Dynamic Control Method for Urban Flood Control and Drainage System Based on Multi-source Hydrological Data

Zhifeng Wang

Anhui Transport Consulting & Design Institute Co., Ltd., Hefei, Anhui, 230000, China

Keywords: Multi source hydrological data; Urban flood control; Drainage system; Dynamic regulation; Big data technology

Abstract: In recent years, extreme rainfall has occurred frequently, and flood disasters have become increasingly severe, posing new challenges to the flood control and disaster reduction system of high-density cities in northern China. The traditional early warning and emergency response mechanisms are no longer able to cope with the complex and ever-changing urban waterlogging risks. It is urgent to integrate new generation information technologies such as big data and artificial intelligence (AI) to build a smart flood control decision-making system with "full information integration, full process control, and full social participation". This article focuses on the practical problems of high-density urban areas with dense buildings, concentrated population, multiple causes of floods, and diverse and heterogeneous data sources. It innovatively proposes a multi-source heterogeneous data management scheme for flood control and drainage based on the "data lake" architecture. This architecture effectively integrates multi-source hydrological data, systematically studies data collection, cleaning, standardized storage, and real-time processing processes, breaks through the bottleneck of traditional data silos, and achieves efficient aggregation and intelligent analysis of cross departmental and cross platform data. The research results provide a solid data foundation and decision-making support for the dynamic perception of urban flood risks, accurate early warning and forecasting, and scientific emergency dispatch.

1. Introduction

With the urban expansion, land hardening, drainage system lag and climate change superimposed effects, the traditional hydrological cycle process has been significantly reconstructed: the rainfall runoff response has accelerated, the convergence path has changed, the surface storage capacity has declined, the "rain island effect" has intensified, extreme rainstorm events have occurred frequently, and urban waterlogging has become increasingly serious [1]. This not only seriously threatens the safety of residents' lives and property, but also triggers secondary disasters such as traffic paralysis, infrastructure damage, and environmental pollution, becoming a key bottleneck restricting the construction of urban public safety system and high-quality economic and social development [2]. In this context, how to scientifically construct a flood disaster warning mechanism, accurately assess risk levels, and optimize emergency dispatch decisions has become a core issue that urgently needs to be addressed in modern urban governance [3]. The urban flood model, as the scientific basis for disaster prevention and control, engineering design, risk assessment, and early warning and forecasting, has long been a focus of academic research both domestically and internationally [4].

Research has mainly focused on the impact of urbanization on hydrological response mechanisms, simulation of surface runoff, evaluation of pipeline drainage capacity, and fine modeling of coupled meteorological hydrological hydraulic processes [5]. However, the effectiveness of the model highly relies on high-quality, multidimensional, and timely foundational data support [6]. In reality, urban flood related data is often scattered across multiple departments such as meteorology, water conservancy, housing and emergency management, with heterogeneous data formats, inconsistent standards, and lagging updates, forming a serious "data island" that greatly restricts model accuracy and decision-making efficiency [7]. In addition, although the widely

  DOI: 10.25236/icacel.2025.096

used indicator system evaluation method has certain universality at the municipal level, its weight setting mostly relies on expert experience or static assignment, lacking dynamic adaptability; Due to differences in geography, climate, and infrastructure, it is difficult to unify the selection and weight allocation of indicators in different cities, resulting in poor comparability and weak guidance of evaluation results [8]. It is particularly crucial that most existing flood resilience assessment frameworks fail to systematically characterize the resilience evolution characteristics of the entire process from pre disaster to mid disaster to post disaster, resulting in a single assessment dimension and fragmented temporal sequence.

At the same time, weight calculation methods often use single weighting techniques such as entropy weighting, which fail to integrate multi-source information, dynamic feedback mechanisms, and machine learning (ML) optimization strategies, making it difficult to comprehensively reflect the differences in importance of various indicators in different disaster stages. With big data AI、The rapid development of new generation information technologies such as the Internet of Things (IoT) and intelligent analysis and decision support systems based on multi-source heterogeneous data fusion are becoming the forefront direction in the field of urban flood control and disaster reduction. This article innovatively proposes a multi-source heterogeneous data management scheme for flood control and drainage based on the "data lake" architecture. This architecture breaks through the structural limitations of traditional databases and supports elastic storage and efficient calling of massive structured and unstructured data; Through standardized collection, intelligent cleaning, semantic alignment, and real-time stream processing, seamless integration and dynamic updates of cross departmental and cross platform data can be achieved; The ultimate goal is to build an intelligent data center that covers the entire chain of "perception analysis warning decision-making feedback", providing a solid data foundation and intelligent engine for dynamic assessment of urban flood risks, quantification of resilience capabilities, and precise scheduling of emergency resources.

2. Difficulties in Building Flood Models and Flood Disaster Indicators

2.1 Building Difficulties

The construction of flood models faces multiple systemic difficulties, rooted in the fact that urban flood control involves two major administrative systems: water conservancy and municipal administration. There are significant differences in industry norms, design standards, and management objectives, which often lead to "each working independently" in engineering planning and implementation, lacking overall coordination at the watershed scale. In reality, flood control projects (such as dams and reservoirs) and drainage facilities (such as pipelines and pumping stations) are often designed independently and scheduled separately, which can easily lead to a vicious cycle of "flooding caused by floods and flooding promoted by floods", weakening the overall disaster prevention effectiveness [9]. At the same time, the complex spatial structure of high-density cities, highly heterogeneous underlying surfaces, and intensive human activities have led to the characteristic of "multi-source heterogeneity, multi-dimensional massive" flood causing factors [10]. Multiple heterogeneous data information are scattered and have different formats, which not only increases the difficulty of identifying disaster mechanisms and modeling multi factor coupling, but also puts higher demands on intelligent mining and dynamic response of warning precursor information.

In the plain river network embankment area, flood control and drainage mainly rely on the "joint discharge and adjustment" of water gates, pump station groups, and regulating water systems, and risk control is achieved through the exchange of water bodies in the internal and external river networks. However, due to the dense distribution of rivers, the large number of gates and pumps, and the highly nonlinear water flow dynamics, the macro "water balance method" is currently widely used to estimate the scale of drainage. Although this method is convenient for overall planning, it is difficult to depict the local hydrodynamic response, the coordinated scheduling process of gates and pumps, and the spatiotemporal evolution of storage facilities, resulting in a disconnect between engineering design and actual operation, which restricts the scientific and practical implementation of the "flood control" strategy. The deeper challenge lies in the fact that traditional flood warning models often focus on natural factors such as rainfall intensity and river water levels, neglecting social dimensions such as human activities, infrastructure resilience, and emergency management capabilities, and lacking a systematic warning framework that couples "nature society". In addition, the scientific setting of flood control water levels, start stop thresholds for drainage pump stations, and the highest warning water level in inland rivers in plain areas are influenced by multiple factors such as river network topology, ground elevation, development intensity, and engineering systems. Currently, there is a lack of a unified, dynamic, and adaptable water level control mechanism.

2.2 Indicator System

Urban flood disasters are essentially the product of the interaction between natural extreme events (such as heavy rainfall and storm surges) and the urban socio-economic system, and are a concrete manifestation of the "human environment relationship" in the dimension of disasters. It is not only a hydro meteorological process, but also a complex system that integrates natural environment, human activities, infrastructure, and emergency management capabilities. According to the theory of natural disaster systems, flood disasters are the result of the dynamic coupling and interaction of "disaster prone environment, disaster causing factors, and disaster bearing bodies" (as shown in Figure 1), and the construction of a scientific and systematic evaluation index system needs to be carried out in coordination from these three dimensions. Firstly, the sensitivity of the disaster prone environment reflects the natural susceptibility of the region to flood disasters, and is the fundamental carrier of risk incubation. The evaluation should focus on natural factors such as terrain undulation, river network density, surface permeability, vegetation coverage, etc., to characterize the regional hydrological response capacity and storage and discharge potential, and identify highly sensitive and vulnerable areas. Secondly, the risk of disaster causing factors reflects the intensity and frequency of disaster occurrence.

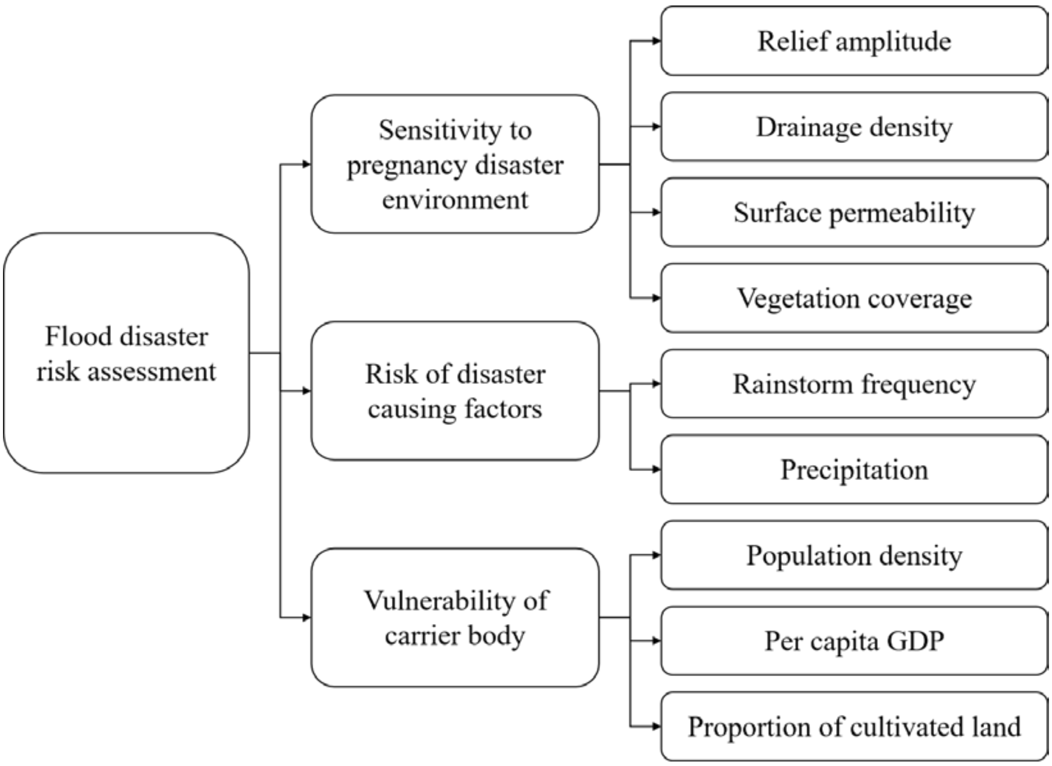

Figure 1 Flood disaster indicator system

Rainstorm and flood have both sudden and continuous characteristics, so the indicator design needs to take into account the dual dimensions of "total amount" and "intensity". It is recommended to select "annual maximum daily rainfall", "hourly maximum rainfall intensity" to represent the impact of extreme precipitation, supplemented by "annual rainstorm days" and "return period rainstorm frequency" to reflect the space-time law of disaster occurrence, so as to comprehensively describe the disaster causing dynamics. Thirdly, the vulnerability and resilience of disaster bearing bodies constitute the core expression of disaster impact. The disaster bearing body is the spatial carrier of human activities such as urban population, economy, and infrastructure. Its vulnerability refers to the degree of loss that may be suffered under specific flood impacts. Indicators such as "population density", "per capita GDP", "proportion of built-up areas", and "proportion of cultivated land" are commonly used to measure exposure and vulnerability. At the same time, urbanization not only exacerbates exposure risks, but also enhances disaster resilience by improving infrastructure levels and emergency response capabilities. In summary, the indicator system for urban flood disasters should break through a single natural dimension and construct a four-dimensional linkage framework of "environmental sensitivity disaster drive exposure vulnerability resilience support", taking into account both static characteristics and dynamic capabilities, providing systematic support for scientific assessment, accurate warning, and differentiated governance.

3. System Construction Based on the "Data Lake" Architecture

3.1 System Construction

This article focuses on the characteristics of multi-source heterogeneous data, strong real-time performance, complex structure, and high demand for cross departmental collaboration in flood control and drainage business scenarios. It innovatively constructs a "data lake" architecture management system for urban flood real-time dispatch decision support (as shown in Figure 2). This architecture is divided from bottom to top into a data collection layer, a data storage layer, a data transformation layer, and a data service layer, forming a data center platform covering the entire chain of "collection storage governance application". The data collection layer serves as the system entrance and supports integration with multiple heterogeneous data sources such as meteorological radar, hydrological monitoring stations, IoT sensors, and municipal pipeline GIS. It adopts a "native loading" strategy to preserve the original data semantics and structure to the maximum extent possible, avoiding information loss caused by preprocessing. At the same time, by automatically extracting metadata and structurally storing it in a metadata database or configuration file, automatic registration and dynamic tracking of data assets can be achieved, greatly improving data access efficiency and manageability.

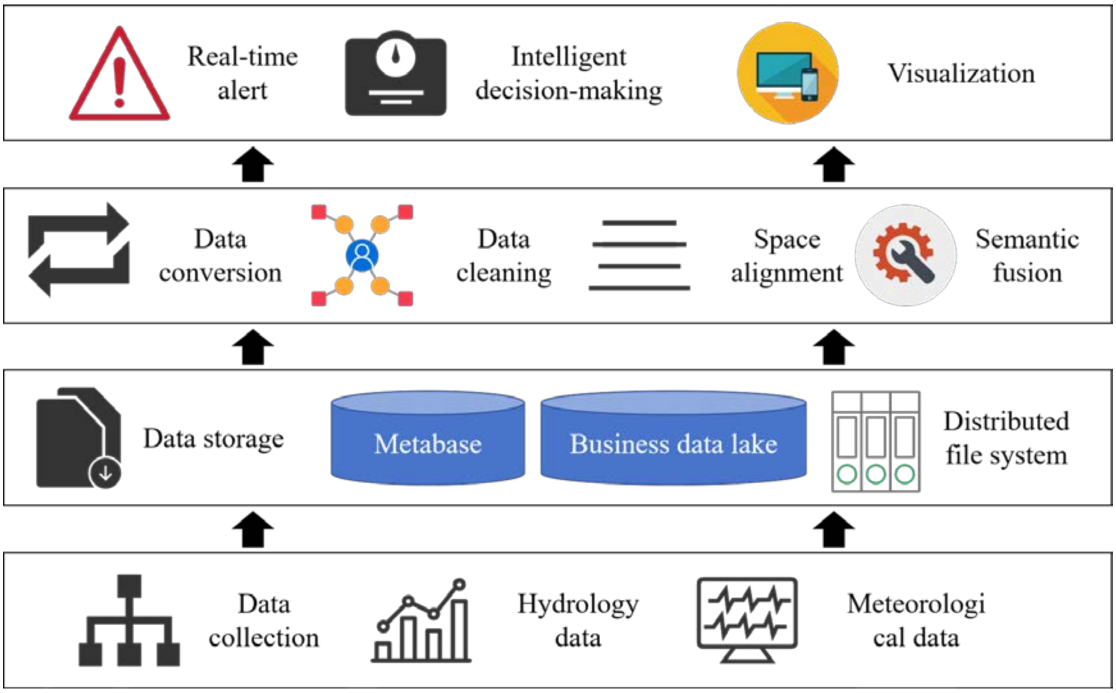

Figure 2 System structure

The data storage layer adopts a hybrid storage architecture, with the core consisting of a "meta database" and a "business data lake". The meta database centrally manages descriptive information of all data sources, supporting data lineage tracing and permission control; The business data lake is compatible with structured, semi-structured, and unstructured data, and relies on distributed file systems to achieve elastic storage of massive data. By encapsulating a unified JSON API interface, shielding underlying storage differences, providing standardized and high-performance data query and access capabilities for upper layers, and supporting complex spatiotemporal correlation analysis. The data conversion layer performs operations such as cleaning, standardization, spatial alignment, and semantic fusion on the raw data, achieving efficient conversion from raw data to commercially available data based on the needs of flood control dispatch business. This layer supports integrated stream batch processing, ensuring dual capabilities of real-time warning and historical backtracking. The data service layer serves as the system outlet, targeting multiple types of users such as flood control command, emergency management, urban planning, etc. It provides diversified services such as visual components and decision model interfaces, realizing on-demand data calling, intelligent recommendation, and collaborative sharing.

3.2 Algorithm Support

Based on the relevant theories of the formation mechanism of urban flood disasters, the occurrence of flood disasters is not a single event, but a three-stage evolutionary process driven by both natural processes and social responses: the first stage is the concentrated occurrence of heavy rainfall - specific areas experience precipitation intensity or total amount exceeding their carrying capacity in a short period of time, exceeding the comprehensive drainage capacity of surface infiltration, pipeline discharge, and river flood discharge; The second stage is waterlogging caused by water accumulation - rainwater cannot be promptly absorbed, resulting in surface stagnant water or internal flooding, marking the formation of the physical form of the disaster; Based on the above evolutionary logic, this article starts from the inherent mechanism of disaster formation, systematically deconstructs the chain of flood disaster generation, and constructs a risk assessment framework that integrates "risk of waterlogging - sensitivity to waterlogging - vulnerability to waterlogging". Flood causing factors:

$$M(X)=\sum_{j=1}^{i}\left[W_j \times M_{ji}(X)\right] \tag{1}$$

Pregnancy flood factor:

$$B(X)=\sum_{j=1}^{i}\left[W_j \times B_{ji}(X)\right] \tag{2}$$

Flood bearing factor:

$$S(X)=\sum_{j=1}^{i}\left[W_j \times S_{ji}(X)\right] \tag{3}$$

Flood risk coefficient:

$$R(X)=W_M M(X)+W_B B(X)+W_S S(X) \tag{4}$$

In the formula, $M_{ji}(X), B_{ji}(X), S_{ji}(X)$ is the normalized value of each indicator; $M(X), B(X), S(X)$ represents the risk of waterlogging, sensitivity to waterlogging, and vulnerability to waterlogging; W_M, W_B, W_S is its corresponding weight, $R(X)$ is the comprehensive risk coefficient of regional floods, reflecting the overall risk level after weighted integration of the three.

4. Conclusions

Flood disasters, as one of the most frequent and destructive natural disasters in the world, seriously threaten the safety of urban residents' lives and property, as well as the stable operation of the social economy. Its risk warning and intelligent decision-making have become the core issues of the national flood control security system and the modernization of urban public governance. This paper focuses on the practical bottlenecks faced by high-density cities in the process of flood control and drainage, such as "data islands, heterogeneous formats, lagging updates, and disjointed dispatching". It systematically analyzes the multi-source, real-time, and complexity characteristics of urban flood data, and innovatively proposes a data lake architecture solution oriented to practical dispatching needs. This system achieves efficient aggregation, intelligent cleaning, and standardized storage of multi-dimensional heterogeneous data such as meteorology, hydrology, pipeline networks, remote sensing, municipal administration, and public opinion by constructing a data management process covering the entire chain of "collection cleaning storage service"; Relying on a distributed elastic architecture, it supports high-speed writing and horizontal expansion of massive data, effectively breaking the limitations of traditional databases in terms of scale and flexibility. The research results not only provide a solid, unified, and traceable data foundation for real-time warning, dynamic simulation, and emergency dispatch of urban floods.

References

[1] Chen Shu, Yuan Weikang, Yan Kesheng, et al. Comprehensive assessment of urban flood disaster resilience in Hubei Province by integrating multi-source data[J]. Water Resources and Hydropower Technology (Chinese and English), 2024, 55 (03): 51-60.

[2] Cheng Penggen, Huang Yi, Guo Fusheng, et al. Risk assessment of urban flood disasters based on multi-source data[J]. Journal of Disaster Science, 2022, 37 (03): 69-76.

[3] Li Mengqi, Li Gongquan, Xie Zhihui. Flood disaster assessment and recovery by integrating multi-source remote sensing data -- Taking the rainstorm disaster of "July 20" in Henan Province as an example[J]. Remote Sensing of Natural Resources, 2024,36 (01): 250-266.

[4] Li Guanyi, Yu Zexing. Mountain torrent disaster early warning research based on multi-source rainfall data integration[J]. People's the Pearl River, 2022,43 (10): 90-95.

[5] Xing Ziyao, Dong Xinrui, Yang Xiaoli, et al. A method for extracting and simulating flood ranges by integrating multi-source data such as VGI and remote sensing[J]. Journal of Earth Information Science, 2023, 25 (09): 1869-1881.

[6] Cao Yong, Zhang Hengde, Guo Yunqian, et al. Short term quantitative precipitation forecasting based on multimodal forecasting optimization fusion technology[J]. Journal of Hohai University (Natural Science Edition), 2022, 50 (05): 10-15.

[7] Wang Huimin, Huang Jing, Liu Gaofeng, et al. Research paradigm of big data-driven urban flood risk perception and early warning decision-making[J]. Frontiers of Engineering Management Technology, 2022, 41 (01): 35-41.

[8] Cao Xuejian, Qi Youcun, Li Mengdi, et al. Research on urban waterlogging risk early warning system under the threat of extreme rainstorm[J]. Atmospheric Science, 2022,46 (04): 953-964.

[9] Luan Qinghua, Qin Zhiyu, Wang Dong, et al. Urban rainstorm road ponding monitoring technology and its application progress[J]. Water Resources Protection, 2022,38 (01): 106-116+140.

[10] Zhuang Huifu, Wang Peng, Su Yanan, et al. Dynamic monitoring of floods and inundation in Zhuozhou based on multi-source time-series SAR data[J]. Natural Resources Remote Sensing, 2024, 36 (04): 218-228.