Jurnal Sains dan Teknologi Lingkungan Volume 18, Number 2, June 2026 Hal. **155-170**
p-ISSN:2085-1227 dan e-ISSN:2502-6119

ESTIMATION OF THE EMPIRICAL MODEL PARAMETERS GOVERNING LONGITUDINAL DISPERSION COEFFICIENT IN 1D STREAM FLOW USING STATISTICAL METHODS

Mohammed Mawat
University of Basrah, Basrah, Iraq
* Korespondensi: mohammed.mawat@uobasrah.edu.iq

Abstract

The objective of this study is to find the values of the empirical parameters of a functional relationship for the longitudinal dispersion coefficient (D) in 1D open channel flow, in order to have the best prediction of waste concentration and propagation in a long well mixed estuary. Linear and non-linear regression and Bayesian inference are employed for the estimation of the model parameters α, b and c, using field data. The paper compares simulated and actual dispersion coefficients and evaluates model accuracy by using statistical measures such as R-squared, Mean Absolute Error (MAE), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), and Log-Likelihood. The results prove the capability of statistical models in accurately representing the degree of dispersion in real-world channels. The OLS regression model has α of 0.1045, b of -0.26275, c of 1.38056, and R2 of 0.682. The model fits well, however there are no strong predictors. This is probably because the sample size is limited and there may be multicollinearity. The nonlinear regression analysis, based on data observations, produced optimal parameter estimates of $\alpha = 0.015914$, $b = 1.299749$, and $c = 1.361944$. The model demonstrated a coefficient of determination (R^2) of 0.83, indicating a strong agreement between predicted and observed dispersion coefficients. Bayesian Inference gave coefficient estimates with confidence intervals. The c variable is the most important, followed by alpha, while b may be statistically insignificant because its confidence interval includes 0. The model is stable, as evidenced by R_hat (Gelman–Rubin diagnostic) ≈ 1 and high ESS (Effective Sample Size) values. The differences between the real and predicted D from the Bayesian model are reasonable, indicating a good fit to the model.

Keywords: Longitudinal Dispersion, Empirical Model, Bayesian Inference, Curve Fitting, Nonlinear Regression, Open Channel Flow.

1. INTRODUCTION

One dimensional (1D) advection dispersion equation is widely used in water quality modelling in rivers (M. Mawat & A. N. Hamdan, 2023; Mawat & Hamdan, 2024a; Zeng & Huai, 2014). The longitudinal dispersion coefficient (D) is one of the important parameters in this case, which describes the dispersion of a solute in the longitudinal direction as a result of the processes of advection and diffusion (M. J. Mawat & A. N. A. Hamdan, 2023a).

The longitudinal dispersion coefficient is an important variable that must be known in predicting the fate and transport of contaminants in natural and open channels (Mawat, 2025). This coefficient is affected by a number of hydraulic variables including, depth, channel width, shear velocity, and mean flow velocity. Several analytical and statistical techniques have been derived and used to provide an accurate prediction of this coefficient (Baek & Seo, 2013), according to the type of data and flow conditions, where each has its advantages and disadvantages.

P-ISSN: 2085-1227 I E-ISSN: 2502-6119

Several mathematical equations, such as the Fischer model, are derived from theories of transport and dispersion and used in the analytical technique (Fischer, 1967), which uses numerical integration of the dispersion equation, allowing the value of D to be calculated in terms of hydraulic properties. While they perform admirably in a perfect environment, they could struggle to handle complicated or missing data in the actual world (Mawat, 2025). Empirical methods and regression (approximate) based on linear and non-linear regressions (curve fitting) are needed to provide suitable equations that can be used directly. Regression analysis, especially nonlinear regression, is an essential tool to analyze biological and other data (Dalatu et al., 2016; Motulsky & Christopoulos, 2004; Ukpaka & Agunwamba, 2023). Many researchers studied the parameter estimation of the empirical dispersion coefficient equation (Fischer, 1979; Liu, 1977; Seo & Cheong, 1998). (Baek & Seo, 2013) used the least-square iterative method, as a nonlinear model to estimate dispersion coefficient. (Zeng & Huai, 2014) presented regression analysis based on the collected data sets to study effects of some parameters on precision results both for natural rivers and artificial trapezoidal flumes. (Lee & Park, 2024) in this study, the transverse dispersion coefficient was estimated using machine learning regression methods applied to oversampled datasets. (Baek & Lee, 2023) empirically proposed a simplified expression for the transverse dispersion coefficient and the regression coefficients are determined by using the dispersion and hydraulic data sets. (Nezaratian et al., 2021) used genetic algorithm-based support vector machine to estimate transverse mixing coefficient in streams. (Zhao et al., 2021) designed an interpretable machine learning method called evolutionary symbolic regression network to predict accurate of Longitudinal Dispersion Coefficient. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) used to compare different models. If the same data is used, the model with the lowest AIC/BIC is the best.

This paper focuses on the longitudinal dispersion coefficient equation parameters estimation for the Shatt al-Arab River based on a hybrid of mathematical derivation, statistical inference, and Bayesian inference to achieve a reasonable and precise longitudinal dispersion coefficient estimation. In addition, applying indices such as R^2, RMSE, AIC, BIC, and Log-Likelihood for statistically validating models makes it convenient to compare different methodologies.

2. MATERAILS AND METHODS

2.1. THEORETICAL BACKGROUND

A number of empirical formulas can be used to predict the longitudinal dispersion coefficient for different flow conditions, particularly for rivers, streams, and ground water. Usually, the dispersion coefficient is expressed as a function of velocity, hydraulic radius, and depth of flow. Some of the empirical formulas are presented in the following Table 1. The nature of the fluid, type of the channel, and the flow regime are usually involved in the selection of the empirical formulas. The required accuracy and the availability of data dictate the selection of an empirical formula.

Table 1. Evaluating D using Popular Empirical Equations

Author	Derived equation	simple
Taylor's Dispersion Equation (Taylor, 1954)	$D_L = \frac{U^2 d^2}{210 D_m}$	• U = Mean flow velocity • d = Depth of flow • Dm = Molecular diffusivity
Elder's Equation (Elder, 1959)	$D_L = 5.93 \left(\frac{U d^2}{H}\right)$	• U = Flow velocity • d = Flow depth • H = Hydraulic radius (which can be approximated by depth in wide channels)
Fischer's Equation (Fischer, 1979)	$D_L = 0.011.U.W.\left(\frac{H}{W}\right)^{0.5}$	• U = Mean flow velocity • W = Width of the stream • H = Depth of flow
Liu's Equation (Liu, 1977)	$D_L = 0.018.\left(\frac{UHL}{gS}\right)^{0.5}$	• U = Mean flow velocity • H = Depth of flow • L = Length of reach • g = Acceleration due to gravity • S = Channel slope
Seo and Cheong's Equation (Seo & Cheong, 1998)	$D_L = 5.93 \left(\frac{UHW}{HgS}\right)^{0.5}$	• U = Mean flow velocity • H = Depth of flow • W = Width of the stream • S = Channel slope g = Acceleration due to gravity
Sukhija's Equation (1996)	$D_L = \alpha_L.U.d$	• αL = Longitudinal dispersivity (a scale-dependent factor) • U = Flow velocity • d = Characteristic length scale
Kashefipour and Falconer's Equation (Kashefipour & Falconer, 2002)	$D_L = 0.15 \left(\frac{U^2 d^2}{D_m}\right)$	• U = Flow velocity • d = Water depth • Dm = Molecular diffusivity
Wilson's Equation (1965)	$D_L = \alpha.U.d$	Where α is an empirical coefficient based on channel roughness and flow conditions

2.2 METHOD STATEMENT

Several key variables influence longitudinal dispersion in open channel flows, and all are incorporated into equation (1) (Zhao et al., 2021). This model can be used to describe solute dispersion in both natural and artificial streams. The equation highlights that the transport of contaminants through a stream is a function of flow velocity, channel geometry, and bed shear stress.

$$D = \alpha(Hu_*)\left(\frac{U}{u_*}\right)^b\left(\frac{W}{H}\right)^c \tag{1}$$

Explanation of Terms:

- D: Longitudinal dispersion coefficient.
- α: An empirical coefficient.
- H: Flow depth (mean depth of the water).
- u∗: Shear velocity, often calculated as $u_* = \sqrt{\frac{\tau_o}{\rho}}$, where τ_o is the bed shear stress and ρ is the fluid density.
- U: Mean flow velocity of the water.
- W: Flow width (or channel width).
- b and c: Empirical exponents that adjust the influence of the velocity ratio $\frac{U}{u_*}$ and the width-to-depth ratio $\frac{W}{H}$ on the dispersion coefficient.

Statistical optimization methods like nonlinear regression or least squares analysis are often used to find the empirical coefficients α, b, and c using data collected in the field or in experiments. Finding the coefficients α, b, and c in the equation generally means utilizing experimental or field data to calibrate the equation. Here are some particular ways to do it:

2.2.1. Log-Linear Regression:

Rearrange the equation into a log-linear form to make it easier to apply regression techniques:

$$\ln(D) = \ln(\alpha) + \ln(Hu_*) + b.\ln\left(\frac{U}{u_*}\right) + c.\ln\left(\frac{W}{H}\right) \tag{2}$$

Or

$$\ln(y) = \ln(\alpha) + X_1 + b.X_2 + c.X_3 \tag{3}$$

Where

$$X_1 = ln(Hu_*), \quad X_2 = ln\left(\frac{U}{u_*}\right) \text{ and } X_3 = ln\left(\frac{W}{H}\right) \tag{4}$$

Now we may consider equation (2) as a multiple linear regression problem. The dependent variable is $\ln(D)$ and the independent variables are$ln(Hu_*)$, $ln\left(\frac{U}{u_*}\right)$ and $\ln\left(\frac{W}{H}\right)$. By employing the statistical tools, such as Python's stats models, the multiple linear regressions can be done. The regression yields α, b, and c resulting in estimations of ln(α), b, and c. The initial connection between the variables was nonlinear due to the equation's multiplicative structure. This equation is intrinsically nonlinear because of the exponents b and c. By applying the logarithm to both sides of the equation, we converted the nonlinear model into a linear format. Subsequent to this change, the connection attains linearity in the logarithms of the variables, hence facilitating the use of Ordinary Least Squares (OLS) regression. Thus, while the original relationship is nonlinear, the method used to estimate the parameters involved linear regression on the log-transformed variables, which is a common approach to handling certain types of nonlinear regression problems.

2.2.2. Nonlinear Regression

In situations when the connection between the dependent and independent variables is not linear, nonlinear regression is employed for model fitting. Using an iterative numerical optimization process, the parameters were updated from physically plausible beginning assumptions until convergence was reached in accordance with predetermined tolerance conditions. Goodness-of-fit statistics and residual analysis were used to evaluate the model's performance in order to confirm that the fit was adequate and that the statistical assumptions were consistent. This method guarantees physically relevant parameter estimates while maintaining the model's original nonlinear structure. Suppose you have data points for D, U, u∗, W, and H. You want to fit the model:

$$D_i = \alpha(H_i u_{*i}) \left(\frac{U_i}{u_{*i}}\right)^b \left(\frac{W_i}{H_i}\right)^c \quad (5)$$

The nonlinear regression method will begin with initial estimates for α, b, and c and use the model with these starting values to find the expected values of $\widehat{D_i}$. Find the differences between D_i and $\widehat{D_i}$ to get the residuals. The loss function or error function is usually the sum of the squares of the residuals:

$$SSE = \sum_{i=1}^{n} (D_i - f(x_i, \alpha, b, c))^2 \quad (6)$$

The model's parameters are estimated to optimize the fit to empirical data, and statistical inference is used to assess the model's goodness-of-fit and to quantify the uncertainty of the parameter estimates. Adjust the parameters to reduce the residuals. Repeat the steps until the parameters have converged or the error function is minimized.

2.2.3. Bayesian Inference PyMC3

The Bayesian approach is useful when there is variability or uncertainty in empirical data (Salvatier et al., 2016). PyMC (formerly PyMC3) is a Python package for Bayesian statistical modeling and model fitting which focuses on advanced Markov chain Monte Carlo (MCMC) and variational inference (VI) algorithms. The models and algorithms are extensible, allowing users to write their own model classes, samplers and other functions for very specific models and tasks. The posterior probability distribution for each parameter is derived by combining prior information with current data using Bayes' theorem. This allows for obtaining an estimated value and knowing the degree of confidence in it (through credibility intervals), thereby enhancing the model's reliability in environmental and hydrological applications. This method calculates a posterior distribution for each parameter instead of just one value, depending on the data plus prior information. Gives probabilities instead of a single value (e.g. $\alpha = 0.75 \pm 0.05$), can handle noisy or sparse data and very useful in physical or engineering models. We model the dispersion coefficient D using a model in the following form:

$$D_i = \alpha(H_i u_{*i}) \left(\frac{U_i}{u_{*i}}\right)^b \left(\frac{W_i}{H_i}\right)^c + \varepsilon_i \quad (7)$$

$\varepsilon_i \sim N(0, \sigma^2)$ is the random error is assumed to follow a normal distribution. We assume that the error $\varepsilon_i$is normal

$D_i \sim N(\mu_i, \sigma^2)$

2.3. STUDY AREA AND DATA SETS

The Tigris and Euphrates Rivers converge at the Al-Qurnah area in southern Iraq, creating the Shatt Al-Arab River(M. J. Mawat & A. N. A. Hamdan, 2023b). The Shatt Al-Arab River extends about 192 kilometers, running southeast through Basrah City before flowing into the Arabian Gulf (Abd-El-Mooty et al., 2016; Hamdan et al., 2016). The seven creeks of the province—Jubyla, Muftya, Robat, Khandek, Ashar, Khora, and Saraji—are linked to the Shatt Al-Arab River and are affected by tidal phenomena (Mawat & Hamdan, 2024a). The width of the Shatt Al Arab River varies along its course, measuring 250-300 meters near the confluence of the Euphrates and Tigris, 600 meters around Basrah City Center, and 2000 meters at the estuary (Hamdan, 2016). For the remaining 95 kilometers of its path, adjacent to Umm Al Rasas Island, the river delineates the boundary between Iraq and Iran (Abdullah, 2016; Kadhim, 2018). Besides transportation, the Shatt Al Arab River plays a crucial role in supplying water for home purposes, agriculture, and industrial processes (Hamdan et al., 2018). Numerous tributaries of the Shatt Al-Arab River, such as Al-Sweeb, Ezz, Garmat Ali, Karkheh, and Karun

Rivers, contributed to its flow along its length. The contributions from the Tigris and Euphrates Rivers, along with their low flow rates, have diminished due to the policies of neighboring governments, leading to a substantial increase in Total Dissolved Solids (TDS) in the Shatt Al Arab River, attributed to salinity intrusion from the Arabian Gulf (Hamdan et al., 2020; Mohamed[1] & Abood, 2017; Yaseen et al., 2016). The Shatt Al-Arab River is of considerable significance in a region that is characterized by aridity and a hot, humid environment, as it enables agricultural productivity (Mawat & Hamdan, 2024b). The water system of the Shatt Al Arab River is now under increasing strain, both in terms of water quantity and quality (Abdullah, 2016).

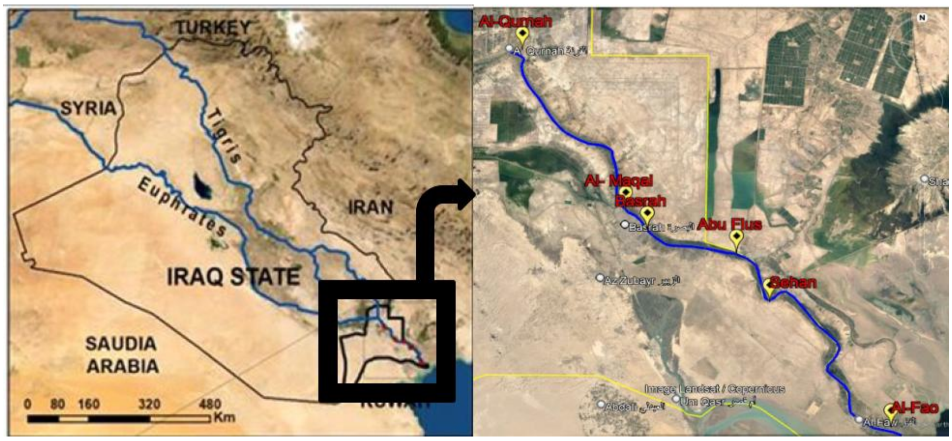

Figure 1. Shatt Al Arab River

The daily discharge measurements collected from May to November 2020 were utilized to gather the required data. Velocity and the geometric data (depth and width) were obtained from hydrodynamic model output.

3. RESULTS

3.1 HYDRAULIC RESULTS

Table 2 presents the values of depth, width, area, velocity, shear velocity, and the estimated dispersion coefficient from available pairs of data. As shown in the table, the estimated dispersion coefficient is calculated at five stations along the river, where the positions are recorded from the upstream end. The hydraulic parameters are extracted from the HEC-RAS model. The value of D is affected by these parameters and ranges between 26.25 and 55.95 m^2/s; this range matches the results of previous works (M. J. Mawat & A. N. A. Hamdan, 2023a). The table also shows that the highest value of D corresponds to the lowest value of R, and vice versa

Table 2. The hydraulic variable of cross-sections of Shatt Al-Arab River at the multi stations

Station	*H* (m)	*W* (m)	*A* (m^2)	*R* (m)	*U* (m/s)	u^* (m/s)	*D* (m^2/s)
3000	5.01	436.15	1294.99	2.96	0.500	0.0170	55.95
7200	7.94	222.57	1054.96	4.7	0.533	0.0215	26.25
11300	9.18	331.45	1239.55	3.72	0.483	0.0191	37.26
17240	8.61	501.67	1662.82	3.31	0.533	0.0180	44.67
26670	10.29	478.6	2076.36	4.33	0.522	0.0206	29.79

3.2 OLS REGRESSION STATISTICAL RESULTS

The regression analysis was conducted using OLS to examine the relationship between the dispersion coefficient D and three predictors: X1, X2, and X3 and the results are listed in table2 and 3. The model's R^2 value was 0.682, which means that the predictors explained around 68.2% of the variation in D. The adjusted R^2 value was 0.527, which shows that the small sample size (n = 5) impact the model. The total F-test for model significance was not statistically significant (F = 2.486, p = 0.426), suggesting that there is insufficient evidence to infer that all variables jointly explain the dependent variable. The estimated coefficients were as follows: constant = -2.2585 (p = 0.790), x1 = -0.2628 (p = 0.673), x2 = 1.3806 (p = 0.678), and x3 = 0.2192 (p = 0.751). None of these coefficients were statistically significant at the 5% level. Diagnostic statistics showed that the residuals appeared approximately normally distributed (Jarque-Bera p = 0.673), there was no evidence of autocorrelation (Durbin–Watson = 2.667), and the condition number was 416, suggesting potential multicollinearity between predictors. Overall, the model provides a moderate fit but with no significant predictors, likely due to the small sample size and possible multicollinearity, and it is recommended to use a larger dataset for more reliable results. The estimated scaling coefficient (α\alpha) was 0.1045, b equal to -0.26275 and the c equal to 1.38056. Standard Errors assume that the covariance matrix of the errors is correctly specified. Omni_normtest is not valid with less than 8 observations; 5 samples were given.

Table 3. OLS Regression Statistical Summary

Prob(F-statistic)	0.426	R-squared	0.682
Log-Likelihood	4.7462	Adj. R-squared	0.527
AIC	-1.492	F-statistic	2.486
Df Residuals	1	BIC	-3.055

Df Model		3		Covariance Type		nonrobust

Variable	Coef	Std Err	t	P>|t|	[0.025	0.975]
const	-2.2585	6.611	-0.342	0.79	-86.259	81.742
x1	-0.2628	0.466	-0.564	0.673	-6.186	5.66
x2	1.3806	2.496	0.553	0.678	-8.306	33.097
x3	0.2192	0.532	0.412	0.751	-6.534	6.973

Omnibus	nan	Durbin-Watson	2.667
Prob(Omnibus)	nan	Jarque-Bera(JB)	0.791
Skew	0.957	Prob(JB)	0.673
Kurtosis	2.637	Cond. No.	416

Table 4. Model Variables and Coefficients

Symbol	Description	Value / Expression
y	Dispersion coefficient	D (m^2/s)
x_1	ln(H × u*)	Log of product of depth and shear velocity
x_2	ln(U / u*)	Log of velocity ratio
x_3	ln(W / H)	Log of aspect ratio
α	Scaling coefficient	0.1045
b	Exponent on U/u*	-0.26275
c	Exponent on W/H	1.38056
R^2	Coefficient of determination	0.682

3.3 OPTIMIZED NONLINEAR MODEL RESULTS

The nonlinear regression analysis, conducted with data observations, yielded optimized parameter estimates of α = 0.015914, b = 1.299749, and c = 1.361944. With a R^2 of 0.83, the model showed that the predicted and observed dispersion coefficients were highly concordant. A moderate level of inaccuracy was indicated by the predictions, with an RMSE of 10.0897 and an MAE of 9.7943. The model's goodness-of-fit statistics included a log-likelihood of −18.6522, an Akaike Information Criterion (AIC) value of 43.3045, and a Bayesian Information Criterion (BIC) value of 42.1328, which together suggest a reasonably well-fitting model for the given dataset as shown in table 5

Table 5. Statistical of Optimized Nonlinear Model

Item	Value
Method	Nonlinear Regression
Optimized α	0.015914
Optimized b	1.299749
Optimized c	1.361944
R^2	0.83
RMSE	10.0897
MAE	9.7943
Log-Likelihood	-18.6522
AIC	43.3045
BIC	42.1328

This plot compares the measured and predicted values of the dispersion coefficient D across data points are plotted in Figure 2. The predicted D obtained from the nonlinear regression model. In general, the predicted curve seems to generally match the trend of the measured data and it hits both the peaks and the troughs though it does vary a bit.

The close alignment in peaks and troughs reflects the model's ability to replicate observed variations, consistent with the $R^2 = 0.83$ value obtained earlier, indicating a strong fit. The differences visible in certain points correspond to the RMSE and MAE values reported in the statistical summary.

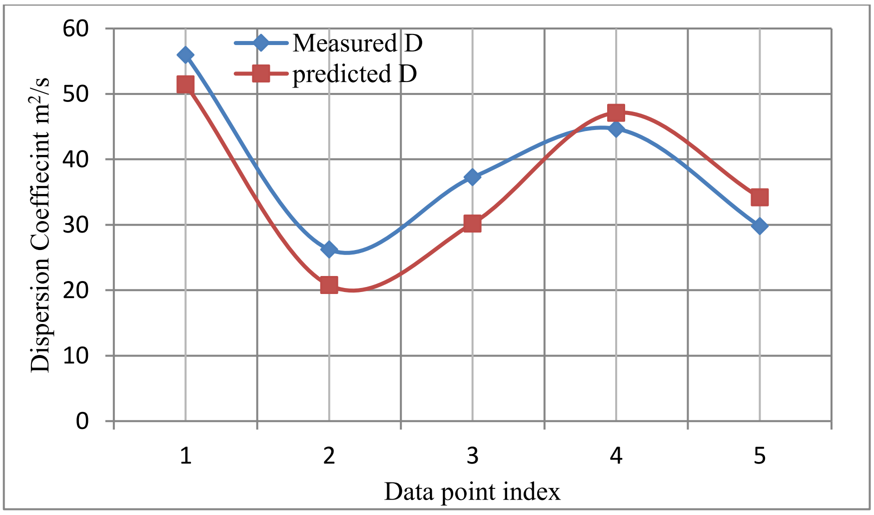

Figure 2. The measured and predicted (nonlinear regression model) values of the dispersion coefficient D

3.4 BAYESIAN INFERENCE PyMC3

The results of this method give the probability distribution of each coefficient Mean, standard deviation, 95% Credible Interval (CI), it defines the interval containing 95% of the posterior probability mass, confidence and uncertainty indices (R-hat, ESS). R-hat (Gelman–Rubin diagnostic) evaluates convergence of the (MCMC) sampling process. It compares within-chain and between-chain variance. Values close to 1.00 (typically < 1.01) indicate that the chains have converged to the target posterior distribution. ESS (Effective Sample Size) measures the number of independent samples equivalent to the correlated MCMC draws. Because MCMC samples are autocorrelated, the effective sample size is smaller than the total number of draws. Larger ESS values indicate more reliable posterior estimation. This method is more powerful than curve_fit because it takes uncertainty into account.

The outputs of this method are shown in Figure 3. Left of the figure, the top plots for each parameter represent its probability distribution (posterior). These plots show multiple copies (Markov chains), and the more identical they are, the more confident the results are. Right of the figure, trace plots showing the evolution of the value of each parameter across samples. A good random distribution with no clustering or repeating pattern indicates good convergence of the samples.

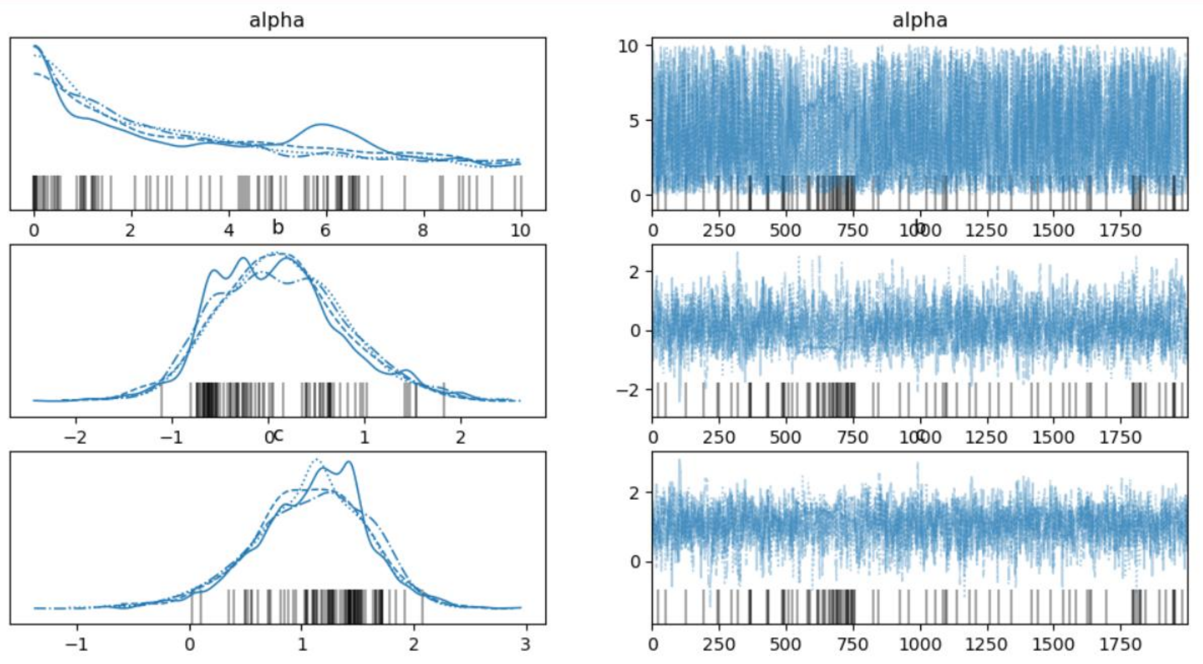

Figure 3.

The trace summary results are illustrated in table 6. α has a positive mean (≈ 3.9) but a large standard deviation, indicating high uncertainty. b is close to zero with a high probability of being insignificant (because the confidence interval contains zero). c appears to be more persistent with a positive value (≈ 1.07) with a relatively narrow confidence interval, indicating that it is more significant. sigma is the deviation of the model error, which is normal to have at this magnitude (≈ 13.7). r_hat (convergence diagnostics): All values ≈ 1, which is excellent. It indicates that the chains have converged and are not stuck in different regions of the distribution. Interpretation: RMSE = 4.7344, the radical deviation between the predicted and real values is considered acceptable if we know that the real values of D range from ~26 to ~56. MAE = 3.8720, It means that the mean absolute error is about 5 m²/s, which is almost similar to the RMSE indicating that there are no very large errors. R^2 = 0.815, The model explains about 81.5% of the variance in the true values, which is very good and indicates the quality of the prediction. Bayesian Inference gave coefficient estimates with confidence intervals. The c variable is the most important, then alpha, while b may be statistically insignificant (because its interval contains 0). The model is stable, as evidenced by r_hat ≈ 1 and high ESS values.

Table 6. Trace Summary

Model	PyMC	R-squared	0.815
Method	Bayesian Inference	RMSE	4.7344
Date	Fri,13 June 2025	MAE	3.8720
Time	14:36:22	Log-Likelhood	-15.4010
No. Observations	5	AIC	36.8020
		BIC	35.6303

Variable	Mean value	Sd	Hdi-3%	Hdi-97%	r-hat
α	3.8966	2.8994	0.0011	8.8757	1.0042
b	0.1261	0.6688	-1.0430	1.4438	1.0082
c	1.0741	0.5144	0.0728	2.0034	1.0023
sigma	13.7695	0.532	6.1066	21.1704	1.0143

The differences between the measured D (within data collected) and predicted D (calculated from the Bayesian model) are reasonable, indicating a good fit to the model as indicated in Figure 4. The maximum value of the difference is at state 1 (around 6.9), and the rest are very close. Accuracy can be further improved by adding more data or choosing more specialized distributions. The Bayesian model adheres to a general pattern of real values, albeit with some deviations. Differences may be due to variability in the data, the small number of models, and asymmetric distribution of some coefficients (as seen from HDI).

Figure 4. The measured and predicted (Bayesian model) values of the dispersion coefficient D

4. CONCLUSION

This paper estimates the parameters of an empirical model for longitudinal dispersion coefficient (D) in one-dimensional open channel flow. The parameters α, b, and c are fitted statistically by using linear and non-linear regression and Bayesian methods. The comparison between the predicted and observed dispersion coefficient was carried out and the model performance was evaluated by using the statistical indexes, namely, R-squared, RMSE, MAE, AIC, BIC and Log-Likelihood. The estimated dispersion coefficient was computed at five stations on the Shatt Al-arab river. Hydraulic parameters are obtained from HEC-RAS model. These parameters are used for calculating the D, which varied between 26.25 and 55.95 m2/s. The estimated α values by OLS Regression model was 0.1045, b -0.26275 and c 1.38056, with R2 0.682. The model was moderate but insignificant predictors can be explained by the fewer data points and multicollinearity. Nonlinear regression analysis is performed on data observations and estimated α, b and c values are 0.015914, 1.299749 and 1.361944 respectively. It can be seen that the R^2 for the model is 0.83, meaning the agreement between the observed dispersion coefficients and those predicted by the model is strong. Bayesian Inference provides us with the estimated values of the coefficients along with their confidence intervals. The most significant variable is c, followed by alpha, and b may be statistically insignificant (as the interval covers 0). The model is stable, indicated by R_hat ≈ 1 and high ESS. The differences between actual and predicted D from the Bayesian model are within reasonable bounds, suggesting a decent fit of the model.

REFERENCES

Abd-El-Mooty, M., Kansoh, R., & Abdulhadi, A. (2016). Challenges of water resources in Iraq. Hydrology Current Research, 7(4), 1-8.

Abdullah, A. D. (2016). Modelling approaches to understand salinity variations in a highly dynamic tidal river. The case of the Shatt Al-Arab River (Delft, the Netherlands: Delft University of Technology and of the Academic Board of the UNESCO-IHE) PhD thesis.

Baek, K. O., & Lee, D. Y. (2023). Development of simple formula for transverse dispersion coefficient in meandering rivers. Water, 15(17), 3120.

Baek, K. O., & Seo, I. W. (2013). Empirical equation for transverse dispersion coefficient based on theoretical background in river bends. Environmental fluid mechanics, 13(5), 465-477.

Dalatu, P. I., Fitrianto, A., & Mustapha, A. (2016). A comparative study of linear and nonlinear regression models for outlier detection. International conference on soft computing and data mining,

Elder, J. (1959). The dispersion of marked fluid in turbulent shear flow. Journal of fluid mechanics, 5(4), 544-560.

Fischer, H. B. (1967). The mechanics of dispersion in natural streams. Journal of the Hydraulics division, 93(6), 187-216. http://dx.doi.org/10.1061/JYCEAJ.0001706

Fischer, H. B. (1979). Mixing in inland and coastal waters. Academic Press.

Hamdan, A. (2016). Simulation of salinity intrusion from Arabian gulf to Shatt Al-Arab River. Basrah Journal for Engineering Sciences, 16(1), 28-32.

Hamdan, A., Dawood, A., & Naeem, D. (2018). Assessment study of water quality index (WQI) for Shatt Al-arab River and its branches, Iraq. MATEC web of conferences,

Hamdan, A. N. A., Al-Mahdi, A. A. J., & Mahmood, A. B. (2020). Modeling the effect of sea water intrusion into Shatt Al-Arab River (Iraq). Journal of University of Babylon for Engineering Sciences, 28, 210-224.

Hamdan, A. N. A., Dawood, A. S. J. E., & Journal, T. (2016). Neural network modelling of TDS concentrations in Shatt Al-Arab River water. 34(2 Part A), 334-345.

Kadhim, A. A. (2018). Measuring and modeling the effects of sea level rise on near-coastal riverine regions: A geospatial comparison of the Shatt Al-Arab River in southern iraq with the Mississippi River Delta in southern Louisiana, USA. Michigan State University.

Kashefipour, S. M., & Falconer, R. A. (2002). Longitudinal dispersion coefficients in natural channels. Water Research, 36(6), 1596-1608.

Lee, S., & Park, I. (2024). Application of Oversampling Techniques for Enhanced Transverse Dispersion Coefficient Estimation Performance Using Machine Learning Regression. Water, 16(10), 1359.

Liu, H. (1977). Predicting dispersion coefficient of streams. Journal of the Environmental Engineering Division, 103(1), 59-69.

Mawat, M. (2025). Comparative analysis of analytical and empirical methods for estimating the longitudinal dispersion coefficient in open-channel flows. J. Civ. Hydraul. Eng, 3(2), 91-99.

Mawat, M., & Hamdan, A. N. (2023). Review of Mathematical Surface Water's Hydrodynamic/Water Quality Models with Their Application on the Shatt Al Arab River Southern Iraq. Euro J Eng Sci Tech, 6(1), 30-49. http://DOI:10.33422/ejest.v6i1.1057

Mawat, M. J., & Hamdan, A. N. A. (2023a). Integration of numerical models to simulate 2D hydrodynamic/water quality model of contaminant concentration in Shatt Al-Arab River with WRDB calibration tools. Open Engineering, 13(1), 20220416. http://dx.doi.org/10.1515/eng-2022-0416

Mawat, M. J., & Hamdan, A. N. A. (2023b). Simulation of 2D depth averaged saint Venant model of Shatt Al Arab river south of Iraq. International Journal of Design and Nature and Ecodynamics, 18(3), 583-592. http://dx.doi.org/10.18280/ijdne.180310

Mawat, M. J., & Hamdan, A. N. A. (2024a). 2D Hydrodynamic and Eutrophication Modeling in the Shatt Al-Arab River, Basrah, Iraq. Environmental & Earth Sciences Research Journal, 11(1). http://dx.doi.org/10.18280/eesrj.110101

Mawat, M. J., & Hamdan, A. N. A. (2024b). Assessing the Impact of Proposed Regulator's Construction on Water Quality of Shatt Al-Arab River-Iraq. Journal of Engineering, 30(8), 101-119. http://dx.doi.org/10.31026/j.eng.2024.08.07

Mohamed[1], A.-R. M., & Abood, A. N. (2017). Ecological health assessment of the Shatt Al-Arab river, Iraq.

Motulsky, H., & Christopoulos, A. (2004). Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting. Oxford University Press.

Nezaratian, H., Zahiri, J., Peykani, M. F., Haghiabi, A., & Parsaie, A. (2021). A genetic algorithm-based support vector machine to estimate the transverse mixing coefficient in streams. Water Quality Research Journal, 56(3), 127-142.

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. PeerJ Computer Science, 2, e55.

Seo, I. W., & Cheong, T. S. (1998). Predicting longitudinal dispersion coefficient in natural streams. Journal of hydraulic Engineering, 124(1), 25-32.

Taylor, G. I. (1954). The dispersion of matter in turbulent flow through a pipe. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 223(1155), 446-468.

Ukpaka, C., & Agunwamba, J. C. (2023). Performance of equations for the longitudinal dispersion coefficient: a case study in the Orashi River. Water Practice & Technology, 18(11), 2775-2784.

Yaseen, B. R., Al Asaady, K. A., Kazem, A. A., & Chaichan, M. T. (2016). Environmental impacts of salt tide in Shatt al-Arab-Basra/Iraq. IOSR Journal of Environmental Science, Toxicology and Food Technology, 10(1-2), 35-43.

Zeng, Y., & Huai, W. (2014). Estimation of longitudinal dispersion coefficient in rivers. Journal of hydro-environment research, 8(1), 2-8. http://dx.doi.org/10.1016/j.jher.2013.02.005

Zhao, Y., Liu, Z., Zhang, P., Li, S. Z., & Galindo-Torres, S. A. (2021). An interpretable prediction model for longitudinal dispersion coefficient in natural streams based on evolutionary symbolic regression network. CoRR.