Scholar Publishing Group
Water Pollution Prevention and Control Project
https://doi.org/10.38007/WPPCP.2021.020306
ISSN 2790-2730 Vol. 2, Issue 3: 51-60

Linear Regression for Water Pollution Control Planning under Support Vectors

Lizunov Sae*

Monash Univ Malaysia, Selangor Darul Ehsan 47500, Malaysia

**corresponding author*

Keywords: Support Vectors, Water Pollution, Prevention and Control Planning, Linear Regression

Abstract: China is facing a serious situation in water pollution prevention and control, and the uncertainty of regional development should be fully considered when formulating water pollution prevention and control planning, and its scientificity should be improved through effective mathematical methods. Water resources and water environment are closely related to human health, and China has not yet established a comprehensive set of water resources and environment monitoring network, so the water pollution prevention and control planning process is crucial to the implementation of water pollution control planning. Due to the relatively low level of economic development in China, the state attaches less importance to water pollution prevention and control issues, and therefore also has a certain impact on the water environment. In order to reduce the problems caused by water pollution, the Water Resources Management Committee has organised the preparation of water pollution control plans. This paper presents an empirical analysis of the support vector model for water pollution prevention and control planning in China. The results of the study show that the introduction of the support vector approach to water pollution prevention and control planning in China can effectively improve the scientific nature of the planning. This paper firstly introduces the support vector model, which is a multivariate model designed based on the support vector method, and can be used to analyse multiple indicators and conduct comprehensive analysis in the prediction process; finally, the planning scheme should be designed with full consideration of water pollution prevention and control planning to avoid conflicts between different water resources and environmental issues. However, there are many problems with the use of the support vector approach in water pollution prevention and control planning that warrant further research.

1. Introduction

Water pollution is a natural phenomenon that causes a series of ecological problems such as eutrophication of water bodies and destruction of water ecosystems due to pollution, which seriously affects the health and function of the ecosystem. Since China's water pollution prevention and control work, water pollution management has been effective, but there are still many problems, such as small-scale water pollution management, resulting in serious water pollution; water quality is unstable, some rivers water quality deterioration; urban sewage treatment capacity is insufficient, sewage directly into the river; industrial waste high cost of water treatment, low cost of pollution; dense population, limited environmental capacity; public awareness of environmental protection is not strong, public participation channels The public is not aware of environmental protection and the channels for public participation are not smooth. In the context of China's current ecological civilization construction, the integration of water pollution prevention and control into national economic and social development planning has become one of the indispensable components of China's socio-economic development [1-2].

In a related study, Swati et al. proposed an efficient river water pollution monitoring system using deep neural networks and remote communication techniques [3]. A compressed deep neural network for monitoring river water pollution was first designed. Next, a knowledge distillation technique was used to train the compressed deep neural network. In addition, a game theory-based approach is proposed to estimate the duration of river water data transmission using an appropriate extension factor of the remote network. Finally, the effectiveness of the proposed system is demonstrated.Amal et al. propose a principled non-parametric weighted network model based on an exponential family random graph model and local likelihood estimation, and investigate their model-based clustering method and its application to large-scale water pollution network analysis [4]. The method greatly extends the methodology and applicability of statistical network models.

This paper first introduces the support vector model, and then carries out the planning scheme design should fully consider water pollution control planning to avoid conflicts between different water resources environmental issues. This paper takes the study of how linear regression models are considered in water environment planning and the application of support vector methods to water resources governance planning as an example to explore what issues to pay attention to in planning for the implementation of water pollution control planning under the water resources management strategy under support vectors. By analysing the interrelationships between the constraints in the support vector model it is possible to visualise the proportion of each factor in the water environment factors.

2. Design Studies

2.1. Problems with Water Pollution Prevention and Control

Water pollution prevention and control is an important part of the construction of ecological civilization and a major issue related to the people's livelihood, so the task of water management in China is very arduous [5-6]. However, there are still some problems with water pollution prevention and control in China, mainly in.

(1) The spatial layout and quantitative characteristics of water pollution control have not achieved a high degree of integration.

(2) The impact of each water pollutant on the water environment exhibits different properties and characteristics.

(3) A complete match between environmental quality objectives and environmental quality standards has not been achieved.

(4) The environmental governance needs of the water governance process are not being adequately addressed.

(5) The government and the public lack awareness of relevant laws and regulations and technical means.

To address the above issues, this paper proposes the "support vector" theory for water pollution control planning and decision-making. A support vector is a mathematical model consisting of parameter functions with vectorial significance, which mainly includes regression functions, multiple linear regression and mathematical equations [7-8]. Improvements to the model are required for a given target situation.

2.2. The "Three Waters" Conversion Relationship in A River Basin

Atmospheric precipitation, surface water and groundwater are referred to as the "three waters", which can be transformed into each other under certain conditions and are closely related to each other, and are important components of the water cycle system [9-10].

Atmospheric precipitation is an important source of recharge for both surface water and groundwater. Atmospheric precipitation infiltrates into the aquifer through the air pocket (Figure 1 pathway ①) and becomes one of the important sources of groundwater recharge in a river basin. Groundwater receives recharge from precipitation and also recharges the Jing River and its tributaries in the form of lateral runoff (Figure 1 pathway ②), and groundwater discharge from the river is transformed into river runoff, which is important for safeguarding river runoff to maintain the ecological health of the Jing River and its tributaries [11-12].

In addition to the infiltration of atmospheric precipitation into the air pocket, the superinfiltrated flow production partly forms surface runoff (Figure 1 route ③), which eventually converges into the Weihe main stream and its tributaries (Figure 1 route ④) and becomes an important part of river runoff. Surface water and groundwater are transformed by atmospheric precipitation, and the two can also be transformed into each other through a certain hydrological cycle. Evaporation also takes place at all times during the water cycle in the "three waters", although the amount of evaporation is significantly reduced compared to the "three waters". A small part of the water vapour formed by evaporation will condense again into condensate and enter the air pocket or surface water bodies, a part will form precipitation and eventually be transformed into surface water and groundwater again, and a part will evaporate through transpiration; the above water transport process constitutes a mutual replenishment and transformation, cyclic "three waters" transformation relationship [13 -14].

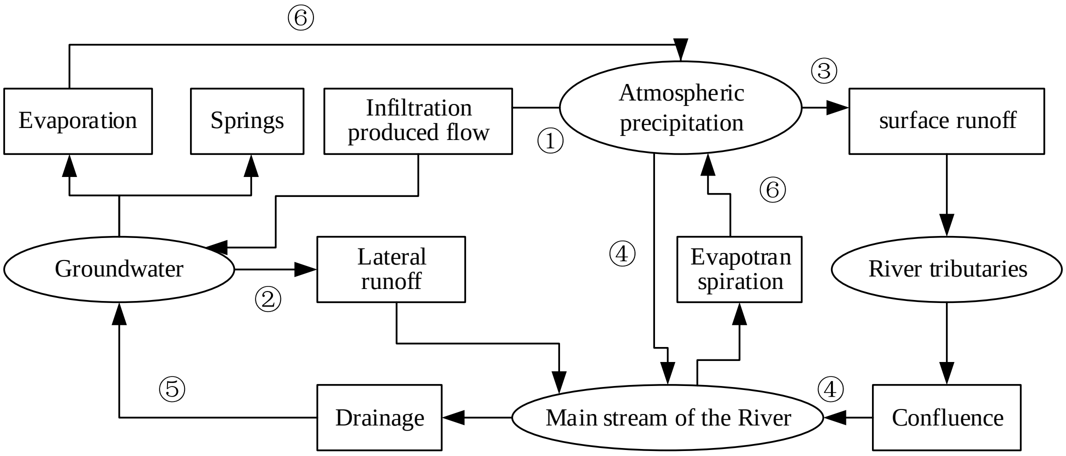

Figure 1. River basin "three waters" transformation map

This shows that the 'three waters' of a river basin are both correlated and constrained by each other. The correlation is reflected in the fact that precipitation acts as a common source of recharge for both surface water and groundwater, while the source of recharge for precipitation comes from evaporation, the same discharge method for both surface water and groundwater, thus bringing the 'three waters' into the same cyclic relationship. The constraint system is that there are checks and balances on the amount and timing of transformations between the three waters [15-16].

2.3. Support Vector Model

The model is a multivariate model designed based on the support vector method. Multiple indicators can be applied and combined in the analysis of the forecasting process [17-18]. The model has several main components.

(1) Predicting different water quality types based on trends in water environment quality and water use status.

(2) Delineation of control units based on pollutant emission intensity targets.

(3) Planning objectives and programmes based on the degree of control objective completion after the introduction of a multi-objective function in the model.

(4) Planning based on planning indicators to predict trends in the quality of the water environment in the basin.

2.4. Lasso Model

Consider a multiple linear regression model, as in equation (1).

$$Y = X\beta + \varepsilon \tag{1}$$

where X=(x1,x2,... ,xp)T, β=(β1,β2,... ,βp). The Lasso method is a good solution to this problem and can be used for the selection of variables with cointegration.

The coefficients are estimated using the Lasso method, i.e. by finding the minimum value of equation (2).

$$\hat{\beta}^{\text{Lasso}} = \arg\min_{\beta}\left\{\frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1\right\} \tag{2}$$

where $\|y - X\beta\|_2^2 = \sum_{i=1}^{N}(y_i - (X\beta)_i)^2$, is used to judge the model fit, $\|\beta\|_1 = \sum_{j=1}^{p}|\beta_j|$ is the penalty term, and λ>0 is the reconciliation parameter that determines the strength of the penalty and affects the number of variables selected. Using the Donoho-Johnstone's Lemma and progressing to solve for β, equation (3) can be obtained.

$$\hat{\beta}_k = \left(\frac{Y_{(k)}{}^T X_k}{\|X_k\|_2^2} - \beta_k\right)_+ \bullet \operatorname{sgn}(Y_{(k)}{}^T X_k) \tag{3}$$

In the above equation, the $(\mathrm{a})_+ = \begin{cases} 0, a < 0 \\ a, a \geq 0 \end{cases}$ $\operatorname{sgn}(a) = \begin{cases} 1, a > 0 \\ 0, a = 0 \\ -1, a < 0 \end{cases}$

2.5. Model Evaluation Indicators

(1) Correlation coefficient

The correlation coefficient (r) is a measure used to characterize the relationship between two variables and is calculated as

$$r=\frac{\sum_{i=1}^{n}(Q_{O_i}-\overline{Q}_O)(Q_{S_i}-\overline{Q}_S)}{\sqrt{\sum_{i=1}^{n}(Q_{O_i}-\overline{Q}_O)^2\sum_{i=1}^{n}(Q_{S_i}-\overline{Q}_S)^2}} \tag{4}$$

Where r denotes the correlation coefficient; QS is the mean value of the simulated runoff series. The larger the correlation coefficient, the stronger the correlation, which takes values in the range [-1,1].

(2) Relative error

The relative error in runoff is the absolute value of the percentage difference between the sum of the simulated runoff and the sum of the measured runoff during the simulation period, and is expressed as

$$RE=\frac{\left|\sum_{i=1}^{n}Q_{S_i}-\sum_{i=1}^{n}Q_{O_i}\right|}{\sum_{i=1}^{n}Q_{O_i}}\times 100\% \tag{5}$$

The smaller the value, the more reliable the model simulation will be.

3. Experimental Research

3.1. Analysis of Water Pollution Control Planning Based on Multiple Linear Regression Models

Traditional linear regression models use non-linear matrices to describe the correlation between matrices. This method performs linear calculations of the correlation coefficients, although it allows better analysis of various parameters and yields better results. However, there are also multiple linear regression models where multiple data do not correspond exactly to each other, making it impossible to carry out effective predictive analysis. The support vector method is a good solution to this problem. The method can be well used in practical applications for prediction analysis, but there are great problems in some complex regional analysis. This paper uses a multiple linear regression model to carry out an empirical analysis of water pollution control planning in China.

3.2. Water Environment Constraints

The water quality constraints include: COD, total nitrogen, total phosphorus, ammonia nitrogen and total phosphorus. Based on the study of each water environment constraint factor, it can be seen that there is a close relationship between all factors. These factors include.

(1) Total nitrogen in water quality has relative importance under the pollution control objectives.

(2) Has a significant impact under the total phosphorus pollution control target.

(3) Other influences have a strong control role under the pollution control objectives.

When water quality objectives are not met, there is a negative impact on the environment. The importance of water quality constraints can be seen from previous research on water quality objectives and the literature on these constraints. These include: current and potential risk factors and pollutant discharge intensity factors. With economic development and urbanisation, the pollution of water bodies will tend to increase.

4. Experimental Analysis

4.1. Waste Water and Main Pollutant Emissions

The amount of wastewater discharged and the amount of major pollutants entering the river within a river basin over a six-year period (about 70% of the study area) was calculated from a provincial water resources bulletin. The details are shown in Table 1.

Table 1. Statistics on wastewater discharge and main pollutants entering the river

	1	2	3	4	5	6
Waste water discharge (billions of tonnes)	0.6	0.4	0.5	0.6	0.55	0.9
Chemical oxygen demand(Tonnes)	10,500	13100	12600	4900	5000	1000
Ammonia nitrogen (Tonnes)	1000	1100	1200	500	450	0

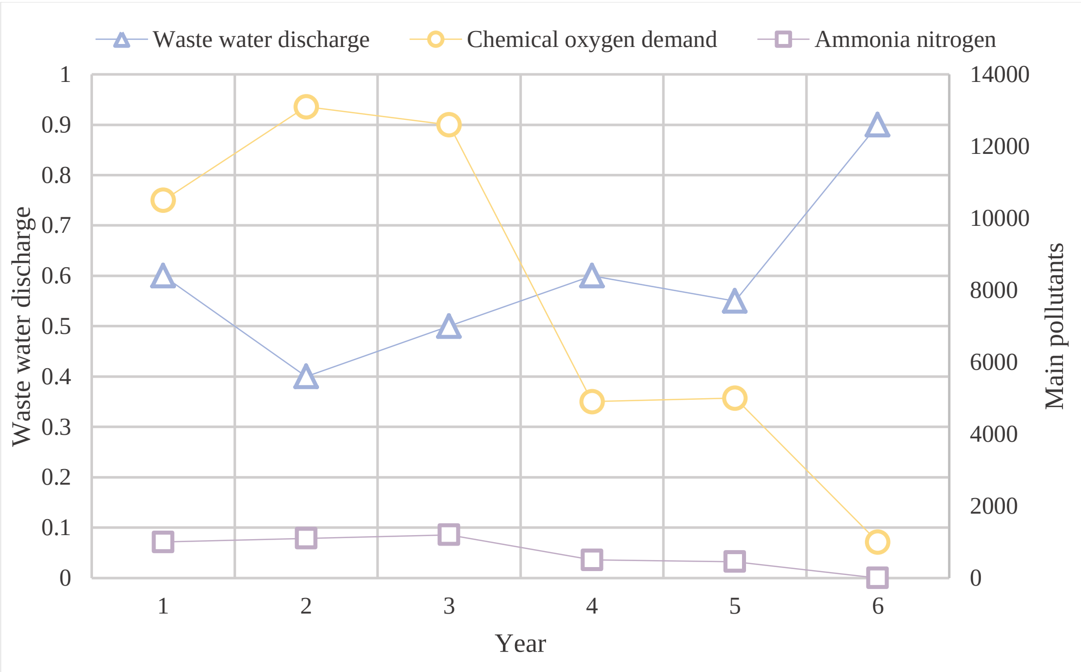

Figure 2. Analysis of wastewater discharge and main pollutants entering the river

As can be seen in Figure 2, wastewater discharges show a steady variation within a small range, ranging from 0.57 to 0.71 billion tonnes, while compliance discharges account for 43.2% of total discharges. In the second year, wastewater discharges reach a minimum at the basin scale, after which they show an increasing trend. The COD discharged to the river shows an increase followed by a decrease, while ammonia nitrogen is the one that shows a gradual decrease over time.

4.2. Construction of the Model Database

(1) Type of land use

The land use data used in this study was reclassified by applying land use classification criteria derived from the Vegetation Atlas. In the WetSpa model, land use types were reclassified based on seven indicators: root depth (D), Manning's coefficient (M), maximum foliage index (L_Max), minimum foliage index (L_Min), maximum field water holding capacity (I_Max), minimum field water holding capacity (I_Max) and imperviousness (F) (Table 2).

Table 2. Criteria for reclassification of land use types

Land use type	D(m)	F(%)	M	Field water holding capacity (mm)		Leaf area index	
				I_Max	I_Max	L_Max	L_Min
Mixed forest	1.0	5	5.5	30.0	5.0	60.0	30.0
Depressional scrub	0.8	5	4.0	25.0	5.0	60.0	10.0
Sparse scrub	0.8	5	4.0	20.0	5.0	60.0	10.0
savanna	1.0	5	5.0	30.0	5.0	60.0	8.0
grassland	0.8	5	3.0	20.0	5.0	20.0	5.0
Swamp	0.5	100	5.0	10.0	2.0	60.0	5.0
Agricultural land	0.8	5	3.5	20.0	5.0	60.0	5.0
Towns	0.5	50	0.5	0.0	0.0	0.0	0.0
Snowfields	0.1	100	0.5	0.0	0.0	0.0	0.0
wasteland	0.5	20	1.0	10.0	2.0	20.0	5.0
Water	0.1	100	0.5	0.0	0.0	0.0	0.0

(2) Soil information

Soil texture characteristics directly influence the infiltration process of the circulation and are a very critical piece of soil information. Based on the soil type data integrated from the sand and clay content at different depths, the soil related parameters were set for each grid cell in the basin according to Figure 3 below. These fixed grid cell parameters are only used in the WetSpa model.

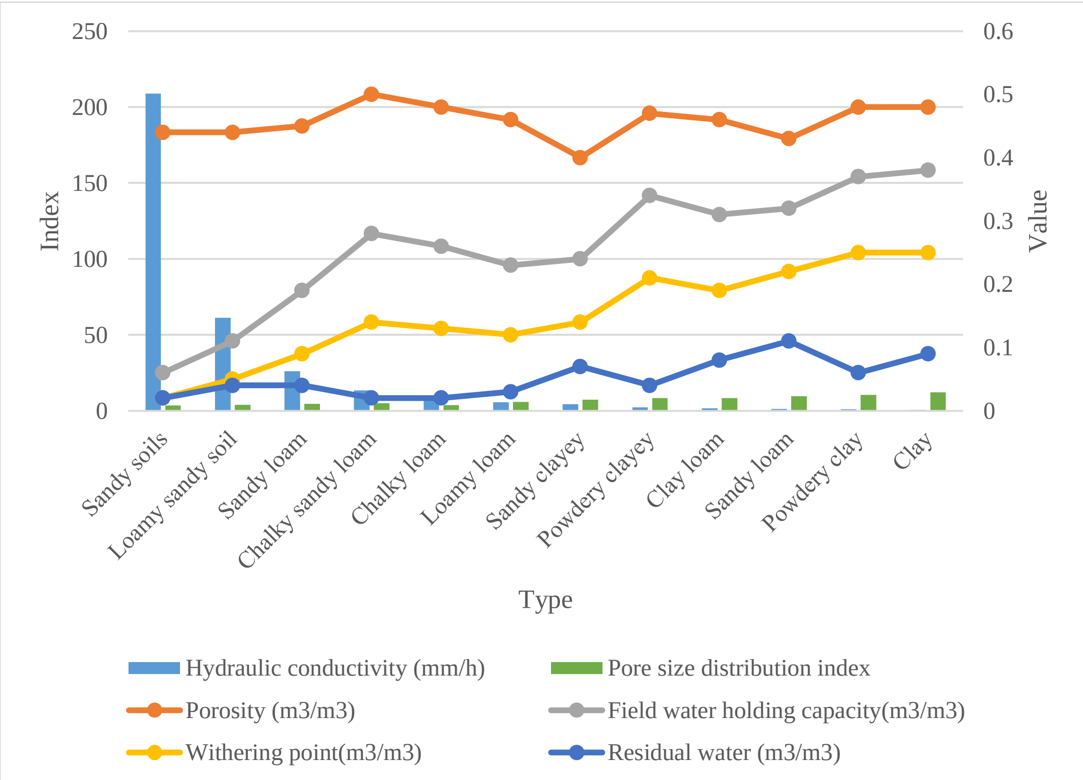

Figure 3. Wetland model soil type reclassification

4.3. Hydrological Model Parameter Rate and Validation

In the simulation of a hydrological model, some of the parameters or parameters that have no direct physical meaning cannot be obtained through actual observations or experimental measurements, and therefore some of the parameters of the model need to be rate determined, i.e. "the combination of simulated and observed values that is most consistent". Model validation is the assessment of the confidence that the parameters of the model's rate reflect the true situation, and the degree of agreement over the validation period is usually used to reflect the validity of the model. In order to reduce the workload of model validation, a sensitivity analysis of the model parameters is required before the model can be fully validated. A sensitivity analysis of the parameters of a river basin hydrological model was carried out and eight sensitivity parameters were calculated, with the range of variation, meaning and optimal values shown in Figure 4.

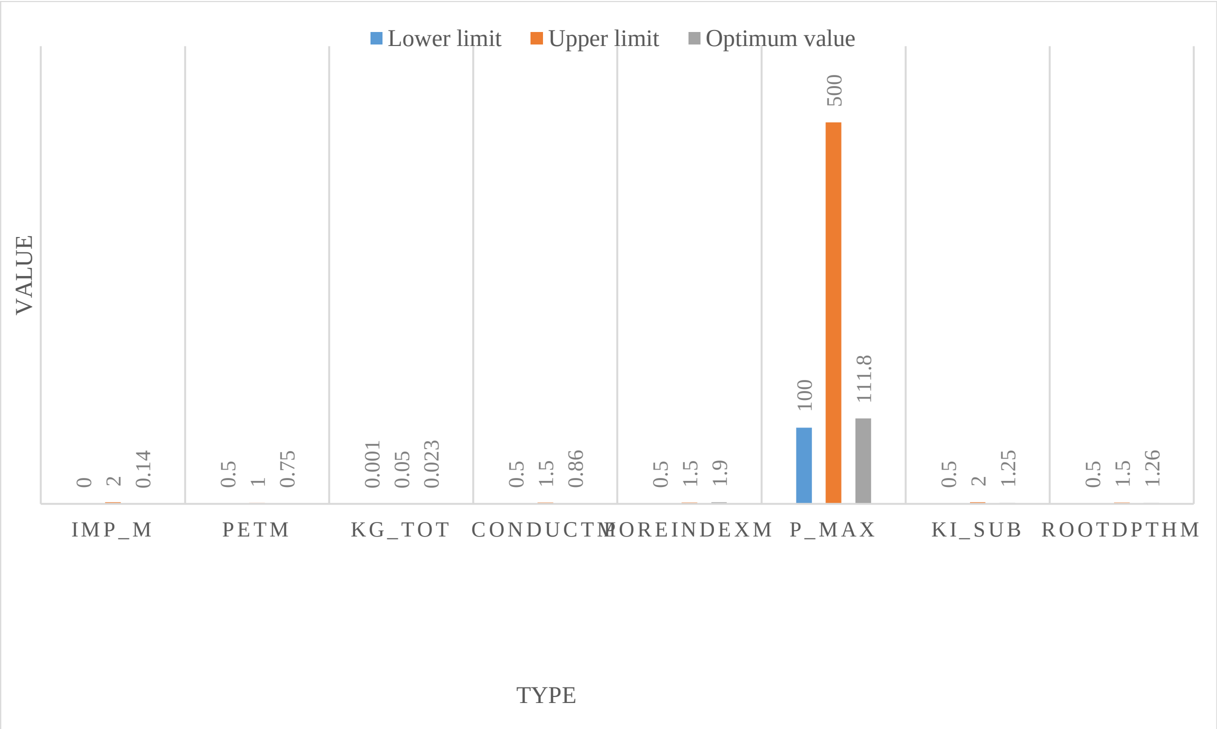

Figure 4. Sensitivity parameters in the WetSpa model for a river basin

5. Conclusion

Water pollution prevention and control is one of the priorities in China and water resources management needs to be continuously strengthened. As one of the most widely researched methods, the support vector model has good feasibility and can effectively reduce the uncertainty in the implementation of water pollution planning in China. The introduction of support vector methods in water resource management can therefore effectively improve the scientific nature of water resource management planning. At present, the use of water resources in all regions of China is not ideal, and the water environment capacity indicator is one of the most significant indicators that affects the quality of water resources and water environment. It is not an absolute optimum for each region and therefore requires comprehensive analysis to arrive at the optimum value. There are also significant differences in the level of water use between regions. By analysing the interdependence between the structure of water resources use and the indicators of water pollution prevention and control plans, it is possible to visualise the differences between regions and to analyse them together.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Taufik D A , Setiawan I . Integration of linear regression and aggregate planning for Hino OW 190/200 Leaf Spring production planning and control in the automotive component industry. Operations Excellence Journal of Applied Industrial Engineering, 2021, 13(2):245-254.

[2] Saedi R , Verma R , Zockaie A , et al. Comparison of Support Vector and Non-Linear Regression Models for Estimating Large-Scale Vehicular Emissions, Incorporating Network-Wide Fundamental Diagram for Heterogeneous Vehicles:. Transportation Research Record, 2020, 2674(5):70-84.

[3] Swati Chopade, Hari Prabhat Gupta, Rahul Mishra, Preti Kumari, Tanima Dutta: An Energy-Efficient River Water Pollution Monitoring System in Internet of Things. IEEE Trans. Green Commun. Netw. 5(2): 693-702 (2021).

[4] Amal Agarwal, Lingzhou Xue: Model-Based Clustering of Nonparametric Weighted Networks With Application to Water Pollution Analysis. Technometrics 62(2): 161-172 (2020).

[5] Shakeel Ahmad, Muhammad Aslam: Another proposal about the new two-parameter estimator for linear regression model with correlated regressors. Commun. Stat. Simul. Comput. 51(6): 3054-3072 (2021).

[6] Lars Erik Gangsei, Trygve Almøy, Solve Sæbø: Linear regression with bivariate response variable containing missing data. Strategies to increase prediction precision. Commun. Stat. Simul. Comput. 51(2): 527-538 (2021).

[7] Senay Özdemir, Olçay Arslan: Combining empirical likelihood and robust estimation methods for linear regression models. Commun. Stat. Simul. Comput. 51(3): 941-954 (2021).

[8] Daniela Rodriguez, Marina Valdora, Pablo Vena: Robust estimation in partially linear regression models with monotonicity constraints. Commun. Stat. Simul. Comput. 51(4): 2039-2052 (2021).

[9] Onur Toka, Meral Çetin, Olçay Arslan: Robust estimation in restricted linear regression. Commun. Stat. Simul. Comput. 51(3): 1015-1029 (2021).

[10] Elena Dumitrescu, Sullivan Hué, Christophe Hurlin, Sessi Tokpavi: Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. Eur. J. Oper. Res. 297(3): 1178-1192 (2021).

[11] T. Syed Akheel, V. Usha Shree, S. Aruna Mastani: Stochastic gradient descent linear collaborative discriminant regression classification based face recognition. Evol. Intell. 15(3): 1729-1743 (2021).

[12] Abdelkrim Bouasria, Khalid Ibno Namr, Abdelmejid Rahimi, El Mostafa Ettachfini, Badr Rerhou: Evaluation of Landsat 8 image pansharpening in estimating soil organic matter using

multiple linear regression and artificial neural networks. Geo spatial Inf. Sci. 25(3): 353-364 (2021).

[13] Taesoo Kwon, Moon-Sik Lee, Youngil Jeon, HyeonWoo LEE: Linear-regression based performance approximation for millimeter-wave multicell networks with β-Ginibre deployed base stations. ICT Express 8(2): 302-308 (2021).

[14] Carmen Biedma-Rdguez, María José Gacto, Augusto Anguita-Ruiz, Jesús Alcalá-Fdez, Rafael Alcalá: Transparent but Accurate Evolutionary Regression Combining New Linguistic Fuzzy Grammar and a Novel Interpretable Linear Extension. Int. J. Fuzzy Syst. 24(7): 3082-3103 (2021).

[15] Udora N. Nwawelu, Mamilus Aginwa Ahaneku, Benjamin O. Ezurike: Improving Weighted Multiple Linear Regression Algorithm for Radiolocation Estimation in LoRaWAN. Int. J. Interdiscip. Telecommun. Netw. 14(1): 1-12 (2021).

[16] Razeef Mohd, Muheet Ahmed Butt, Majid Zaman Baba: Grey Wolf-Based Linear Regression Model for Rainfall Prediction. Int. J. Inf. Technol. Syst. Approach 15(1): 1-18 (2021).

[17] Oluwaseun Omodemi, Martina Kaledin, Alexey L. Kaledin: Permutationally invariant polynomial representation of polarizability tensor surfaces for linear regression analysis. J. Comput. Chem. 43(22): 1495-1503 (2021).

[18] Surajit Nandi, Jonas Busk, Peter Bjørn Jørgensen, Tejs Vegge, Arghya Bhowmik: Cheap Turns Superior: A Linear Regression-Based Correction Method to Reaction Energy from the DFT. J. Chem. Inf. Model. 62(19): 4727-4735 (2021).