Scholar Publishing Group
Water Pollution Prevention and Control Project
https://doi.org/10.38007/WPPCP.2023.040202
ISSN 2790-2730 Vol. 4, Issue 2: 10-17

The Early Warning Model of Sudden Water Pollution Based on the Latent Factor

Inyeol Yoon*

Keio University, Yokohama 223-8522, Japan

**corresponding author*

Keywords: Latent Factor Model, Sudden Water Pollution, Early Warning Model, Pollution Source

Abstract: At present, the treatment process for sudden water pollution(WP) problems generally includes water quality exceeding warning, artificial investigation of warning reasons, artificial search for pollution source(PS), risk assessment by experts, and formulation and implementation of emergency plans. However, sudden WP events usually cause huge harm in a short time, and emergency measures often fail to respond in a timely manner. Therefore, in order to give early warning and quickly deal with sudden pollution, this paper constructs a recurrent WP early warning model(EWM). By comparing the prediction and simulation of total phosphorus concentration in water body by the latent factor model and Mike model, it is found that the prediction value based on the LFM is very consistent with the simulation value, It can be used to predict the distribution characteristics of water pollutants from PS and help people deal with WP incidents quickly.

1. Introduction

With the rapid development of information technology such as the Internet of Things and cloud computing, intelligent solutions to WP problems have become a focus of attention. In response to this problem, the construction of a practical and simple EWM for sudden WP can effectively reduce the task of manually dealing with WP and improve the efficiency of WP treatment. The need for intelligent treatment of sudden WP problems is becoming increasingly strong.

Research on early warning of sudden WP has achieved good results. For example, some scholars argue that the scope of emergency response to sudden WP in rivers mainly includes the design of emergency monitoring networks, WP early warning, WP traceability, WP risk assessment and analysis and emergency response, while early warning is initiated before a WP incident occurs and it is necessary to trace the source of pollution because it can guide the subsequent assessment and analysis of WP risk and emergency response [1]. Some scholars have analysed a large amount of water quality data, combined with predictive models and anomaly detection algorithms to form anomaly warning techniques to support the response to unexpected WP events and improve the efficient management of the water environment [2]. Some scholars have connected the EFD water quality simulation model with the information system through the interface between the Java programming model and the derivation system, which greatly improves the calculation speed and accuracy of the water quality model and realizes the online operation of the Cali model [3]. In short, the EWM to monitor water quality data and issue alerts when water quality is abnormal is a key link in the early warning of WP.

This paper first analyzes the research status of the sudden WP EWM, proposes the risk sources of WP accidents, then introduces the design principles of the EWM and the meaning of the LFM, uses the LFM to build the model logic structure, and finally analyzes the application of the EWM in TP prediction by comparing the predictive value of the LFM and the analog value of Mike model.

2. Basic Overview

2.1. Research Status of Sudden WP Warning Model

In WP EWMs, it is necessary to identify water quality anomalies to warn of water quality, so detection of water quality anomalies has been a focus of society. Anomaly detection can be divided into two categories: statistical methods and machine learning methods [4]. The first type of anomaly detection algorithm is based on statistical principles and determines whether an unknown sample is anomalous based on the statistical characteristics of sample information examples; Gaussian anomaly detection methods, histogram anomaly detection methods and nuclear function anomaly detection methods are all developed on the basis of statistics. This class of algorithms allows sharing of anomalies in unknown statistical distribution models and is easy to implement, but is limited to applications with large sample data [5-6]. The latter class is divided into supervised, semi-supervised and unsupervised learning based on the dataset labels. Supervised is a way to label normal and abnormal water quality, semi-supervised only labels normal water quality conditions, and unsupervised learning does not label all data [7]. According to the actual situation, water quality anomalies are more difficult to make labeling, so unsupervised and semi-supervised are more commonly used in water quality anomaly detection.

2.2. Risk Sources of Sudden WP Accidents

WP risk sources are factors that may lead to WP incidents, including people, objects and events with hazardous and uncertain characteristics [8]. Risk sources can be divided into two categories in terms of source, namely man-made risk sources and natural risk sources. The former is a series of risk sources that cause WP accidents due to human conscious or unintentional moves, including urban sewage discharges, industrial wastewater discharges, agricultural wastewater discharges and pollution discharges caused by transportation accidents [9]; natural risk sources are those that cause WP due to sudden climate change or natural disasters such as earthquakes [10]. Based on distribution characteristics, risk sources can be classified as fixed sources, watershed pollutants and mobile pollutants. Fixed PSs refer to the source of pollutants is in a place, that is, the discharge of pollutants in the same geographical location, fixed sources of pollution can usually be regarded as point sources in the WP EWM, the spread of its pollution usually gradually from point source pollution to surface source pollution, the scope of pollution is generally small, such as industrial pollution, domestic sewage pollution and so on [11]; watershed pollution mainly from the watershed near the residents of household-generated Pollutants, which can usually be regarded as distributed sources in the model, urban domestic sewage, pesticide residues, agricultural drainage and other pollutants enter river water bodies through rainfall [12]; mobile PSs are PSs characterized by the discharge of large amounts of pollutants and unknown discharge points, such as tanker traffic accidents on bridges across rivers, leakage accidents from ships, etc [13].

When a WP accident occurs, the transfer of water pollutants follows the convective diffusion equation. The basis of the inversion problem for the source term of river WP is a coupled hydrodynamic-water quality system with the following control equations:

$$\nabla \cdot u = 0 \tag{1}$$

$$\frac{\partial u}{\partial t} + u \cdot \nabla u = -\frac{\nabla p}{\rho} + v \nabla^2 u \tag{2}$$

$$p = \gamma \cdot h \tag{3}$$

Where u is the fluid velocity vector, m/s; T is the time, s; P is the fluid pressure, Pa; ρ Is the fluid density, kg/m3; V is the viscosity coefficient of fluid motion, Pas; h is the water height; γ is the parameter term.

3. Overall Design of EWM for Sudden WP

3.1. Design Objectives and Principles

A sudden WP emergency warning model is created which simulates and quickly predicts sudden WP and displays simulation results, while emergency information can be viewed through a database linked to the system, providing technical support for government environmental agencies to develop sudden emergency plans for the purpose of reducing the hazard of sudden WP incidents [14]. In order to create an EWM system with a simple interactive interface and powerful simulation calculations, its overall design as well as functional implementation should follow the following principles.

(1) Principle of rapidity

WP emergencies affect a large area and it is hoped that the WP EWM will be able to make response decisions as quickly as possible, so the first principle of EWM design is efficiency, which can also be called speed. Want the system to be fast need to meet two requirements: first, the system simulation speed, need to complete the accident simulation within a few minutes, such as parameter input, simulation calculation and graphics display total time; second, in order to support the rapid query information after the accident, decision makers should be able to quickly retrieve decision support information, including laws and regulations, pollution confidential information, etc., and make decisions based on this information and simulation prediction graphics [15- 16].

(2) Principle of fluency

As the end-users of the model are not GIS software experts and have no experience in GIS, the software is required to follow the fluency principle during the development process to minimise parameter configuration and simplify operational steps while meeting the basic requirements of the accident simulation parameters, so that non-specialist end-users can master the use of the system in a very short time [17]. The user database is then designed and operated as required, thereby simplifying data editing and avoiding data redundancy due to wasted resources.

(3) Openness principle

Sudden WP EWM itself is an innovative research topic, the model database function module and content ambiguous, not perfect, so the system in the development design and use of the test process to follow the principle of openness, adopt the user given modification suggestions, and constantly test the system performance, find out the system failure and deficiencies, and gradually improve the system function.

3.2. EWM of Sudden WP Based on Latent Factor

(1) Latent Factor Model (LFM)

LFM is a machine learning algorithm in which there is an implicit relationship between the user and the object, subject to implicit factors, which are difficult to interpret in terms of implicit factors. In order to find suitable implicit factors so that the objective function can be optimal, the process of actually making recommendations has to be trained continuously using data [18].

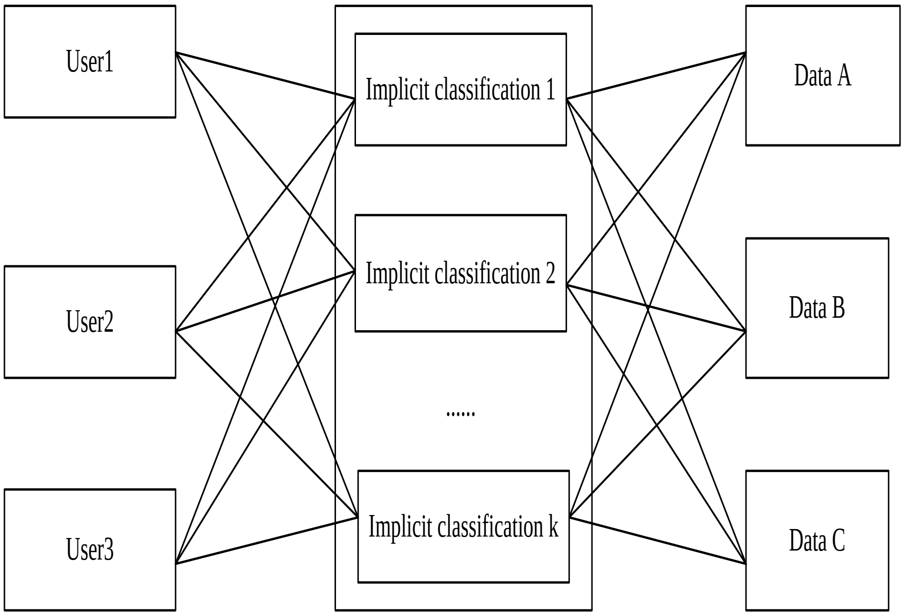

Figure 1. User - implicit classification - data relationship

A diagram of the user-implicit classification-data relationship is shown in Figure 1. Appropriate positive and negative feedback data is selected for both explicit and implicit feedback data to construct positive and negative sample ratios. Positive feedback data is better chosen, for example, user ratings of items, and products that users frequently browse can be used as positive feedback data. When users do not frequently perform browsing actions on products of interest, it is obviously impractical to select negative feedback data. In terms of negative sample selection, a typical negative sample is something that is popular, but users do not explicitly express their preferences for such things, as many people buy some popular products because they appear frequently, but users do not express their preferences and it is not necessarily assumed that users like such items [19].

(2) Model logic structure based on LFM

The sudden WP EWM is the second development of GIS based on the characteristics of WP emergencies, the process of using the model involves a large amount of basic geographical information, basic environmental information and mathematical model calculations, for the design objectives and principles of the model, the concept of layered design, the system is designed as an organic whole consisting of data layer, business logic layer, user layer, these three logical layers respectively These three logical layers correspond to the database, functional links, user interface three structural layers, the three interact with each other to complete the technical support for sudden WP emergency decision-making. The logical structure of the model is shown in Figure 2.

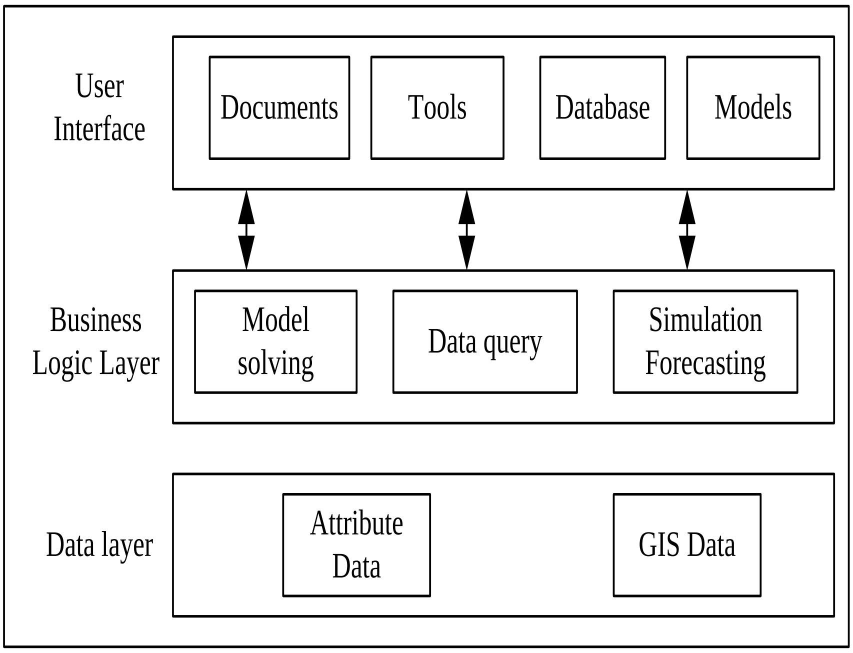

Figure 2. Logic structure of sudden WP EWM

4. Application of Sudden WP Warning Model

4.1. TP EWM Result Analysis

In this experiment, an EWM was used to predict the concentration of TP in water. There are 623 data of total phosphorus (TP) in WP, 500 of which are randomly selected to train the warning model based on the implicit meaning. When the result of the implicit meaning is within the error range, the training is stopped. In order to verify the accuracy of the model, the remaining 123 data were simulated and predicted using the EWM based on the implicit meaning, and the predicted value(PV) based on the implicit meaning was compared with Mike's simulated value(SV).

Table 1. Simulated and predicted values of peak concentration

Distance from the point of accident	LFM PV	Mike's SVs
0	1.52	1.49
10	0.34	0.35
20	0.23	0.24
30	0.16	0.16
40	0.10	0.11
50	0.09	0.085
60	0.084	0.083
70	0.082	0.083

It is known from Table 1 that the PV of peak concentration is consistent with the SV. When it is at the point of WP accident, the PV of the cryptic model is 1.52, and the SV of Mike model is 1.49. The farther away from the accident place, the smaller the PV and SV are. The value after 40m is basically 0.08-0.09.

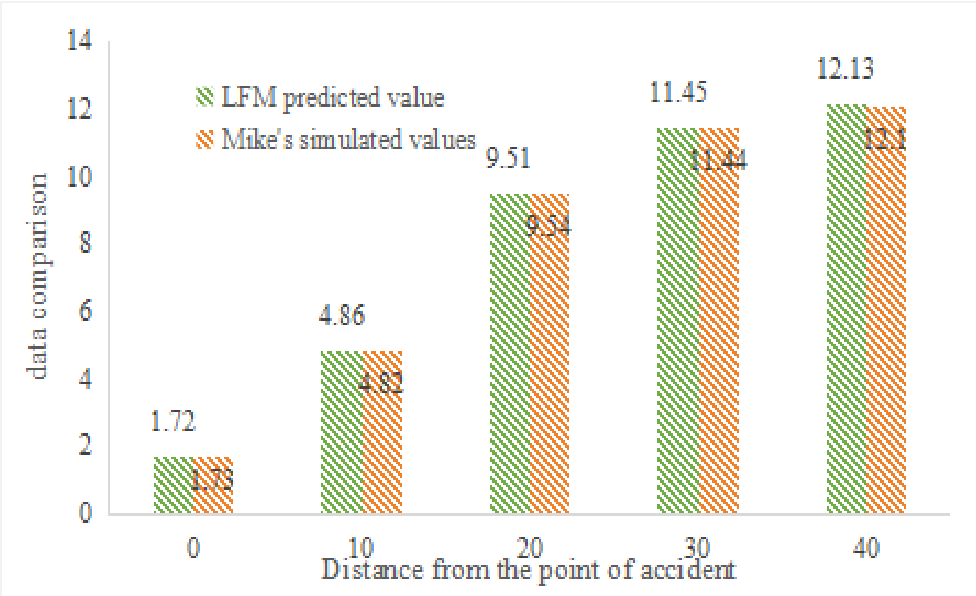

Figure 3. Simulated and predicted values of peak density distance from source

It can be seen from Figure 3 that the PV of the distance between the peak concentration and the PS is basically consistent with the SV. The result of the training data of the distance between the peak density and the PS in the cryptic model and Mike model increases with the increase of the distance from the accident occurrence point.

4.2. Error Comparison

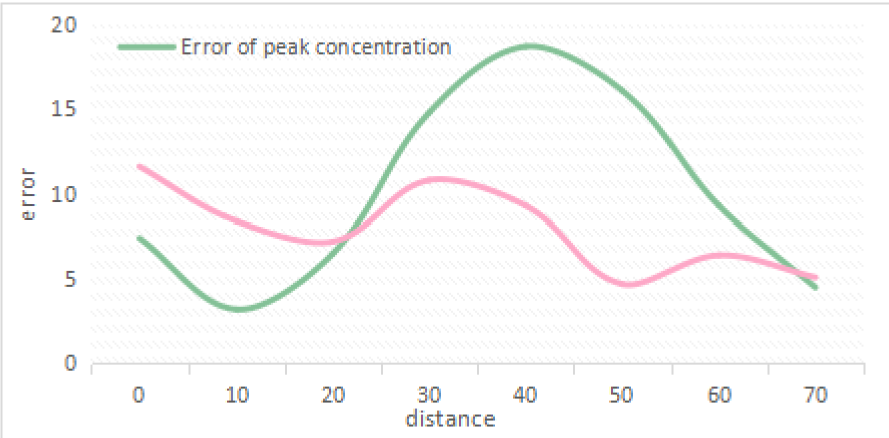

Figure 4. Error results (%)

As shown in Figure 4, the error results between the PV and the SV of the peak concentration and the distance from the peak concentration to the PS at different accident points are shown. The error between the PV of peak concentration and the SV is basically kept at 10%, and the error between the SV and the PV is more than 10%, or even close to 20%, at the data mutation (30-50m), but it meets the accuracy requirements on the whole. The maximum error between the PV and the SV of the distance between the peak concentration and the PS is 11.6%. Therefore, it is believed that the EWM based on the implicit meaning can quickly and effectively predict the distribution characteristics of pollutants after the occurrence of sudden WP accidents, provide a basis for emergency treatment of accidents, and reduce the losses caused by accidents.

5. Conclusion

The sudden WP EWM established in this paper can detect water quality abnormalities in a timely manner through water quality prediction and dynamic early warning, improve the efficiency of emergency response, reduce the risk of WP, and prevent WP from spreading over a large area through continuous improvement of monitoring technology. With the continuous improvement of water quality monitoring technology in China, conventional water quality indicators can be monitored online. In addition, the crypto semantic model provides a basis for data mining and has a good application prospect in water quality prediction. Therefore, this paper studies the water quality prediction and early warning function based on the argot meaning, which is of great significance for dealing with sudden WP events, helping people prevent WP events and make emergency decisions after WP events.

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] Lily Popova Zhuhadar, Mark Ciampa. Novel findings of hidden relationships in offshore tax-sheltered firms: a semantically enriched decision support system. J. Ambient Intell. Humaniz. Comput. (2021) 12(4): 4377-4394. https://doi.org/10.1007/s12652-019-01392-1

[2] Samuel A. Ajila, Chung-Horng Lung, Anurag Das. Analysis of error-based machine learning algorithms in network anomaly detection and categorization. Ann. des Telecommunications. (2022) 77(5-6): 359-370. https://doi.org/10.1007/s12243-021-00836-0

[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. (2021) 5(2): 693- 702. https://doi.org/10.1109/TGCN.2021.3062470

[4] Amal Agarwal, Lingzhou Xue. Model-Based Clustering of Nonparametric Weighted Networks With Application to Water Pollution Analysis. Technometrics. (2020) 62(2): 161-172. https://doi.org/10.1080/00401706.2019.1623076

[5] Mohamed S. Abdalzaher, M. Sami Soliman, Sherif M. EI-Hady, Abderrahim Benslimane, Mohamed EIwekeil. A Deep Learning Model for Earthquake Parameters Observation in IoT System-Based Earthquake Early Warning. IEEE Internet Things J. (2022) 9(11): 8412-8424. https://doi.org/10.1109/JIOT.2021.3114420

[6] Luis Otero-Gonzalez, Dildora Ibragimova, Onofre Martorell Cunil, Luis-lgnacio Rodriguez Gil. EWMs for European banks. J Intell. Fuzzy Syst. (2020) 38(5): 5455-5462. https://doi.org/10.3233/JIFS-179637

[7] R. Saravana Ram, M. Vinoth Kumar, N. Krishnamoorthy, A. Baseera, D. Mansoor Hussain, N. Susila. Industrial Centric Node Localization and Pollution Prediction Using Hybrid Swarm Techniques. Comput. Syst. Sci. Eng. (2022) 42(2): 545-460. https://doi.org/10.32604/csse.2022.021681

[8] Javier de Frutos, Victor Gaton, Paula M. Lopez-Perez, Guiomar Martin-Herran. Investment in Cleaner Technologies in a Transboundary Pollution Dynamic Game: A Numerical Investigation. Dyn. Games Appl. (2022) 12(3): 813-843. https://doi.org/10.1007/s13235-022-00445-z

[9] Ratul Lahkar, Vinay Ramani. An Evolutionary Approach to Pollution Control in Competitive Markets. Dyn. Games Appl. (2022) 12(3): 872-896. https://doi.org/10.1007/s13235-021-00412-0

[10] Akihiko Yanase, Keita Kamei. Dynamic Game of International Pollution Control with General Oligopolistic Equilibrium: Neary Meets Dockner and Long. Dyn. Games Appl. (2022) 12(3): 751-783. https://doi.org/10.1007/s13235-022-00434-2

[11] Sandeep Mathur. Modeling the stubble burning generated airborne contamination with air pollution components through MATLAB. Earth Sci. Informatics. (2022) 15(2): 1137-1146. https://doi.org/10.1007/s12145-022-00787-7

[12] Paul D. Rosero-Montalvo, Vivian F. Lopez Batista, Ricardo P. Arciniega-Rocha, Diego Hernan Peluffo-Ordonez. Air Pollution Monitoring Using WSN Nodes with Machine Learning Techniques: A Case Study. Log. J. IGPL. (2022) 30(4): 599-610. https://doi.org/10.1093/jigpal/jzab005

[13] Sudan Jha, Sidheswar Routray, Sultan Ahmad. An expert system-based loT system for minimisation of air pollution in developing countries. Int. J. Comput. Appl. Technol. (2022) 68(3): 277-285. https://doi.org/10.1504/IJCAT.2022.10049754

[14] Wael K. Hanna, Rasha Abdel sattar EI Stohy, Nouran M. Radwan. Prediction of air pollution and analysis of its effects on the pollution dispersion of PM10 in Egypt using machine learning algorithms. Int. J. Data Min. Model. Manag. (2022) 14(4): 358-371. https://doi.org/10.1504/IJDMMM.2022.10051600

[15] Angelika Zube, Dominik Kleiser, Alexander Albrecht, Philipp Woock, Thomas Emter, Boitumelo Ruf, lgor Tchouchenkov, Aleksej Buller, Boris Wagner, Ganzorig Baatar, Janko Petereit. Autonomously mapping shallow water environments under and above the water surface. Autom. (2022) 70(5): 482-495. https://doi.org/10.1515/auto-2021-0145

[16] Victor Manuel Zezatti, Alberto Ochoa, Gustavo Urquiza, Miguel Basurto, Laura Castro, Juan Garcia. The Implementation of a Nickel-Electroless Coating in Heat Exchanger Pipes Considering the Problem of the Environmental Conditions ofthe Cooling Water Without Recirculation to Increase the Effectiveness Under Uncertainty. Int. J. Comb. Optim. Probl. Informatics. (2022) 13(4): 73-82.

[17] Totan Garai, Harish Garg. Possibilistic multiattribute decision making for water resource management problem under single-valued bipolar neutrosophic environment. Int. J. Intell. Syst. (2022) 37(8): 5031-5058. https://doi.org/10.1002/int.22750

[18] Suresh Muthulingam, Suvrat S. Dhanorkar, Charles J. Corbett. Does Water Scarcity Affect Environmental Performance? Evidence from Manufacturing Facilities in Texas. Manag. Sci. (2022) 68(4): 2785-2805. https://doi.org/10.1287/mnsc.2021.4013

[19] Tristan Thielmann. Environmental conditioning: Mobile geomedia and their lines of becoming in the air, on land, and on water. New Media Soc. (2022) 24(11): 2438-2467. https://doi.org/10.1177/14614448221122190