosion, microbial activity, acid stimulation, or oxygen ingress. (see Tab.1) [6, 7].

Table 1. Representative examples of common oilfield scales.

Type of scale	Formula	Acid soluble
Calcium Carbonate	Calcite $(CaCO_3)$	soluble
	Aragonite $(CaCO_3)$	soluble
	Vaterite $(CaCO_3)$	soluble
Calcium Sulfate	Anhydrite $(CaSO_4)$	not soluble
	Hemihydrate $\left(CaSO_4.\frac{1}{2}H_2O\right)$	not soluble
	Gypsum $(CaSO_4.2H_2O)$	slightly
Strontium Sulfate	Celestite $(SrSO_4)$	not soluble
Barium Sulfate	Barite $(BaSO_4)$	not soluble
Iron Oxo Species	FeO_4, $FeOOH$, $Fe(OH)_2$, etc.	soluble
Iron Sulfides	FeS_y	highly dependent on $Fe{:}S$ ratio
Iron Carbonates	e.g., Magnesite $(MgCO_3)$	soluble
other Hydroxides	e.g., Brucite $(Mg(OH)_2)$	soluble
Halite	$NaCl$	soluble in fresh water
Lead Sulfide	Galena (PbS)	soluble
Silica and Metal	Amorphous (SiO_2)	not soluble
Silicates	Various metal silicates	-

Predicting scale formation is a complex and ongoing challenge. Thermodynamic models remain widely used to assess scale-forming tendencies, using the Saturation Index SI as a core metric:

$$SI = \log\left(\frac{IAP}{K_{sp}}\right). \tag{1.1}$$

Where IAP is the ion activity product and K_{sp} is the solubility product of the mineral. A positive SI indicates supersaturation and a tendency for scale to form; $SI = 0$ implies equilibrium; and a negative SI suggests undersaturation, where dissolution may occur. However, thermodynamic predictions do not account for the rate of scale formation, which is influenced by kinetic factors such as temperature, flow dynamics, and turbulence. For instance, high shear zones (like valve restrictions or sharp bends) often serve as scale nucleation sites, especially when fluids are heated or mixed rapidly [8]. To overcome these limitations,

researchers have adopted kinetic modeling approaches and experimental techniques such as dynamic coreflood tests, which simulate scale deposition under flow conditions representative of subsurface environments. These studies have highlighted the roles of contact time, ionic concentration ratios, and residence time in governing the dynamics of precipitation and the associated reduction in permeability [9]. In parallel, data-driven approaches, particularly machine learning (ML), have gained traction for their ability to model complex, nonlinear relationships between operational parameters and scale-induced damage. Techniques such as artificial neural networks (ANNs) and support vector machines have been applied to predict permeability impairment with high accuracy, identifying variables such as initial permeability, temperature, pressure drop, ionic concentrations, and injected volume as critical predictors [10, 11].

This study evaluates four ML models (GPR), decision trees, support vector machines (SVMs), and neural network architectures) to identify the most robust predictor of permeability damage (K_d) caused by barium sulfate scaling. The dataset, comprising *431* experimental tests, includes ion concentrations, temperature, pressure, and pore volume variables. We hypothesize that GPR will outperform other models due to its ability to:

- Model nonlinear relationships between scaling drivers and permeability loss.
- Quantify prediction uncertainty (critical for decision-making in high-risk environments).
- Avoid overfitting through Bayesian regularization, even with limited data.

By addressing these challenges, this work aims to establish GPR as a state-of-the-art tool for proactive scaling management, reducing dependence on costly remedial measures and improving long-term reservoir performance.

2. Gaussian process

A Gaussian Process (GP) is a non-parametric, probabilistic machine learning method that models an unknown function as a distribution over functions rather than optimizing parameters within a predefined function class. Unlike traditional parametric models, GPs make predictions by leveraging the observed data to infer a posterior distribution over all plausible functions consistent with the data. [12, 13].

A Gaussian Process (GP) is completely defined by two fundamental elements: the mean function, $m(x)$, which specifies the expected value of the process at each input point *x*, and the covariance function or kernel, $k(x,x')$, which measures the degree of similarity or statistical correlation between any two input points x and x'. Together, these functions determine the distribution of the corresponding function values.

$$\begin{bmatrix} \\ f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(x_1) \\ m(x_2) \\ \vdots \\ m(x_3) \end{bmatrix} , \begin{bmatrix} k(x_1,x_1) & \dots & k(x_1,x_n) \\ \vdots & \ddots & \vdots \\ k(x_n,x_1) & \dots & k(x_n,x_n) \end{bmatrix} \right), \tag{2.1}$$

where

$f(X)$: the values at training points X,

$m(X)$: the mean function evaluated at X,

$K(X,X)$: the covariance between training points.

The choice of the kernel function is essential, as it determines how well the model can capture the underlying patterns in the data. It allows GPs to model complex, nonlinear relationships while also providing uncertainty estimates with each prediction. Given training inputs X and corresponding outputs y, GPR models the joint distribution of the function values at these training points and a new test point x^*. From this, it derives a

conditional Gaussian distribution for the prediction at X^*. Specifically, the vector containing the function values at the training points $f(X)$ and the value at the new input $f(x^*)$ follows a multivariate normal distribution [14].

$$\begin{bmatrix} f(X) \\ f(x^*) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(X) \\ m(x^*) \end{bmatrix}, \begin{bmatrix} K(X,X)+\sigma_n^2 I & K(X,x^*) \\ K(x^*,X) & K(x^*,x^*) \end{bmatrix} \right), \tag{2.2}$$

where

$f(x^*)$: the values at the test points x^*,

$m(x^*)$: the mean function evaluated at x^*,

$K(X,x^*)$: the covariance between training and test,

$K(x^*,x^*)$: the covariance at the test point.

From multivariate normal theory, the conditional distribution of $f(x^*)$ given $f(x)=y$ is also Gaussian:

$$f(x^*) \mid X, y \sim \mathcal{N}\left(\mu(x^*), \sigma^2(x^*)\right) \tag{2.3}$$

where the conditional mean and variance are:

$$\mu(x^*) = m(x^*) + K(x^*_*, X)\left[K(X,X)+\sigma_n^2 I\right]^{-1}(y - m(X)), \tag{2.4}$$

$$\sigma^2(x^*) = k(x^*,x^*) - K(x^*,X)\left[K(X,X)+\sigma_n^2 I\right]^{-1} K(X,x^*). \tag{2.5}$$

These predictive formulas involve computing the kernel matrix between training points, the cross-kernel between the test and training points, and solving a linear system involving the noise variance. The result is a robust prediction method that adapts to the data and provides confidence bounds for each prediction. However, the performance of Gaussian Process Regression heavily depends on the choice of the kernel (or covariance) function, which encodes assumptions about the function being modelled. It is crucial to select an appropriate kernel and optimize its hyperparameters (via marginal likelihood maximization). Poor choices can lead to underfitting or overfitting, while well-tuned kernels can significantly enhance predictive accuracy and uncertainty estimation.

3. Data gathering

A comprehensive dataset of *431* experimental trials was compiled from open-source studies [15, 16]. The dataset integrates two complementary experimental approaches designed to simulate sulfate scaling under controlled laboratory conditions: beaker tests (thermodynamic analysis) and core flooding tests (hydrodynamic and permeability analysis). Synthetic formation water (rich in Ba^{2+}, Sr^{2+}, Ca^{2+}) and seawater (high in SO_4^{2-}) were mixed at predefined ratios and heated to temperatures varying from *50*°C to *90°C*. Solutions were equilibrated for *24* hrs, and residual ion concentrations were measured using atomic absorption spectrometry (AAS). Core flooding tests simulated permeability damage during seawater injection into sandstone reservoirs. The key experimental parameters included temperature, differential pressure, ionic composition of the brines, permeability damage, pore volume, and flow rate.

4. Modelling and results

In injection systems, pressure drop (ΔP) and flow rate (q) are physically dependent, meaning they cannot both be used as independent inputs for predicting scaling. These systems operate in either pressure-controlled or rate-controlled modes [17] (fixing one determines the other). $K_d = f(q/\Delta P)$ is more suitable for monitoring or controlling injection operations in real time, but not for reliably predicting scaling, which depends on chemical and thermodynamic factors beyond flow dynamics.

Therefore, we select input variables that reflect the underlying chemistry and thermodynamics of the system: ion concentrations, differential pressure , temperature , pore volume , and initial permeability . These inputs represent the conditions that govern saturation, nucleation, and growth of scales in the porous medium. The target output is the damaged permeability K_d, which reflects the degree of flow impairment due to scaling.

$$K_d = f\left(\left[Ca^{2+}\right], \left[Sr^{2+}\right], \left[Ba^{2+}\right], \left[SO_4^{2-}\right], \Delta p, T, \Delta t, K_i\right). \tag{4.1}$$

To implement Gaussian Process Regression (GPR) for modelling permeability damage due to barium sulfate scaling, MATLAB was used with both the Regression Learner App and custom scripts from the Statistics and Machine Learning Toolbox [18]. The dataset, comprising *431* samples, was first preprocessed by normalizing all input features using z-score transformation to ensure uniform scaling, as GPR is sensitive to feature magnitudes.

To assess the presence of outliers in the K_d dataset, both a histogram with a fitted Gaussian distribution and a $\mathrm{Q}-\mathrm{Q}$ (quantile-quantile) plot were analyzed [19, 20]. The histogram (Fig.1) indicated that all data points lie within the $\pm 3\sigma$ range, with only a few values extending slightly beyond $\pm 2\sigma$ – well within the expected limits of a normally distributed dataset. This suggests that the data do not contain any statistically significant outliers based on standard deviation thresholds. Furthermore, the $\mathrm{Q}-\mathrm{Q}$ plot (Fig.2) confirmed the

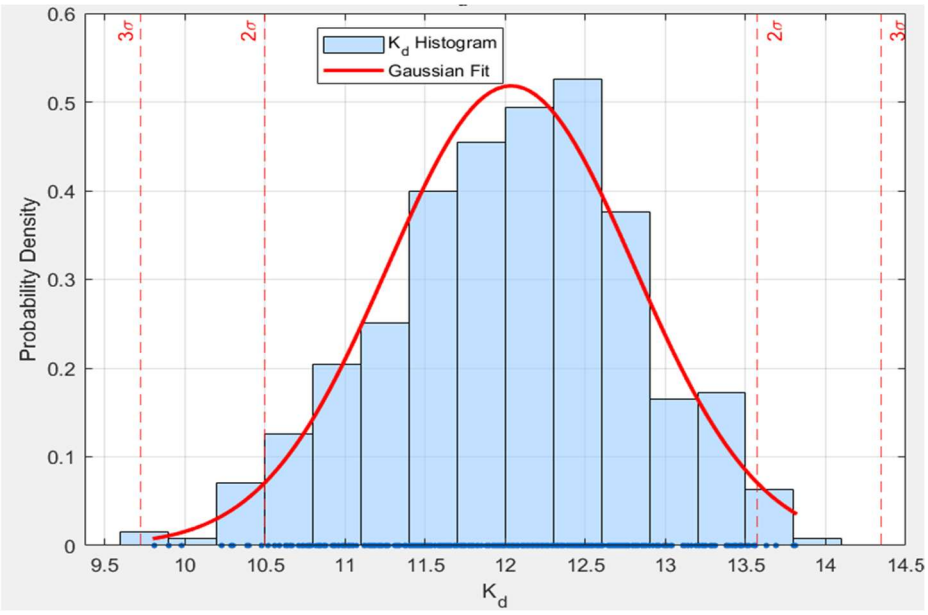


Fig.1. Histogram of permeability damage with Gaussian curve.

approximate normality of the distribution, with most points closely following the reference line, except for minor deviations at the upper tail. These small departures indicate mild skewness but do not justify the removal of any values. Therefore, no data points were excluded, and the full dataset was retained for subsequent analysis.

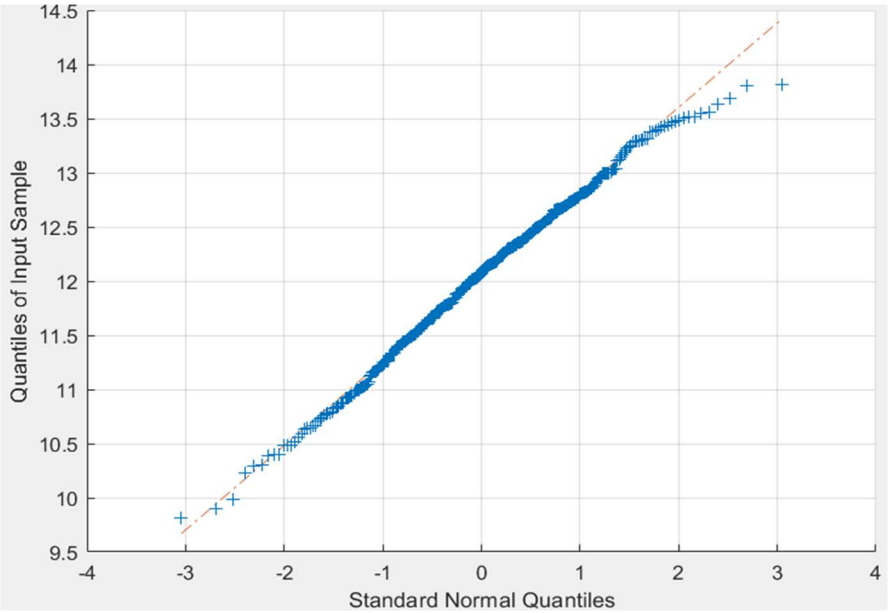


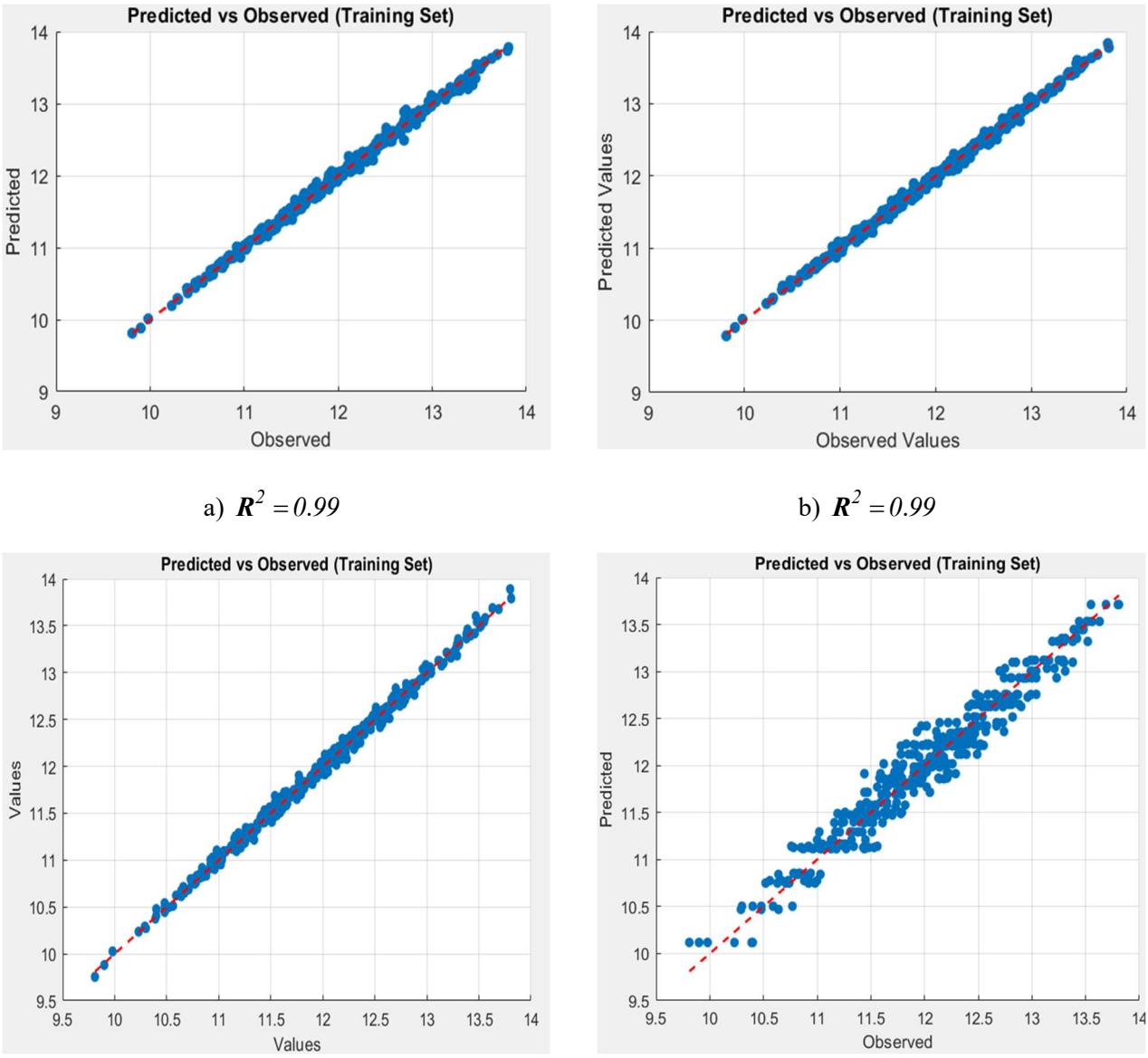
Fig.2. Q – Q plot for permeability damage.

The dataset was divided into two subsets: *80%* for training and *20%* for testing, using the holdout method of cross-validation. Hyperparameters, including kernel parameters and noise variance, were optimized using Bayesian optimization to maximize the log marginal likelihood. Model performance was evaluated on the test set using RMSE, MAE, and R^2 metrics. Predictive uncertainty was quantified using *95%* confidence intervals derived from the posterior distribution. , several machine learning models were evaluated, including Gaussian Process Regression (GPR), neural networks with different architectures, decision trees, and Support Vector Machines (SVM).

The quality of fit achieved during model training is illustrated in Fig.3, which shows the predicted versus actual values for the training subset. The Gaussian Process, Support Vector Machine $\left(R^2 = 0.99\right)$, and Artificial Neural Network $\left(R^2 = 0.99\right)$ models exhibit an excellent match with the experimental data, confirming their strong capability to capture the nonlinear behavior of scaling-induced permeability damage. The Decision Tree model also performs well with $R^2 = 0.95$, though slightly lower than the other methods. These results demonstrate the ability of all four models to effectively learn the underlying patterns in the training data, providing a solid basis for subsequent validation on unseen test samples.
The developed models and their corresponding performance indicators are reported in Tab.2.

Table 2. Model performance summary.

Model	**Type**	***RMSE***	***R²***	***MAE***	***MAPE***
Model 1	Optimizable GPR	*0.0839*	*0.99*	*0.0529*	*0.4%*
Model 2	Optimizable Tree	*0.2951*	*0.85*	*0.2369*	*2.0%*
Model 3	Optimizable SVM	*0.2266*	*0.91*	*0.1702*	*1.4%*
Model 4	Neural Network	*0.1164*	*0.98*	*0.0715*	*0.6%*



a) $R^2 = 0.99$

b) $R^2 = 0.99$

c) $R^2 = 0.99$

d) $R^2 = 0.95$

Fig.3. Training predicted vs actual plot: (a) Gaussian process model, (b) SVM model, (c) ANN model, (d) tree model.

From Tab.2, it is evident that the Optimizable Gaussian Process Regression (GPR) model (Model 1) outperformed all other models, achieving the lowest RMSE (0.0839), the highest $R^2 = 0.99$, and the smallest MAE (0.0529) and MAPE (0.4%). These metrics indicate that the GPR model provides highly accurate predictions with excellent generalization across various operating conditions. The high R^2 value suggests that the model explains nearly all the variability in the data, while the low error metrics indicate minimal prediction errors. This performance is critical in applications where precise prediction of permeability damage can prevent operational inefficiencies and economic losses in oil reservoirs.

The Neural Network (Model 4) also demonstrated strong performance with an R^2 of 0.98 and RMSE of 0.1164, slightly below that of the GPR model. However, its greater structural complexity and higher

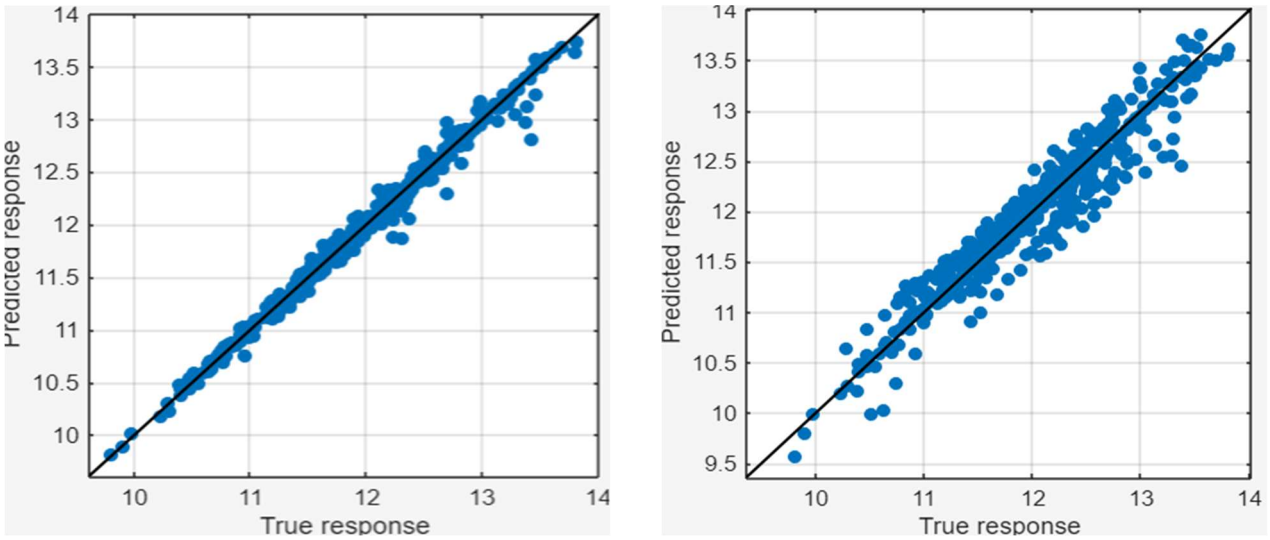
computational requirements make it less practical for some applications. In contrast, the Optimizable Tree model (Model 2) had the weakest performance, with an R^2 of 0.85 and higher error metrics, indicating it is less suitable for this nonlinear prediction task due to its limited ability to capture complex data patterns. The Optimizable SVM (Model 3) provided a compromise between interpretability and accuracy but still underperformed compared to the GPR and neural network models, likely due to its sensitivity to hyperparameter tuning.

The optimal hyperparameters and model architectures were determined for each machine learning method. The Gaussian Process Regression (GPR) model was configured with a zero basis function and a nonisotropic Matern 3/2 kernel, while the Neural Network (NN), Decision Tree (DT), and Support Vector Machine (SVM) models were tuned with their respective optimal settings. The complete configurations are summarized in Tab.3.

Table. 3. Architecture and Optimal Hyperparameters of the Best Models.

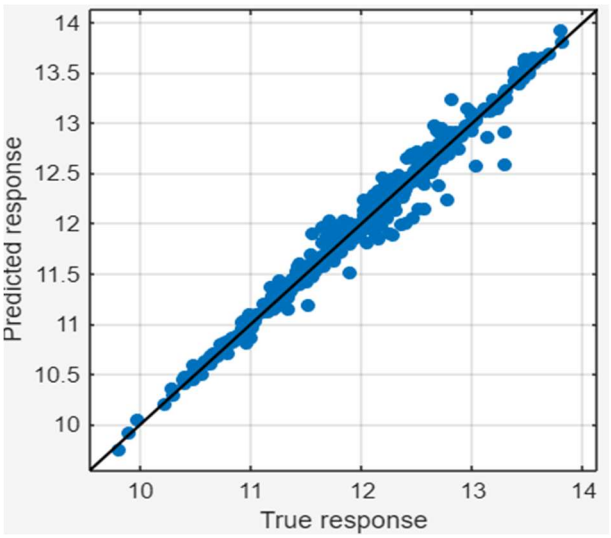
Model	Architecture and Optimal Hyperparameters
Gaussian Process Regression (GPR)	Basis function: Zero, Kernel function: Nonisotropic Matern *3/2*, Kernel scale: *532.11*, Sigma (Noise level): *0.00139*.
Neural Network (NN)	Number of hidden layers: *1*, Neurons in hidden layer: *18*, Activation function: Sigmoid, Regularization: $\lambda = 4.052 \times 10^{-6}$.
Decision Tree (DT)	Minimum leaf size: *2*
Support Vector Machine (SVM)	Kernel function: Quadratic, Box constraint: *0.4416*, Epsilon: *0.0456*.

The choice of the Nonisotropic Matern $3/2$ kernel is particularly suitable for this task, as it can model smooth variations in permeability damage while accommodating local fluctuations caused by scaling.

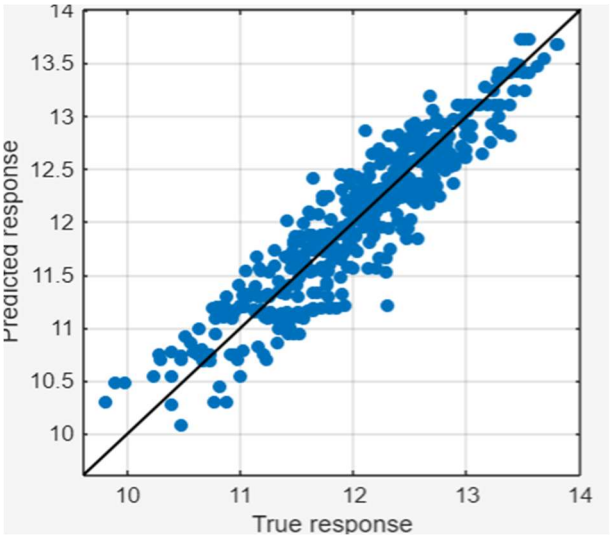


a) $\boldsymbol{R^2 = 0.99}$ b) $\boldsymbol{R^2 = 0.91}$

Fig.4. Validation predicted vs actual plot: (a) Gaussian process model, (b) SVM model, (c) ANN model, (d) tree model.



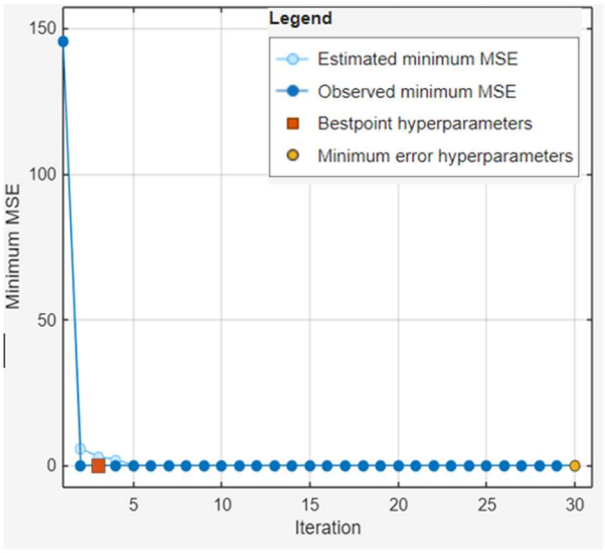
c) $R^2 = 0.98$



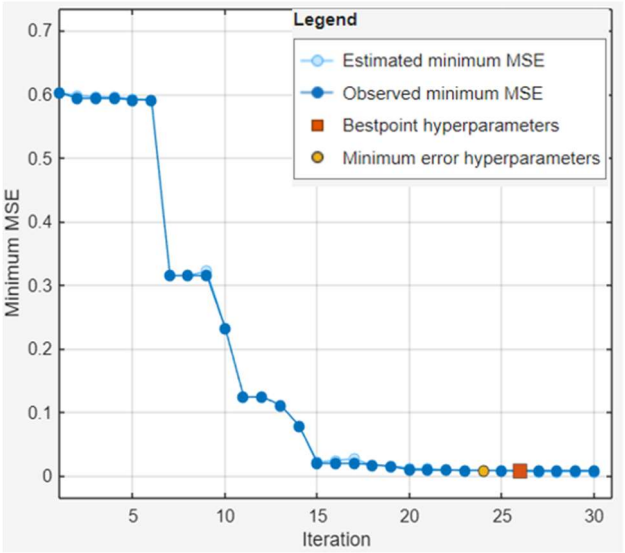
d) $R^2 = 0.85$

Fig.4 cont. Validation predicted vs actual plot: (a) Gaussian process model, (b) SVM model, (c) ANN model, (d) tree model.

Figure 4(a) displays a scatter plot comparing the predicted values with the actual observations obtained from the GPR model. The data points are tightly clustered around the diagonal line, indicating that the predicted values closely match the true values. This tight clustering demonstrates the high accuracy of the model and suggests that it has low bias, as there is no systematic deviation from the diagonal. Without suffering from heteroscedasticity, which is critical for reliable predictions across diverse operating conditions.

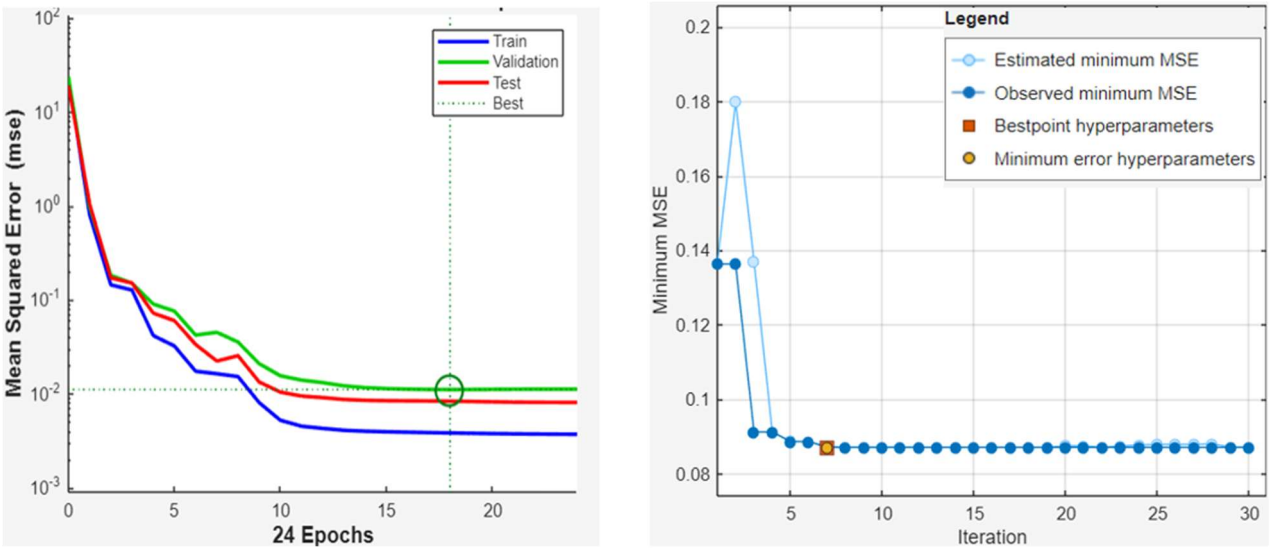


a)



b)

Fig.5. Validation predicted vs actual plot: (a) Gaussian process model, (b) SVM model,(c) ANN model, (d) tree model.



c) d)

Fig.5 cont. Validation predicted vs actual plot: (a) Gaussian process model, (b) SVM model,(c) ANN model, (d) tree model.

Figure 5 shows the convergence behavior of the four models. The Gaussian Process (a) and Decision Tree (d) reached stable performance within the first few iterations, while the SVM (b) required more iterations before stabilizing. The Neural Network (c) showed a typical learning curve with training, validation, and test errors decreasing smoothly until convergence around 24 epochs. Overall, all models converged properly with optimized hyperparameters, ensuring reliable performance without signs of overfitting.

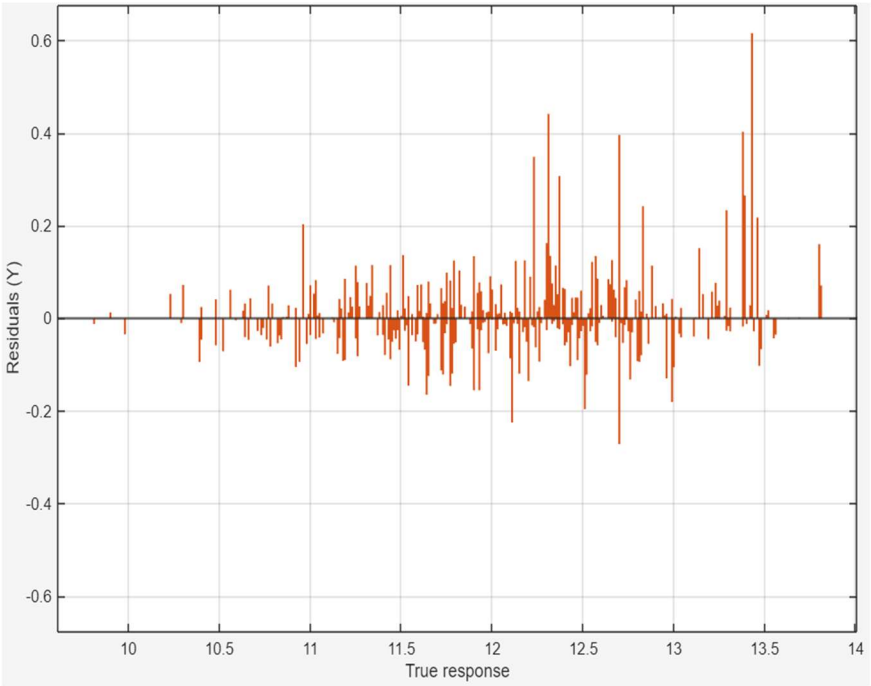


Fig.6. Validation residuals plot for prediction: GPR model.

Furthermore, Fig.6 shows the residuals plot for the GPR model's predictions. The residuals are randomly scattered around zero, with no apparent pattern, which confirms that the model's errors are random and that there is no systematic bias in the predictions. This randomness in residuals is a desirable property, indicating that the model has effectively captured the underlying data patterns without overfitting or underfitting. The residuals plot complements the findings from Fig.5(a), providing additional evidence of the model's reliability.

Figure 7 clearly demonstrates that the Gaussian Process Regression (GPR) model outperforms the other models in terms of predictive accuracy, achieving the lowest validation RMSE. In contrast, the decision tree model exhibits the highest error, indicating limited generalization capability.

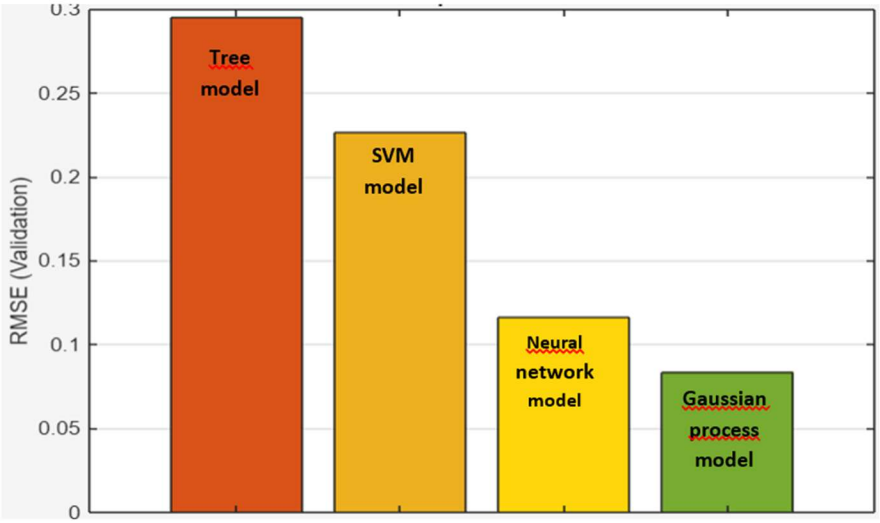


Fig.7. Compare results: RMSE (validation).

Conclusion

This study demonstrates the effectiveness of Gaussian Process Regression (GPR) as a robust and interpretable probabilistic machine learning method for predicting permeability damage resulting from sulfate scale deposition. The GPR model exhibited outstanding predictive performance, achieving a coefficient of determination $\left(R^2\right)$ of 0.99, along with low error metrics $\left(\text{RMSE} = 0.0839;\ \text{MAE} = 0.0529;\ \text{MAPE} = 0.4\%\right)$. It outperformed alternative machine learning approaches, including decision trees, support vector machines $\left(\text{SVMs}\right)$, and artificial neural networks $\left(\text{ANNs}\right)$, while also offering uncertainty quantification, a key advantage for risk-informed operational decision-making. The findings suggest that the Optimizable GPR model is the most effective approach for modeling permeability impairment due to barium sulfate scaling. While previous studies have applied models such as ANNs and SVMs to similar problems, the results of this investigation underscore the superior accuracy, generalization capability, and interpretability of GPR, particularly when enhanced with appropriately selected kernel functions. Unlike traditional deterministic models, the probabilistic nature of GPR provides confidence intervals and predictive distributions, offering valuable insights for uncertainty-aware forecasting in scale-prone reservoir environments.

Acknowledgment

This study was funded by the Directorate General for Scientific Research and Technological Development (DGRSDT) of the Algerian Ministry of Higher Education and Scientific Research, as part of the

National Research Project (PNR) titled "Development of decision-making tools for managing and improving production from petroleum fields".

Nomenclature

$\left[Ca^{2+}\right]$ – calcium ion concentration (ppm).

$\left[Ba^{2+}\right]$ – barium ion concentration (ppm).

$f(x)$ – function value at input x .

IAP – Ion Activity Product.

K_d – damaged permeability (after scaling).

K_i – initial permeability (before scaling).

K_{sp} – solubility product of the mineral.

$K(X,X)$ – covariance matrix of training inputs.

$K\left(X,x^*\right)$ – cross-covariance matrix between training and test inputs.

$K\left(x^*,x^*\right)$ – covariance at test point.

$k(x,x')$ – covariance function (kernel) between x and x' .

MAE – Mean Absolute Error.

MAPE – Mean Absolute Percentage Error.

$m(x)$ – mean function of the Gaussian Process at input x .

PV – Pore Volume.

Q – flow rate (cm^3 / min).

R^2 – coefficient of determination.

RMSE – Root Mean Square Error.

SI – Saturation Index.

$\left[SO_4^{2-}\right]$ – sulfate ion concentration (ppm).

$\left[Sr^{2+}\right]$ – strontium ion concentration (ppm).

T – temperature $\left(°C\right)$.

X – training input matrix.

x^* – new input (test point).

y – output vector (training labels).

Δp – differential pressure $\left(psi\right)$.

$\mu\left(x^*\right)$ – predictive mean of the GP at test input x^* .

σ_n^2 – noise variance.

$\sigma^2\left(x^*\right)$ – predictive variance of the GP at x^* .

References

[1] Crabtree M., Eslinger D., Fletcher P., Miller M., Johnson A. and King G. (1999): *Fighting scale-removal and prevention.*– Oilfield Review, vol.11, No.3, pp.30-45.

[2] Mackay E.J., Collins I.R., Jordan M.M. and Feasey N. (2003): *PWRI: Scale formation risk assessment and management.*– Proc. SPE International Symposium on Oilfield Scale, Aberdeen (UK), SPE-80385-MS, pp.1-18, https://doi.org/10.2118/80385-MS.

[3] Moghadasi J., Jamialahmadi M., Müller-Steinhagen H., Sharif A., Ghalambor A., Izadpanah M.R. and Motaie E. (2003): *Scale formation in Iranian oil reservoir and production equipment during water injection.*– Proc. SPE International Symposium on Oilfield Scale, Aberdeen (UK), SPE-80406-MS, pp.1-14, https://doi.org/10.2118/80406-MS.

[4] Zhang P. (2020): *Review of synthesis and evaluation of inhibitor nanomaterials for oilfield mineral scale control.*– Frontiers in Chemistry, vol.8, No.576055, pp.1-14, https://doi.org/10.3389/fchem.2020.576055.

[5] Mitchell R.W., Grist D.M. and Boyle M.J. (1980): *Chemical treatments associated with North Sea projects.*– Journal of Petroleum Technology, vol.32, No.5, pp.904-912, https://doi.org/10.2118/7880-PA.

[6] Mace R.E., Nicot J.P., Chowdhury A.H., Dutton A.R. and Kalaswad S. (2006): *Please Pass the Salt: Using Oil Fields for the Disposal of Concentrate From Desalination Plants.*– Texas Water Development Board Report, No.366, pp.1-198, https://www.twdb.texas.gov/publications/reports/numbered_reports/doc/R366.pdf.

[7] Graham G.M. and Frigo D.M. (2022): *Inorganic Mineral Scale Mitigation.*– Flow Assurance, Gulf Professional Publishing, pp.287-442, https://doi.org/10.1016/B978-0-12-823021-3.00008-0.

[8] Mura F. and Zaccone A. (2016): *Effects of shear flow on phase nucleation and crystallization.*– Physical Review E, vol.93, No.4, No.042803, https://doi.org/10.1103/PhysRevE.93.042803.

[9] Okhovat M.R., Hassani K., Rostami B. and Khosravi M. (2020): *Experimental studies of CO_2-brine-rock interaction effects on permeability alteration during CO_2-EOR.*– Journal of Petroleum Exploration and Production Technology, vol.10, No.6, pp.2293-2301, https://doi.org/10.1007/s13202-020-00883-8.

[10] Khodabakhshi M.J. and Bijani M. (2024): *Predicting scale deposition in oil reservoirs using machine learning optimization algorithms.*– Results in Engineering, vol.22, No.102263, https://doi.org/10.1016/j.rineng.2024.102263.

[11] Larestani A., Mousavi S.P., Hadavimoghaddam F. and Hemmati-Sarapardeh A. (2022): *Predicting formation damage of oil fields due to mineral scaling during water-flooding operations: gradient boosting decision tree and cascade-forward back-propagation network.*– Journal of Petroleum Science and Engineering, vol.208, No.109315, https://doi.org/10.1016/j.petrol.2021.109315.

[12] Williams C.K. and Rasmussen C.E. (2006): *Gaussian Processes for Machine Learning.*– MIT Press, Cambridge, MA, vol.2, No.3, p.4, https://doi.org/10.7551/mitpress/3206.001.0001.

[13] MacKay D.J. (1998): *Introduction to Gaussian processes.*– NATO ASI Series F: Computer and Systems Sciences, vol.168, pp.133-166, http://www.inference.org.uk/mackay/gp.pdf.

[14] Su G., Yu B., Xiao Y. and Yan L. (2014): *Gaussian process machine-learning method for structural reliability analysis.*– Advances in Structural Engineering, vol.17, No.9, pp.1257-1270, https://doi.org/10.1260/1369-4332.17.9.1257.

[15] BinMerdhah A.B., Yassin A.A.M. and Muherei M.A. (2010): *Laboratory and prediction of barium sulfate scaling at high-barium formation water.*– Journal of Petroleum Science and Engineering, vol.70, No.1-2, pp.79-88, https://doi.org/10.1016/j.petrol.2009.10.001.

[16] Rostami A., Shokrollahi A., Shahbazi K. and Ghazanfari M.H. (2019): *Application of a new approach for modeling the oil field formation damage due to mineral scaling.*– Oil & Gas Science and Technology - Revue d'IFP Energies Nouvelles, vol.74, No.62, https://doi.org/10.2516/ogst/2019032.

[17] Civan F. (2023): *Reservoir Formation Damage: Fundamentals, Modeling, Assessment, and Mitigation.*– Gulf Professional Publishing, https://doi.org/10.1016/C2020-0-03823-7.

[18] Al-Malah K.I. (2023): *Machine and Deep Learning Using MATLAB: Algorithms and Tools for Scientists and Engineers.*– John Wiley & Sons, https://doi.org/10.1002/9781119904267.

[19] Wilk M.B. and Gnanadesikan R. (1968): *Probability plotting methods for the analysis of data.*– Biometrika, vol.55, No.1, pp.1-17, https://doi.org/10.1093/biomet/55.1.1.

[20] Sheskin D.J. (2003): *Handbook of Parametric and Nonparametric Statistical Procedures.*– Chapman and Hall/CRC, https://doi.org/10.1201/9781420036268.

Received: June 22, 2025
Revised: December 2, 2025