International Journal of Applied Mathematics and Numerical Research

Using Vector Support Mechanism to Model and Forecast Water Flows into the Mosul Dam Lake

Dhafer M Jabur Allela
Department of Community Health Techniques, Mosul medical Technical Institute, Northern Technical University, Mosul, Iraq

* Corresponding Author: **Dhafer M Jabur Allela**

Article Info

Impact Factor (RSIF): 8.62
ISSN (Online): 3107-7110
Volume: 02
Issue: 03
Received: 23-02-2026
Accepted: 25-03-2026
Published: 27-04-2026
Page No: 28-32

Abstract
Modelling water flowing series of the Mosul Dam Lake for the year 2023-2024(730 days) using support vector mechanism for forecasting in order to train the SVM and determine its most accurate parameters, the first 657 days (0.90 of the Time series data) were utilized for both validation and training. The test set, which was used to determine the prediction potential of SVM, consisted of the remaining 73 days (0.10 of the series). There are numerous varieties of SVM models., These models depending on the different parameters (□, c,□) which are experimentally selected for a limited number of values, the appropriate range for C=1,10,100,1000, In order to obtain a more comprehensive analysis, we have expanded the maximum range to 10,000. the appropriate range for γ =0.0001 TO 100, The best model is C=100, γ = 100, which contains 92 support vectors that model has the fewest support vectors and the lowest training error.

DOI: https://doi.org/10.54660/IJAMNR.2026.2.3.28-32

Keywords: Time Series, Support Vector Machine

1. Introduction

Mosul Dam Lake is a vital water resource in Iraq, and finding a predictive model for water inflows to the lake is important for managing stored water and controlling floods. This study proposes using the Support Vector Machine (SVM) mechanism for modeling and predicting water inflows to Mosul Dam Lake. The SVM model will be applied to data of water inflows for two previous years. The performance of the SVM model will be evaluated using metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The results of this study can help improve the accuracy of predicting water inflows to Mosul Dam Lake, enabling better decision-making for water resource management.

2. Time Series

Time series are considered one of the most used statistical methods in daily life fields where it is desired to analyze their phenomena for a certain period of time and then predict them in the future. Among the most important fields of time series usage are in commerce (sequential weekly or monthly sales data), in economics (monthly unemployment and employment statistics), in sociology (crime statistics), in nature and engineering (rain statistics and chemical batch operation data), and in medicine and public health (epidemic disease statistics and vaccination statistics). [1, 4]

3. Support Vector Machines (SVM)

Support Vector Machines are based on Statistical Learning Theory. SVM is considered a supervised learning algorithm, aiming to find the best hyperplane that classifies the data as accurately as possible "[11].
Support Vector Machines (SVM) were employed for a variety of tasks, particularly classification and liner and nonlinear

regression issues. In particular, time series forecasting, stock price prediction, and weather forecasting "[2, 10].

" Support Vector Machines operate in time series using a similar technique to classification: A maximum-margin hyperplane is used to split the data after it has been mapped to a higher-dimensional space. the main goal is to find a function that can strictly predict future values"[3].

Examine a training set of n data points □xi, yi □n with input data xi□□p, where p represents the total number of patterns in the data, and the output is yi□□.building SVM to converge a function often entails executing a linear regression in the feature space after nonlinearly mapping the data x into a high-dimensional feature space.

Support Vector Machines takes the following form to approximate the function:

$$y(x) = w^T\varphi(X) + b \tag{1}$$

When the input spacex is nonlinearly translated to a high-dimensional feature space denoted by φ(x). [7] [15].

The following function is minimized in order to estimate the coefficients w and b:

$$Min\left(\tfrac{1}{2}w^t w\right) \tag{2}$$

Subject to the following constraints:

$$\begin{aligned} &yi - wt\varphi(xi) - b \le \varepsilon \\ &wt\ \varphi(xi) + b - yi \le \varepsilon \end{aligned} \tag{3}$$

Minimize

$$R(w,\xi,\xi^*) = \tfrac{1}{2}W^T W + C\sum_{i=1}^{N}(\xi_i,\xi_i^*) \tag{4}$$

Subject to the constraints:

$$\begin{aligned} &yi - wtxi - b \le \varepsilon + \xi i \\ &wt\ xi + b - yi \le \varepsilon + \xi i \\ &\xi i,\ \xi i* \ge 0 \end{aligned} \tag{5}$$

Equation (4)'s initial term (1/2) In order to determine the tradeoff between the regularized term and the empirical error, wTw stands for the weights vector norm, yi for the desired value, and C for the regularized constant. The approximation accuracy applied to the training data points is represented by ε, which is also known as the tube size of SVM. Here, we introduce the slack variables ξ and ξ*. The decision function provided by Eq. 4 adopts the following explicit form by utilizing Lagrange multipliers and taking advantage of the optimality constraints.:

$$\begin{aligned} L = \tfrac{1}{2}W^T W + C\sum_{i=1}^{N}(\xi_i,\xi_i^*) - \sum_{i=1}^{N} a_i(\varepsilon + \xi_i - y_i + w^T x_i + b) \\ - \sum_{i=1}^{N} a^*_i(\varepsilon + \xi^*_i - y_i + w^T x_i + b) - \sum_{i=1}^{N}(\eta_i\xi_i + \eta^*_i\xi^*_i) \end{aligned} \tag{6}$$

With the constraints:

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} a^*_i \qquad 0 \le a_i^* \le C, \quad i = 1,2,\dots,n$$
$$0 \le a_i \le C, \quad i = 1,2,\dots,n \tag{7}$$

This can also be expressed in the form:

$$y(x) = \sum_{i=1}^{N}(a_i - a^*_i).(\varphi(x_i)) + b \tag{8}$$

in generally:

$$y(x) = \sum_{i=1}^{N}(a_i - a_i^*).K(x_i, x) + b \tag{9}$$

where K□ x, xi □ represents kernel function[7, 8].

Kernel value is equal to the interior product of two vectors, Xi and Xj, in the area of feature φ(xi) and φ(xj), that is,

$$K(x,x_i) = \varphi(x_i)\times\varphi(x_i) \tag{10}$$

The kernel function examples include:

linear: $K(x, x_i) = x_i^T x_j$

sigmoid: $K(x, x_i) = \tanh(\gamma x_i^T x_j + r)$

polynomial: $K(x, x_i) = (\gamma x_i^T x_j + r)^d, \gamma > 0$

Gaussian function: $K(x, x_i) = exp\left(-\frac{\|x-x_i\|^2}{2\gamma}\right)$, $\gamma > 0$

Exponential RBF: $K(x, x_i) = exp\left(-\frac{\|x-x_i\|}{2\gamma}\right)$

There are different types of SVM models, These models depending on the different parameters (□,c,□) which are experimentally selected for a limited number of values, the appropriate range for C =1,10,100,1000 [6] [12]. In order to obtain a more comprehensive analysis, we have expanded the maximum range to 10,000. the appropriate range for γ =0.0001 TO 100 [9, 14].

(γ, r, d)are represent kernel parameters; kernel parameter should be chosen carefully, It is responsible for determining the structure of the high-dimensional feature space φ(x) and controls the complexity of the final solution. [13].

4. Data Fitting for SVM

The time series Included 730 daily incomes to the lake of al-Mosul dam from 2023-2024, Figure 1. Show the Autocorrelation Function of the series it can be observed that it stable, Ten percent of the series point were used for SVM testing, and the remaining ninety percent were used to training. Addition, ten-fold cross-validation was carried out.

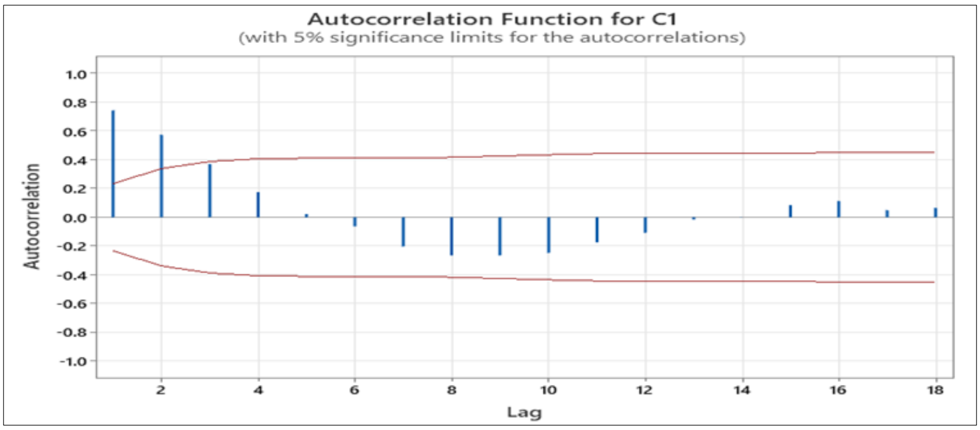

Fig 1: Observed ACF of the water flow Series

Utilizing SVM forecasting a time series {x1, x2,., xn},Time series should be transferred into an autocorrelated dataset. This implies that the correlated variables of the input should be the prior value $\overline{xt} = \{xt{-}1, xt{-}2, \ldots, xt{-}p\}$ if {xt }is the forecasting target value, y={ xt } p is an embedding dimension in this case.

We took into account how the predicting horizon and the embedding dimension affected SVM performance. Practical issues have to be taken into consideration while choosing the embedding dimension. We were able to collect time series data that was appropriate for SVM learning by transferring the data in this manner. The RMSE was used to assess the prediction performance. We had to first figure out the ideal embedding dimension p because we were unsure of it. We conducted our initial trials with p= {2,3,4,5,6,7,8,9}, with all other circumstances held constant. We determined that five-day-lag daily indices were best suited for predicting the following day's index based on these studies, which showed that the RMSE was lowest when p=5. The time series observation is divided in half. In the first, 657 observations (90 percent of the series) were employed for both validation and training, helping to determine the best parameters for the SVM. The final 73 data observations (10% of the series) made up the test set, which was used to assess SVM's prediction ability. The number of support vectors and the generalization error regard to C and ε were investigated because there is no systematic way to choose the free parameters of SVMs. The validation set was used to determine the kernel parameters C and γ. The number of support vectors and RMSE in relation to the free parameters were examined. The kernel function of the SVMs in this study was the Gaussian function. The following table shows several models of SVM Different depending on their parameters, these models were compared using two basic criteria: the training error and the number of support vectors. The parameter values used were as follows by using (Minitab program) :

C: 1, 10, 100, 1000, 10000

γ:0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 10, 30, 50, 70, 100

Table 1: SVM model for water flowing series of the Mosul Dam Lake

model	C	γ	Number of Support Vectors	Training Error	model	C	γ	Number of Support Vectors	Training Error
M1	1	0.0001	300	845.365791	M8	1	0.5	245	165.288704
M2		0.0005	294	657.214782	M9		1	215	142.338214
M3		0.001	291	561.284103	M10		10	154	77.204198
M4		0.005	287	432.221867	M11		30	127	53.884059
M5		0.01	285	414.100287	M12		50	121	44.387946
M6		0.05	282	384.510237	M13		70	112	37.199531
M7		0.1	279	340.881046	M14		100	108	19.349981
M15	10	0.0001	286	577.102273	M22	10	0.5	234	153.883619
M16		0.0005	281	480.748102	M23		1	228	150.441028
M17		0.001	279	458.210996	M24		10	140	96.351098
M18		0.005	276	460.744401	M25		30	111	63.882493
M19		0.01	269	366.308591	M26		50	107	38.559304
M20		0.05	265	320.407716	M27		70	102	21.260336
M21		0.1	255	277.518409	M28		100	97	17.511872
M29	100	0.0001	279	425.205518	M36	100	0.5	216	148.344861
M30		0.0005	274	407.347100	M37		1	213	142.359713
M31		0.001	268	403.633108	M38		10	131	86.315479
M32		0.005	259	387.205779	M39		30	110	42.196842
M33		0.01	255	343.874412	M40		50	103	28.128746
M34		0.05	249	297.298671	M41		70	98	17.351983
M35		0.1	247	247.344812	M42		100	92	10.249851
M43	1000	0.0001	282	418.746103	M50	1000	0.5	218	140.387762
M44		0.0005	277	413.559143	M51		1	209	136.658736
M45		0.001	271	407.288934	M52		10	118	66.322846
M46		0.005	266	383.977105	M53		30	112	49.736874

M47		0.01	263	369.544601	M54		50	103	38.354183
M48		0.05	254	281.349751	M55		70	96	26.354987
M49		0.1	248	231.910057	M56		100	94	17.339428
M57	10000	0.0001	283	406.566302	M64	10000	0.5	214	141.669713
M58		0.0005	273	428.699107	M65		1	190	124.610038
M59		0.001	264	419.488503	M66		10	116	46.217763
M60		0.005	255	357.446309	M67		30	105	16.331874
M61		0.01	249	271.010087	M68		50	103	14.001874
M62		0.05	246	244.896604	M69		70	97	12.957213
M63		0.1	241	190.337504	M70		100	95	10.318732

The best model is C=100, $\gamma = 100$, which contains 92 support vectors through Table (1). It was found that Model 42 has the least training error and the least number of support vectors. Therefore, it will be used in predicting future values (Test Set).

Table 2: Observed and forecast value of the daily incomes to the lake of al-Mosul dam from 2023-2024 for 73 observed data points at end of 2024 using the optimal SVM

No.	Observed	Forecasted	No.	Observed	Forecasted
1	183	191	38	78	81
2	115	111	39	69	73
3	131	136	40	81	84
4	118	122	41	119	124
5	123	124	42	114	118
6	147	147	43	121	120
7	185	196	44	121	122
8	104	108	45	107	109
9	109	113	46	92	95
10	101	108	47	107	108
11	80	87	48	88	90
12	80	82	49	86	93
13	41	45	50	78	80
14	48	51	51	82	84
15	42	42	52	87	89
16	118	123	53	129	131
17	164	170	54	152	153
18	187	186	55	180	184
19	166	164	56	193	194
20	177	179	57	192	196
21	175	178	58	159	163
22	226	225	59	143	145
23	195	197	60	122	123
24	137	140	61	118	118
25	189	188	62	104	108
26	73	77	63	169	171
27	45	46	64	128	133
28	55	58	65	131	129
29	75	77	66	107	108
30	59	62	67	95	98
31	83	86	68	132	135
32	71	74	69	148	152
33	60	63	70	129	130
34	60	63	71	149	149
35	61	64	72	160	165
36	68	72	73	167	168
37	71	73			

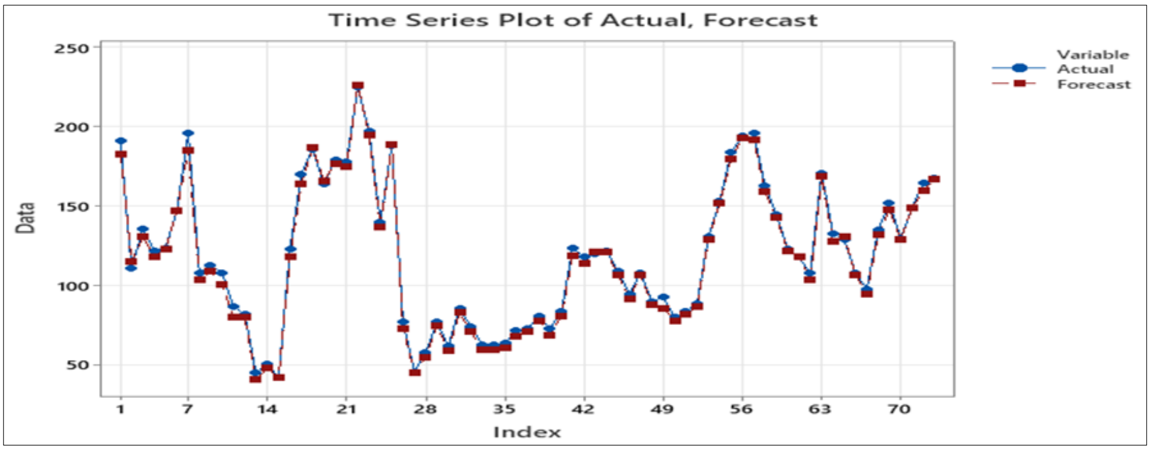

Fig 2: Shows the original values of the time series and the forecast values

Conclusion

Time series data of water quantities flowing into Mosul Dam Lake were analyzed during the time period from the beginning of 2023 to the end of 2024 (730 days), used Support Vector Machines, in order to train the SVM and identify the optimal parameter, the first 657 observed points were used for validation, the remaining items constitute the test set, 73 data points were detected (10% of the series) which was utilized to test SVM's predictive ability, the kernel parameters γ and C values used(C=1-10000) & (γ=0.0001-100), The best model is C=100, γ = 100, which contains 92 support vectors. It was found that Model 42 has the least training error and the least number of support vectors.

References

1. Al-Obaidi AGJ. Analysis and modeling of time series of temperatures in Mosul City [master's thesis]. Mosul: University of Mosul, College of Science; 1989.
2. Cao DZ, Pang SL, Bai YH. Forecasting exchange rates using support vector machines. In: Proceedings of the 4th International Conference on Machine Learning and Cybernetics; 2005 Aug 18–21; Guangzhou.
3. Huang H, Tian Y. A novel visual modeling system for time series forecast: application to the domain of hydrology. J Hydroinform. 2013;15(1):21–37.
4. Zaki AH. Using neural networks in forecasting time series by application to electricity consumption in Nineveh Governorate [master's thesis]. Mosul: University of Mosul, College of Computer Sciences and Mathematics; 2005.
5. Mellit A, Pavan A, Benghanem M. Least squares support vector machine for short-term prediction of meteorological time series. Theor Appl Climatol. 2013;111(1):297–307.
6. Meyer D. Support vector machines—the interface to libsvm in package e1071. Austria: FH Technikum Wien; 2018.
7. Zhang L, Zhou WD, Yang JW, Li FZ. Iterated time series prediction with multiple support vector regression models. Neurocomputing. 2013;99:411–22.
8. Mustaffa Z, Yusof Y. A hybridization of enhanced artificial bee colony–least squares support vector machines for price forecasting. J Comput Sci. 2012;8(10):1680–90.
9. Okasha MK. Using support vector machines in financial time series forecasting. Int J Stat Appl. 2014;4(1):28–39.
10. Pai PF, Hong WC. Support vector machines with simulated annealing algorithms in electricity load forecasting. Energy Convers Manag. 2005;46(17):2669–88.
11. Qasim O. Application of intelligent techniques in bioinformatics [PhD thesis]. Mosul: University of Mosul, College of Computer Sciences and Mathematics, Department of Mathematics; 2010.
12. Suganyadevi MV, Babulal CK. Support vector regression model for the prediction of loadability margin of a power system. Appl Soft Comput. 2014;24:304–15.
13. Wang J, Li L, Niu D, Tan Z. An annual load forecasting model based on support vector regression with differential evolution algorithm. Appl Energy. 2012;94:65–70.
14. Wang LY, Zhao WG. Daily discharge forecasting based on support vector regression. Adv Mater Res. 2010;113–116:386–9.
15. Zhang Z, Ye S. Long term load forecasting and recommendations for China based on support vector regression. In: Proceedings of the International Conference on Information Management; 2011. p. 597–602.

How to Cite This Article

Allela DMJ. Using vector support mechanism to model and forecast water flows into the Mosul Dam Lake. Int J Appl Math Numer Res. 2026;2(3):28–32. doi:10.54660/IJAMNR.2026.2.3.28-32

Creative Commons (CC) License