sustainability

MDPI

Article

A Bidirectional Closed-Loop Coupled SD–BN Model for Metro Rainstorm Disaster Chains: Supporting Sustainable Urban Rail Infrastructure Resilience

Hongxia Tang [1], Zhixuan Liu [1,*], Yang Liu [2], Kai Song [2] and Zhiguo Shao [2]

[1] School of Management Engineering, Qingdao University of Technology, Qingdao 266525, China; happy18023202@126.com
[2] School of Business and Logistics, Qingdao University of Technology (Linyi), Linyi 273400, China; liuyang@qut.edu.cn (Y.L.); songkai1105@126.com (K.S.); shaozhiguo@qut.edu.cn (Z.S.)
* Correspondence: liu73412@163.com

Abstract

As cities advance toward the UN Sustainable Development Goals (SDGs) for resilient, low-carbon urban transport systems, protecting underground metro infrastructure from climate-driven flood risks has become a core priority for sustainable urban operation. Given the complex nature of metro rainstorm disaster chains—characterized by the interplay of continuous evolution, discrete transitions, and probabilistic uncertainty—a single model is insufficient to fully capture their evolutionary process. This study proposes a bidirectional closed-loop coupled model integrating System Dynamics (SD) with Bayesian Networks (BNs), where SD variables are discretized and fed into the BN at each time step, and BN inference results dynamically adjust SD control parameters in the form of risk levels. A three-tier recursive CPT calibration method, combining maximum likelihood estimation, Laplace smoothing, and physical logical constraints, is developed based on 47 historical accident cases with a 13-node network. The model is validated using the Qingdao Metro "July 23" flooding incident. Results show that, compared to the unidirectional coupling model, the closed-loop model dramatically reduces peak station hall water accumulation, shortens operational disruption, and lowers peak passenger retention under idealized instantaneous-response conditions. Specifically, peak station hall water accumulation is reduced by 14.7%, operational disruption duration is shortened from 5.55 h to 0.05 h (99.1% reduction), and peak passenger retention falls from 2790 to 40 passengers (98.6% reduction). These values are controlled model-internal scenario outputs, and are not directly generalizable to other metro systems without local recalibration and empirical validation. The closed-loop mechanism primarily accelerates post-peak risk decline and shortens high-risk exposure duration, rather than suppressing peak risk levels. The core contribution of this work is the transferable bidirectional coupling framework and small-sample CPT calibration workflow, which provides a quantitative basis for dynamic risk assessment and adaptive emergency response in metro flood management—supporting sustainable urban mobility aligned with SDG 11 (sustainable cities and communities) and SDG 13 (climate action).

Keywords: metro rainstorm disaster chain; system dynamics; Bayesian network; bidirectional closed-loop coupling; dynamic simulation

Check for updates

Academic Editor: Dariusz Młyński

Received: 14 September 2026
Revised: 14 October 2026
Accepted: 19 October 2026
Published: 22 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Urban rail transit forms the backbone of low-carbon, sustainable public transport, and its reliable operation is central to achieving SDG 11 (inclusive, safe, resilient and sustainable urban transport) and SDG 13 (improving adaptive capacity to climate-related disasters). Driven by both global climate change and rapid urbanization, the frequency and intensity of extreme heavy rainfall events have been steadily increasing [1]. As the backbone of metropolitan public transportation systems, urban rail transit systems' underground spaces possess inherent characteristics—such as large burial depths, constrained drainage conditions, and high passenger density—which render them particularly vulnerable under extreme rainstorm conditions [2]. During flood disasters, the tight coupling between various subsystems within the system can easily trigger a chain reaction, ranging from surface water accumulation and backflow at stations and exits to tunnel water ingress and track area inundation, ultimately leading to operational disruptions, passenger retention, or even casualties; the consequences of such disasters are dramatically amplified by the enclosed nature of the underground environment [3]. During the severe rainstorm on 20 July 2021, in Zhengzhou, backflow of floodwater into the Line 5 tunnel caused trains to be forced to stop midway and resulted in a large number of passengers being trapped, fully illustrating the typical disaster chain pathway of "heavy rainfall–urban flooding and backflow–train forced stop–operational disruption–personnel casualties" [4]; subsequently, the heavy rainfall events in Shenzhen (2023) and Changsha (2024) both led to the suspension of metro services and damage to infrastructure. Beyond China, metro flood incidents have also been reported in major cities worldwide: the 2021 New York subway flooding caused by Hurricane Ida [5], the 2022 Seoul metro inundation during record rainfall [6], the 2024 Dubai flood that disrupted metro services [7], and the 2025 Mumbai waterlogging event [8] all illustrate that metro flood risk is a global urban challenge. The 2021 New York City pluvial flooding caused by Hurricane Ida severely disrupted the subway system, the 2022 Seoul metro inundation during record rainfall, the 2024 Tokyo metro disruption, and the 2025 Mumbai waterlogging event all illustrate that metro flood risk is a global urban challenge. By the end of 2024, the total operating mileage of China's urban rail transit networks had exceeded 10,000 km [9]. Scientifically elucidating the evolutionary mechanisms of the metro disaster chain triggered by heavy rainfall and dynamically quantifying system risk levels has become a core scientific challenge in ensuring the operational safety of urban rail transit systems [10]. As urban rail transit constitutes a core component of sustainable urban infrastructure under UN Sustainable Development Goal 11 (Sustainable Cities and Communities), enhancing metro flood resilience is integral to climate adaptation under SDG 13 (Climate Action).

Existing research on the analysis and risk assessment of metro disaster chains has been conducted along three main lines of inquiry. One approach employs methods based on disaster chain topology and complex network theory; by extracting risk nodes and transmission pathways from historical incidents, a static network is constructed, and topological indicators are used to quantify the criticality of each node in the disaster propagation process [11]. Other studies, drawing on the catastrophe chain theory, have revealed the structural characteristics of metro disaster chains—namely, their high inherent intensity and strong propagation [12]; additionally, some studies have conducted quantitative ranking of vulnerable nodes and identified critical paths within urban flood disaster chains [13], finding that traffic disruption, emergency material shortage, and command failure are the most critical nodes, and that the propagation edges connecting these nodes play a dominant role in the overall network structure. However, such methods abstract disaster events into discrete binary nodes, making it difficult to address dynamic issues involving the accumulation of physical quantities or variations in transmission rates [14].

The BN approach shifts the analytical focus from "structure" to "causal probability" using directed acyclic graphs to structurally represent the dependencies among variables and quantifying causal strength through conditional probability tables; this enables the dynamic updating of the risk probability distribution across the entire network upon the input of initial evidence [15]. Research in this direction has advanced the application of BNs to metro flooding scenarios from multiple dimensions. Some studies have focused on probabilistic estimation of human casualties and facility damage under heavy rainfall scenarios [16]; other studies have developed risk early-warning models that integrate early-warning information with emergency response measures [17]; still other studies have established systematic risk analysis frameworks based on the modeling of disaster-causing factors [18], demonstrating that integrating early-warning information into the BN model can effectively reduce predicted accident consequences. Compared to static networks, BNs offer an advantage by enabling the handling of uncertainty in disaster propagation; however, their fundamental limitation lies in the fact that they remain, by nature, discrete, static frameworks —where node states are predefined as finite levels, and once the conditional probability tables are calibrated, they no longer update in response to the evolution of the physical process. Although some studies have attempted to incorporate dynamic Bayesian networks [19], their temporal dimension remains centered on discrete state transitions, limiting their ability to capture continuous physical processes driven by differential equations [20].

The SD-based approach shifts the analytical focus from "discrete events" to "continuous processes" centering on stock–flow structures and information feedback loops, and uses systems of differential equations to characterize the continuous accumulation and decay of state variables [21]. The contribution of this method lies in its ability to simulate the temporal continuity of physical quantities during disaster events—such as how water depth increases over time as rainfall continues or how passenger retention accumulates over time—and to assess the system's response trajectory under different management and control measures. In urban stormwater management and the resilience assessment of metro systems against flooding, SD has demonstrated its advantages in continuous dynamic simulation [22]. However, traditional SD models assume that control parameters remain constant throughout the entire simulation; they cannot dynamically adjust strategy intensity based on real-time risk levels, nor do they possess the capability to account for the probability of event occurrence [23].

In the existing literature, complex network approaches excel at static topology identification but lack the ability to characterize continuous dynamics; Bayesian network methods are superior for probabilistic inference, but their conditional probability tables do not update in accordance with the evolution of physical processes; system dynamics methods are more effective for continuous simulation, but their control and management parameters remain constant and they lack mechanisms for handling probabilistic uncertainty. Internationally, metro flood risk research has been conducted in European, North American, and Asian contexts, including BN-based flood risk assessment for urban rail systems and SD-based resilience modeling for transit networks; however, most existing studies focus on either static risk analysis or continuous simulation alone. Most prior studies combining SD and BN approaches involve a unidirectional, loose integration without a strategic closed feedback loop. In the risk analysis of drilling operations by Zhang et al. [24], BN's static probabilistic inference is combined with SD's dynamic simulation to perform a static–dynamic coupled analysis of risk factors; however, the BN inference results are not used to feed back and adjust the SD control parameters. In the dynamic risk assessment of fuel replenishment operations by Khan et al. [25], SD simulation provides time-varying inputs, while BN performs probabilistic inference; the two systems operate in series across time

steps, yet again without establishing a BN→SD strategic feedback channel. Both studies demonstrated the value of static–dynamic coupling for industrial safety risk assessment, but their unidirectional architecture cannot capture the adaptive feedback between risk perception and operational response. None of these studies have implemented a bidirectional closed-loop interaction—namely, real-time SD→BN evidence input and BN→SD strategic feedback—making it difficult to fully capture the complete emergency response process of "perception–decision–action." The metro disaster chain under heavy rainfall represents a typical scenario in which continuous evolution, discrete transitions, and probabilistic uncertainty are intricately intertwined [26]; therefore, a single method or a unidirectional coupling is insufficient to fully characterize this scenario. While international studies have explored coupled SD–BN frameworks for drilling operations and marine fuel replenishment, their application to metro flood disaster chains–and particularly bidirectional closed-loop coupling–remains underexplored.

Derived from the above-mentioned research gap, this study addresses three interrelated research questions: How can a bidirectional closed-loop coupling mechanism between SD and BN capture the two-way feedback between physical flood evolution and operational disruption in metro systems? To what extent does the closed-loop coupling mode reduce operational disruption compared with open-loop and one-way coupling under identical rainstorm scenarios? And what are the critical nodes and optimal response windows in the metro rainstorm disaster chain, and how can they support sustainable resilience management? Correspondingly, three hypotheses are proposed to be examined in the subsequent modeling and case-study evaluation: closing the SD→BN→SD feedback loop reduces simulated operational disruption more effectively than open-loop or one-way coupling; critical nodes identified by the entropy-weighted BN align with physically bottlenecked elements in the metro flood system; and interventions activated within the identified response window yield greater resilience benefits than delayed interventions.

To address the aforementioned limitations, this study develops a bidirectional closed-loop coupled dynamic simulation model for the metro disaster chain under heavy rainfall conditions, integrating SD with BN, with core innovations manifested in three aspects. First, an SD–BN bidirectional closed-loop coupling mechanism is proposed; this mechanism facilitates bidirectional data exchange—where the continuously simulated SD variables, after discretization, are fed into the BN at each simulation step, and the BN's inference results are fed back as risk levels to adjust the SD control parameters—thereby achieving real-time synchronous mapping of both the physical evolution of the disaster and the assessment of risk probabilities. This approach overcomes the limitations inherent in previous studies, which often employed either a unidirectional, loose integration or a static series connection between SD and BN. Second, a three-tier recursive conditional probability table calibration method is introduced for scenarios with small sample sizes; this method organically combines maximum likelihood estimation, Laplace smoothing, and physical logical constraints. Using 47 historical metro flood accident cases as the datasets, the parameters for a 13-node network are calibrated. The calibration process successfully passed both the normalization validation based on 109 parent–node combinations and the monotonicity verification; this addresses the shortcoming in existing research where the calibration of conditional probability tables often relied heavily on expert judgment and lacked empirical data support, thereby providing a reusable methodological framework for the parameter calibration of Bayesian networks under small-sample conditions. Third, a disaster chain evolution analysis framework based on critical threshold identification is constructed; this framework utilizes the closed-loop model to quantitatively identify critical transition nodes at each stage of the disaster chain's temporal sequence, revealing the stepwise transmission characteristic of the disaster chain—namely, its progression from

continuous accumulation to discrete transitions—thus providing a quantitative basis for determining the optimal timing of emergency response, and overcoming the limitation in existing disaster chain research, which predominantly relies on static topological analysis and lacks dynamic, quantitative identification of temporal sequences. The research findings can provide a quantitative basis for decision-making in the dynamic risk assessment and emergency response to metro flooding caused by heavy rainfall. While the model is demonstrated in the Qingdao case, the coupling methodology and CPT calibration framework are transferable to other urban metro systems after local parameter recalibration, supporting sustainable flood resilience management aligned with SDG 11 and SDG 13.

2. Data and Model

This study develops a bidirectional closed-loop coupled model for the metro storm-induced disaster chain that integrates system dynamics with Bayesian networks; the overall technical approach is illustrated in Figure 1. The SD module, implemented in Vensim PLE, constructs a stock-flow model and verifies the numerical accuracy of the differential equations; the BN module, implemented in GeNIe, constructs a 13-node network topology and validates the soundness of the CPT-based probabilistic reasoning. The bidirectional data exchange and rolling-horizon iteration between these two modules are implemented using Python programming 3.14.

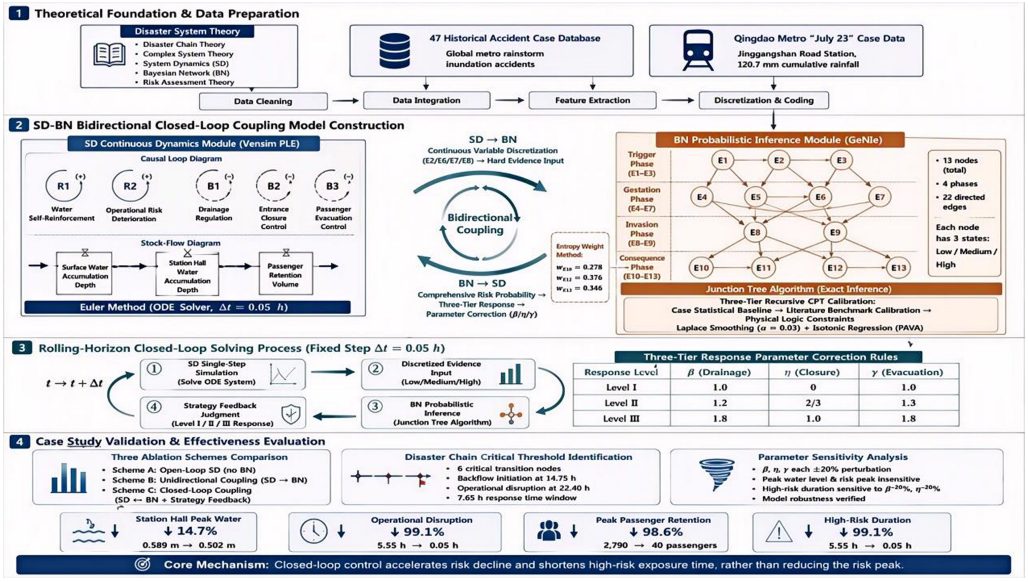


Figure 1. Technical approach.

2.1. SD Continuous Dynamics Module

2.1.1. Cause-And-Effect Diagram

To elucidate the systematic evolution patterns of metro flooding disasters induced by heavy rainfall, a causal loop diagram of the disaster is constructed, as shown in Figure 2. Rainfall intensity, serving as an exogenous driving variable for the system, drives changes in the depth of water accumulation in station halls through a chain-like transmission process involving rainfall runoff generation, surface water accumulation, and backflow of stormwater at station entrances and exits, thereby forming the initial disaster-causing pathway. The model identifies two positive feedback amplification loops and three negative feedback balancing loops, providing a comprehensive depiction of the interplay between the self-evolution of the disaster and human intervention.

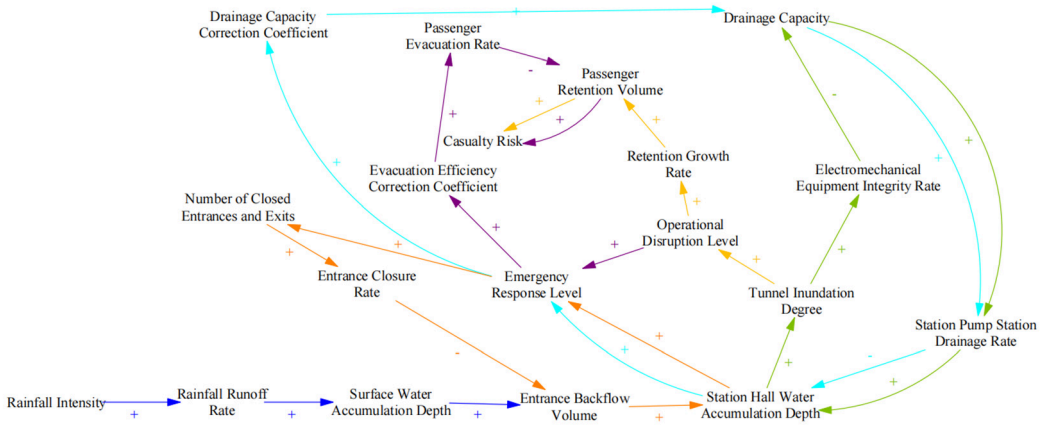


Figure 2. Causal loop model for metro flooding disasters during heavy rainfall. Note: Arrows indicate causal linkages between variables; "+" and "−" denote positive and negative feedback polarities, respectively.

As shown in Figure 2, the positive feedback loop represents the endogenous disaster amplification mechanism. Within the water accumulation self-amplification loop, an increase in the water depth in the station halls exacerbates the extent of tunnel inundation, leading to a decline in the operational performance of electromechanical equipment and a reduction in the system's drainage capacity; as a result, the drainage volume from the station's pumping stations decreases, further aggravating water accumulation and sedimentation within the station and creating a vicious cycle of progressively worsening internal flooding. In the operational risk escalation loop, increased tunnel inundation raises the line operation disruption severity level, accelerates the rate of passenger retention growth, and leads to the continuous accumulation of stranded passengers within the station, thereby progressively increasing the risk of personnel casualties and achieving self-amplification of secondary disaster risks. The negative feedback loop embodies the risk mitigation mechanism inherent in emergency management and control. The drainage regulation loop triggers an emergency response level upon detection of rising water levels, involving the deployment of additional drainage equipment to enhance system drainage efficiency and accelerate the removal rate of water from the station, thereby effectively curbing the continuous rise in water levels; the entrance/exit closure regulation loop utilizes emergency management measures to seal off entrances and exits and block rainwater backflow pathways, reducing the influx of external water and slowing the accumulation of water within the station; the personnel evacuation regulation loop addresses passenger retention resulting from operational disruptions by intensifying emergency evacuation operations to accelerate passenger evacuation efficiency, reduce the number of stranded passengers within the station, and significantly lower the risk of personnel casualties.

2.1.2. Stock-Flow Diagram and Differential Equations

Based on the feedback logic revealed by causal networks, a system dynamics stock-flow diagram for metro flooding disasters under heavy rainfall conditions is constructed, as shown in Figure 3.

Following standard system-dynamics modeling conventions, the causal-loop diagram (Figure 2) provides qualitative feedback logic, which is further translated into quantitative stock-flow formulation (Figure 3): state accumulators are represented as stock variables; material and information change rates derived from causal links are defined as corresponding flow variables; intermediate causal factors that do not accumulate mass are implemented as auxiliary variables. These stocks, flows and auxiliary variables are then mathematically

encoded into the set of differential and algebraic governing equations presented in the manuscript and Supplementary Material S1.

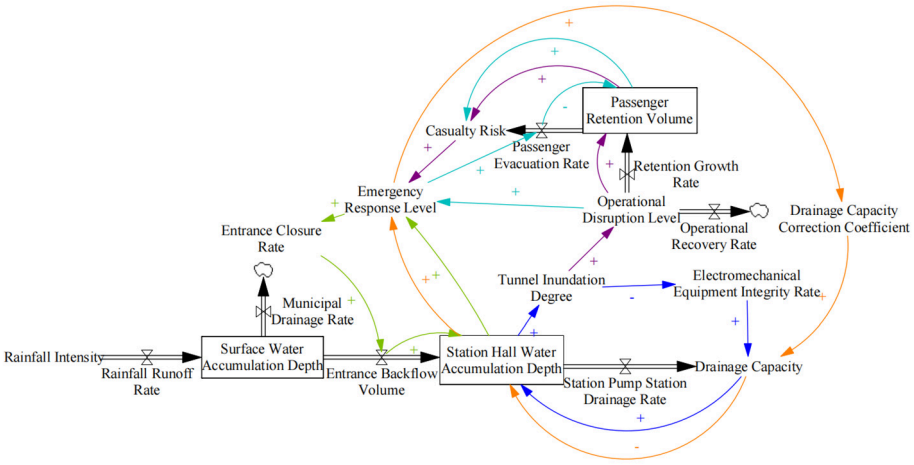


Figure 3. Stock-flow diagram for metro flooding disasters during heavy rainfall Events. Note: Arrows indicate the direction of material flow (valves/flows) and information flow (thin arrows); boxes represent stocks, and double-lined arrows represent conserved flows.

The model uses three stock variables—surface water accumulation depth, station hall water accumulation depth, and passenger retention volume—to represent the system's cumulative state; the rainfall runoff rate, ingress/egress backflow volume, station pump station drainage rate, passenger retention growth rate, and passenger evacuation rate serve as flow variables to characterize the instantaneous rate of change of the system state; all other variables are auxiliary variables; surface runoff forms and triggers backflow of rainwater at entrances and exits, providing continuous replenishment to the water accumulation in the station hall.

Based on the urban hydrology runoff generation and convergence theory and the stock–flow modeling method using system dynamics [27], differential dynamic equations for stock variables are derived, as shown in Equations (1)–(3).

$$\frac{dh_g(t)}{dt} = \frac{I_{rain}(t)}{1000} \cdot \alpha - h_g(t) - R_{drain} - \frac{Q_{in(t)}}{A_g} \tag{1}$$

$$\frac{dh_s(t)}{dt} = \frac{Q_{in}(t) - Q_{leak}(t) - Q_{pump,0} \cdot \beta}{A_s} \tag{2}$$

$$\frac{dN(t)}{dt} = L_{op}(t) \cdot g_0 - V_{evac,0} \cdot \gamma \tag{3}$$

The inlet/exit backflow rate is derived from the basic principle of thin-walled weir flow [28] and is lumped and simplified into a piecewise linear function as shown in Equation (4).

$$Q_{in}(t) = \begin{cases} k \cdot [h_g(t) - h_{th}] \cdot (1 - \eta_{close}), h_g(t) > h_{th} \\ 0, h_g(t) \le h_{th} \end{cases} \tag{4}$$

where $h_g(t)$ is the surface water accumulation depth at time t; $h_s(t)$ is the station hall water accumulation depth at time t; $N(t)$ is the passenger retention volume within the station at time t; $I_{rain}(t)$ is the rainfall intensity at time t; α is the rainfall runoff coefficient; C_r is the surface runoff coefficient; R_{drain} is the municipal drainage rate; A_g is the equivalent surface water accumulation area; h_{th} is the entrance water threshold height; $Q_{in}(t)$ is the entrance backflow rate; k is the backflow coefficient; $\eta_{close}(t)$ is the entrance closure rate; A_s is the equivalent station hall water accumulation area; $Q_{pump,0}$ is the pumping station design

discharge; β is the drainage capacity correction coefficient; $Q_{leak}(t)$ is the seepage volume; g_0 is the baseline retention growth rate per unit disruption level; $L_{op}(t)$ is the operational disruption level; $V_{evac,o}$ is the baseline evacuation rate; γ is the evacuation efficiency correction coefficient. $Q_{in}(t)$, $Q_{leak}(t)$, and $Q_{pump,0}$ are volumetric flow rates (m^3/h), consistent with the thin-walled weir flow formulation in Equation (4); dividing by the corresponding area yields the equivalent depth rates (m/h) applied in the stock equations.

Tunnel inundation depth and track-area water depth are formulated as auxiliary variables derived from station hall water accumulation depth: tunnel inundation depth equals zero until station hall water depth exceeds the 0.10 m threshold, after which water propagates into the tunnel; track-area water depth then equals tunnel inundation depth minus a 0.05 m freeboard. Operational disruption level is determined by track-area water depth thresholds (0.20 m and 0.35 m) specified in the Code for Design of Metro (GB 50157-2013) [29]. This auxiliary mapping captures the stepwise propagation from station hall to tunnel to track area without requiring a full 3D hydrodynamic model, which is appropriate for the single-station disaster-chain risk inference focus. The complete variable-equation list and unit specifications are provided in the Supplementary Material (Section S1). Full governing equations and parameter sets for tunnel water depth and track-area water depth (key variables determining BN evidence and service-disruption status), together with explicit units for all flow-related physical quantities, are compiled in Supplementary Material Section S1. In addition, the non-negative constraint for passenger-stock variables is enforced within simulation implementation to avoid physically-meaningless negative passenger values.

The aforementioned system of differential equations constitutes a system of first-order ordinary differential equations, which is solved using the Euler method [30] for numerical integration. Based on the state variables and their rates of change at the current time step, the state at the next time step is approximated using a first-order Taylor expansion; this approach involves low computational cost and exhibits good stability, making it suitable for long-term time-series simulation of stock-flow models. The simulation time step is determined through a series of sensitivity tests with different time step sizes. When the time step is reduced from 0.10 h to 0.05 h, the rate of change of the peak water accumulation depth is less than 1.5%, satisfying the numerical convergence requirement; when the time step is further reduced to 0.025 h, the computational cost increases by approximately 60%, while the improvement in accuracy is less than 0.3%; therefore, a fixed time step is adopted.

2.2. BN Probability Reasoning Module

Building upon a disaster node system derived from an analysis of 47 metro accident cases under heavy rainfall conditions, this study constructs a Bayesian disaster chain network to enable probabilistic risk inference under discrete state scenarios.

2.2.1. Network Topology

Using grounded theory [31], we conducted a coding and systematization process based on 47 publicly reported and news-reported metro accidents caused by heavy rainfall; we read the risk-chain texts sentence by sentence and assigned initial conceptual labels to each independent event. For example, occurrences such as “rainwater flooding the tunnel” and “flood surging into the tunnel section” were uniformly coded as “tunnel water ingress,” resulting in a total of 46 initial concepts. These initial concepts were then grouped according to the principle of “same process, same attribute”; for instance, “water ingress at entrances/exit points” and “backflow into ventilation shafts” were grouped under the category “rainwater backflow,” yielding 17 secondary categories. From these secondary categories, core categories were extracted, and details that appeared only in isolated cases

were removed. Using the grouping criteria of "frequency ≥ 5" and "clear causal linkage," 13 universally applicable key disaster event nodes were identified; the definitions of these nodes are provided in Table 1.

Table 1. Definition of network nodes for the metro heavy rainfall disaster chain network.

Number	Node Name	Node Description
E1	Extreme rainfall	Hourly rainfall exceeds the metro flood control design standard or is accompanied by extreme weather events such as typhoons or hurricanes.
E2	Surface water accumulation	The urban drainage system is overloaded, leading to urban flooding or widespread flooding on the ground.
E3	Drainage overload	The drainage network and pump station capacity are insufficient, resulting in water accumulation exceeding the design standards.
E4	Rainwater backflow	Water accumulation occurs through entrances and exits, evacuation routes, construction joints, drainage ditches, etc.
E5	Entrance/exit water intake	Metro entrance/exit flooded or forced temporary closure
E6	Tunnel water ingress	Rainwater enters the tunnel section or the track area.
E7	Station hall flooding	Water accumulation has occurred in the station hall and platform areas.
E8	Track area water accumulation	When the water level in the track area exceeds the track surface, it directly impacts train operations.
E9	Underground space flooded	Underground halls, passageways, underground parking garages, and other ancillary spaces have been flooded.
E10	Operational disruption	Temporary metro service suspensions, full-line service suspensions, section closures, station closures, or passenger flow restrictions
E11	Station closed	Metro station temporarily closed
E12	Facility damage	Mechanical and electrical systems, track infrastructure, signaling facilities, and other installations have been damaged by flooding.
E13	Train damaged and stranded	Train carriage flooded, vehicle body damaged, or passengers trapped inside the train

Based on the principles of disaster hydrology and the operational characteristics of metro systems, and by incorporating the evolutionary patterns observed in historical case studies, causal transmission relationships between disaster nodes are identified and compiled to form the edge set of the disaster network. These causal relationships reflect the temporal triggering and propagation logic of disaster events, with the transmission occurring in a unidirectional manner along the evolutionary sequence: "triggering phase $\rightarrow$ hazard incubation phase $\rightarrow$ invasion phase $\rightarrow$ consequence phase." The resulting directed disaster network comprises 22 directed edges, contains no isolated nodes or self-loop edges, and constitutes a single weakly connected network; its topological structure fully aligns with the disaster evolution logic revealed by historical case studies.

2.2.2. Conditional Probability Table Determination

1. Node State Classification Criteria

All 13 BN nodes are categorized into three state levels: low, medium, and high. For nodes for which industry standards or specifications exist, thresholds are determined in

accordance with national and industry standards; for nodes without a unified quantitative assessment criterion, a qualitative assessment is conducted based on descriptions provided in accident investigation reports and authoritative news reports. The specific classification criteria are presented in Table 2.

Table 2. Node status classification criteria and basis.

Node Numbering	Low Risk	Medium Risk	High Risk
E1	Total 24 h rainfall < 50 mm	50~100 mm	>100 mm
E2	Water accumulation depth < 0.15 m	0.15~0.30 m	>0.30 m
E3	Drainage capacity meets demand	Inadequate local drainage	Complete drainage system failure
E4	No backflow occurred	Minor backflow of water at the entrance/exit point	Large amount of backflow water ingress
E5	Entrance/exit drying	Minor water accumulation at entrances and exits	Water flows into the station hall.
E6	Tunnel drying (<0.10 m)	Tunnel local water accumulation (0.10–0.30 m)	Large-scale tunnel water ingress (>0.30 m)
E7	Platform dryness (<0.05 m)	Minor water accumulation in the station hall (0.05–0.20 m)	Large-scale water accumulation in the station hall (>0.20 m)
E8	<0.25 m	0.25~0.35 m	>0.35 m
E9	Not flooded	Adjacent space partially flooded	Large areas of the main space are flooded.
E10	Normal operation	Partial service suspension or passing through a station	Full service suspension
E11	Normal operation	Some entrances and exits are closed.	Site Closed
E12	No damage	Some equipment is submerged in water	Severe damage to mechanical and electrical equipment
E13	No influence	Passenger evacuation	Train stranded or personnel injured or killed

Discretization thresholds are synthesized from metro flood-protection manuals, published metro-flood case literature, and underground station physical characteristics. The classification of node statuses in Table 2 is as follows: E1 is determined according to the 24 h rainfall classification standard specified in the "Grade of precipitation" (GB/T 28592-2012) [32]; E2 is determined based on the internal flooding prevention design standards outlined in the "Technical code for urban flooding prevention and control" (GB 51222-2017) [33] and the verification requirements for internal flooding prevention specified in the "Standard for design of outdoor wastewater engineering" (GB 50014-2021) [34]; for E8, measured relative to the tunnel bottom, the medium/high-risk threshold of 0.35 m is determined based on the minimum installation height of train running gear equipment, which ranges approximately from 0.30 to 0.40 m; water accumulation exceeding this height may cause short-circuit damage to electrical equipment; therefore, according to the safety redundancy principle, system shutdown is triggered before water reaches the track surface height—this approach aligns with the provisions of the "Urban Rail Transit Operation

Management Measures" and conservative practical operational practices; nodes such as E3, E4, E6, E9, and E12 are qualitatively assessed based on accident news reports; E5 and E7 are qualitatively assessed based on case descriptions; E10 is assessed based on official metro operation announcements; E11 and E13 are assessed based on official announcements and case descriptions. We note that thresholds are set at the conservative lower bound of recommended ranges, applied identically across open-loop and closed-loop simulations, and do not artificially inflate closed-loop performance.

E2, E7, E6, and E8 are interface variables between the SD module and the BN module; their continuous state values are simulated by the SD module and, after discretization, are used as evidence inputs to the BN module; the remaining nodes serve as inference nodes for the BN module, where the posterior probabilities are updated via Bayesian inference.

2. CPT Construction

The CPT is constructed using a three-tier iterative methodology comprising "Case-Based Statistical Baseline–Literature-referenced Calibration–Physical Logical Constraints"; the specific steps are as follows.

Step 1: Case-based statistical baseline. Using the maximum likelihood estimation method [35], we statistically calculate the co-occurrence frequency of each sub-node state under every parent–node state combination across the 47 cases, and then compute the baseline conditional probability as shown in Equation (5).

$$P_0(Y = y_j | X = x_i) = \frac{N(X = x_i, Y = y_i)}{\sum_{j=1}^{r} N(X = x_i, Y = y_i)} \quad (5)$$

where $N(X = x_i, Y = y_i)$ is the co-occurrence frequency of the parent node in state x_i and the child node in state y_i in the case base; r is the total number of child node states.

For extreme state combinations not encountered in the case study, maximum likelihood estimation yields a zero probability, rendering subsequent Bayesian inference invalid. To address this, the Laplace smoothing method [36] is introduced; a smoothing parameter α is added to the frequency of each state to provide a uniform prior distribution for zero-sample combinations; the calculation formula is given by Equation (6).

$$P_s(Y = y_j | X = x_i) = \frac{N(X = x_i, Y = y_i) + \alpha}{\sum_{j=1}^{r} N(X = x_i, Y = y_i) + \alpha \cdot r} \quad (6)$$

where α is the Laplace smoothing parameter. Comparative tests are conducted with α taking values of 0.01, 0.02, 0.03, 0.05, 0.1, 0.5, and 1.0. Taking node E8 as an example, for combinations with samples, when α varies within the interval [0.01, 0.1], the high-risk probability fluctuation is less than 1%; for zero-sample combinations, the uniform prior probability assigned by Laplace smoothing is constantly 1/K = 0.333, independent of the α value. An excessively small α provides insufficient smoothing for small-sample combinations, while an excessively large α significantly deviates from the sample statistical frequency. Considering both smoothing effect and sample information retention, α = 0.03 is selected. This value is far smaller than the minimum non-zero sample size n = 1, exerting an influence of less than 0.7% on the probabilities of combinations with n $\geq$ 9, while providing a uniform prior probability of 1/K for each state in zero-sample combinations.

We distinguish two estimation strategies for CPT entries according to sample availability: maximum likelihood estimation is adopted for parent–node combinations with observed records (n > 0), while Laplace smoothing within the three-tier recursive calibration framework is applied for zero-sample (n = 0) combinations to address data sparsity. Our α sensitivity tests show that Laplace smoothing assigns a conservative uniform prior

(1/K = 0.333) for zero-sample parent–node combinations and avoids generating extreme probability values.

Taking E8 as an example, its parent nodes are E6 and E3. Co-occurrence frequency analysis based on 47 cases reveals that when E6 = Low and E3 = Medium (n = 23), all instances of E8 are classified as low-risk, indicating that the risk of water accumulation in the track area is extremely low when the tunnel is not flooded; when E6 = High and E3 = High (n = 9), all instances of E8 are classified as high-risk, which aligns with the physical logic that the higher the risk of the parent node, the greater the probability that its child node will be classified as high-risk. Among the 47 cases, there are two zero-sample parent–node combinations for E8; after applying Laplace smoothing, these combinations are assigned a uniform prior probability, which is then adjusted to a non-uniform distribution using isotonic regression constraints.

Step 2: Literature-based calibration. To address the CPT sparsity and zero-probability issues in small-sample scenarios, this step involves retrieving existing research on BN modeling for metro flood disasters and extracting conditional probabilities from similar scenarios as a literature-based prior reference. Due to the limitations imposed by the accuracy of publicly available accident data, this step is primarily intended to assist in assessing the plausibility of zero-sample combination probabilities; it does not involve a rigorous 6:4 quantitative weighted fusion approach. The actual CPT calibration is primarily based on case-study statistical baselines and physical logical constraints; the literature-based prior serves as a qualitative reference to ensure the physical plausibility of the probability distribution.

After applying maximum likelihood estimation and Laplace smoothing, three constraints are introduced to ensure the physical plausibility and numerical stability of the CPT. For zero-sample parent–node combinations (n = 0) absent from the 47 historical accident cases, we adopt a two-stage processing workflow. First, Laplace smoothing assigns a fixed conservative uniform prior (1/K = 0.333) to prevent reasoning failure caused by zero probabilities; our α-sensitivity test shows that this uniform prior is independent of α values. Second, isotonic regression is applied to adjust these flat uniform probabilities to satisfy physical monotonic constraints. It should be noted that this adjusted output represents a physically plausible constrained estimate rather than empirically-observed frequency. Although the uniform prior avoids extreme probability values, residual uncertainty persists for these unobserved combinations, because no real-world accident cases are available to constrain their conditional probabilities. This source of uncertainty is further acknowledged in Section 5.4 Limitations. It is important to distinguish two groups of CPT entries in this study: entries for parent–state combinations with observed accident records (n > 0) are primarily derived from case frequencies using maximum-likelihood estimation, where real-world accident observations provide empirical support; by contrast, zero-sample parent–state combinations (n = 0) have no corresponding accident records in the dataset, and their final probabilities are mainly determined by Laplace smoothing uniform priors and subsequent isotonic regression monotonicity constraints rather than empirical case evidence.

① Maintaining orderly regression with monotonicity constraints—To avoid subjective bias introduced by manually setting upper or lower probability bounds, isotonic regression is employed to perform a weighted least-squares adjustment on the coarse-grained probabilities after Laplace smoothing. The objective is to minimize the weighted squared deviation between the adjusted probabilities and the smoothed frequency distributions, subject to a strict non-decreasing monotonicity constraint: when the state of any parent node changes from low → medium → high, the posterior probability of the high-risk state for the corresponding child node must be non-decreasing, with all other parent node states

remaining unchanged. The weight assigned to each parent–node combination is the sample size n, ensuring that the adjusted CPT deviates as little as possible from the original sample frequency distribution. For two-dimensional parent–node combinations, the iterative Pool Adjacent Violators Algorithm (PAVA) is applied, performing one-dimensional isotonic regression alternately along the row and column directions until convergence is achieved. This method is entirely data-driven and requires no arbitrary specification of subjective probability bounds, thereby fundamentally preventing the counterintuitive phenomenon where the risk probability of a child node decreases despite an increase in the state of its parent node. After applying isotonic regression to the high-risk probability distribution, the remaining probability is allocated proportionally to the original frequencies (after Laplace smoothing), ensuring that the sum of the three-state probabilities equals 1.

② Zero-sample processing constraint—For parent–node combinations with sample size n = 0, a uniform prior probability of 1/K is initially assigned via Laplace smoothing; this value is then adjusted using the aforementioned isotonic regression method to bring it to a reasonable level, thereby avoiding the reasoning failure associated with zero probability and eliminating the disruption of monotonicity caused by the uniform prior.

③ Normalization constraint—The sum of the state probabilities for all states of a given node must always equal 1; after applying isotonic regression and proportional allocation, the normalization error is checked on a per-combination basis to ensure that the error is less than 10^{-6}.

All CPT entries have been re-checked; the probability sum for each parent–state combination equals 1.000 after rounding adjustment, and monotonicity has been verified across all parent–node combinations.

After a three-level recursive calibration process, the CPT for E8 is presented in Table 3. E8 serves as a key intermediate node in the disaster chain—linking "tunnel water ingress–water accumulation in the track area–operational disruption"—and it has two parent nodes; thus, its CPT calibration process is representative. Taking the combination E6 = High and E3 = High as an example (with a sample size of n = 9, all of which are classified as high-risk), after applying Laplace smoothing, the high-risk probability is 0.993; the adjustment magnitude for this combination using isotonic regression is 0; and after final normalization, the high-risk probability is 0.994, thereby preserving the sample's statistical information to the greatest extent possible.

Table 3. Final conditional probability for node E8.

E6	E3	n	P(E8 = Low)	P(E8 = Medium)	P(E8 = High)
Low	Low	1	0.969	0.028	0.003
Low	Medium	23	0.996	0.001	0.003
Low	High	5	0.988	0.006	0.006
Medium	Low	0	0.496	0.496	0.008
Medium	Medium	3	0.010	0.983	0.007
Medium	High	5	0.006	0.987	0.007
High	Low	0	0.333	0.333	0.334
High	Medium	1	0.028	0.028	0.944
High	High	9	0.003	0.003	0.994

3. Probability Reasoning Mechanism

The CPT, constructed using the three-level recursive method, remains constant during model execution; the posterior probability of each node is inferred based on Bayes' theorem,

and precise inference is implemented using the Joint Tree Algorithm; the update formula is given by Equation (7).

$$P(Y|E) = \frac{P(E|Y) \cdot P(Y)}{P(E)} \quad (7)$$

where E is the discretized evidence supplied by the SD module; $P(Y)$ is the node prior probability; $P(Y \mid E)$ is the posterior probability after evidence input.

The junction tree algorithm achieves global exact probabilistic reasoning by converting the BN into a clique tree structure and passing messages between cliques. Its computational complexity is $O(n \cdot \omega \cdot d^{\omega})$, where n is the number of nodes, ω is the maximum clique size, and k is the number of node states. The maximum clique size of the 13-node network is 3, and the number of node states k = 3, so the computational cost of a single reasoning is extremely small, satisfying real-time simulation requirements.

2.3. SD–BN Closed-Loop Coupling Mechanism

2.3.1. SD→BN Variable Discretization and Evidence Input

Four core physical state variables are selected as the coupling interface for the SD→BN system, each corresponding to one of the four critical nodes in the BN network: the ground surface water depth corresponds to E2, which directly determines whether backflow is triggered; the station hall water depth corresponds to E7, serving as a direct measure of the extent of flooding within the station; the tunnel inundation depth corresponds to E6, acting as a key intermediate variable in the propagation of the disaster from the station to the tunnel section; and the track area water depth corresponds to E8, functioning as a critical node that links the physical disaster to the operational consequences and directly determines whether trains will be suspended. These four variables are all direct outputs of the SD differential equation system and possess clear physical meaning and causal relationships within the BN network, enabling an effective mapping between physical states and risk reasoning.

Each variable is categorized into three levels—low, medium, and high—where the discretization threshold aligns perfectly with the state classification criteria specified for the corresponding nodes in Table 2. At each simulation time step, discretization and evidence input are performed according to the following steps.

1. Continuous variable values: Extract the continuous values of the four interface variables (corresponding to E2, E7, E6, and E8) from the calculation results at the current step size of the SD module.
2. Threshold comparison and state determination: Compare each continuous value with the two-level thresholds corresponding to the respective node in Table 2 to determine the discrete state.
3. Probability update. The four discrete states are fed as hard evidence into the BN module; the observed states of the four interface nodes are fixed, and the remaining nodes perform global probability reasoning.

The low–medium–high classification thresholds for four coupling variables are derived from metro flood-protection engineering manuals, existing published metro flood case studies, and physical characteristics of underground station facilities. These empirical–physical thresholds are adopted to map continuous simulation outputs into discrete BN evidence states. Full quantitative threshold values are listed in Table 2. Future work may conduct targeted threshold sensitivity analysis to further test the robustness of coupling outputs.

2.3.2. BN→SD Strategy Feedback and Parameter Adjustment

The traditional SD model features fixed control parameters, making it inadequate for capturing the intervention effects of dynamic emergency response efforts. By establishing a

BN→SD strategy feedback mechanism, real-time risk assessments conducted by the BN can be converted into control commands to dynamically adjust the SD parameters.

1. Comprehensive disaster risk probability synthesis

The comprehensive disaster risk probability is synthesized using the weighted risk index method; the three core consequence nodes—E10, E12, and E13—are selected as the evaluation dimensions for the comprehensive risk. The posterior probability of each node in its high-risk state is used; the synthesis formula is given by Equation (8).

$$R(t) = \omega_{E10} \cdot P(E_{10} = High|E) + \omega_{E12} \cdot P(E_{12} = High|E) + \omega_{E13} \cdot P(E_{13} = High|E) \quad (8)$$

where $R(t)$ is the system's comprehensive disaster risk probability at time t, with a value range of [0, 1]; $P(E_i = High|E)$ is the posterior probability of node E_i in the high-risk state; ω_{E_i} is the node weight coefficient, satisfying 1.

The weights are determined using the entropy weight method, which assigns weights objectively based on the variability of indicator data, thereby avoiding the arbitrariness inherent in subjective weighting. Based on 47 historical accident cases, whether each of the three consequence nodes was in a high-risk state in each case is counted, constructing a 47 × 3 binary (0–1) indicator matrix; the information entropy of each indicator is then calculated as $e_j = -\frac{1}{\ln 47}\sum_{i=1}^{47} p_{ij} \ln p_{ij}$, where a lower information entropy value indicates stronger discriminative power for that indicator, and thus a higher weight should be assigned; the weights are subsequently derived from $\omega_j = \frac{1-e_j}{\sum(1-e_j)}$. According to the calculations, among the 47 cases, E10 high-risk occurred 9 times, E12 occurred 5 times, and E13 occurred 6 times, with corresponding information entropies of 0.571, 0.418, and 0.465; substituting these values yields $\omega_{10} = 0.278$, $\omega_{12} = 0.376$, $\omega_{13} = 0.346$. This weight distribution reflects the relatively high discriminative power of equipment damage within the consequences of metro rainstorm disasters, which aligns with the actual observation that equipment damage falls into the high-risk category with the lowest frequency of occurrence in the case statistics, yet carries severe consequences should it occur. Some values of β, η and γ are derived following the tiered-response framework stated in national metro flood emergency management guideline documents; several values are determined via practical engineering empirical judgment; a subset of key parameters are inverse-calibrated against the Qingdao "July 23" flooding event. It should be noted that inverse-calibrated parameters are case-specific for this station scenario. Without local parameter adjustment and re-calibration according to station-specific design data and local emergency response rules, these tuned values cannot be directly transferred to other metro stations or different flood events.

Regarding parameter origins: Level-I baseline values follow national metro emergency-response guidelines; Level-II coefficients rely on engineering empirical judgment for typical on-site emergency performance; Level-III coefficients are inverse-calibrated specifically for the Qingdao "July 23" case and are not universally transferable, requiring local recalibration for other stations.

2. Tiered Response Mechanism

The strategic feedback mechanism employs a three-tiered response framework, using the comprehensive disaster risk probability as the triggering criterion to map risk levels into three distinct response tiers, each corresponding to a set of adjusted control parameter values. The risk probability threshold is determined by referencing the widely adopted three-tiered risk matrix classification method in the field of risk assessment, tailored to the practical context of emergency management for metro disasters caused by heavy rainfall. The adjusted values of the three control parameters increase monotonically as the response

tier rises. The values of these control parameters are derived from the tiered response framework outlined in the National Urban Rail Transit Operational Emergency Response Plan and are calibrated in accordance with the typical intensity of flood prevention and control measures employed in metro emergency response scenarios; the specific values are provided in the case study section.

Parameter adjustment is achieved by directly incorporating the correction coefficient into the SD differential equation; the specific adjustment formulas are given by Equations (9)–(11).

$$Q_{pump}(t) = Q_{pump,0} \cdot \beta(t) \tag{9}$$

where $Q_{pump,0}$ is the baseline value of pumping station design discharge; $\beta(t)$ is the drainage capacity correction coefficient at time t. The physical meaning of this equation is that the actual drainage capacity equals the baseline drainage capacity multiplied by the correction coefficient, with $\beta(t) > 1$ indicating increased total drainage through additional pumps or mobile equipment.

$$Q_{in}(t) = k \cdot \left[h_g(t) - h_{th}\right]^+ \cdot [1 - \eta_{close}(t)] \tag{10}$$

where $\left[h_g(t) - h_{th}\right]^+ = \max\left(h_g(t) - h_{th}, 0\right)$ is the positive part of surface water accumulation exceeding the threshold height; $\eta_{close}(t)$ is the entrance closure rate at time t. The physical meaning of this equation is that the actual backflow equals the baseline backflow volume multiplied by the proportion of unclosed entrances, with $\eta_{close} = 1$ indicating that backflow is completely blocked.

$$v_{evac}(t) = v_{evac,0} \cdot \gamma(t) \tag{11}$$

where $v_{evac,0}$ is the baseline evacuation rate; γ(t) is the evacuation efficiency correction coefficient at time t. The physical meaning of this equation is that the actual evacuation rate equals the baseline evacuation rate multiplied by the correction coefficient, with $\gamma > 1$ indicating accelerated passenger evacuation through enhanced guidance and additional evacuation channels.

The equations correspond to the parametric expression of three types of emergency measures—drainage enhancement, source blocking, and passenger evacuation—in the SD model, achieving quantitative mapping from BN risk assessment to SD physical intervention.

3. Rolling Solution

The coupled model employs a rolling time-domain iterative solution mechanism; within each simulation time step, the "SD Simulation → Discretization → BN Reasoning total T → Strategy Feedback" cycle is executed, gradually advancing the simulation time t until the preset total simulation duration is reached. During simulation execution, the process begins with initialization—setting the total simulation duration, initial state variables, and baseline parameters, as well as loading the rainfall process curve and the CPT for the BN network; next, the SD differential equation system is solved using the current time step, all continuous state variables are updated, and the current values of the four core interface variables are output; then, the discretization and evidence input steps are performed according to the procedure described in Section 2.3.1, and BN probabilistic reasoning along with policy feedback are executed according to the mechanism outlined in Section 2.3.2 to dynamically adjust the control parameters; if the total simulation duration has not been reached, the simulation proceeds to the next time step, and this cycle is repeated until the simulation concludes.

2.4. Use of Generative AI

During the preparation of this work, the authors used the AI-assisted diagramming website PicDoc to generate Figure 1. All other figures were generated using Python 3.14 with Matplotlib 3.11.1 based on simulation data.

3. Case Study

3.1. Study Area and Data Sources

The model was validated using the Qingdao Metro's "July 23" flooding incident as a case study. From 22–23 July 2020, Qingdao experienced continuous heavy rainfall; according to statistics from the Qingdao Meteorological Observatory, the cumulative average precipitation across the city between 08:00 on 22 July and 12:00 on 23 July reached 120.7 mm, with the Xihai Coast New Area recording a cumulative rainfall of 152.1 mm. At 23:10 on 22 July, the Qingdao Meteorological Observatory upgraded the heavy rain yellow alert to a heavy rain red alert. In the early hours of 23 July, rainwater inundated the under-construction Qingdao Metro Line 1; the accumulated water breached the barrier wall separating Line 1 from The West Coast Line and spilled into the The West Coast Line track area, resulting in the temporary closure of Jinggangshan Road Station and Jimiya Station. As a consequence, operations on The West Coast Line were adjusted to a short-turn service running between Lingshanwei Station and Dongjiakou Railway Station; after emergency response measures were implemented, full-line operations were restored on 24 July. This incident fully illustrates the typical evolution path of a metro heavy-rain disaster chain: "extreme rainfall → surface water accumulation → breach of barrier walls → backflow of rainwater → tunnel flooding → track area water accumulation → operational disruption." The route alignment of the West Coast Express Line and locations of affected stations are shown in Figure 4.

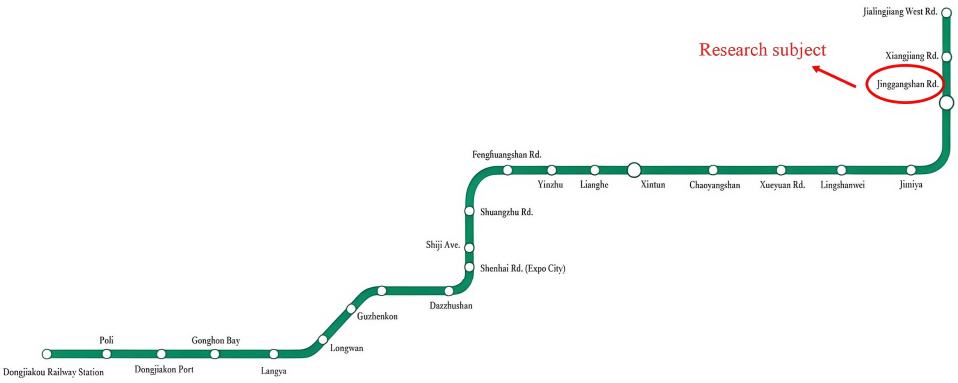


Figure 4. Route map of the West Coast Express Line (Qingdao), with Jinggangshan Road Station (study site) marked.

The rainfall data used in the simulation were reconstructed based on publicly available information officially issued by the Qingdao Meteorological Observatory; the cumulative rainfall amount was 120.7 mm, the peak rainfall intensity was 28.6 mm/h, and the time of the heavy rainfall red alert upgrade was 23:10 on July 22. The reconstructed 28 h rainfall process curve matches the rainfall input used in the Vensim model (cumulative rainfall: 120.8 mm), which is largely consistent with the officially reported value. The peak rainfall occurred at the 15th hour, which coincided with the time of the red alert upgrade.

3.2. Model Configuration and Parameter Calibration

3.2.1. Comparative Solution Design

The simulation duration is 28 h, which corresponds to the actual rainfall event; the initial state is specified as, $N(0) = 0$, with the emergency response level set at Level I. To assess the independent contribution of each component within the system model, three control scenarios were designed; these scenarios employed a stepwise superposition approach to isolate the performance of different modules. All three scenarios utilized identical rainfall inputs, initial conditions, and baseline parameters; the only differences lay in the configuration of the BN module and the activation switch for policy feedback. The component configurations are presented in Table 4.

Table 4. Component composition of the three ablation scenarios.

Plan Number	Scenario Name	SD Module	BN Module	Strategy Feedback
Scheme A	Open-loop SD Model	√	×	×
Scheme B	Unidirectional Coupling Model	√	Only SD → BN input	×
Scheme C	Closed-loop coupling model	√	Two-way closed-loop	√

As shown in Table 4, the comparison between Scheme A → B demonstrates the independent value of the BN probabilistic reasoning module; the comparison between Scheme B → C verifies the disaster mitigation effectiveness of the BN → SD strategy feedback mechanism.

3.2.2. Parameter Calibration

The physical parameters are determined based on the actual design data and industry standards for Jinggangshan Road Station of Qingdao Metro. Specifically, the entrance water threshold height is set to 0.15 m, as specified in the "Code for design of metro" (GB 50157-2013) [36] for underground station entrances. The remaining physical parameters are lumped-parameter values determined through inverse calibration based on the disaster evolution characteristics observed in the Qingdao Metro "July 23" incident; the calibration objective was to align macro-temporal characteristics—such as the peak station hall water accumulation level during open-loop simulation and the duration of operational disruption—with the descriptions provided in the accident report. The parameter values are as follows:

The parameter $\alpha = 3.50$ incorporates the ratio effect between the catchment area surrounding the station and the surface water accumulation area, rather than simply using a runoff coefficient; it accounts for the rainwater convergence effect in the low-lying areas surrounding the station entrances. The parameter $\lambda = 0.06\ \mathrm{h}^{-1}$ characterizes the discharge behavior of the municipal sewer network for surface water accumulation; this value is determined based on the design flow calculation formula $Q = \varphi q F$ for stormwater drainage networks specified in the "Standard for design of outdoor wastewater engineering" (GB 50014-2021), where φ represents the runoff coefficient, q represents the design storm intensity, and F represents the catchment area; the magnitude of this parameter is derived through inverse calculation using the storm intensity formula $q = \frac{1919.009 \times (1 + 0.997 \lg P)}{(t + 10.740)^{0.378}}$ applicable to the Qingdao region, followed by calibration via simulation inversion. The parameter $V_{drain} = 0.003$ m/h reflects the runoff characteristics of the urban surface; this value is determined through parameter inversion and calibration. $k = 1300\ \mathrm{m}^2/\mathrm{h}$ is simplified to $Q_{in} = k \cdot (h_g - h_{th})$, and its value is determined through simulation inversion and calibration.

Jinggangshan Road Station is a two-level underground island platform station with a total station hall floor area of approximately 2000 m^2; since water accumulation is primarily concentrated in the entrance passages and low-lying areas within the station hall, the equivalent water accumulation area As is set to 800 m^2. $Q_{pump,0} = 100$ m^3/h is determined in accordance with the "Code for design of metro" (GB 50157-2013) and considering the typical drainage volume for a two-level underground island platform station. The parameter $Q_{leak}(t) = 0.005$ m/h represents the drainage volume through other pathways, such as construction joints and ventilation shafts; this value is estimated to be 5–10% of the station's total drainage volume, as specified in the Code for design of metro (GB 50157-2013). According to the Construction Standard for Urban Rail Transit Projects (Jianbiao 104-2008), the peak-hour evacuation capacity for intermediate stations is set at 800 passengers/h; however, given the absence of emergency guidance in an open-loop scenario and the relatively low efficiency of spontaneous passenger evacuation, an effective evacuation rate $V_{evac,o}$ 300 passengers/h is adopted.

The three-level values of β, η, and γ are determined based on the actual measure intensity under the emergency response framework, referencing the hierarchical response framework outlined in the National Urban Rail Transit Operational Emergency Plan (Guobanhan [2015] No. 32), and established in conjunction with the typical intensity of flood control emergency response measures for metro systems, as well as relevant engineering experience. To improve reproducibility, the basis for each tier of parameters is explicitly clarified:

Level I corresponds to normal operational conditions without emergency activation. All correction coefficients take baseline no-adjustment values (β = 1.0, η = 0, γ = 1.0), representing unmodified station physical performance without intervention.

Level II corresponds to early warning response. The coefficients (β = 1.2, η = 2/3, γ = 1.3) reflect typical moderate-intensity metro flood emergency measures: temporary mobile pump deployment delivering approximately 20% extra drainage capacity, two-thirds of entrance flood barriers deployed, and on-site passenger guidance improving evacuation efficiency by roughly 30%, consistent with common practice in metro flood emergency response.

Level III corresponds to peak emergency response. The coefficients (β = 1.8, η = 1.0, γ = 1.8) are calibrated against actual on-site measures deployed during the Qingdao "July 23" event, including full entrance closure, full mobile pump fleet deployment, and multi-channel evacuation guidance.

As reported in Section 3.2.2, dedicated ±20% perturbation tests on these parameters confirm that core model conclusions remain stable within a reasonable parameter range, with no change to the central finding that closed-loop control primarily accelerates post-peak risk decline.

The adjustment rules for three-level hierarchical response parameters are presented in Table 5.

To assess the impact of uncertainty in hierarchical response control parameters on simulation results, the results are summarized in Table 6. ±20% perturbation tests were conducted separately on the drainage capacity correction coefficient, entrance/exit closure rate, and evacuation efficiency correction coefficient for Level II and Level III response scenarios. The ±20% perturbation interval is selected as a standard, widely used range for sensitivity testing in system dynamics-based infrastructure emergency modeling, and reflects realistic on-site deployment variability of emergency response measures. In practical metro flood emergency operations, the actual temporary drainage capacity, entrance closure ratio, and evacuation guidance efficiency typically fluctuate within ±20% of the planned design values, due to factors such as mobile equipment performance differences, on-site command execution deviations, and real-time passenger flow variation. This in-

terval covers the common engineering uncertainty range for such emergency response parameters. It should be acknowledged that this single-parameter perturbation range is reasonable but not exhaustive; more comprehensive sensitivity analysis, including joint multi-parameter perturbation and wider interval testing, is recommended as future work to further verify model robustness. While keeping the rainfall input, initial conditions, and physical parameters unchanged, the variation ranges of three core indicators—peak water accumulation in the station hall, duration of operational disruption, and peak passenger retention—were observed; the results are presented in Table 6.

Table 5. Tiered response parameter adjustment rules.

Response Level	Risk Probability Interval	β	η	γ	Source
Level I	R < 0.30	1.0	0	1.0	Normal operation baseline, no emergency intervention
Level II	0.30 ≤ R < 0.70	1.2	2/3	1.3	Derived from the National Urban Rail Transit Operational Emergency Plan (Guobanhan [2015] No.32) framework, combined with typical metro flood emergency response engineering experience
Level III	R ≥ 0.70	1.8	1.0	1.8	Calibrated against actual on-site emergency measures deployed during the Qingdao "July 23" flooding event

Note: β = 1.8 represents the maximum short-duration emergency pumping capacity activated during Level III response, reflecting combined operation of primary and backup pumps rather than sustained long-term pump operation, consistent with metro emergency drainage protocols.

Table 6. Sensitivity analysis results.

Disturbance Scheme	Platform Peak Water Accumulation (m)	Operational Downtime Duration (h)	Peak Passenger Retention (People)	Comprehensive Risk Peak
benchmark	0.502	0.05	40	0.602
β + 20%	0.502	0.05	40	0.602
β − 20%	0.502	0.10	60	0.602
η + 20%	0.502	0.05	40	0.602
η − 20%	0.502	0.10	60	0.602
γ + 20%	0.502	0.05	40	0.602
γ − 20%	0.502	0.05	40	0.602

The ±20% perturbation range adopted for response coefficients β, η, and γ is a commonly-used uncertainty interval within the infrastructure emergency-response modeling literature, representing reasonable fluctuation of emergency implementation efficiency under practical engineering conditions. We acknowledge that this interval is reasonable but not exhaustive; multi-factor combined sensitivity analysis is suggested for future research.

As shown in Table 6, both the peak station hall water accumulation level and the comprehensive risk peak are insensitive to disturbances of ±20% in the parameters β, η, and γ. The Level II response under closed-loop control was triggered at 22.45 h, whereas the peak station hall water accumulation occurred at 22.50 h; the two events occurred only 0.05 h apart. By the time the control parameters took effect, the water accumulation had already approached its peak value. Therefore, the parameter adjustment primarily accelerates the decline phase following the peak, having a limited impact on the already established peak water level. Specifically, β + 20% increases the Level II response drainage capacity from 1.2 to 1.44, while η + 20% raises the entrance/exit closure rate from 0.667 to 0.8; however, the Level II response is activated at 22.45 h, only one fixed time step (0.05 h)

before the peak at 22.50 h; the parameters therefore take effect essentially as the peak is reached, so the peak water level in the station hall remains unchanged at 0.502 m. The parameter γ only affects the passenger evacuation rate and does not participate in the physical water accumulation process; consequently, $\gamma \pm 20\%$ has no impact on either the peak station hall water level or the comprehensive risk peak. This result validates the core mechanism of closed-loop control: its ability to adjust the peak floodwater level—which is primarily driven by rainfall—remains limited. Although $\beta - 20\%$ and $\eta - 20\%$ do not alter the peak water level in the station hall, they increase the duration of operational disruption from 0.05 h to 0.10 h and raise the peak passenger retention from 40 passengers to 60 passengers; this indicates that the control parameters indeed influence system outcomes by extending the duration of high-risk exposure. The core mitigation mechanism of the closed-loop feedback is to shorten the duration of high-risk periods, rather than to reduce the peak values of physical quantities. Thus, the uncertainty associated with the tiered response parameters within the $\pm 20\%$ range does not alter the central conclusion—that closed-loop control accelerates the decline of risk levels—and demonstrates the robustness of the model.

3.3. Model Verification and Implementation Consistency Check

This section performs model verification to check the internal consistency between Python implementation and the original Vensim model, including mathematical constraints, probabilistic logic and physical plausibility. It should be emphasized that this is implementation-oriented verification rather than empirical validation against real-world observed inundation or service-disruption records. Quantitative empirical validation is hindered by the scarcity of high-resolution field monitoring data for metro flood accidents, and this limitation is further elaborated in Section 5.4 Limitations.

The verification focuses on checking the correctness of the coupled model implementation—namely, ensuring that the Python-based SD equation implementation aligns with the Vensim model, that the BN inference logic is correct, and that all system physical constraints are satisfied. Due to limitations imposed by the accuracy of publicly available accident data, a quantitative comparison between simulation results and actual observations is not yet feasible; the soundness of the three sets of ablation experiments described in Section 3.2 instead provides internal consistency and plausibility evidence for the model, rather than empirical validation of its predictive accuracy. The independent verification results for the SD module demonstrate that the peak relative errors for all four variables are less than 0.001%, and the coefficient of determination (R^2) for the four time-series curves is equal to 1.000; all results satisfy the criteria of a relative error < 5% and $R^2 > 0.95$, indicating a high degree of consistency between the Python-implemented SD differential equation solver and the Vensim PLE stock-flow model. Specifically, the peak floor-level water depth in the Python simulation is 0.5888 m, while in the Vensim model, it is also 0.5888 m, yielding a relative error of only 0.0004%; the peak ground-level water depth is 0.2821 m for both models; the peak track area water depth is 0.439 m for both models; and the peak passenger retention is 2790 passengers in the Python simulation, which matches the output from the Vensim model. We also note that the Qingdao "July 23" case is excluded from the 47-case CPT training dataset, eliminating circular validation concerns.

The BN module verifies the fundamental correctness of the parameters and the reasoning process through the CPT normalization check and the reasoning monotonicity test. The CPT normalization check results indicate that, across all 109 parent–node combinations across the 13 nodes, the sum of the three state probabilities equals 1.000; there are no outliers, and all combinations pass the check. Taking node E8—whose parent nodes are E6 and E3—as an example for the reasoning monotonicity test, the test examines the monotonicity

of the high-risk probability for node E8 as the state of one parent node remains unchanged while the state of the other parent node varies from low to medium to high; all six test cases satisfy a strict monotonically increasing trend. For instance, when E6 = High, as E3's state changes from low → medium → high, the high-risk probabilities for E8 are 0.334, 0.944, and 0.994, respectively— representing a strict increasing sequence; similarly, when E3 = High, as E6's state changes from low → medium → high, the high-risk probabilities are 0.006, 0.007, and 0.994, respectively—also exhibiting a strict increasing trend. After applying isotonic regression, the CPTs for all 13 nodes pass the monotonicity test. All five physical plausibility tests are successfully passed. All state variables remain non-negative throughout the entire process; the posterior probabilities for all nodes lie within the range [0, 1], and the overall disaster risk probability ranges from [0.013,0.602]; the control parameters increase monotonically as the response level rises; the peak station hall flooding level of 0.502 m in the closed-loop model is lower than the 0.589 m observed in the open-loop model (a reduction of 14.7%), consistent with the expected direction of the strategy feedback mechanism; the average absolute error for the mass conservation of station hall flooding is 0.00%. It should be clarified that the above tests constitute implementation verification rather than empirical validation. The Python implementation is checked against the Vensim reference model to ensure mathematical and computational consistency; however, quantitative validation against observed flood depth and operational disruption records remains limited by the scarcity of publicly available high-resolution accident data. The model parameters are calibrated based on the Code for Design of Metro(GB 50157-2013), national emergency response guidelines, and engineering experience values documented in previous studies. Physically, the modeled entrance backflow mechanism corresponds to overtopping of the entrance water-retaining curb (0.15 m), which is consistent with the observed barrier-breach process in the Qingdao case. Importantly, the 47 historical cases used for CPT calibration do not include the Qingdao "July 23" incident; the case study is used only for scenario demonstration and is independent from the training dataset, avoiding data overlap. All the aforementioned verification results satisfy the predefined verification criteria, ensuring the correctness of the coupled model's implementation across all levels—from module-level to system-level.

4. Results

4.1. Open-Loop–Unidirectional Coupling–Closed-Loop Comparative Analysis

Comparison of core efficacy endpoints among the three treatment groups is shown in Table 7; comparison of the evolution of core physical endpoints across the three treatment groups is shown in Figure 5. All comparative results below are generated under identical rainfall input, initial water levels, passenger flow inputs, and baseline physical parameters, with the unidirectional coupling scenario (Plan B) as the reference baseline. Percentage reductions are calculated as (Plan B value−Plan C closed-loop value)/Plan B value × 100%.

As shown in Figure 5a, the three curves largely coincide during the initial phase of rainfall; at this stage, the surface water accumulation depth has not yet exceeded the 0.15 m water retention threshold, and backflow has not been triggered, resulting in zero water accumulation in the station hall. During the peak rainfall period, once the surface water accumulation depth exceeds the water retention threshold, backflow is initiated; under Plan A and Plan B, due to the absence of a strategic feedback mechanism, the water accumulation in the station hall continues to rise to a peak of 0.589 m before gradually declining. Under Plan C, since the comprehensive risk probability reaches the 0.30 threshold at 22.45 h, a Level II response is triggered, the drainage capacity is increased to 1.2 times, two-thirds of the entrances/exit points are closed, and the evacuation efficiency is increased to 1.3 times; as a result, the peak water accumulation in the station hall drops to 0.502 m and then

declines rapidly. Compared to Plan B, Plan C reduces peak water accumulation in the station hall by 14.7% and brings the time of peak occurrence forward by approximately 2.8 h. As shown in Figure 5b, the evolution of water accumulation in the track area follows a similar pattern to that of the station hall water accumulation: the peak water accumulation in the track area for Plans A and B is 0.439 m; for Plan C, due to enhanced drainage measures and the closure of entrances/exit points, the peak water accumulation in the track area is reduced to 0.352 m, a decrease of 19.7%. After the water accumulation in the track area exceeded the 0.35 m service disruption threshold at 22.40 h, it declined rapidly upon the activation of the Level II response at 22.45 h, with the service disruption lasting only 0.05 h; in contrast, the service disruption under Plans A and B lasted for 5.55 h. As shown in Figure 5c, regarding the number of stranded passengers: after the service disruption, the number of stranded passengers under Plans A and B continued to rise to 2790; however, under Plan C, upon the activation of the Level II response, the evacuation efficiency increased to 1.3 times, and the peak number of stranded passengers was only 40, representing a reduction of 98.6%, thereby effectively preventing large-scale passenger retention.

Table 7. Comparison of core performance indicators between open-loop–unidirectional coupling–closed-loop models.

Evaluating Indicator	Scheme A	Scheme B	Scheme C	B → C Optimize
Platform peak water accumulation depth (m)	0.589	0.589	0.502	↓ 14.7%
Platform peak time (h)	25.3	25.3	22.45	2.8 h in advance
Track area peak water accumulation (m)	0.439	0.439	0.352	↓ 19.7%
Peak passenger retention (people)	2790	2790	40	↓ 98.6%
Comprehensive risk peak probability	—	0.602	0.602	Flat
High-risk duration (h)	—	5.55	0.05	↓ 99.1%
Operational disruption duration (h)	5.55	5.55	0.05	↓ 99.1%
Operational disruption trigger time (h)	22.40	22.40	22.40	—
Level II response triggering time (h)	—	—	22.45	—
Level III response triggering time (h)	—	—	Not triggered	—

Note: All reported percentage reductions are calculated relative to the unidirectional coupling (Plan B) baseline, under strictly controlled simulation conditions where rainfall forcing, initial conditions, drainage, and evacuation parameters remain identical across all three scenarios. These values are internal simulation outputs specific to the Qingdao Jinggangshan Road Station case, representing controlled model scenario comparisons rather than empirically measured outcomes from real-world emergency events. The absolute reduction magnitudes cannot be directly generalized to other metro systems without local parameter recalibration. Note: "↓" denotes a reduction relative to the Plan B (unidirectional coupling) baseline.

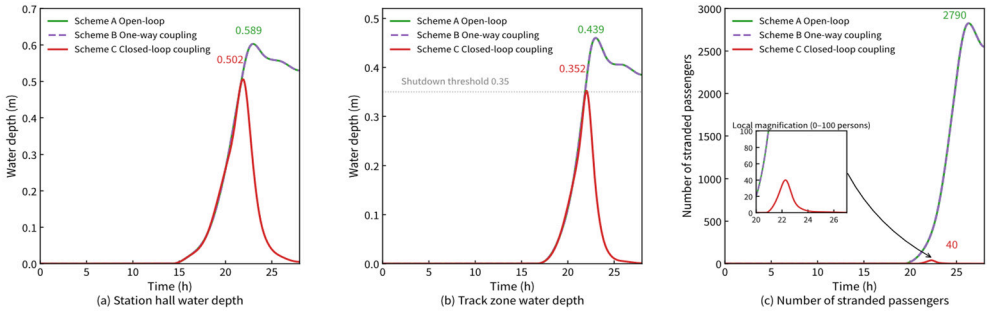


Figure 5. Comparative analysis of core physical performance variables across the three scenarios. Note: Schemes A and B produce nearly identical trajectories under the same controlled conditions and therefore overlap; this is expected, as one-way coupling (B) provides no closed-loop feedback benefit over open-loop operation (A).

When further analyzed in conjunction with the core indicators presented in Table 7, although Plan C and Plan B exhibit nearly identical overall peak risk levels, Plan C reduces the operational disruption duration from 5.55 h to 0.05 h, lowers the peak passenger retention count from 2790 to 40, and shortens the duration of high-risk periods by 99.1%. This comparison demonstrates that the core value of closed-loop management lies not in suppressing peak risk levels, but in leveraging a rapid "perception → assessment → response" closed-loop process to significantly compress the high-risk exposure window and minimize the potential impact of disasters to the lowest possible level.

There is a 0.05 h time offset between Level-II response activation (22.45 h) and the occurrence of station hall water peak (22.50 h), which originates from the fixed simulation time step of 0.05 h adopted in this study. In this manuscript, high-risk duration is defined as the total simulation time during which track-area water depth exceeds the service-disruption threshold of 0.35 m. Operational-disruption duration is counted based on continuous time steps satisfying the disruption condition under the 0.05 h time-step setting.

4.2. Disaster Chain Critical Threshold Identification

Using Scheme C to derive time-series data, the critical transition thresholds during the disaster chain evolution process are identified, as shown in Table 8.

Table 8. Key critical thresholds for the metro heavy rainfall disaster chain.

Disaster Event	Critical Threshold	Triggering Time (h)	Disaster Chain Phase Transition
Rainwater backflow activation	Surface water accumulation $\geq$ 0.15 m	14.75	Hazard incubation → Invasion
Station hall flooding risk	Platform water accumulation $\geq$ 0.05 m	17.00	The station has been flooded.
Tunnel intake water risk	Tunnel water level $\geq$ 0.10 m	18.75	Tunnel space is flooded.
High risk of water accumulation in the station hall	Platform water accumulation $\geq$ 0.20 m	18.80	Increased risk associated with mechanical and electrical equipment
Track-area enters medium-risk state	Water accumulation in the track area $\geq$ 0.25 m	20.95	Track area transitions from low to medium risk
Operational disruption trigger	Track area water accumulation $\geq$ 0.35 m	22.40	Full-service suspension
Level II Emergency Response Triggered	Overall risk probability $\geq$ 0.30	22.45	Emergency Response Initiated

As shown in Table 8, the disaster chain exhibits a distinct stepwise propagation pattern, involving six critical transition nodes—from the initiation of backflow to the occurrence of operational disruption—where each node is triggered sequentially along the propagation path: "surface water accumulation → station hall water accumulation → tunnel inundation → track area water accumulation → operational disruption." The time intervals between adjacent critical nodes vary: the transition from medium-risk to high-risk water accumulation in the station hall takes only 1.8 h, representing the fastest propagation rate. Based on the attribution analysis using Equations (2) and (4), the physical mechanism underlying this rapid transition is as follows: between 17.00 h and 18.80 h, rainfall has just passed its peak; surface water accumulation remains at a high level and consistently exceeds the water threshold height of 0.15 m, and the backflow rate Qin remains at a relatively high value. At this stage, the tunnel inundation depth has not yet reached the

medium-risk threshold of 0.10 m; the electromechanical equipment is not submerged, the performance of the drainage system has not deteriorated, and the pumping station discharge rate Qpump remains constant at the design value of 100 m^3/h. The equivalent area of the station hall, As, is only 800 m^2; this relatively small area causes the net inflow rate to translate into a rapid water level rise rate. The combined effect of these factors leads to a swift increase in water accumulation within the station hall during this phase—rising from 0.05 m to 0.20 m within 1.8 h. This mechanistic explanation indicates that during the window period when backflow has already commenced but the drainage system has not yet been compromised, the water accumulation rises at its fastest rate, making this the most critical time node for emergency response. In contrast, the interval from the high-risk state in the station hall to the medium-risk state in the track area lasts 2.15 h, representing the slowest propagation rate; this reflects the varying rates at different stages of the disaster's propagation—namely, the rate of water accumulation increase within the station hall is faster than the rate at which the water spreads to the track area. The total duration from the backflow initiation to the operational disruption is approximately 7.65 h, providing a critical time window for the emergency response to metro disasters caused by heavy rainfall. We note that response benefits vary with exact activation timing; this window indicates the available lead time for effective response, rather than implying equivalent effectiveness for action at arbitrary points within the interval. Systematic activation-timing optimization is identified as future work.

5. Discussion

5.1. Overall Findings and Broader Context

This study developed a bidirectional closed-loop SD–BN coupling model for the metro rainstorm disaster chain, applied to the Qingdao Metro "July 23" event. Against the broader background of growing climate-driven urban flood risks and the need for improved underground infrastructure resilience, three key findings emerge that advance existing scholarly discourse on metro disaster-chain simulation. Three key findings emerge.

First, closing the SD→BN→SD feedback loop substantially reduces operational disruption and passenger retention compared with open-loop and one-way coupling under identical rainfall input. The closed-loop mode shortens operational disruption duration by 99.1% and lowers peak passenger retention by 98.6%. Notably, the mechanism primarily accelerates post-peak risk decline and shortens high-risk exposure duration, rather than suppressing peak risk levels. This suggests that operational feedback—disruption-aware emergency response—matters as much as physical flood propagation in determining metro system outcomes, a finding consistent with the broader urban flood resilience literature emphasizing the role of adaptive management.

Second, the entropy-weighted BN identifies critical nodes along the disaster chain, which align with physically bottlenecked elements such as entrance backflow and station hall inundation. The 7.65 h interval between backflow initiation and operational disruption provides a quantifiable early-warning window, and six critical transition nodes mark the stepwise propagation from continuous accumulation to discrete service failure. This adds empirical quantitative evidence to existing discussions on temporal early-warning windows for underground-space flood disasters.

Third, the three-tier recursive CPT calibration method—combining maximum likelihood estimation, Laplace smoothing, and physical monotonicity constraints—successfully parameterizes a 13-node network from 47 historical cases, passing both normalization (109 parent–node combinations) and monotonicity checks. This provides a reusable pathway for small-sample BN calibration in urban infrastructure risk analysis. It responds to a

widely acknowledged methodological gap: many BN risk models for infrastructure heavily rely on subjective expert elicitation when empirical case records are scarce.

These findings contribute to the broader discourse on sustainable urban rail infrastructure resilience under climate change, aligning with SDG 11 and SDG 13.

5.2. Comparison with Existing SD–BN Coupling Studies

Existing SD–BN coupling studies have addressed risk analysis in several domains. Zhang et al. [20] combined BN static probabilistic inference with SD dynamic simulation for drilling construction risk analysis, but the BN inference results were not fed back to adjust SD control parameters; the model operated in a static series fashion. Khan et al. [21] applied SD simulation and BN inference to marine fuel replenishment operations, where SD provided time-varying inputs to BN but again without a BN→SD strategic feedback channel. The bidirectional closed-loop coupling model proposed in this study differs from prior SD–BN integration approaches in two key aspects. First, existing studies [20,21] typically adopt a unidirectional architecture where SD provides time-varying inputs to BN but BN results do not feed back to adjust SD parameters. In this study, the BN→SD feedback channel dynamically adjusts drainage capacity, entrance closure rate, and evacuation efficiency based on real-time risk levels, forming a complete "perception–decision–action" loop. Second, the three-tier recursive CPT calibration method addresses small-sample sparsity through maximum likelihood estimation, Laplace smoothing, and physical monotonicity constraints, reducing reliance on subjective expert judgment. In this study, entropy weighting is adopted because it objectively reflects the distinguishing ability of each consequence node calculated from the 47 historical accident records, avoiding subjective expert-driven weight assignment. Consequence-severity weighting would heavily rely on manual expert scoring, which is less suitable for this small-sample dataset. The 14.7% peak water depth reduction is modest, which is expected because the Level II response is triggered only 0.05 h before the natural peak; this finding suggests that metro flood management should prioritize rapid response activation over peak suppression. This study was motivated by the Qingdao Metro "July 23" event, which exposed a practical gap: existing SD models simulate physical flood evolution but cannot dynamically adjust emergency response based on real-time risk probability, while BN models infer risk probabilistically but do not capture continuous physical dynamics. The bidirectional closed-loop design directly addresses this gap.

In this study, entropy weighting is adopted because it objectively reflects the distinguishing ability of each consequence node calculated from the 47 historical accident records, avoiding subjective expert-driven weight assignment. Consequence-severity weighting would heavily rely on manual expert scoring, which is less suitable for this small-sample dataset.

5.3. Practical Implications for Metro Flood Management

The Qingdao case study yields several practical insights. The 7.65 h interval between backflow initiation and operational disruption provides a critical early-warning and response lead time. Under the modeled trigger timing, activating emergency measures within this window reduces disruption duration by 99.1%; as discussed in Section 4.2, the realized benefit depends on the exact activation point and is not uniform across the entire interval. The critical threshold analysis further identifies the transition from medium-risk to high-risk station hall water depth as the fastest bottleneck (1.8 h), highlighting the need for early intervention at this stage. These findings can support metro operators in optimizing response trigger thresholds and resource deployment timing.

5.4. Limitations

This study has several limitations. First, the model operates at the single-station level and does not account for passenger flow transfer between multiple stations or network-level cascading effects. Second, the 47 historical cases are predominantly from Chinese metro systems; generalizability to other urban contexts requires further validation. Third, the tiered response parameters are based on national emergency guidelines and engineering experience, and could be improved with field calibration. A related simplification is that emergency control parameters (β, η, γ) are implemented as instantaneous step changes once thresholds are crossed, rather than gradual ramp-up, as occurs in real emergency response; this on–off control approximation does overstate short-term response speed, though it does not change the long-term system behavior of faster post-peak risk decline. Fourth, the model assumes instantaneous BN inference and response activation, without considering decision-making and implementation delays. We explicitly note that the reported 0.05 h (3 min) operational disruption duration is an idealized lower bound under this instantaneous-response assumption: accounting for even a conservative 5 min on-site response delay, the simulated disruption duration would increase to approximately 0.13 h (8 min), corresponding to a 97.7% reduction rather than 99.1%, though the qualitative conclusion that closed-loop coupling substantially reduces disruption duration remains robust. Fifth, several parent–node combinations within CPT have zero observed samples and their probabilities are estimated by Laplace smoothing and isotonic regression rather than direct empirical observation; these smoothed CPT entries introduce additional uncertainty for risk inference, and expanding the case database is recommended in future work, and case-subset-based sensitivity simulation is also suggested for future investigation. Accordingly, the simulated improvement magnitudes (99.1% for operational-disruption reduction and 98.6% for peak passenger-retention reduction) are specific to the Qingdao case under the set simulation scenario. These values cannot be generalized to other metro systems without further multi-case empirical validation.

6. Conclusions

To advance sustainable urban rail infrastructure resilience under climate change, this study developed a bidirectional closed-loop coupled model integrating SD and BN for simulating the metro rainstorm disaster chain, addressing the challenge that single-method approaches cannot simultaneously capture continuous physical evolution, discrete state transitions, and probabilistic uncertainty, with the Qingdao Metro "July 23" flooding incident adopted as a case study. The model shows that, relative to the open-loop baseline under identical rainfall input, the closed-loop coupling shortens operational disruption duration by 99.1% and lowers peak passenger retention by 98.6%. These are model-internal scenario results under controlled simulation conditions; the absolute magnitudes should not be directly generalized to other metro systems without further empirical validation. These simulated improvement values are also affected by uncertainty originating from the limited 47-case observational dataset adopted for BN parameter calibration.

It should be explicitly clarified that these quantitative reduction values are specific to the Qingdao Jinggangshan Road Station case under the set simulation scenario, and are not universal performance indicators applicable to all metro systems. The generalizable contribution of this study lies in the proposed bidirectional closed-loop SD–BN coupling framework and the three-tier recursive CPT calibration approach, rather than the specific percentage magnitudes reported for this single case. Consistent with the limitations discussed in Section 5.4, these case-specific results are constrained by the single-station setup, the China-dominated historical accident dataset, engineering-estimated response parameters, and the assumption of instantaneous response activation; direct generalization

of these numerical values to other metro systems or cities requires additional multi-site empirical validation and local parameter recalibration.

Three key findings emerge from the Qingdao case. First, the SD–BN bidirectional closed-loop coupling mechanism synchronizes physical disaster evolution with real-time risk assessment, reducing operational disruption duration by accelerating post-peak risk decline rather than suppressing peak levels. Second, the three-tier recursive CPT calibration method, combining maximum likelihood estimation, Laplace smoothing, and physical monotonicity constraints, is validated across 109 parent–node combinations and provides a reusable framework for small-sample BN parameter calibration. Third, the critical-threshold-based disaster chain analysis framework quantitatively identifies six transition nodes along the temporal propagation path, revealing a 7.65 h early-warning window between backflow initiation and operational disruption to support optimal emergency response timing. Beyond this case, the coupling methodology and CPT calibration framework are transferable to other urban metro systems after local parameter recalibration. The bidirectional SD–BN coupling mechanism can also be adapted to other small-sample disaster-chain scenarios beyond metro flooding, and the critical threshold identification approach supports proactive emergency response timing in broader urban waterlogging and underground space safety contexts. These findings contribute to sustainable urban rail infrastructure resilience management under climate change, aligned with UN SDG 11 and SDG 13. Specifically, the dynamic risk assessment and early warning framework supports SDG 11 by enhancing the safety and resilience of urban public transport infrastructure, and the closed-loop emergency response mechanism supports SDG 13 by improving adaptive capacity to extreme rainfall and flood events.

Future research directions include expanding the model from single-station to multi-station network-level simulation—based on published metro emergency literature, typical on-site decision and implementation delays for station flood response range from 5 to 15 min; incorporating such delays would reduce the magnitude of observed closed-loop improvements, though the core post-peak risk decline mechanism would remain qualitatively valid; incorporating decision-making delay into the BN → SD feedback loop; validating the model across multiple cities and flood events; and integrating real-time rainfall forecasting for proactive emergency response.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su182110713/s1.

Author Contributions: Writing—review and editing: H.T. and Z.L.; visualization: H.T.; methodology: H.T.; funding acquisition: H.T.; investigation: H.T. and Z.L.; formal analysis: H.T., Z.L. and Z.S.; writing—original draft: Z.L.; validation: Z.L.; conceptualization: Z.L.; supervision: Y.L., K.S. and Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Shandong Province, grant numbers ZR2024QG020 and ZR2026MS0942.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: During the preparation of this manuscript/study, the author(s) used the AI-assisted diagramming website PicDoc, and Python 3.14 with Matplotlib 3.11.1 for the pur-poses of generating Figure 1 and all other figures based on simulation data. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Markolf, S.A.; Hoehne, C.; Fraser, A.; Chester, M.V.; Underwood, B.S. Transportation resilience to climate change and extreme weather events—Beyond risk and robustness. *Transp. Policy* **2019**, *74*, 174–186. [CrossRef]
2. Zhang, D.-M.; Du, F.; Huang, H.; Zhang, F.; Ayyub, B.M.; Beer, M. Resiliency assessment of urban rail transit networks: Shanghai metro as an example. *Saf. Sci.* **2018**, *106*, 230–243. [CrossRef]
3. Kappes, M.S.; Keiler, M.; von Elverfeldt, K.; Glade, T. Challenges of analyzing multi-hazard risk: A review. *Nat. Hazards* **2012**, *64*, 1925–1958. [CrossRef]
4. Liu, G.; Li, J.; Wang, H.; Gong, Y.; Tao, F. Identification of urban rainstorm and flood disaster chains and systematic risk assessment. *Chin. J. Manag. Sci.* **2025**, *33*, 222–231.
5. Kasaei, S.; Orton, P.M.; Ralston, D.K.; Warner, J.C. Pluvial and potential compound flooding in a coupled coastal modeling framework: New York City during post-tropical Cyclone Ida (2021). *Hydrol. Earth Syst. Sci.* **2025**, *29*, 2043–2058. [CrossRef]
6. Jang, M.; Yoon, J.; Yun, Y.; Kim, J.; Jang, H.; Benmarhnia, T.; Han, C. Changes in healthcare utilization after the 2022 Seoul metropolitan flood: Applying a generalized synthetic control approach. *GeoHealth* **2024**, *8*, e2024GH001084. [CrossRef] [PubMed]
7. Sian, K.T.C.L.K.; Sagero, P.; Kebacho, L.L.; Ongoma, V. Dubai's record precipitation event of 16 April 2024-A diagnosis. *Atmos. Res.* **2025**, *315*, 107924. [CrossRef]
8. Shukla, K.K.; Mishra, A.K.; Kumar, G.; Sahu, D.K.; Bhute, S.; Adhikary, S. Atmospheric drivers of the 26 May 2025 heavy rainfall event over Mumbai: Insights from observations and reanalysis. *Nat. Hazards* **2026**, *122*, 380. [CrossRef]
9. Tang, H.X.; Zheng, J.X.; Li, M.D.; Wang, F.; Shao, Z.G. Network characteristics and resilience enhancement strategies for the heavy rainfall risk evolution network in urban rail transit systems. *Railw. Stand. Des.* **2025**, *69*, 64–71.
10. Huggins, T.J.; E, F.; Chen, K.; Gong, W.; Yang, L. Infrastructural aspects of rain-related cascading disasters: A systematic literature review. *Int. J. Environ. Res. Public Health* **2020**, *17*, 5175. [CrossRef]
11. Jiao, L.D.; Cai, Z.P.; Huo, X.S.; Wang, H. A risk assessment study on the multi-layer disaster chain caused by heavy rainfall and flooding in urban rail transit spatial systems. *J. Railw. Sci. Eng.* **2026**, *23*, 2881–2892.
12. Li, H.R.; Ouyang, Z.L.; Jiang, J.; Yang, Q.F.; Liu, B.; Xi, Y.H. Urban rail transit disaster chain evolution network model and its risk analysis: A case study of metro flooding. *Railw. Stand. Des.* **2020**, *64*, 153–157.
13. Liu, Y.Y.; Li, J.W.; Chen, R.S.; Wu, H.Y.; Wnag, C.; Du, S.Q. A study on the disaster chain of extreme rainstorm-induced flooding in Zhengzhou City based on complex networks. *Disaster Sci.* **2024**, *39*, 227–234.
14. Wang, W.; Qu, Z.N.; Guo, X.D.; Ma, D.H. Construction approach and methodology for an urban multi-disaster correlation framework and chain network. *China Saf. Sci. Technol.* **2023**, *19*, 5–12.
15. Bakhtiari, S.; Najafi, M.R.; Goda, K.; Peerhossaini, H. A dynamic Bayesian network approach to characterize multi-hazard risks and resilience in interconnected critical infrastructures. *Reliab. Eng. Syst. Saf.* **2025**, *257*, 110815. [CrossRef]
16. Ren, Y.C.; Zhang, R.; Zhang, Y.S.; Sun, H.; Li, M.; Liu, Q.H.; Xu, J. Scenario analysis and simulation exercise for the "Zhengzhou heavy rainfall-induced metro disaster event" based on Bayesian networks. *Chin. J. Atmos. Sci.* **2023**, *46*, 904–916.
17. Zhang, C.; Weng, W.G.; Chen, Y.; Dai, B.Q.; Qin, T.X. Research on a risk early warning model and collaborative response mechanism for subway flooding caused by heavy rainfall based on Bayesian networks. *China Saf. Sci. Technol.* **2023**, *19*, 179–186.
18. Wang, L.; Zhang, C.G.; Huo, Y.N. Scenario construction and evolutionary analysis of metro flood event scenarios based on Bayesian networks. *Saf. Environ. Eng.* **2023**, *30*, 101–106+172.
19. Jiang, B.; Zhang, C.; Chen, T.; Yuan, H.Y.; Fan, W.C. A Bayes network-based method for constructing and evolving heavy rainfall scenarios. *J. Tsinghua Univ. (Nat. Sci. Ed.)* **2021**, *61*, 509–517.
20. Murphy, K.P. Dynamic Bayesian Networks: Representation, Inference and Learning. Ph.D. Thesis, University of California, Berkeley, CA, USA, 2002.
21. Sterman, J. *Business Dynamics: Systems Thinking and Modeling for a Complex World*; Irwin/McGraw-Hill: Boston, MA, USA, 2000.
22. Li, K.; Xiahou, X.; Wu, Z.; Shi, P.; Tang, L.; Li, Q. Influencing factor identification and simulation for urban metro system operation processes—A resilience enhancement perspective. *Systems* **2024**, *12*, 43. [CrossRef]
23. Wang, N.; Wu, M.; Yuen, K.F. Modelling and assessing long-term urban transportation system resilience based on system dynamics. *Sustain. Cities Soc.* **2024**, *109*, 105548. [CrossRef]
24. Zhang, J.; Wu, L.; Zhang, T.; Wu, D. Systemic and dynamic risk analysis of drilling construction based on Bayesian network and system dynamics model. *Sci. Rep.* **2025**, *15*, 27989. [CrossRef] [PubMed]
25. Khan, H.B.; Zhang, T.; Pan, Z.S. Dynamic risk analysis of bunkering operations based on system dynamics simulation and Bayesian network. *Reliab. Eng. Syst. Saf.* **2026**, 112564. [CrossRef]
26. Tilloy, A.; Malamud, B.D.; Winter, H.; Joly-Laugel, A. A review of quantification methodologies for multi-hazard interrelationships. *Earth-Sci. Rev.* **2019**, *196*, 102881. [CrossRef]
27. Xin, S.; Li, Z.; Nong, J.; Wu, J.; Zou, X.; Wu, R.L.; Dong, S.; Wang, S.; Wu, R.J. Environmental system dynamics: Current development and applications. *Ecol. Model.* **2025**, *506*, 111135. [CrossRef]

28. Li, J.X.; Wang, Y.R.; Chang, L.; Li, Z.; Zhao, T.L.; Wang, R.F. Experimental study and theoretical analysis of the discharge formula for rectangular thin-walled weirs under free-flow conditions. *J. Hexi Univ.* **2022**, *38*, 25–36.
29. *GB 50157-2013*; Code for Design of Metro. Ministry of Housing and Urban-Rural Development of the People's Republic of China: Beijing, China, 2013.
30. Shampine, L.F. *Numerical Solution of Ordinary Differential Equations*; Routledge: London, UK, 2018.
31. Yang, T.Y.; Cui, Y.M.; Wang, H. A dual-driven resolution mechanism for breakthrough innovation risks: Based on the context of Jiaokong technology. *Chin. J. Manag.* **2024**, *21*, 1121–1131.
32. *GB/T 28592-2012*; Grade of Precipitation. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China: Beijing, China, 2012.
33. *GB 51222-2017*; Technical Code for Urban Flooding Prevention and Control. Ministry of Housing and Urban-Rural Development of the People's Republic of China: Beijing, China, 2017.
34. *GB 50014-2021*; Standard for Design of Outdoor Wastewater Engineering. Ministry of Housing and Urban-Rural Development of the People's Republic of China: Beijing, China, 2021.
35. Ahmadi Nadi, A.; Afshari, R.; Sadeghpour Gildeh, B. Control charts for monitoring relative risk rate in the presence of Weibull competing risks with censored and masked data. *Qual. Technol. Quant. Manag.* **2024**, *21*, 340–362. [CrossRef]
36. Sakai, T. The probability smoothing problem: Characterizations of the Laplace method. *Math. Soc. Sci.* **2025**, *135*, 102409. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.