HYDRO-BIOCHEMICAL PROCESSES AND RIVER WATER QUALITY SIMULATION USING MULTIVARIATE STATISTICS AND SOFT COMPUTING TECHNIQUES

SAMEER ARORA

DEPARTMENT OF CIVIL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

NOVEMBER 2021

HYDRO-BIOCHEMICAL PROCESSES AND RIVER WATER QUALITY SIMULATION USING MULTIVARIATE STATISTICS AND SOFT COMPUTING TECHNIQUES

by

SAMEER ARORA

Department of Civil Engineering

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

NOVEMBER 2021

CERTIFICATE

This is to certify that the thesis entitled **“Hydro-biochemical processes and river water quality simulation using multivariate statistics and soft computing techniques”** submitted by **Sameer Arora (2013CEZ8073)** to the Indian Institute of Technology Delhi for the award of the degree of **Doctor of Philosophy** is a bonafide record of research work carried out by him under my supervision and guidance. The thesis work, in my opinion, has reached the requisite standard fulfilling the requirements for the degree of **Doctor of Philosophy**.

The results contained in this thesis have not been submitted, in part or full, to any other University or Institute for the award of any degree or diploma.

Prof. (Dr.) Ashok K. Keshari
Professor
Department of Civil Engineering
Indian Institute of Technology Delhi
Hauz Khas, New Delhi 110016, India

ACKNOWLEDGMENTS

The research work is a collaboration of encouragement, motivation, and hard work. I have been inspired and motivated from various people and would like to express my gratitude to all of them. First and foremost, I would like to express my deepest gratitude to my guide and mentor Prof. Ashok K. Keshari, Department of Civil Engineering who stimulated me to work on this topic. His constant support and clear vision, inspired and motivated me throughout the journey. His resourcefulness and critical feedbacks have greatly assisted me in this work. The vast experience of my supervisor in the field of water resource engineering, enhanced my knowledge and confidence to perform in this field and undertaking research.

My thanks are also due to the student research committee (SRC) members, Prof. Manoj Datta (Chairperson), Prof Arvind Nema (Departmental Expert) and Prof. V Haridas (Subject Expert) for their valuable comments and suggestions.

I want to record my gratitude to the faculty members of Civil Engineering Department for their help and giving me an opportunity to carry out this research work. I would also like to appreciate the staff of Civil Engineering Department especially Mr. N R Gehlot for providing me helpful co-operation throughout the completion of this research work.

IIT Delhi become second home during this tenure and met with family members like friends since I joined here. I would like to acknowledge all PhD fellows with whom I shared social life and experience. I am also very grateful to my friends, Dr. Nitesh Patidar, Mr. Gopinadh Rongali and Mr. Amar Landage, for providing rejuvenating tea sessions and period of productive thinking.

Lastly, I am fortunate and grateful to my parents, Mrs. Savita Arora and Mr. J S Arora for everything that I have accomplished in life with their blessings, inspiration and support. Without their encouragement, support and sacrifice my study would not have been possible. I am thankful to my sisters and wife for her continuous support in my ups and downs. Thanks to all of you, I really appreciate everything that you have done for me.

Sameer Arora

ABSTRACT

River dynamics involve composite interactions among flow, nutrients, transported material and channel type. Spatial and temporal variations are generally observed in biochemical parameters and concentration of dissolved chemicals in freshwater bodies, corresponding to its flow through different climatic, geographical and topographic environment. Discharge of domestic and industrial effluents from multiple sources accelerate the rate of variation in biochemical characterisation of water body. Inherent ability of water flow to assimilate organic load allow the self-purification of streams. Self-purification is an intricate phenomenon, comprising of physical, chemical and biological processes occur simultaneously to recover the native state of water body. The re-aeration capacity of the stream defines the net rate of oxygen transfer from atmosphere to air-water interface and depends significantly upon the degree of turbulence of the flowing stream, as the interface is constantly changing with the flow rate due to the mixing of the flowing water. The indirect measurement of re-oxygenation is based on the oxygen deficit of river, which is calculated as the variance between the dissolved oxygen content and the native oxygen concentration at specific temperature.

The assimilation capacity of the stream can be measured by estimating several indices like biota of the river, dissolved oxygen, temperature, and amount of flow, velocity, concentration of suspended and dissolved solids. Determination of re-aeration coefficient during re-oxygenation process in water body should depends on reasonable valuation of oxygen transfer rate across air-water interface. For years, several attempts have been made by researchers for envisaging the rate of oxygen transfer from defined physical and hydraulic properties of the river. However, the functional relationships between the hydraulic variables and the re-aeration coefficient are limited to specific boundary conditions. The relative

significance of variables used to design the re-aeration equations such as depth, velocity, slope, and width are also not found reliable in the literature and applicable to the streams free from any organic load. The several predictive re-aeration models have been developed and applied on confined study areas having tremendous variability of channel types causes significant difference in the results of each model.

The improvement in the estimation of an adequate re-aeration coefficient is a vital requirement to simulate the river water quality. The hierarchical agglomerative cluster analysis (HACA), principal component analysis (PCA) and structural equation modelling (SEM) are employed to examine the similarities in water quality parameters, critical parameters and their indirect relationships. The combination of these statistical approaches helps to identify the role of different parameters in variation of water quality individually and to identify the significant source of pollutants. In this study, 17 predictive re-aeration models have been compared to reveal their applicability to the study area. Yamuna River has been selected as the study area, receiving the high organic load from anthropogenically influenced multiple point sources, which significantly affects the flow rate of the river. The river receives the wastewater from 17 major and minor drains, which causes significant fluctuations in the river physio-chemical properties. The error in the predictive re-aeration models is estimated using MAE, MAPE, MSE and MME; however, none of the -reaeration model gives promising results with the estimated values of the re-aeration coefficient.

The artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS) and autoregressive integrated moving average (ARIMA) have been applied to develop a new model for the estimation of the re-aeration coefficient. The ten different configurations of input parameters were used to design various ANN and ANFIS models, and the performance of models are assessed using R, R^2 and RMSE. The model with least RMSE and highest R^2 from ANN and ANFIS is further integrated with ARIMA to improve the estimation of the re-aeration coefficient. The ANFIS-ARIMA model improves the estimation of the re-aeration coefficient for the degraded river in comparison to predictive re-aeration models.

The simulation of river water quality is performed using QUAL2Kw integrated with the ANFIS-ARIMA model to identify the impact of newly developed model for the estimation of re-aeration coefficient. The sensitivity analysis of the DO and the BOD is also carried out to examine the variations in parameters concentration and related river characteristics. The study reveals that the developed ANFIS-ARIMA model improves the estimation of re-aeration coefficient significantly and can be used in the prediction of water quality. The study would be useful to the decision-makers while structuring the pollution abatement methods, developing river water management policies and defining the level of treatment that can be focused on the removal of sensitive contaminants or critical species that severely affect the river health.

बहुभिन्नरूपी सांख्यिकी और सॉफ्ट कंप्यूटिंग तकनीकों का उपयोग करते हुए जल-जैव रासायनिक प्रक्रियाएं और नदी जल गुणवत्ता सिमुलेशन

सार

नदी की गतिशीलता में प्रवाह, पोषक तत्वों, परिवहन सामग्री और चैनल प्रकार के बीच समग्र परस्पर क्रिया शामिल है। विभिन्न जलवायु, भौगोलिक और स्थलाकृतिक वातावरण के माध्यम से इसके प्रवाह के अनुरूप, मीठे पानी के निकायों में जैव रासायनिक मानकों और भंग रसायनों की एकाग्रता में स्थानिक और अस्थायी भिन्नताएं आम तौर पर देखी जाती हैं। कई स्रोतों से घरेलू और औद्योगिक अपशिष्टों का निर्वहन जल निकाय के जैव रासायनिक लक्षण वर्णन में भिन्नता की दर को तेज करता है। कार्बनिक भार को आत्मसात करने के लिए जल प्रवाह की अंतर्निहित क्षमता धाराओं की आत्म-शुद्धि की अनुमति देती है। आत्म-शुद्धि एक जटिल घटना है, जिसमें भौतिक, रासायनिक और जैविक प्रक्रियाएं शामिल हैं जो जल निकाय की मूल स्थिति को ठीक करने के लिए एक साथ होती हैं। धारा की पुन: वातन क्षमता वायुमंडल से वायु-जल अंतरापृष्ठ में ऑक्सीजन स्थानांतरण की शुद्ध दर को परिभाषित करती है और बहने वाली धारा की अशांति की डिग्री पर महत्वपूर्ण रूप से निर्भर करती है, क्योंकि मिश्रण के कारण प्रवाह दर के साथ इंटरफ़ेस लगातार बदल रहा है बहता पानी। पुन: ऑक्सीकरण का अप्रत्यक्ष माप नदी के ऑक्सीजन की कमी पर आधारित है, जिसकी गणना विशिष्ट तापमान पर घुलित ऑक्सीजन सामग्री और देशी ऑक्सीजन एकाग्रता के बीच अंतर के रूप में की जाती है।

नदी के बायोटा, घुलित ऑक्सीजन, तापमान और प्रवाह की मात्रा, वेग, निलंबित और घुलित ठोस पदार्थों की सांद्रता जैसे कई सूचकांकों का अनुमान लगाकर धारा की आत्मसात क्षमता को मापा जा

सकता है। जल निकाय में पुन: ऑक्सीकरण प्रक्रिया के दौरान पुन: वातन गुणांक का निर्धारण वायु-जल इंटरफेस में ऑक्सीजन अंतरण दर के उचित मूल्यांकन पर निर्भर होना चाहिए। वर्षों से, शोधकर्ताओं द्वारा नदी के परिभाषित भौतिक और हाइड्रोलिक गुणों से ऑक्सीजन हस्तांतरण की दर की परिकल्पना के लिए कई प्रयास किए गए हैं। हालांकि, हाइड्रोलिक चर और पुन: वातन गुणांक के बीच कार्यात्मक संबंध विशिष्ट सीमा स्थितियों तक सीमित हैं। गहराई, वेग, ढलान और चौड़ाई जैसे पुन: वातन समीकरणों को डिजाइन करने के लिए उपयोग किए जाने वाले चर के सापेक्ष महत्व भी साहित्य में विश्वसनीय नहीं पाए जाते हैं और किसी भी कार्बनिक भार से मुक्त धाराओं पर लागू होते हैं। कई भविष्य कहनेवाला पुन: वातन मॉडल विकसित किए गए हैं और सीमित अध्ययन क्षेत्रों पर लागू किए गए हैं जिनमें चैनल प्रकारों की जबरदस्त परिवर्तनशीलता प्रत्येक मॉडल के परिणामों में महत्वपूर्ण अंतर का कारण बनती है।

नदी के पानी की गुणवत्ता का अनुकरण करने के लिए पर्याप्त पुन: वातन गुणांक के आकलन में सुधार एक महत्वपूर्ण आवश्यकता है। जल गुणवत्ता मापदंडों, महत्वपूर्ण मापदंडों और उनके अप्रत्यक्ष संबंधों में समानता की जांच करने के लिए पदानु क्रमित समूह विश्लेषण (एचएसीए), प्रमुख घटक विश्लेषण (पीसीए) और संरचनात्मक समीकरण मॉडलिंग (एसईएम) को नियोजित किया जाता है। इन सांख्यिकीय दृष्टिकोणों का संयोजन व्यक्तिगत रूप से पानी की गुणवत्ता की भिन्नता में विभिन्न मापदंडों की भूमिका की पहचान करने और प्रदूषकों के महत्वपूर्ण स्रोत की पहचान करने में मदद करता है। इस अध्ययन में, अध्ययन क्षेत्र में उनकी प्रयोज्यता को प्रकट करने के लिए 17 भविष्य कहनेवाला पुन: वातन मॉडल की तुलना की गई है। यमुना नदी को अध्ययन क्षेत्र के रूप में चुना गया है, जो मानवजनित रूप से प्रभावित कई बिंदु स्रोतों से उच्च जैविक भार प्राप्त कर रहा है, जो नदी की प्रवाह दर को महत्वपूर्ण रूप से प्रभावित करता है। नदी 17 बड़े और छोटे नालों से अपशिष्ट जल प्राप्त करती है, जिससे नदी के भौतिक-रासायनिक गुणों में महत्वपूर्ण उतार-चढ़ाव होता है। माध्य पूर्ण त्रुटि (एमएई), माध्य निरपेक्ष प्रतिशत त्रुटि (एमएपीई), औसत वर्ग त्रुटि (एमएसई) और माध्य गुणक त्रुटि (एमएमई) का उपयोग करके

भविष्य कहनेवाला पुनः वातन मॉडल में त्रुटि का अनुमान लगाया गया है; हालांकि, कोई भी रीएरेशन मॉडल पुनः वातन गुणांक के प्रेक्षित मूल्यों के साथ आशाजनक परिणाम नहीं देता है।

कृत्रिम तंत्रिका नेटवर्क (एएनएन), अनुकूली तंत्रिका फजी अनुमान प्रणाली (एएनएफआईएस) और स्वप्रतिगामी एकीकृत चलती औसत (एआरआईएमए) को पुनः वातन गुणांक के आकलन के लिए एक नया मॉडल विकसित करने के लिए लागू किया गया है। विभिन्न एएनएन और एएनएफआईएस मॉडल को डिजाइन करने के लिए इनपुट पैरामीटर के दस अलग-अलग कॉन्फ़िगरेशन का उपयोग किया गया था, और मॉडल के प्रदर्शन का मूल्यांकन आर, आर^2 और आरएमएसई का उपयोग करके किया जाता है। एएनएन और एएनएफआईएस से कम से कम आरएमएसई और उच्चतम आर^2 वाले मॉडल को पुनः वातन गुणांक के अनुमान में सुधार के लिए एआरआईएमए के साथ एकीकृत किया गया है। एएनएफआईएस - एआरआईएमए मॉडल भविष्य कहनेवाला पुनः वातन मॉडल की तुलना में अवक्रमित नदी के लिए पुनः वातन गुणांक के अनुमान में सुधार करता है।

पुनः वातन गुणांक के आकलन के लिए नए विकसित मॉडल के प्रभाव की पहचान करने के लिए एएनएफआईएस - एआरआईएमए मॉडल के साथ एकीकृत क्यूयूएएल2केडब्ल्यू का उपयोग करके नदी के पानी की गुणवत्ता का अनुकरण किया जाता है। पैरामीटर एकाग्रता और संबंधित नदी विशेषताओं में भिन्नता की जांच के लिए डीओ और बीओडी का संवेदनशीलता विश्लेषण भी किया जाता है। अध्ययन से पता चलता है कि विकसित एएनएफआईएस - एआरआईएमए मॉडल पुनः वातन गुणांक के अनुमान में काफी सुधार करता है और इसका उपयोग पानी की गुणवत्ता की भविष्यवाणी में किया जा सकता है। प्रदूषण उपशमन विधियों की संरचना, नदी जल प्रबंधन नीतियों को विकसित करने और उपचार के स्तर को परिभाषित करने के दौरान निर्णय लेने वालों के लिए अध्ययन उपयोगी होगा जो संवेदनशील दूषित पदार्थों या महत्वपूर्ण प्रजातियों को हटाने पर केंद्रित हो सकता है जो नदी के स्वास्थ्य को गंभीर रूप से प्रभावित करते हैं।

***

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DO	Dissolved Oxygen
BOD	Biochemical Oxygen Demand
CPCB	Central Pollution Control Board
USEPA	United States Environmental Protection Agency
DOBT	Dissolved oxygen balance technique
ANN	Artificial neural network
ANFIS	Adaptive neuro-fuzzy inference system
SVM	Support vector machine
ARIMA	Auto regressive integrated moving average
KNN	K-nearest neighbours
MLR	Multivariate linear regression
SEM	Structural equation modelling
PCA	Principal Component analysis
CA	Cluster analysis
DA	Discriminant analysis
FA	Factor analysis
COD	Chemical oxygen demand
TDS	Total dissolved solid
ANOVA	Analysis of variance
AI	Artificial Intelligence
GA	Genetic algorithm
FFBP	Feed forward back propagation
RBF	Radial basis function

ARMAX	Autoregressive–moving-average model with exogenous inputs
NARX	Nonlinear autoregressive network with exogenous inputs
NH_3-N	Ammonia nitrogen
WNN	Wavelet neural network
FNN	Feedforward neural network
RNN	Recurrent neural network
LM	Levenberg Marquardt
BR	Bayesian regularization
GDX	Gradient descent with momentum and adaptive learning rate
RMSE	Root mean square error
ARMA	Auto regressive moving average
DENFIS	Dynamic evolving neural fuzzy inference system
MLP	Multilayer perceptron
ANFIS-GP	Adaptive neuro-fuzzy inference system – Grid Partitioning
ANFIS-SC	Adaptive neuro-fuzzy inference system – Subtractive Clustering
FFA	Firefly optimization algorithm
R^2	Coefficient of determination
NSE	Nash-Sutcliffe Efficiency
WI	Willmott's Index
MAE	Mean Absolute error
SSA	Singular spectrum analysis
SA	Sensitivity analysis
MCS	Monte Carlo simulation
SAFE	Sensitivity analysis for everybody
HACA	Hierarchical aligned cluster analysis

FV	Feature Vector
NME	normalized mean error
SE	standard error
MME	mean multiplicative error
MSE	mean square error
MAPE	mean absolute percentage error
M-FIS	Mamdani fuzzy inference system
TS-FIS	Takagi-Sugeno fuzzy inference system
ACF	autocorrelation function
PACF	partial autocorrelation function
DJB	Delhi Jal Board

LIST OF NOMENCLATURE

D	Final DO deficit (mg/l)
D_0	Initial DO deficit (mg/l)
K_1	De-oxygenation coefficient (1/day),
K_2	Re-aeration coefficient (1/day),
K_3	Settling rate coefficient (1/day),
K_r	K_1+K_3
L_0	Initial BOD (mg/l)
L	BOD (mg/l)
t	time
c	oxygen concentration at time t
c_s	oxygen concentration at saturation point
H	depth (m)
V	velocity (m/s)
S	slope (1/S)
S_B	Sediment oxygen demand
L_x	longitudinal distance
F_r	Froude's number
E_d	Square Euclidean distance
Δ	Merging factor
$\boldsymbol{\lambda}$	Eigenvector
σ_Z^2	Variance of new variable
σ_{rs}^2	Variance of translated variable
SS	Square symmetric matrix

E	Square orthogonal covariance matrix
D_m	Diagonal matrix
MP	Measured parameter
LV	Latent variable
α	error of measured parameter
β	coefficient between latent and measured parameter
ζ	coefficient between latent variables
φ	Variance of latent variable
E_p	normalized error
μ_{A_i}	Membership function
w_i	Antecedent function
$\overline{w_i} f_i$	Coefficient of function
Q_{in}	Influent discharge from river
Q_{ab}	Effluent discharge from river
R_n	Reach of river