

Article

Predicting Typhoon Flood in Macau Using Dynamic Gaussian Bayesian Network and Surface Confluence Analysis

Shujie Zou ¹, Chiawei Chu ^{1,*} , Weijun Dai ², Ning Shen ³, Jia Ren ⁴ and Weiping Ding ⁵ ¹ Faculty of Data Science, City University of Macau, Macau 999078, China; d21092100231@cityu.mo² Artificial Intelligence College, Guangdong Polytechnic Institute, Guangzhou 510091, China; wjdai@gdrtvu.edu.cn³ Department of Innovation, Technology and Entrepreneurship, United Arab Emirates University, Al Ain 15551, United Arab Emirates; ningshen@uaeu.ac.ae⁴ School of Information and Communication Engineering, Hainan University, Haikou 570100, China; renjia@hainanu.edu.cn⁵ School of Information Science and Technology, Nantong University, Nantong 226000, China; ding.wp@ntu.edu.cn

* Correspondence: cwchu@cityu.mo

Abstract: A typhoon passing through or making landfall in a coastal city may result in seawater intrusion and continuous rainfall, which may cause urban flooding. The urban flood disaster caused by a typhoon is a dynamic process that changes over time, and a dynamic Gaussian Bayesian network (DGBN) is used to model the time series events in this paper. The scene data generated by each typhoon are different, which means that each typhoon has different characteristics. This paper establishes multiple DGBNs based on the historical data of Macau flooding caused by multiple typhoons, and similar analysis is made between the scene data related to the current flooding to be predicted and the scene data of historical flooding. The DGBN most similar to the scene characteristics of the current flooding is selected as the predicting network of the current flooding. According to the topography, the influence of the surface confluence is considered, and the Manning formula analysis method is proposed. The Manning formula is combined with the DGBN to obtain the final prediction model, DGBN-m, which takes into account the effects of time series and non-time-series factors. The flooding data provided by the Macau Meteorological Bureau are used to carry out experiments, and it is proved that the proposed model can predict the flooding depth well in a specific area of Macau under the condition of a small amount of data and that the best predicting accuracy can reach 84%. Finally, generalization analysis is performed to further confirm the validity of the proposed model.

Keywords: dynamic Gaussian Bayesian network; Manning formula; flood prediction; surface confluence**MSC:** 68T09

check for updates

Citation: Zou, S.; Chu, C.; Dai, W.; Shen, N.; Ren, J.; Ding, W. Predicting Typhoon Flood in Macau Using Dynamic Gaussian Bayesian Network and Surface Confluence Analysis. *Mathematics* **2024**, *12*, 340. <https://doi.org/10.3390/math12020340>

Academic Editor: Jose Luis Vicente Villardon

Received: 13 December 2023

Revised: 9 January 2024

Accepted: 17 January 2024

Published: 19 January 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nowadays, flood disaster is still a major problem faced by many cities. The probability of flood disaster is very high, especially in coastal cities. In coastal cities, when the typhoon comes, it will cause heavy rainfall, seawater backflow and other phenomena. There will be surface confluence in areas with larger terrain fluctuations, which will lead to severe flood disaster in areas with lower terrain. According to statistics, flood disaster in various regions of the world will cause huge economic losses and casualties every year [1]. For example, several provinces in China were hit by floods in 2017, resulting in tens of millions of people being affected, of which Macau was heavily hit. If the trend of flooding can be predicted in advance, some protective measures can be taken in advance to avoid casualties and reduce economic losses.

Flood disaster prediction is a concern of many researchers, and many prediction schemes have been proposed. Among them, machine learning methods have been successfully applied to flood disaster prediction by many researchers, and have achieved good results in terms of computational speed and accuracy [2]. For example, Tehrani et al. [3] combined a support vector machine and decision tree to analyze the correlation between flood risk level and various influencing factors. Satria et al. [4] developed a system to predict flood depth in Manila city using K-nearest neighbors (KNNs) and inverse-distance-weighted interpolation (IDW). Wasiq et al. [5] combined machine learning models and data analysis methods with Internet of Things (IoT) sensor data to predict flood risk levels, but the prediction results were intervals, which can only provide vague reference results. Du et al. [6] used Soil Moisture Active Passive (SMAP) and Landsat monitoring to evaluate and predict flood inundation, and achieved good results, but the model used was complex and difficult to implement. Due to the rapid development of deep learning, some researchers use the advantages of artificial neural networks to describe the linear and nonlinear characteristics of data to establish flood prediction models. For example, Ramil et al. [7] combined a back propagation (BP) neural network with a Kalman filter to model the upstream and downstream data of the water flow to predict the water level downstream of the water flow. Sunita et al. [8] conducted a study of water catchments in the UK using an improved artificial neural network. Imrie et al. [9] built a complex flood early warning model using neural networks and achieved good prediction results, which provided help for flood prevention. Kim et al. [10] used an artificial neural network to predict after-runner storm surges on the coast of Tottori, Japan. The prediction of storm surges can indirectly prevent flooding disasters. Gude et al. [11] combined autoregressive integrated moving average (ARIMA) and a long short-term memory network (LSTM) to build the model for predicting flood depth. Kourgialas et al. [12] built a neural network to predict extreme water flow in a small agricultural watershed in the Mediterranean, demonstrating the predictive advantage of neural networks for such scenarios. In addition, Dai et al. [13,14] used a neural network and ensemble learning methods to predict the flood depth of Macau during a typhoon period, respectively, conducting many experiments and exploring a variety of research paths to obtain effective research and prediction results, and also provided a good research foundation for the work of the paper.

Although the above various machine learning algorithms (neural network, ensemble learning, etc.) have achieved certain results in flood disaster prediction, these algorithms require a large amount of training data to obtain a better prediction ability. Taking Macau flooding as an example, the data provided by the Macau Meteorological Bureau show that, due to equipment conditions, the amount of data generated by the flood disaster in Macau caused by typhoons is relatively small. There are currently only six recorded typhoons, of which the typhoon with the longest duration of flooding (Mangkhut) produced only 759 pieces of data on the flooding depth (data recorded every one minute), and the data volume of other typhoons causing flooding in Macau is below 300. More importantly, the prediction model established by a neural network cannot know the relationship between each flooding factor and the relationship between each flooding factor and the flooding depth (direct or indirect influence), which is the reason for why the neural network is called a black box. In most cases, it is necessary not only to predict the flooding depth but also to know which flooding factors have a direct or indirect impact on the flooding depth so that appropriate and accurate measures can be taken to prevent the flooding disaster. Therefore, a DGBN is adopted as the predicting network of flooding depth in this paper, which has the following advantages in Macau's flooding depth prediction scenario:

- When the amount of data is small, expert experience can be added in the construction of the Bayesian network to guide the direction of algorithm learning. And, the characteristic information of the flooding factor is further enhanced, and the learned network structure will also better fit the characteristics of the data.

- Bayesian networks have multiple network structures (stationary, non-stationary, high-order, low-order) to deal with changing scenarios. According to different scene characteristics, the appropriate network structure is selected to process the data.
- After the Bayesian network is established, the relationship between various flooding factors and the relationship between flooding factors and flooding depth can be clearly known, which will help relevant departments to take effective flood control measures.

The content of this paper is arranged as follows: Section 2 introduces the literature on the use of Bayesian networks for related research; Section 3 introduces the principle and construction process of the DGBN; Section 4 presents the surface confluence analysis, which leads to the Manning formula analysis method; Section 5 designs the prediction model of the DGBN combined with the Manning formula; Section 6 conducts experiments to prove the effectiveness of the proposed scheme based on the flooding data provided by the Macau Meteorological Bureau; Section 7 analyzes and summarizes the full paper.

2. Related Work

At present, many researchers used a Bayesian network for predictive analysis and obtained better results. For example, Lu used expert knowledge to establish a Bayesian network to assess the flood risk of the Xianghongdian Reservoir, and used the unique attributes of the Bayesian network to make a flexible two-way inference to obtain the probability distribution of each node, more comprehensively understanding and controlling floods [15]. Sebastian used observations from hundreds of tropical cyclones in the Gulf of Mexico to build a non-parametric Bayesian network to simulate storm surges in coastal watersheds and infer possible floods and some uncertain events during storm surges [16]. Sen constructed two dynamic Bayesian networks to assess resilience after natural disasters in various parts of Barak Valley in northeastern India, describing trends over time and differences in resilience among regions [17]. Chen et al. [18] used the uncertainty between expert knowledge and variables to build a dynamic Bayesian network, and used Monte Carlo simulation to provide input data for the dynamic Bayesian network to assess the real-time flood control dispatch risk of a multi-reservoir system in China. In addition, some researchers use Bayesian networks for research in other fields. David et al. [19] established a high-order DGBN to predict the temperature in an industrial furnace for a long time, and the prediction results show that the long-term prediction ability of the DGBN is better than that of a convolutional recursive neural network. Fateme et al. [20] built a dynamic Bayesian belief network to dynamically assess Australia's energy reserves to provide reference and support for power system suppliers' decision making. Dong et al. [21] used a dynamic Bayesian network to model the characteristics of battery degradation during charging and used the established dynamic Bayesian network to predict the health of the battery. Some researchers also used Bayesian networks for risk assessment in different scenarios [22–26]. For example, Zhang et al. [22] proposed fuzzy probabilistic Bayesian networks for network security assessment in industrial control systems; Ma et al. [23] used the dynamic Bayesian network to make a reasonable quantitative assessment of the risks associated with driving, etc. It can be seen that the dynamic Bayesian network has a powerful inference and prediction function. Although much effort has been dedicated toward prediction typhoon flooding, the noted algorithms suffer from the following limitations and challenges:

- i The data studied are basically from the same scene, which is equivalent to the same features involved in the events studied.
- ii These Bayesian network models are built under a large amount of data without considering a small amount of data.
- iii The Bayesian network structure of some of the literature is completely established by experts' experience, which may not capture the uncertainties during the flood.

The data analyzed come from different typhoon scenarios in this paper because the characteristics of each typhoon and the trends and ranges of changes in related weather

attributes caused by typhoons are basically not the same or are even very different. More importantly, the various factors that cause floods in Macau are not exactly the same for each typhoon. For example, some flooding events are caused by heavy rainfall, and some flooding events are caused by a combination of rainfall and storm surge. Therefore, it is necessary to consider how to deal with the different scenarios of flood disasters caused by typhoons before establishing the prediction model. For example, Xu et al. [27] proposed to use similar historical flood alarm sequences to predict the upcoming alarm events of the current flood alarm sequence. In this paper, similarity analysis (Euclidean distance) is used to calculate the similarity between the relevant scene features of the flooding that needs to be predicted and the historical scene features. Then, the DGBN is found as the network model of the flood to be predicted, which is established by the historical flood event most similar to the flood to be predicted. Furthermore, the influence of surface confluence is analyzed, and the surface confluence model is combined with the DGBN (the terrain factor is successfully added to the prediction model), which makes the prediction model of this paper have a convincing and better prediction ability.

3. Dynamic Gaussian Bayesian Network

Bayesian networks are probabilistic graphical models and are directed acyclic, consisting of nodes and directed arcs [19]. Bayesian networks can deal with discrete and continuous variables, and each node in Bayesian networks has a probability distribution: Gaussian distribution (continuous variable) or conditional probability distribution (discrete variable). The expression for the joint probability distribution of the static Gaussian Bayesian network (SGBN) is

$$P(V) = \prod_{i=1}^k f(v_i | pa(v_i)) = \prod_{i=1}^k N(\gamma_{0,i} + \sum_{j=1}^{k_i} \gamma_{j,i} pa_j(v_i); \delta_i^2) \tag{1}$$

where k is the number of network nodes, k_i is the number of parent nodes of the i -th node, $f(\cdot)$ is the probability density function, $V = (v_1, v_2, \dots, v_k)$ is the node set, v_i is the i -th node, $pa(v_i) = \{pa_1(v_i), pa_2(v_i), \dots, pa_{k_i}(v_i)\}$ is the parent node set of the i -th node, $pa_j(v_i)$ is the j -th parent node of the node v_i , δ_i^2 is the variance of the i -th node, $\gamma_{0,i}$ is the intercepts of the i -th node and $\gamma_{j,i}$ is the coefficient of the j -th parent node of the i -th node. Generally speaking, SGBNs become DGBNs after adding the time factor, and the expression of the joint probability distribution of the DGBN is

$$P(V_0, V_1, \dots, V_T) = f(V_0) \prod_{t=0}^{T-1} f(V_{t+1} | V_0, \dots, V_t) \tag{2}$$

where $V_t = (v_{1,t}, \dots, v_{k,t})$ is the node set of the t -th time slice, $v_{i,t}$ is the i -th node in the t -th time slice and T represents the number of time slices. The first-order Markov DGBN represents that the nodes in the current time slice are only affected by the nodes in the current time slice and the previous time slice and have nothing to do with the nodes in the earlier time slice. And, the high-order Markov DGBN represents that the nodes in the current time slice are affected by the nodes in the earlier time slice. For the convenience of description, the expression for the first-order Markov DGBN is given:

$$\begin{aligned} P(V_0, V_1, \dots, V_T) &= f(V_0) \prod_{t=0}^{T-1} f(V_{t+1} | V_t) \\ &= \prod_{i=1}^k N(\gamma_{0,i,0} + \sum_{j=1}^{k_{i,0}} \gamma_{j,i,0} pa_j(v_{i,0}); \delta_{i,0}^2) \\ &\quad \prod_{t=0}^{T-1} \prod_{i=1}^k N(\gamma_{0,i,t+1} + \sum_{j=1}^{k_{i,t+1}} \gamma_{j,i,t+1} pa_j(v_{i,t+1}); \delta_{i,t+1}^2) \end{aligned} \tag{3}$$

where $\gamma_{0,i,t}$ represents the intercept of the i -th node in the Bayesian network of the t -th time slice, $\delta_{i,t}^2$ represents the variance of the i -th node in the Bayesian network of the t -th time slice, $\gamma_{j,i,t}$ represents the coefficient of the j -th parent node of the i -th node in the Bayesian network of the t -th time slice and $k_{i,t}$ represents the number of parent nodes of the i -th node in the Bayesian network of the t -th time slice. The Bayesian network of each time slice can be regarded as an SGBN whose nodes may be affected by the nodes of the previous time slice.

3.1. Data Preprocessing

Before building the DGBNs, the data need to be preprocessed. Some monitoring sensors may malfunction during flooding, resulting in partial data loss. When a small amount of data are lost, linear interpolation is used to fill in the lost data. When a large amount of data are lost, the corresponding data are chosen to be discarded. In this case, the data that need to be filled have lost a large amount of feature information, and the relevant algorithms have difficulty in recovering the lost information. The expression of the linear interpolation is

$$y_c = y_a + \frac{y_b - y_a}{t_b - t_a}(t_c - t_a) \quad (4)$$

where $t_a < t_c < t_b$, y_c is the missing value at time t_c and y_a and y_b are the known values at time t_a and t_b , respectively.

3.2. Network Structure Learning

The DGBN's structure G_D consists of the initial network G_0 and the transition network G_{\rightarrow} , $G_D = (G_0, G_{\rightarrow})$. Here, the initial network G_0 is taken as an example. Firstly, stochastic Bayesian networks are established by using the uniform random acyclic directed graph algorithm. The uniform random acyclic directed graph algorithm is shown in Algorithm 1. The algorithm complexity of Algorithm 1 is lower than $O(|V|^4)$ and the details of Algorithm 1 can be found in [28]. When the Markov chain MC in Algorithm 1 converges to a uniform distribution, MC is defined to generate an acyclic digraph every i iterations. A total of N acyclic digraphs are generated, which can ensure the diversity of the generated random acyclic digraph [29]. Then, the randomly generated N Bayesian networks are used as the starting network of the tabu search algorithm combined with the Bayesian information criterion (BIC) scoring algorithm. The tabu search algorithm is shown in Algorithm 2, where the computational complexity of Algorithm 2 is lower than $O(Np'M^2)$. The expression of the BIC scoring algorithm is

$$Score(G, D) = \sum_{i=1}^{|V|} [\log f(v_i | pa(v_i)) - \frac{|d_i|}{2} \log n] \quad (5)$$

where n is the number of samples, $f(\cdot)$ is the probability density function, $|V|$ is the number of network nodes, v_i is the i -th network node, $pa(v_i)$ is the parent node of node v_i , $|d_i|$ is the number of parameters of the node, G is a directed acyclic graph and D is the dataset. It is worth noting that the blacklist variables in Algorithm 2 are combined with expert experience to optimize the algorithm, so that the network structure learned by Algorithm 2 fits the data better. After repeated analysis and comparison, the blacklist set in this paper includes:

- i Other flooding factors and flooding depths are prohibited from becoming the parent node of the typhoon track because the typhoon track is mainly affected by factors such as gravity and subtropical high pressure.
- ii It is forbidden for the flooding depth to become the parent node of the flooding factor because the flooding factor affects the change in the flooding depth.

Algorithm 1 Uniform Random Acyclic Directed Graph Algorithm

Input: network node $V = (v_1, v_2, \dots, v_k)$; a Markov Chain MC whose state space $ST = \{s_1, \dots, s_t, \dots\}$ is all acyclic directed graphs composed of network nodes V , s_t is the state of the Markov chain MC at time t .

- 1: **Initialization:** the initial state of MC is an empty graph.
- 2: set the state transition function $f(s_t)$ of the Markov chain:
- 3: uniformly randomly set ordered pairs (v_i, v_j) , $v_i, v_j \in V$, $v_i \neq v_j$.
- 4: if there is an arc e between the ordered pair (v_i, v_j) in s_t , then $s_{t+1} = s_t \setminus e$, which means that the arc e is removed from the state s_t .
- 5: if there is no arc e between the ordered pair (v_i, v_j) in s_t , then s_{t+1} is equal to adding arc e in s_t , and check whether s_{t+1} is acyclic, if s_{t+1} is a cyclic graph, then $s_{t+1} = s_t$.
- 6: **for** $t = 1$ to p **do**
- 7: uniform random selection of number i from $1 : k$ to get v_i .
- 8: uniform random selection of number j from $1 : k \setminus i$ to get v_j .
- 9: $s_{t+1} = f(s_t)$
- 10: **end for**
- 11: after p iterations, the Markov chain MC converges to a uniform distribution.

After learning from Algorithm 2, N Bayesian networks are obtained. Two thresholds, α and β ($0 < \alpha < 1, 0 < \beta < 1$), are set to represent the connection strength and arc direction strength of the arc between any two vertices (v_i, v_j) , respectively. Then, we compute the ratio α' of occurrence of each arc and the ratio β' of occurrence of its direction separately in N Bayesian networks. If the conditions of $\alpha' \geq \alpha, \beta' \geq \beta$ are satisfied at the same time, the corresponding arc with direction is recorded, and the recorded arcs are defined as the set $Ar = (a_1, \dots, a_\epsilon), \epsilon \geq 1$. Finally, an initial Bayesian network G_0 is established by the set Ar . Similarly, the network structure of the transfer network G_{\rightarrow} and other time slices is jointly established by Algorithms 1 and 2.

It is worth noting that the network structure of each time slice and the transition network G_{\rightarrow} are allowed to change, which means that the number of parent nodes of each node may change in the Bayesian network of each time slice, as shown in Figure 1. Figure 1 shows the schematic of the simple DGBN with three time slices, where there are three network nodes in each time slice. The DGBN established in this way is realistic, and some flooding factors do not always have an effect on flooding depth. For example, early flooding depths may be affected by rainfall, while later flooding depths may be affected by storm surges.

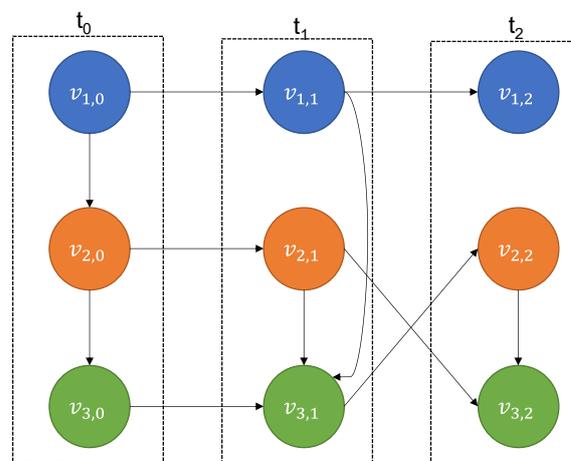


Figure 1. Schematic diagram of the one-order unsteady DGBN. Each dotted box represents a time slice, solid circles represent network nodes, the different colored nodes represent different variables, initial network $G_0 = \{v_{1,0}, v_{2,0}, v_{3,0}\}$, the network structure of the three time slices (t_0, t_1, t_2) is different and the structure of the two transfer networks $G_{\rightarrow} = \{G_{t_0 \rightarrow t_1}; G_{t_1 \rightarrow t_2}\}$ is also different.

Algorithm 2 Tabu Search Algorithm

Input: the starting Bayesian network $G = (G_1, \dots, G_N)$, dataset D , blacklist $blacklist = (e'_1, \dots, e'_i, \dots)$.

```

1: Initialization: tabu list  $tabulist = \emptyset$ , tabu length  $tabulong = TL$ , the number of
   iterations  $p'$ , neighborhood solution network  $G_O = (G_{O,1}, \dots, G_{O,N})$ , optimal scoring
   network  $G_B = (G_{B,1}, \dots, G_{B,N})$ ,  $G_B = G_O = G$ , empty list  $list = zeros(1, M)$ .
2: Definition: the neighborhood solution network  $G_O$  is equal to the operation of adding,
   subtracting, and reversing the direction of the arc to the current network  $G_B$ , and the
   set of operated arcs  $eo = (e_1, \dots, e_M)$ .
3: for  $i = 1$  to  $N$  do
4:   for  $j = 1$  to  $p'$  do
5:     for  $k = 1$  to  $M$  do
6:        $list(1, k) = BIC(G_{O,i}, e_k, D)$ 
7:     end for
8:     while  $length(list) \neq 0$  do
9:        $[a, b] = find(list == max(list))$ 
10:      if  $e_b \in blacklist$  then
11:        remove  $b$  from  $list$ 
12:        continue
13:      end if
14:      if  $e_b \in tabulist$  then
15:        if  $max(list) > BIC(G_{B,i}, D)$  then
16:          remove  $e_b$  from  $tabulist$ 
17:           $G_{B,i} = (G_{O,i}, e_b)$ 
18:           $G_{O,i} = G_{B,i}$ 
19:          add  $e_b$  to the tail of  $tabulist$ 
20:          break
21:        else
22:          remove  $b$  from  $list$ 
23:          continue
24:        end if
25:      else
26:         $G_{B,i} = (G_{O,i}, e_b)$ 
27:         $G_{O,i} = G_{B,i}$ 
28:        add  $e_b$  to the tail of  $tabulist$ 
29:        if  $length(tabulist) > TL$  then
30:          remove the first element in the  $tabulist$ 
31:        end if
32:        break
33:      end if
34:    end while
35:  end for
36: end for
Output:  $G_B$ 

```

3.3. Network Parameter Learning

If the variable is discretized, the flooding depth will be discretized into multiple intervals. However, the variation range of the flooding depth caused by each typhoon is different. When the current predicted flooding depth exceeds or is lower than the historical flooding depth interval, the discrete Bayesian network cannot predict. Therefore, the use of continuous data to build Bayesian networks (Gaussian Bayesian networks) facilitates predictive reasoning. The maximum likelihood parameter estimation is used to learn

DGBN parameters to characterize the degree of influence between nodes. Taking a single node $v_{i,t}$ as an example, the expression for maximum likelihood parameter estimation is

$$L(\theta|D, G_D) = P(D, G_D|\theta) = \prod_{j=1}^n f(v_{j,i,t}|\theta) = \prod_{j=1}^n N(\gamma_{0,i,t} + \sum_{j'=1}^{k_i} \gamma_{j',i,t} p a_{j'}(v_{i,t}); \delta_{i,t}^2) \quad (6)$$

$$l(\theta|D, G_D) = \ln L(\theta|D, G_D) = \sum_{j=1}^n \ln N(\gamma_{0,i,t} + \sum_{j'=1}^{k_i} \gamma_{j',i,t} p a_{j'}(v_{i,t}); \delta_{i,t}^2) \quad (7)$$

$$\hat{\theta} = \arg \max_{\theta} l(\theta|D, G_D) \quad (8)$$

where $\theta = (\gamma_{0,i,t}, \gamma_{j',i,t}, \delta_{i,t}^2), v_{j,i,t}$ represents the j -th sample point of node $v_{i,t}$, n is the sample size, D is the dataset, G_D is DGBN network structure, $\gamma_{j',i,t}$ represents the coefficient of the j' -th parent node of the i -th node in the Bayesian network of the t -th time slice, $\gamma_{0,i,t}$ represents the intercept of the i -th node in the Bayesian network of the t -th time slice, $\delta_{i,t}^2$ represents the variance of the i -th node in the Bayesian network of the t -th time slice, $\hat{\theta}$ represents the maximum likelihood estimator.

3.4. Network Reasoning

In this paper, the approximate reasoning method (likelihood weighting algorithm) is used to perform predictive reasoning. The likelihood weighted algorithm has a better predictive reasoning ability than the logical sampling algorithm. Only the sampling process of the likelihood weighting algorithm is used here, as shown in Algorithm 3. According to the sampled data, the mean value of the sample is calculated to approximate the value of the query node conditioned on the evidence node.

Algorithm 3 Likelihood Weighting Algorithm

Input: DGBN G_D , evidence node set $E = (v_{1,0} = ev_1, \dots, v_{i,t} = ev_i), i \leq k - 1, 0 \leq t \leq T$, query node Q

Sampling:

1. Use the value of the evidence node to instantiate the corresponding node in the Bayesian network, and sample all non-evidence nodes according to the topology and network parameters of the G_D network to obtain N' samples.
2. The sample output of node Q will be queried to obtain a sample set $(q_1, \dots, q_{N'})$.

$$\bar{Q} = \frac{\sum_{i=1}^{N'} q_i}{N'}$$

Output: \bar{Q}

This section briefly introduces network construction, network parameter learning and network inference for DGBNs. In order to understand the DGBN construction and reasoning process, the overall flow chart is given, as shown in Figure 2. The left side of Figure 2 is the construction process of the initial network and the right side is the construction process of the transfer network. The top of Figure 2 is the data preprocessing part and the tail is the process of DGBN parameter learning and predictive inference. Data adjustment represents lagging the data and dividing the lagged data equally according to the number of time slices. Next, we analyze the phenomenon of surface confluence caused by topography.

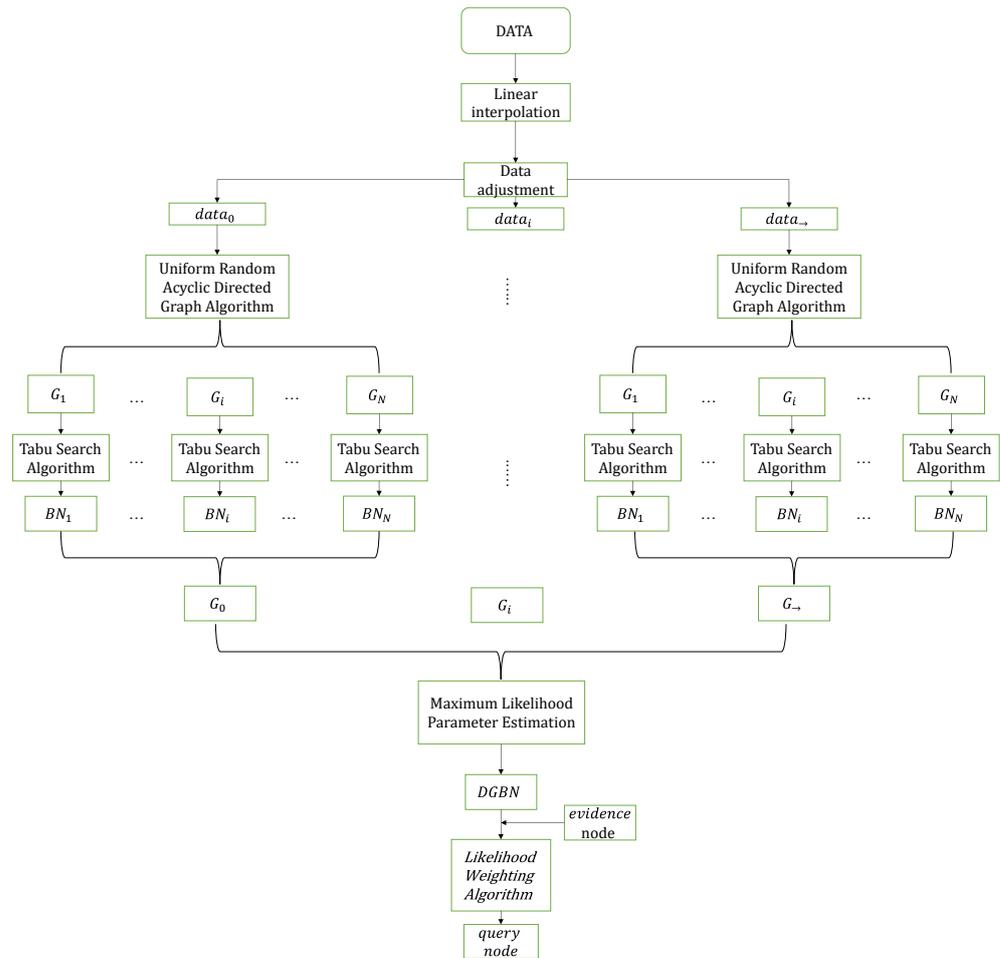


Figure 2. Construction and inference of DGBN. $data_0$ represents the data required to establish the initial network G_0 , $data_{\rightarrow}$ represents the data required to establish the transfer network G_{\rightarrow} , $data_i$ represents the data required to establish the i -th time slice network G_i , BN_i represents Bayesian network structure learned from tabu search algorithm, evidence node represents the value of the characteristic variable (excluding flooding depth) of the flooding to be predicted.

4. Surface Confluence Analysis

The stagnant water in the high-terrain areas will flow to the low-terrain areas in the event of flooding disasters, resulting in a further increase in flooding depth in the low-terrain areas. What we predict is the depth of flooding in a specific area of Macau, and the topographic factor has no time attribute in the flooding in Macau. Thus, the topographic factor is not added when constructing the DGBN. In the past, when the typhoon passed or landed in Macau, flooding would occur at the Inner Harbor Station. The flooding data recorded at the Inner Harbor Station are relatively complete and valuable for research. According to the map provided by the Macau Meteorological Bureau, a number of water level monitoring points in different directions and near the Inner Harbor Station are selected. The Manning formula is used to analyze the surface confluence phenomenon at the Inner Harbor Station. The Manning formula is often used to calculate the open channel flow, which is more reasonable to calculate from high terrain to low terrain. Taking the i -th high terrain as an example, the expression of the Manning formula is

$$W_i = \frac{AR^{2/3}S_i^{1/2}}{\varphi} \tag{9}$$

where W_i (m^3/s) is the flow rate from the i -th high terrain to Inner Harbor Station. A is the conversion coefficient, the international standard is 1 m, R (m) is the hydraulic radius

and S_i is the slope. φ is the Manning coefficient, which is the ground roughness. And, the flooding depth caused by multiple high terrains to low terrains is approximately calculated in the future period, which is defined as $H_G = (H_{G,1}, \dots, H_{G,n'})$, where n' is the number of water level monitoring points in different directions on the high terrain, and

$$H_{G,i} = 60 \frac{W_i}{w} m \tag{10}$$

where $H_{G,i}$ is the flooding depth caused by the i -th high terrain to Inner Harbor Station, w (m^2) is the coverage area of Inner Harbor Station and m (min) is the length of the future period. Because the data provided by the Macau Meteorological Bureau are collected every minute, the above formula needs to be multiplied by a constant 60 when converting minutes into seconds.

5. Prediction Model Construction

In the previous two sections, the DGBN and the surface confluence model are constructed. Combining the DGBN with the surface confluence model yields the prediction model, referred to as DGBN-m.

According to the flooding data provided by the Macau Meteorological Bureau, we select F flooding events at the Inner Harbor Station with relatively complete data, and define $TY = (ty_1, \dots, ty_F)$ as the set of selected flood events in the Inner Harbor Station. According to the data contained in TY , F DGBNs are built, which are defined as $model = \{DGBN_1, \dots, DGBN_F\}$. After a similar analysis is performed between the scene data of the flooding event ty_{pre} of the Inner Harbor Station to be predicted and the scene data of the F flooding events, the $DGBN_i$ model with the most similar characteristics to ty_{pre} is selected from $model$. The scene data for similar analysis are collected before flooding occurs. The Euclidean distance is used for similarity analysis between flooding events, and the expression is

$$dist(B_{ty_{pre}}, C_{ty_i}) = \frac{\sum_{j=1}^{n''} \sqrt{\sum_{i=1}^{k-1} (b_{i,j} - c_{i,j})^2}}{n''} \tag{11}$$

$$DGBN_i \propto \min\{dist(B_{ty_{pre}}, C_{TY})\} \tag{12}$$

where $B_{ty_{pre}} = (b_{1,1}, \dots, b_{k-1,n''})$ and $C_{ty_i} = (c_{1,1}, \dots, c_{k-1,n''})$ are the flooding factor sample (excluding flooding depth) of the current flooding to be predicted and the i -th flooding, respectively, $C_{TY} = (C_{ty_1}, \dots, C_{ty_F})$ is the set of flooding factor samples of F flooding events, $b_{i,j}$ and $c_{i,j}$ are the j -th samples of the i -th flooding factor and n'' is the number of data of the flooding factor before flooding occurred. We define the flooding depth predicted by the $DGBN_i$ as H_B .

Generally speaking, when the ponding in the high terrain exceeds a certain depth, the water flow to the low terrain will be generated. Therefore, a threshold τ is set to decide whether to generate surface confluence. Define the set of flooding depth as $H_H = (H_{h,1}, \dots, H_{h,n'})$ at different azimuth high topographies, where $H_{h,i}$ is the flooding depth of the i -th high topography. Based on the above analysis, the following prediction expression is obtained:

$$H_{NG} = \begin{cases} H_B, & H_H \leq \tau \\ H_B + \sum_{i=1}^{n'_\tau} H_{G,i}, & \exists H_{h,i} \geq \tau, 1 \leq n'_\tau \leq n' \end{cases} \tag{13}$$

where H_{NG} is the predicted flooding depth of the Inner Harbor Station and n'_τ is the number of water level monitoring points whose stagnant water exceeds the threshold τ . The final prediction model DGBN-m of this paper is shown in Figure 3. The prediction model in Figure 3 consists of three modules: the DGBN module, similarity module and surface confluence module. Based on the above analysis and the construction of the DGBN-m,

the performance of the proposed scheme will be tested against the real data provided by the Macau Meteorological Bureau in the next section.

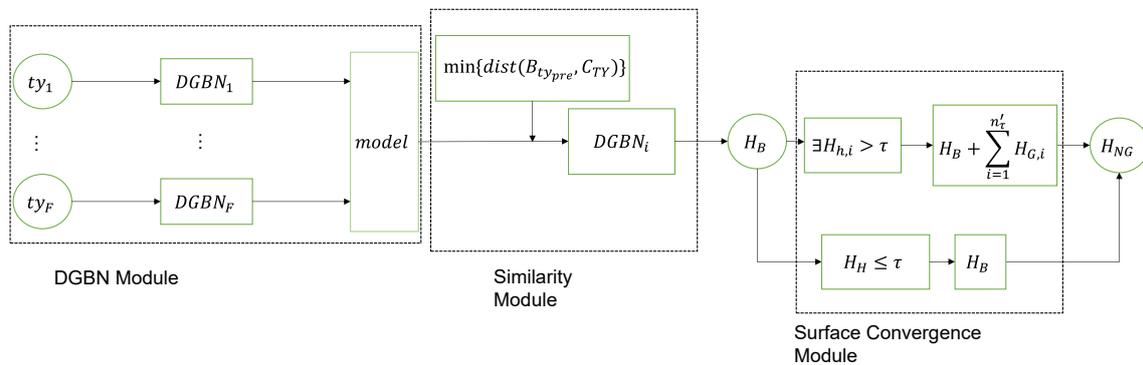


Figure 3. Prediction model DGBN-m. DGBN Module represents the set of DGBNs built from historical flooding disasters, Similarity Module represents the process of selecting the appropriate DGBN from the *model* and performing the initial prediction and Surface Confluence Module represents the process of combining Manning formula with DGBN to predict the flooding depth.

6. Experimental Analysis

6.1. Dataset Selection

According to the data provided by the Macau Meteorological Bureau, three flood events with complete data are selected to establish three DGBNs. The three typhoons associated with the three flooding events are Mangkhut, Nida and Bebinca. These three typhoons have relatively complete flooding data, and the amount of data is relatively large. And, typhoon Hato is selected as the predicted flooding (test set). The flood disaster caused by typhoon Hato to Macau is the most serious in historical records, causing a flooding depth that exceeds 3 m (higher than that of other flooding events). The Macau Meteorological Bureau only recorded data on the rise in flooding depth caused by typhoon Hato but not on the fall in flooding depth because the water level detector was damaged when the flooding depth reached the maximum. Based on the above analysis, the data of typhoon Hato are not suitable for building the DGBN but they are suitable for use as the test set.

6.2. Flooding Factor Selection

The flooding caused by typhoons in Macau is a complicated process. There are many factors that cause flooding in Macau (such as heavy rainfall, storm surge, etc.). Therefore, many factors need to be considered when selecting the flooding factor. The test cases are the flooding events at the Inner Harbor Station in Macau. Based on the above analysis, we select eight factors: the location of the typhoon center (longitude and latitude), the wind speed of the typhoon center, city wind speed (values measured by the Tai Tam Shan Meteorological Observatory in Macau), the rainfall of Macau, the tide (Macau tide and Jiuzhou Port tide) and the flooding depth. These eight factors are used to build the DGBN, referred to as DGBN8. The prediction model is referred to as DGBN8-m. The specific description of the eight factors is shown in Table 1.

Table 1. Abbreviations and units of 8 flooding factors.

Record Data for 8 Factors Every Minute								
Name of Factor	Latitude	Longitude	Typhoon Wind Speed	City Wind Speed	Rainfall in Macau	Macau Tide	Jiuzhou Port Tide	Flooding Depth
Signs and units of factors	La (°N)	Lo (°E)	WD (m/s)	CWD (m/s)	Rain (mm/m)	MC (m)	JZ (m)	D (m)

According to the map of water level monitoring points provided by the Macau Meteorological Bureau, three water level monitoring points are selected in different directions of the Inner Harbor Station, namely Inner Harbor North Station, Kang Kung Temple Station and Xiahuan Street Station.

6.3. Parameter Settings

Set the number of nodes to $k = 8$ in the Bayesian network per time slice, the connection strength of the arc to $\alpha = 0.85$, the direction strength of the arc to $\beta = 0.5$, hydraulic radius to $R = 4/17$ m, Inner Harbor Station coverage area to $w = 4000$ m², Manning coefficient to $\varphi = 0.014$, the number of DGBNs to $F = 3$, the number of high-terrain water level monitoring sites to $n' = 3$ and threshold to $\tau = 0.3$.

6.4. Performance Analysis

Firstly, similarity analysis is made between the data of the flooding factor of typhoon Hato and the data of the flooding factor of Mangkhut, Nida and Bebinca, as shown in Figure 4. It can be seen from Figure 4 that the attributes of typhoon Hato and typhoon Mangkhut are closest to each other and that the difference is about 5. However, the attributes of typhoon Hato and the other two typhoons are far apart, and the difference is about 25 and 27, respectively. Therefore, the DGBN established by typhoon Mangkhut is selected as the network model of typhoon Hato.

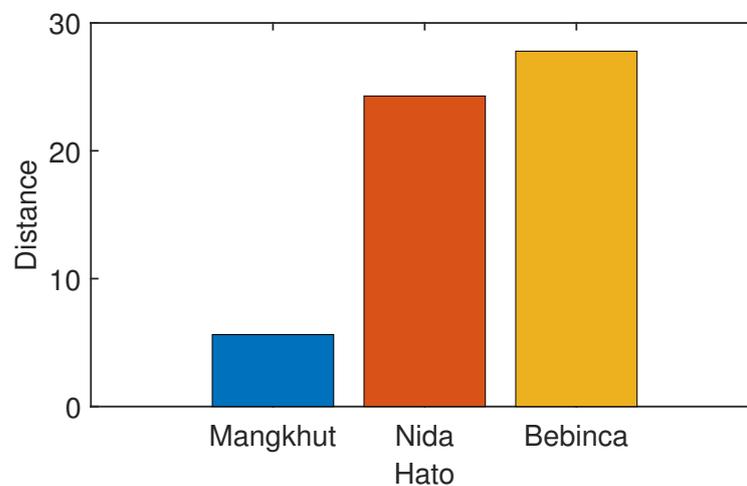


Figure 4. Similarity analysis of Typhoon Hato and Typhoon Mangkhut, Nida and Bebinca.

The flooding depth prediction after 15 min is conducted for the flooding of Macau Inner Harbour Station caused by typhoon Hato, as shown in Figure 5. The abscissa represents the time axis, and the ordinate represents the flooding depth in Figure 5. DGBN8 with two time slices is used for predictive analysis, and the interval between time slices is 15 min. Figure 5 shows the curves of the actual flooding depth, the flooding depth predicted by the DGBN8 model and the flooding depth predicted by the DGBN8-m model. It can be seen from the picture that the predicted curves of DGBN8 and DGBN8-m are coincident before 130 min, which means that the surface confluence has not yet occurred. After 130 min, the predicted curves of DGBN8 and DGBN8-m begin to separate, where the predicted curve of DGBN8-m gradually approaches the real flooding depth curve and the predicted curve of DGBN8 gradually deviates from the real curve. Overall, the prediction curves in Figure 5 can reflect the effectiveness of DGBN8-m proposed in the paper.

Relative error (RE), mean square error (MSE), root mean square error (RMSE) and mean absolute error (MAE) are used to analyze the error in predicting the flooding depth after 15 min for the DGBN8 and DGBN8-m models, as shown in Table 2. It can be seen from Table 2 that the error values of DGBN8-m calculated by various error algorithms are all smaller than the error value of DGBN8. In general, RE is more indicative of the reliability of the predictive model. Therefore, the RE is used to calculate the prediction accuracy of the prediction model. The prediction accuracy of DGBN8-m is 84%, and the prediction accuracy of DGBN8 is 77%, which fully demonstrates the prediction ability of the DGBN8-m model.

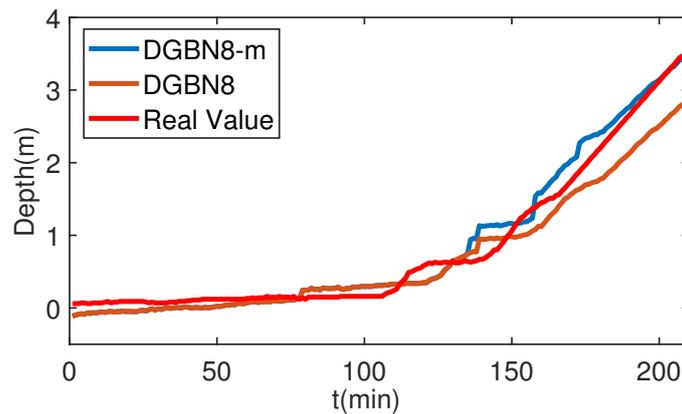


Figure 5. Prediction of flooding depth after 15 min. The abscissa represents the time and the vertical axis represents the flooding depth.

Table 2. Prediction error analysis of DGBN8 and DGBN8-m.

Algorithm Model	RE	MSE	RMSE	MAE
DGBN8-m	0.16	0.026	0.16	0.13
DGBN8	0.23	0.066	0.25	0.19

However, it can be seen from Figure 5 that the predicted values of DGBN8 and DGBN8-m are less than 0 (negative value) before about 50 min, which is contrary to the real phenomenon. To solve this problem, the network structure and network parameters of DGBN8 are carefully studied, and we find that the flooding factor *CWD* can cause the network structure to show unreasonable connections (for example, urban wind speed affects typhoon wind speed). Therefore, we choose to discard the flooding factor *CWD* and establish the seven-variable DGBN to form the new prediction model, referred to as DGBN7-m.

DGBN7-m and DGBN7 models are used to predict the flooding depth after 15 min at Macau Inner Harbor Station under the typhoon Hato scenario, as shown in Figure 6. Similar to the prediction model in Figure 5, a DGBN with two time slices is used for prediction analysis, and the interval between time slices is 15 min. The change trends of the curves of DGBN7-m and DGBN7 models in Figure 6 are basically the same as those in Figure 5. The biggest difference between Figures 5 and 6 is that the predicted values of DGBN7-m and DGBN7 are both positive and close to the real values in the early prediction, which can better reflect the changing trend of flooding depth.

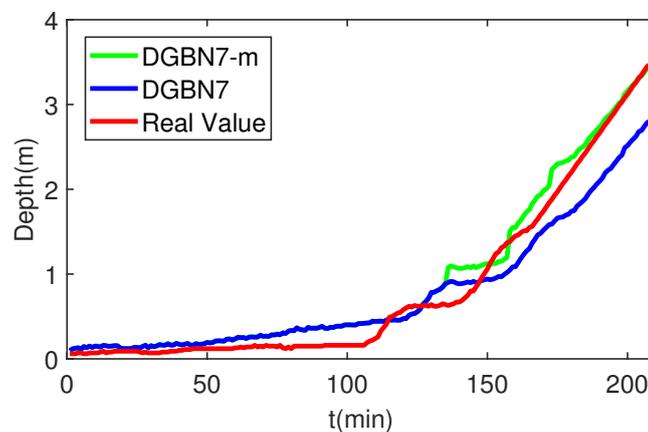


Figure 6. Prediction of flooding depth after 15 min. The abscissa represents the time and the vertical axis represents the flooding depth.

In order to more accurately reflect the prediction performance of the new prediction model DGBN7-m, RE, MSE, RMSE and MAE are also used to analyze the error of DGBN7 and DGBN7-m models in predicting the flooding depth after 15 min, as shown in Table 3. It can be seen that the error values of DGBN7-m calculated by various error algorithms are the same as those of DGBN8-m. The RE is used to calculate the prediction accuracy of the model, and the prediction accuracy of DGBN7-m and DGBN7 is 84% and 75%, respectively. The prediction accuracy of DGBN7 is not much different from that of DGBN8. Further, the number of parameters of DGBN7 and DGBN8 is analyzed, where the number of parameters of the DGBN7 network is 92 and the number of parameters of the DGBN8 network is 106, which shows that the DGBN7-m model runs faster and requires fewer computational resources than the DGBN8-m model under the same prediction accuracy. Therefore, the DGBN7-m model will be used for subsequent experimental analyses.

Table 3. Predicting error analysis of DGBN7 and DGBN7-m.

Model \ Algorithm	RE	MSE	RMSE	MAE
DGBN7-m	0.16	0.026	0.16	0.13
DGBN7	0.25	0.076	0.27	0.21

To further demonstrate the performance of the prediction model DGBN7-m, the flooding depth after 30 min is predicted for the flooding at Macau Inner Harbor Station caused by Typhoon Hato. The DGBN7 of two time slices (the interval between time slices is 30 min), the three-time-slice one-order DGBN7 (the interval between time slices is 15 min) and three-time-slice two-order DGBN7 (the interval between time slices is 15 min) are established, respectively, to predict the flooding depth after 30 min, as shown in Figure 7. The above three prediction models are abbreviated as 2TDGBN7-m, 3T1ODGBN7-m and 3T2ODGBN7-m, respectively. From Figure 7, we can see that all three types of DGBN7-m can approximately predict the changing trend of the flooding depth. Figure 7 shows that 3T2ODGBN7-m has the best prediction performance before about 130 min; 3T1ODGBN7-m has the best prediction performance after about 130 min; 2TDGBN7-m only exhibits a better prediction performance in the later period and its prediction performance is not optimal in the whole prediction period. Although the prediction performance of 3T2ODGBN7-m is slightly better than that of 3T1ODGBN7-m in the early stage, the prediction curves of the two models are close. In the later prediction, the prediction performance of 3T2ODGBN7-m is much worse than that of 3T1ODGBN7-m, and the prediction trend of 3T2ODGBN7-m has deviated from the change trend of the real value.

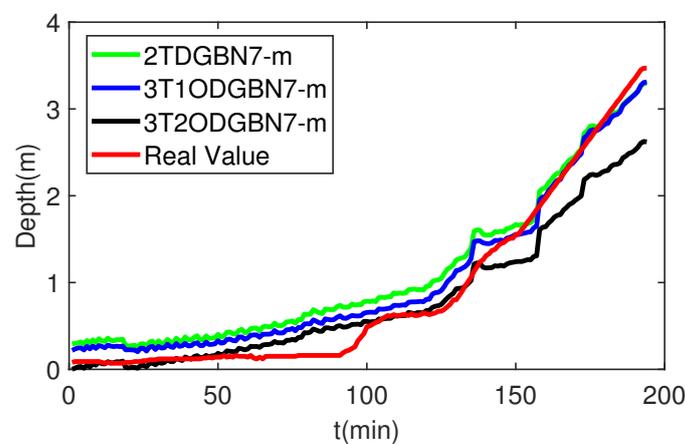


Figure 7. Prediction of flooding depth after 30 min. The abscissa represents the time and the vertical axis represents the flooding depth.

Similarly, RE, MSER, RMSE and MAE are used to analyze the prediction errors of 2TDGBN7-m, 3T1ODGBN7-m and 3T2ODGBN7-m, as shown in Table 4. It can be seen from Table 4 that the prediction errors of the 3T1ODGBN7-m model calculated by various error algorithms are smaller than those of the 2TDGBN7-m model and 3T1ODGBN7-m model. Through RE analysis, we can further obtain that the prediction accuracy of the 3T1ODGBN7-m model is 80%, the prediction accuracy of the 2TDGBN7-m model is 71% and the prediction accuracy of the 3T2ODGBN7-m model is 78%. Compared with the prediction accuracy of the prediction model that predicts the flooding depth after 15 min, the prediction accuracy of the 3T1ODGBN7-m model for predicting the flooding depth after 30 min decreased by only 4%.

Table 4. Prediction error analysis for 2TDGBN7-m, 3T1ODGBN7-m and 3T2ODGBN7-m.

Algorithm Model	RE	MSE	RMSE	MAE
2TDGBN7-m	0.29	0.082	0.28	0.25
3T1ODGBN7-m	0.20	0.042	0.20	0.17
3T2ODGBN7-m	0.22	0.088	0.29	0.20

6.4.1. Robustness Analysis

When the flood disaster occurs, various detectors often fail. Once a device such as the detector fails, data loss will occur. When faced with a large number of missing data and some characteristic variables being completely missing, the corresponding complementary algorithm is useless. But, we still hope that the predictive model can have a certain predictive ability in this case. In this section, the robustness of the DGBN7-m will be tested in the case of a complete loss of data for some flood factors, as shown in Figure 8. Figure 8 represents the flooding depth after 30 min predicted by 3T1ODGBN7-m in the case where each flood factor is lost once (the data of the corresponding flood factor are completely lost). The prediction curves in Figure 8 (except the true value curve) represent the results predicted by the 3T1ODGBN7-m model after the corresponding variable is lost. It can be seen from the picture that the change trend of all predicted curves (except the prediction curve corresponding to variable *D*) is basically similar to the change trend of the real curve, in which the prediction curves corresponding to variables *JZ*, *Lo* and *Rain* are close to the real value about 120 min ago. After 120 min, the predicted curve of the *JZ* variable deviates to a greater extent than the predicted curve of the *Lo* and *Rain* variables deviates from the true curve. This phenomenon shows that 3T1ODGBN7-m still has a well-predictive ability after the data loss of the *Lo* or *Rain* variable, and that 3T1ODGBN7-m can withstand the data loss of the *JZ* variable in the early period. Furthermore, it can be seen that the prediction curves corresponding to *La* and *WD* variables are close to the real curves after 150 min, which indicates that 3T1ODGBN7-m can withstand the data loss of these two variables in the later period. The picture shows that only the prediction curves corresponding to *MC* and *D* variables deviate from the true curve to a large extent during the whole prediction period. The above analysis shows that 3T1ODGBN7-m has strong robustness.

To better describe the predictive ability of the 3T1ODGBN7-m model after the flood factor is missing, the RE analysis graph is given, as shown in Figure 9. It can be seen that when the prediction error of 3T1ODGBN7-m is lower than 0.5, the corresponding variables are *La*, *Lo*, *Rain* and *JZ*; when the prediction error is higher than 0.5, the corresponding variables are *WD*, *MC* and *D*, which is consistent with the prediction result in Figure 8.

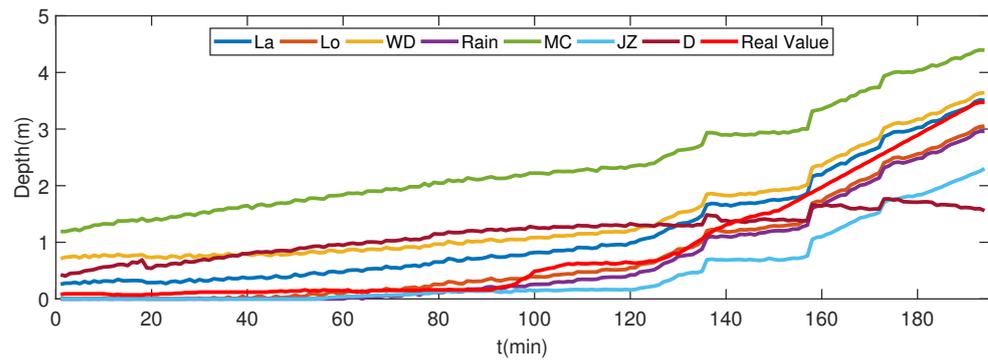


Figure 8. Prediction of flooding depth for 3T1ODGBN7-m with missing flood factor. The abscissa represents the time and the vertical axis represents the flooding depth.

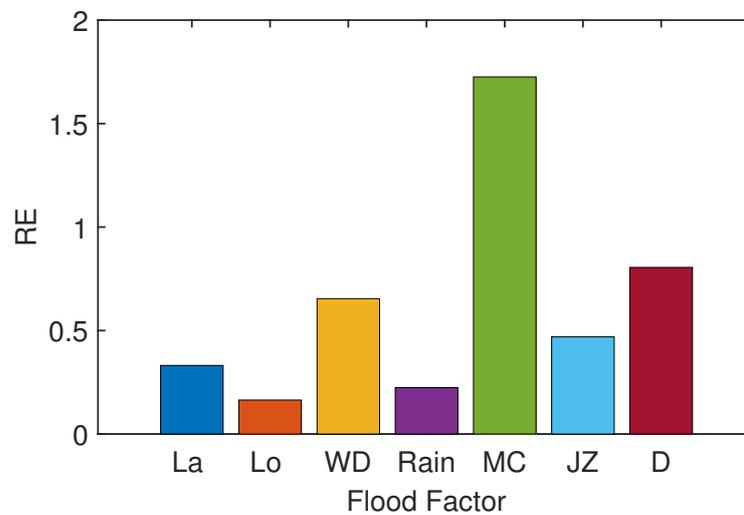


Figure 9. Relative error analysis. The abscissa represents that the corresponding variable is lost and the vertical axis represents the error.

6.4.2. Algorithm Comparison

Due to the rapid development of deep learning, neural networks have achieved better results in prediction. Therefore, the prediction performance of the back propagation (BP) neural network and linear regression is compared with that of the DGBN7-m model. We build the three-layer BP neural network, input layer, hidden layer and output layer. We set the input layer to have 6 nodes (the number of flood factor), the output layer to have 1 node (flooding depth) and the hidden layer to have 13 neurons (according to Kolmogorov’s theorem). The corresponding parameters of the BP neural network are set with a learning rate of 0.01, loss threshold of 0.03 and 10,000 training times. The data of the three typhoons of Mangkhut, Nida and Bebinca are used as the training set of the BP neural network and linear regression, and the data of Typhoon Hato are used as the test set of the BP neural network and linear regression. As shown in Figure 10, the BP neural network and linear regression predict the flooding depth after 15 min at Macau Inner Harbor Station. As can be seen from the picture, the neural network and linear regression cannot accurately predict the changing trend of flooding depth. Through the analysis of RE, the prediction accuracy of the neural network is only 39%, and the prediction accuracy of linear regression is only 22%. The low prediction performance of the BP neural network may be caused by the small amount of data and the difference in flooding scenes, which makes it difficult for the BP neural network to accurately learn the relationship between the flooding factor and the flooding depth. This further reflects the advantages of the prediction model DGBN7-m in scenarios where the amount of data is small and the flooding scenarios are different.

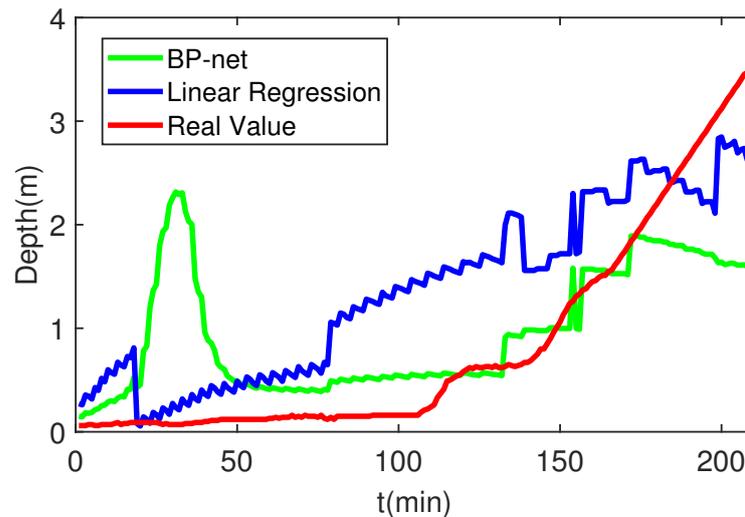


Figure 10. BP neural network and linear regression predict the flooding depth after 15 min. The abscissa represents the time and the vertical axis represents the flooding depth.

6.4.3. Generalization Analysis

In order to verify the validity and reliability of the DGBN7-m model, this section tests the generalization of the DGBN7-m model. Another typhoon (Dianmu) that caused the flooding at the Macau Inner Harbor Station is selected as the test set. The data of the flooding factor of typhoon Dianmu and the data of the flooding factor of typhoon Mangkhut, Nida and Bebinca are analyzed for similarity, as shown in Figure 11. We can see from the picture that typhoon Dianmu is most similar to typhoon Bebinca. However, the gap between typhoon Dianmu and typhoon Bebinca is about 15, which is larger than the gap between typhoon Hato and typhoon Mangkhut. In other words, the DGBN7 established by typhoon Bebinca has a low degree of matching with the characteristics of flooding caused by typhoon Dianmu. Given the limited number of typhoons currently recorded by the Macau Meteorological Bureau, the DGBN7 established by typhoon Bebinca is still selected as the prediction network for typhoon Dianmu.

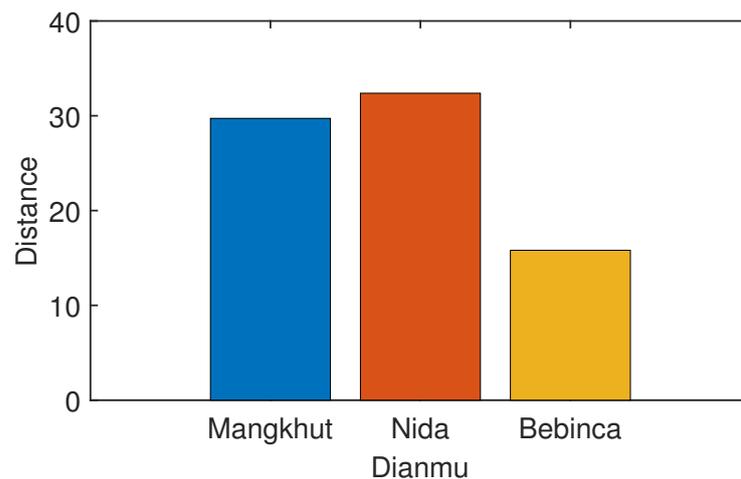


Figure 11. Similarity analysis of typhoon Dianmu and typhoon Mangkhut, Nida and Bebinca.

The flooding depth prediction after 15 min is shown in Figure 12. The change trend of the predicted curve of DGBN7-m in the early and late stages is similar to that of the real flooding depth curve. The change trend of the predicted curve of DGBN7-m in the mid-term has a large deviation from the change trend of the real flooding depth curve. According to the error analysis in Table 5, the difference between the flooding depth predicted by

DGBN7-m and the real flooding depth is small. The prediction accuracy of DGBN7-m calculated by the RE can reach 72%. Therefore, the above analysis shows that the DGBN7-m model still has a good prediction performance in the case of large differences, which further confirms that the DGBN7-m model has good generalization.

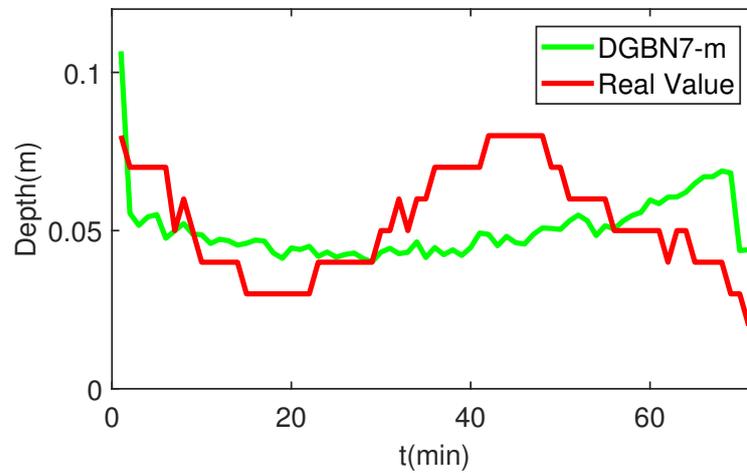


Figure 12. Flooding depth prediction after 15 min. The abscissa represents the time and the vertical axis represents the flooding depth.

Table 5. Error analysis table of DGBN-m.

Algorithm Model	RE	MSE	RMSE	MAE
DGBN7-m	0.28	0.0003	0.018	0.015

7. Conclusions

This paper designs a prediction model for flooding in Macau. A more accurate prediction model, DGBN7-m, is obtained by combining the DGBN with the surface confluence, and achieves better prediction results proved by simulation. The work carried out in this paper can be summarized as follows:

- i In the case of a small amount of data, some expert experience is useful for guiding the learning direction of the algorithm to establish a more reasonable DGBN.
- ii In the case of different flooding scenarios, multiple historical flooding scenarios are used to establish multiple DGBNs, and then a similar analysis is made between the scenarios of flooding to be predicted and the scenarios of multiple historical flooding; finally, the DGBN established by the historical flooding that is most similar to the flooding to be predicted is chosen as the prediction network.
- iii According to the topography of Macau, the influence of surface confluence is studied, and the Manning formula analysis method is put forward. And, the DGBN and Manning formula are combined to obtain the final prediction model, DGBN-m, which has a high prediction ability.

The prediction model that we designed has growth potential: with the accumulation of more scene data of flooding, scene similarity analysis will have more advantages, and the accuracy and effectiveness of DGBN-m model prediction can be further improved. For example, the current flooding can be constructed as the DGBN, which is added to the set of DGBNs to enhance the ability of the prediction model to predict the flood depth of future flooding. Due to the lack of relevant data, this paper does not analyze the effects of infiltration and drainage systems on the flooding areas. In the case of severe typhoons and little changes in the urban drainage system, the impact of infiltration and the drainage system is small in different scenarios. But, the analysis of infiltration and the drainage system can further improve the effectiveness of the prediction model, which can achieve a

more accurate and timely prevention of flooding disasters. Therefore, the more difficult infiltration phenomena and the effects of drainage systems will be worked on in the future. In future research, we will study flooding in different terrains and incorporate scenarios of the confluence and diversion of ponding into the prediction model to achieve a more comprehensive prediction of flooding. In addition, we will explore the characteristics of flooding in different coastal cities to find out the same and different factors that allow the predictive model to be built adaptively for different environments.

Author Contributions: Conceptualization, S.Z. and C.C.; Methodology, S.Z.; Software, S.Z. and C.C.; Validation, S.Z.; Formal analysis, S.Z.; Investigation, S.Z. and W.D. (Weiping Ding); Resources, W.D. (Weijun Dai); Data curation, S.Z. and W.D. (Weijun Dai); writing—original draft, S.Z.; writing—review and editing, C.C., N.S., J.R. and W.D. (Weiping Ding); Supervision, C.C., N.S., J.R. and W.D. (Weiping Ding); Project administration, C.C.; Funding acquisition, C.C. and W.D. (Weijun Dai). All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by: National Natural Science Foundation of China and Macau Science and Technology Development Joint Project (0066/2019/AFJ): Research on Knowledge-oriented Probabilistic Graphical Model Theory based on Multi-source Data. MOST-FDCT Joint Projects (0058/2019/AMJ, 2019YFE0110300): Research and Application of Cooperative Multi-Agent Platform for Zhuhai-Macau Manufacturing Service. Natural Science Characteristic Innovation Project of Guangdong General Universities (2022KTSCX323). Heyuan Social Development Science and Technology Project (2021/62/138).

Data Availability Statement: Macau Meteorological Bureau data are available at <https://www.smg.gov.mo/zh/subpage/355/report/typhoon-yearly-report> (accessed on 1 May 2023).

Conflicts of Interest: We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work and that there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, “Predicting Typhoon Flood in Macau by Dynamic Gaussian Bayesian Network and Surface Confluence Analysis”.

References

- Kellens, W.; Terpstra, T.; De Maeyer, P. Perception and communication of flood risks: A systematic review of empirical research. *Risk Anal. Int. J.* **2013**, *33*, 24–49. [[CrossRef](#)] [[PubMed](#)]
- Mosavi, A.; Ozturk, P.; Chau, K.W. Flood prediction using machine learning models: Literature review. *Water* **2018**, *10*, 1536. [[CrossRef](#)]
- Tehrany, M.S.; Jones, S.; Shabani, F. Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *Catena* **2019**, *175*, 174–192. [[CrossRef](#)]
- Paradilaga, S.N.; Sulistyoningih, M.; Lestari, R.K.; Laksitaningtyas, A.P. Flood Prediction Using Inverse Distance Weighted Interpolation of K-Nearest Neighbor Points. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 4616–4619.
- Khan, W.; Hussain, A.J.; Alaskar, H.; Baker, T.; Ghali, F.; Dhiya, A.; Al-Shamma’a, A. Prediction of Flood Severity Level via Processing IoT Sensor Data Using a Data Science Approach. *IEEE Internet Things Mag.* **2020**, *3*, 10–15. [[CrossRef](#)]
- Du, J.; Kimball, J.S.; Sheffield, J.; Pan, M.; Fisher, C.K.; Beck, H.E.; Wood, E.F. Satellite flood inundation assessment and forecast using SMAP and landsat. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6707–6715. [[CrossRef](#)] [[PubMed](#)]
- Adnan, R.; Ruslan, F.A.; Zain, Z.M. Flood water level modelling and prediction using artificial neural network: Case study of Sungai Batu Pahat in Johor. In Proceedings of the 2012 IEEE Control and System Graduate Research Colloquium, Shah Alam, Malaysia, 16–17 July 2012; pp. 22–25.
- Sinha, S.; Mandal, N. Design and analysis of an intelligent flow transmitter using artificial neural network. *IEEE Sens. Lett.* **2017**, *1*, 1–4. [[CrossRef](#)]
- Imrie, C.E.; Durucan, S.; Korre, A. River flow prediction using artificial neural networks: Generalisation beyond the calibration range. *J. Hydrol.* **2000**, *233*, 138–153. [[CrossRef](#)]
- Kim, S.; Matsumi, Y.; Pan, S.; Mase, H. A real-time forecast model using artificial neural network for after-runner storm surges on the Tottori coast, Japan. *Ocean Eng.* **2016**, *122*, 44–53. [[CrossRef](#)]
- Gude, V.; Corns, S.; Long, S. Flood prediction and uncertainty estimation using deep learning. *Water* **2020**, *12*, 884. [[CrossRef](#)]
- Kourgialas, N.N.; Dokou, Z.; Karatzas, G.P. Statistical analysis and ANN modeling for predicting hydrological extremes under climate change scenarios: The example of a small Mediterranean agro-watershed. *J. Environ. Manag.* **2015**, *154*, 86–101. [[CrossRef](#)]

13. Dai, W.; Cai, Z. Predicting coastal urban floods using artificial neural network: The case study of Macau, China. *Appl. Water Sci.* **2021**, *11*, 161. [[CrossRef](#)]
14. Dai, W.; Tang, Y.; Zhang, Z.; Cai, Z. Ensemble Learning Technology for Coastal Flood Forecasting in Internet-of-Things-Enabled Smart City. *Int. J. Comput. Intell. Syst.* **2021**, *14*, 166. [[CrossRef](#)]
15. Lu, Q.; Zhong, P.A.; Xu, B.; Zhu, F.; Ma, Y.; Wang, H.; Xu, S. Risk analysis for reservoir flood control operation considering two-dimensional uncertainties based on Bayesian network. *J. Hydrol.* **2020**, *589*, 125353. [[CrossRef](#)]
16. Sebastian, A.; Dupuits, E.J.C.; Morales-Nápoles, O. Applying a Bayesian network based on Gaussian copulas to model the hydraulic boundary conditions for hurricane flood risk analysis in a coastal watershed. *Coast. Eng.* **2017**, *125*, 42–50. [[CrossRef](#)]
17. Sen, M.K.; Dutta, S.; Kabir, G. Modelling and quantification of time-varying flood resilience for housing infrastructure using dynamic Bayesian Network. *J. Clean. Prod.* **2022**, *361*, 132266. [[CrossRef](#)]
18. Chen, J.; Zhong, P.A.; An R.; Zhu, F.; Xu, B. Risk analysis for real-time flood control operation of a multi-reservoir system using a dynamic Bayesian network. *Environ. Model. Softw.* **2019**, *111*, 409–420. [[CrossRef](#)]
19. Quesada, D.; Valverde, G.; Larrañaga, P.; Bielza, C. Long-term forecasting of multivariate time series in industrial furnaces with dynamic Gaussian Bayesian networks. *Eng. Appl. Artif. Intell.* **2021**, *103*, 104301. [[CrossRef](#)]
20. Fahiman, F.; Disano, S.; Erfani, S.M.; Mancarella, P.; Leckie, C. Data-driven dynamic probabilistic reserve sizing based on dynamic Bayesian belief networks. *IEEE Trans. Power Syst.* **2018**, *34*, 2281–2291. [[CrossRef](#)]
21. Dong, G.; Han, W.; Wang, Y. Dynamic Bayesian Network-Based Lithium-Ion Battery Health Prognosis for Electric Vehicles. *IEEE Trans. Ind. Electron.* **2020**, *68*, 10949–10958. [[CrossRef](#)]
22. Zhang, Q.; Zhou, C.; Tian, Y.C.; Xiong, N.; Qin, Y.; Hu, B. A fuzzy probability Bayesian network approach for dynamic cybersecurity risk assessment in industrial control systems. *IEEE Trans. Ind. Inform.* **2017**, *14*, 2497–2506. [[CrossRef](#)]
23. Ma, Y.; Qi, S.; Fan, L.; Lu, W.; Chan, C.Y.; Zhang, Y. Dynamic Bayesian network approach to evaluate vehicle driving risk based on on-road experiment driving data. *IEEE Access* **2019**, *7*, 135050–135062. [[CrossRef](#)]
24. Keprate, A.; Ratnayake, R.C. Assessment of Reliability and Remaining Fatigue Life of Topside Piping Using Dynamic Bayesian Network. In Proceedings of the 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Macao, China, 15–18 December 2019; pp. 1114–1118.
25. Wang, J.; Han, M.; Wei, S. Discrete Dynamic Bayesian Network Threat Assessment Method Based on Cloud Parameter Learning. In Proceedings of the 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019; pp. 1–6.
26. Fan, L.-M.; Jia, L.-L.; Ren, Y.; Wang, K.-S.; Yang, D.-Z. Risk Analysis of Discrete Dynamic Event Tree Based on Dynamic Bayesian Network. In Proceedings of the 2019 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE), Zhangjiajie, China, 6–9 August 2019; pp. 141–147.
27. Xu, Y.; Wang, J.; Yu, Y. Alarm event prediction from historical alarm flood sequences based on Bayesian estimators. *Proc. IEEE Trans. Autom. Sci. Eng.* **2019**, *17*, 1070–1075. [[CrossRef](#)]
28. Melançon, G.; Dutour, I.; Bousquet-Mélou, M. Random generation of directed acyclic graphs. *Electron. Notes Discret. Math.* **2001**, *10*, 202–207. [[CrossRef](#)]
29. Scutari, M.; Denis, J.-B. *Bayesian Networks: Examples R*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2021.

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.