Predicting coastal urban floods using artificial neural network: The case study of Macau, China

Weijun Dai¹ · Zhiming Cai²

Received: 25 January 2021 / Accepted: 9 June 2021 / Published online: 18 September 2021
© The Author(s) 2021

Abstract

Using data-driven models to predict floods in advance is one of the current effective methods and hot researches to reduce urban flood disasters. In order to improve the prediction accuracy, it is necessary to select the appropriate flood hazard factors and the number of training samples to construct the prediction model. In our current research, an artificial neural network (i.e., the back-propagation neural network, BPNN) model was developed to predict the flood depth in the next hour. A case study of the urban flood during six typhoons in Macau of China was conducted to prove the performance of the proposed model. The flood depth was collected as output; after analyzing their correlation to the flood typhoon optimum track, urban weather, tides, geographic height and water depth increment of the submerged area were used as input. As a result, four models trained with different sample numbers were developed for training and testing. The model performances were examined using average absolute error, root mean square error and the coefficient of determination. The results show that in this case study, the 30-min scale model provides reliable predictions and can provide useful decision support for the prevention and mitigation of flood disasters in coastal urban.

Keywords Floods prediction · Correlation analysis · BPNN · Time scale · Typhoon event

Introduction

The coastal areas have fertile land, abundant food and convenient access to the sea, and the population density is much higher than the global average (Dhiman et al. 2019). As of 1990, over 1.2 billion people are living within 100 km of the coast in areas where the land height is less than 100 m above sea level (Christopher and Robert 2003). A large part of these areas has a high urbanization process and a developed economy, which has great value in real estate and infrastructure. With that said, the downside is that they are more susceptible to natural disasters, such as coastal floods, typhoons, waves, tsunamis, and so forth. It can cause serious

damage to the urban, take the lives and homes of residents, damage infrastructure, and paralyze operation. Among these disasters, flood disaster is the most destructive and fairly frequent (Easterling et al. 2000; Ostad-Ali-Askari 2018a, b; Javadinejad 2021). Research shows that the 20 cities with the highest risk of flood disasters in the world (such as Guangzhou, Mumbai, New York, New Orleans, etc.) are mainly distributed in the coastal areas of Asia and North America, most of which are in the estuarine deltas and are located in developing countries. Moreover, the losses caused by floods are mainly concentrated in cities with high income but low level of protection (Hallegatte et al. 2013; Ostad-Ali-Askari et al. 2020; Golian et al. 2020).

The land height of coastal urban is many meters lower than sea level; as a consequence, the high-level water (caused by tides and storm surges) and the waves (caused by typhoons) may easily cross coastal protection dams and inundate low-lying areas. Some cities are located in deltas or river mouths, the heavy rainfall brought by the storm caused river runoff to increase, as a result, the river floods and invades the urban. The water level of the river is higher than the water level of drainage pipe, and there is no drainage channel for rainwater on the road, the rainwater gathers on

✉ Zhiming Cai
zhiming_cai@sina.cn

Weijun Dai
weijun_dai@sina.cn

¹ Heyuan Polytechnic College, 1 Donghuan Road, Heyuan 517000, Guangdong, China

² Institute of Data Science, City University of Macau, Avenida Padre Tomás Pereira Taipa, Macau, Macau 999078, China

the lower road to form waterlogging. The flood is also heavily linked to road runoff and drainage infrastructure as they can increase the likelihood of natural disasters in coastal areas. Hence, the flooding mechanism of coastal urban is very complicated, and the study of these processes is of great significance for coastal flood risk assessment (Agnieszka et al. 2017; Ostad-Ali-Askari et al. 2018a, b; Derakhshannia et al. 2020).

The dynamics of coastal flood caused by a specific combination of multiple process driving factors are very complex. Storm surges and waves are stochastic in nature generated by storm events during typhoons. The storm surge and ocean waves caused by typhoons are random in nature. Heavy rainfall and urban drainage facilities also have great uncertainties; accurate prediction of floods is still a challenge and a formidable threat to mankind (Bates et al. 2005). In recent years, data-driven models have been widely developed. This type of model uses a generalized relationship between input and output datasets and uses limited assumptions to define the system (Zhang et al. 2019). The main advantage is that it can be used from real-time, high-frequency data collected by the flood depth gauge to calibrate the model. This is achieved without the need to conduct specific site surveys based on physical methods. In addition, the latest data are continuously collected, and the data implied any changes in the flooded area or the system, thus capturing the dynamic characteristics of the system (Solomantine and Ostfeld 2008). Due to the effectiveness and efficiency of data-driven technologies, methods including autoregressive (AR) models, artificial neural networks (ANN), genetic algorithms, and multivariate analysis have been widely explored and used (Usman et al. 2017; Tiyasha et al. 2020; Ostad-Ali-Askari 2016). For example, Bruen and Yang (2006) combined a physical model and an ANN model to perform flood forecasting of the urban drainage system and compared model results with observation results. Chang et al. (2014) proposed a recurrent neural network to forecast real-time multi-step-ahead water level for urban flood control. Furthermore, Yves et al. (2017) developed a NARX neural network model to predict the water depth of the catchment area in urban drainage systems, and it performed well in both minor and severe storm events. Berkhahn et al. (2019) proposed an artificial neural network-based model for the maximum water levels forecasting during a flash flood event. Simon et al. (2019) proposed an integrated neural network model to predict in real time the maximum water level during a flash flood in urban regions; the accuracy of the results is rather convincing. The flood forecasting model based on artificial neural network is one of the data-driven models, and the input factors and the number of training samples have a great influence on the forecasting effect of the model (Hongtao 2020). In the above-mentioned research, no detailed discussion on this aspect has been carried out.

In this study, a BPNN model will be developed to predict the depth of urban flooding by one hour, and appropriate input factors and number of training samples are discussed and selected to improve model prediction accuracy. The proposed model will be applied to Macau, China. Macau is a coastal city in the South China Sea. Due to active tropical cyclones in the Western Pacific, when typhoons are fast approaching, coastal flooding will be inundated. This study aims to examine the following main objectives: (1) analyze the distribution characteristics of floods in Macau; (2) establish BPNN model to predict the flood depth in the next one hour; analyze the correlation between flood and flood hazard factors, and select appropriate input variables for BPNN model. (3) Based on the varying time scales, four training sets with different sample numbers are established, four prediction models are trained and tested, these models are compared with each other, and the best model is selected.

Methodology

Modeling development

BPNN

ANN is an application based on the principle of the biological neural network. It has characteristics of self-applicability, self-organization and self-learning. When the neural network is in operation, it is based on the forward and reverse movement of the signal flow, and so, there are two basic types and implementations: feedforward and feedback (Sovzil et al. 1997).

BPNN is a common feedforward network. It is a multi-layer network based on the error back-propagation algorithm. It includes an input layer, an output layer, and one or more hidden layers. Neurons in each layer are fully connected, and there is no connection between neurons in the same layer. The operating process of BPNN includes forward transmission of signals and reverse transmission of errors. During forward transmission, the input information is processed by the input layer, the hidden layer, and is then passed to the output layer to output the predicted value. If the predicted value does not match the observed value, then the error reverse transfer phase begins. In reverse transfer, output error is passed back to the hidden layer and then to the input layer in some form of coupling. After that, the error is allocated to all neurons in each layer. In the next time forward transfer, the weight of each layer of neurons is corrected according to this error value. The forward and reverse transmission cycles continue, correcting the weight of each neuron layer and changing the strength of the relationship between connected neurons. The process will stop until the output error of the network is less than a certain value or

reaches a preset number of cycles. This process is known as the learning and training process of BPNN (Wang et al. 2020; Ostad-Ali-Askari and Shayannejad 2021).

The BPNN input vector is denoted by $x = [x_1 \ x_2 \ x_3 \ \dots]^T$, the hidden layer is expressed as $h = [h_1 \ h_2 \ h_3 \ h_4 \ \dots]^T$, and the output layer is represented as $y = [y_1 \ y_2 \ y_3 \ \dots]^T$. Collectively, the signal forward equation can be expressed:

$$\begin{aligned} h &= f(W_1^T x + b_1) \\ y &= f(W_2^T h + b_2) \end{aligned} \quad (1)$$

In the formula, W_1 represents the connection weight of the input layer to the hidden layer, b_1 signifies the threshold value of the hidden layer, W_2 denotes the connection weight of the hidden layer to the output layer, b_2 indicates the threshold value of the output layer, and f symbolizes the excitation function, select function type (such as sigmoid) according to your needs and requirements. A set of input vectors and the corresponding predicted output form a training sample. The number and quality of training samples and the network parameters of Formula 1 directly affect the benefits and drawbacks of network performance (Mondal et al. 2012; Al-Abadi 2016).

In this research, a BPNN model for floods depth prediction in coastal urban was developed. The input data include typhoon optimum track data, urban weather data, urban weather data, tide data, and geographic height data of the submerged area. At the same time, in order to characterize urban surface runoff and sewer drainage performance, flood depth data of the submerged area are also used as input. Some of these inputs are characteristics of the current time, some of which are characteristics of one or two hours in the past, which, in turn, are used to describe the time-series correlation between variables. Consequently, 31 input variables are created. The output is the flood depth of the submerged area in the next hour. Table 1 summarizes the input and output of the model, and the symbols used in this study.

Based on a series of exploratory experiments, a three-layer BPNN was selected as the best and most ideal model structure, consisting of an input layer (with 31 input variables), an output layer (with 1 output variable) and a hidden layer (with 63 neurons). Repeated training and testing are carried out through different parameter settings, and the learning rate is determined to be 0.01; the momentum factor is 0.5; the number of training times is 1000, and the training error is 0.001. These are the best parameters of a BPNN model for training, based on a series of exploratory experiments. Figure 1 details a schematic diagram of the architecture of the developed BPNN flood prediction model.

Correlation analysis

Correlation analysis refers to the analysis of two or more correlated variable elements, so as to measure the closeness of the two variable factors. Taking variable X and variable Y as an example, the linear correlation between them is calculated as follows:

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X]\text{Var}[Y]}} \quad (2)$$

where $\text{Cov}(X, Y)$ is the covariance of X and Y , $\text{Var}[X]$ is the variance of X , and $\text{Var}[Y]$ is the variance of Y . The linear correlation coefficient $r(X, Y)$ is a statistic used to measure the strength of the linear relationship between X and Y . That is, the larger the $|r(X, Y)|$, the greater the degree of correlation; $|r(X, Y)| = 0$ corresponds to the lowest degree of correlation.

In our research, correlation analysis is used to measure the closeness of the correlation between flood depth and flood hazard factors. The purpose of the analysis is to select flood hazard factors with large linear correlation coefficients as the input of the model.

Table 1 A summary of the input and output variables, symbols and definitions used for this research

Symbol	Definition	Use
$D(t+1)$	Flood depth at time $t+1$	Output
$T_{yI}; T_{yLa}; T_{yLo}$	Typhoon intensity, motion longitude and latitude at time t	Input
$\Delta T_{yLa}; \Delta T_{yLo}; \Delta T_y$	Longitude distance, latitude distance, straight line distance between typhoon and city at time t	Input
$T_{yP}; T_{yW}; T_{yP}(t-1); T_{yW}(t-1); T_{yP}(t-2); T_{yW}(t-2)$	Typhoon center pressure and wind speed at time $t, t-1, t-2$	Input
$R; W_s; W_d; R(t-1); W_s(t-1); W_d(t-1); R(t-2); W_s(t-2); W_d(t-2)$	Urban rainfall, wind speed, wind direction at time $t, t-1, t-2$	Input
$T_{iM}; T_{iZ}; T_{iM}(t-1); T_{iZ}(t-1); T_{iM}(t-2); T_{iZ}(t-2)$	Urban tide at time $t, t-1, t-2$	Input
H	Geographic elevation of submerged areas	Input
$D(t); D(t-1); D(t-2)$	Flood depth at time $t, t-1, t-2$	Input

Fig. 1 Architecture of the developed BPNN model

Evaluation of model performance

In order to verify the accuracy of the developed model, mean absolute error (MAE), root of mean square error (RMSE) and coefficient of determination (R^2) are measures used to evaluate the network model according to flood characteristics in coastal urban areas. MAE and RMSE are two statistical indices that are used for measuring the flood depth errors between predictions and observations. As for the model, there exists a correlation regarding the error between the predicted value and observed value, the smaller the error, the more accurate the prediction result. The R^2 value is [0 1], which indicates the closeness of predicted value and observed value. The closer R^2 is to 1, the more it represents the simulation of the developed model and the higher the degree of integration (Ahmed et al. 2002). The calculation formula can be expressed as:

$$MAE(X, h) = \frac{1}{m} \sum_{i=1}^m |h(x_i) - y_i| \quad (3)$$

$$RMSE(X, h) = \sqrt{\frac{1}{m} \sum_{i=1}^m (h(x_i) - y_i)^2} \quad (4)$$

$$R^2(X, h) = 1 - \frac{\sum_{i=1}^m (h(x_i) - y_i)^2}{\sum_{i=1}^m (h(x_i) - \bar{y})^2} \quad (5)$$

where m is the samples number, $h(x_i)$ is the predicted value of the i -th sample, y_i is the observed value of the i -th sample, and $\bar{y}$ is the average value of the observation sample.

Study area

Macau

Macau is a coastal city in the South China Sea; it is located west to the estuary of the Pearl River (the second largest river in China.). It is separated from Hong Kong and Guangzhou on the outer edge of the Pearl River Delta. It consists of two long and narrow islands with a total area of 292 square kilometers and a population 680,000. Macau has a tropical monsoon climate, with an annual rainfall of more than 1,700 mm. Every year, the months between May and September are the active season of tropical cyclones in the Pacific Northwest, and it is also the season with the most rain in Macau. When the distance between typhoon track center and Macau is less than 800 km, the Macau Meteorological Bureau will issue different levels of typhoon warnings (from levels 1–10) and storm surge warnings (from levels 1–5) according to changes in the distance and the typhoon grade. When the meteorological observatory measures rainfall exceeding 50 mm within two hours, and when the rainfall exceeds 100 mm within three hours, a red rainstorm (level 4) and a black rainstorm signal (level 5) will be issued. The higher the value of these warning levels, the more severe

the natural disasters caused by rainfall or typhoon in Macau, such as severe flood disaster (SMG 2020).

Macau has a developed economy, densely distributed buildings and streets, and a high road surface hardening rate. The existing drainage infrastructure is complete, with a large number of sewers and drainage pumping stations; 13 depth meters have been set up in areas that are frequently flooded (please refer to Fig. 2 for more details). The flood depth value of this area is collected at the frequency of one data every minute.

Data collection

The urban flood inundation report issued by the Macau Meteorological Bureau is consistent with the analysis of inundation data recorded by monitoring points. Flood disasters occurred when the typhoon was close to Macau. During the six typhoon periods (Table 2), one piece of flood depth data (measured in minutes) was selected at

Table 2 Typhoon case of model training and testing

Time	Typhoon	Maximum flood depth (m)	Use
201,608	1604 NIDA	0.25	Train
201,608	1608 DIANMU	0.08	Train
201,710	1720 KUANUM	0.05	Train
201,708	1713 HATO	3.47	Train
201,808	1816 BEBINCA	0.24	Train
201,809	1822 MANGKHUT	2.09	Test

eight monitoring points. The remaining five monitoring points have long-term data gaps and data loss due to power failure or damage to the monitoring equipment. As a consequence, they are abandoned because they are useless for training data-driven models. During each typhoon, storm surge water usually crossed coastal dams and inundated streets, but the increase or decrease values of storm surge water were not recorded. So, the project study uses the values in the change of flood depth at the hourly interval at monitoring points, which reflects the inundation situation caused by storm surge water and variations of regional runoff.

Weather data in the same time period include one record per hour of total rainfall, average wind speed and main wind direction. According to multiple automatic weather observation stations, the weather data of TG Station (North Latitude 22°09'36", East Longitude 113.6°33'54") were selected because the record is complete and comprehensive. Considering the small area of Macau, the data from this weather station are sufficient to describe basic meteorological characteristics of this urban area. The tidal height data are one record per hour; they are obtained from two tide monitoring stations located to the west of A-Ma Temple and at Jiuzhou Port, which is located north of Macau.

The typhoon optimum track data come from Tropical Cyclone Data Center of China Meteorological Administration. The data are recorded in three-hour or six-hour intervals, including the typhoon intensity mark, the trajectory latitude and longitude, the lowest central pressure, and the average central maximum wind speed (Ying et al. 2014; CMA 2020). The study selected the optimum track data of six typhoons in 2016, 2017, and 2018. The reason is that during the occurrence of these six typhoons, the data records of flood depth and weather in Macau are accurate, complete and reliable. Around the same time, typhoons, such as "1713 HATO" and "1822 MANGKHUT", caused huge floods in Macau. They were both within-study scope; the implicit relationship between the data is representative.

Fig. 2 Location of depth meter and meteorological station in Macau

Data pre-processing

In order to improve data quality in this study, it is necessary to preprocess the collected data. First, preprocess the flood depth dataset as they tend to have missing data and outliers. By selecting flood depth value in the adjacent time period, linear interpolation is used to fill and replace, so as to comprise of a complete flood depth time series (Yu and Xu 2014). The next step is to change the tide height data (one-hour interval) and typhoon optimum track data (three-hour or six-hour interval) to one-minute interval data. Among them, tidal height data are filled by Cubic Spline Function interpolation, and the typhoon optimum track data are filled by linear interpolation. In the research, urban weather data including rainfall, wind speed and wind direction remained unchanged for one hour when used as model input. In the end, a total of 100,440 sample pairs are formed to train and educate the model.

In order to eliminate the influence of different model input units and different magnitudes, the min–max normalization method is used to map input feature data to a range of $[-1 \ 1]$ to facilitate subsequent data processing (Niu 2020).

Results and discussion

Characteristics of the flood

These six typhoons have appeared in the surrounding areas of Macau; the Meteorological Bureau has issued a typhoon warning of level three or higher, or a storm surge warning of level two or higher. The floods depth data were then captured for the time period until the warning was canceled. There have been flooding events of varying degrees in multiple areas in the urban, and these flood depths in eight areas have been reliably recorded as seen in Fig. 2. As shown in Fig. 3, the distribution frequency of research samples is very uneven, 88,204 records where flood depth is 0, accounting for 97.89% of the sample. There are few records with a flood depth of more than 0.1 m and less than 2 m, indicating that eight areas have different degrees of flooding. There are very few records with a flood depth of more than 2 m; only in a few specific areas, it appeared in a relatively short time frame. The highest record was 3.47 m, which occurred on Neigang Street, a street that is close to the river and the west side of Macau. The distribution of flood depth also shows that the flooding in Macau is very fast and the time frame is very short. This is a typical feature of coastal flood caused by a storm surge or short-term heavy rain during a typhoon.

Fig. 3 Frequency distribution of flood depth

Relationship between flood and impact factors

There are many factors that cause floods in coastal urban areas, such as typhoon, storm surges, local weather, tides, coastal topography, rivers and waves, urban buildings and the condition of roads as well as the underlying surface. There are many uncertain and nonlinear relations between these factors and floods. Limited by technical means and conditions, it is difficult to comprehensively collect data on all influencing factors. Using the linear correlation method, based on the flood data during the six typhoons, we analyzed the relationship between flood depth and related hazards. Flood hazard factors that have no correlation or too low correlation with flood depth are eliminated, and finally, 8 types of factors in Table 1 (a summary of the input and output variables, symbols and definitions used for this research) are retained. As depicted in Fig. 4, the flood depth and the straight-line distance between typhoon and Macau, the flood depth and the maximum wind speed at typhoon center, the flood depth and the rainfall, the flood depth and the tide in Macau are generally positively correlated with a weak relationship. The flood depth and the minimum air pressure at typhoon center, the flood depth and the altitude of the submerged area in Macau are generally negative correlation, and that is also a somewhat weak connection between the two. The flood depth and its increase or decrease in the previous hour, the flood depth and wind speed in Macau have a strong positive correlation.

Results of the BPNN model

The BPNN model was trained and validated using flooding events during the five typhoons, including four typhoons "1604 NIDA", "1608 DIANMU", "1720 KUANUM", and

Fig. 4 Correlation of flood depth and impact factors. **a** Flood depth and straight-line distance between typhoon and Macau; **b** flood depth and minimum air pressure at typhoon center; **c** flood depth and maxi-

mum wind speed at typhoon center; **d** flood depth and rainfall; **e** flood depth and wind speed; **f** flood depth and altitude; **g** flood depth and its increase or decrease

"1816 BEBINCA", which caused a lower flood depth, and one typhoon "1713 HATO", with the largest flood depth in Macau. The flooding event during typhoon "1822 MANGKHUT" was used for further testing of the BPNN model; this specific sample was not used in the training phase to test model promotion ability. On September 15 of 2018, Typhoon MANGKHUT entered 800 km warning range of Macau, and it approached closer to Macau by 60 km on the 16th midday and made landfall near Taishan in Guangdong Province. The Macau Meteorological Bureau issued a typhoon warning (the highest level No. 10) for nine hours. A black storm surge warning was issued. During this period, the highest rainfall in Macau was 148 mm. There was severe flooding in many areas, among which the highest submergence depth of Neigang Street was 2.09 m. Low-lying areas were submerged for a long period of time and the flooding continued for nearly 13 h. The flooding event during

MANGKHUT typhoon has a higher flood depth, which can better test the model's ability to predict changes in floods. It also has a long-term lower flooding depth, which is necessary to test the model's robustness.

Based on the time scale, the four models were trained with different sample numbers. First, the flood inundation depth is one record per minute, and the processed flood hazard factor is also one record per minute. As a result, 89,100 sample pairs with one-minute intervals were constructed for model training. After that, according to 15-min intervals, 30-min intervals and 60-min intervals, 5,940, 2,970 and 1,485 sample pairs were constructed for model training. Finally, the four models are tested with one-minute interval test data, with a total of 10,080 sample pairs. The output of each model is the average of the results per 100 runs, which is used to reduce prediction error caused by ANN disadvantages. These disadvantages include easy to fall into

local optimality, difficult to converge, over-reliance on initial values, and so forth. (Yousef et al. 2019).

In Table 3, as the number of training samples decreased, the MAE value of the training set also decreased. The number of data in the testing set remains unchanged, the MAE value of the 1-min model and the 15-min model has only a small and insignificant change, the MAE value of the 60-min model, on the other hand, has increased significantly. The minimum MAE value of 30-min mode is only 1487.22, which is better than the other three models. Among these four models, the RMSE values of the training set and testing set are relatively close, and the minimum RMSE value of the 30-min model test set is 0.19795, which is better than the other three models. As for the performance of the 30-min model, its testing set RMSE value is 1.72 times the training set RMSE value, which shows that the model has no overfitting problems. In the four models, the R^2 value of the training set is all above 0.91, the R^2 value of the testing set is above 0.67, and the regression results of the four models

are above average. For the 30-min model, the R^2 value of the training set is 0.93357, and the R^2 value of the testing set is 0.83484.

In Fig. 5, the scatter distribution also supports the results, and the distribution trend of model samples is proximately consistent with the diagonal. The BPNN model cannot capture some of the behavior of the system due to the dispersion around the low values of observed depth. This is expected given the imbalanced domain shown in Fig. 2. Among them, the 30-min model sample data have a smaller deviation from diagonal, indicating that the regression model is the most effective and exhibits relatively strong extrapolation ability. The evaluation results indicate that the 30-min model is the optimal model, and the evaluation values affirm the ability of BPNN to express a change of flood depth during a typhoon.

The observation and prediction results obtained through the BPNN model are plotted in Fig. 5. It clearly shows that BPNN has good generalization ability and can better capture the peak of the flood depth. In the six submerged

Table 3 Error statistics of BPNN for training and testing datasets

Time scale	Sample size		MAE		RMSE		R^2	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
1-min	89,100	10,080	369.29	1749.45	0.029555	0.28511	0.91898	0.67159
15-min	5940	10,080	43.82	1772.4	0.033276	0.24615	0.92036	0.74485
30-min	2970	10,080	17.31	1487.22	0.033979	0.19795	0.93357	0.83484
60-min	1485	10,080	13.07	2061.52	0.035082	0.2465	0.95346	0.73957

Fig. 5 The result of multivariate regression analysis of the models. **a** 1-min model; **b** 15-min model; **c** 30-min model; **d** 60-min model

areas, the prediction results of the maximum flood depth tend to be too high. The prediction results of the remaining two regions (region one and region eight) tend to be too low. The reason may be that the drainage pipes in these regions are blocked or the equipment is damaged, which reduces flood drainage capacity. Table 4 compares the maximum flood depth difference of 30-min mode. Among them, the predicted maximum flood depth in area five (Neigang Street) is 0.34 m higher than the observed value. This place is the most severely flooded area in Macau. With respect to the maximum flood depth prediction in other areas, the difference is less than this number, and the minimum result of area seven is 0.1 m,

Table 4 Maximum flood depth difference of the 30-min model

Flooded area	Maximum flood depth (m)		Difference (m)
	Observation	Prediction	
1	0.95	0.64	−0.31
2	1.61	1.90	0.29
3	0.29	0.55	0.26
4	0.22	0.53	0.31
5	2.09	2.43	0.34
6	1.96	2.31	0.32
7	0.86	0.96	0.10
8	1.13	0.97	−0.16

which is substantially lower than that of area five. If the predicted maximum flood depth is too high, it may cause excessive disaster prevention measures and public panic. However, the predicted maximum flood depth is too low, which may cause the disaster prevention departments to relax vigilance or be ignored. In the flooding event during the MANGKHUT typhoon, Macau Meteorological Bureau issued a red storm surge warning, and the maximum depth of flood forecast is 1.5–2.5 m (Area five), the predicted result of the 30-min model is within this range, which can be used to guide the prevention of flood disasters.

Comparing the time series of observation results and prediction results, the changes in predicted values of the three models in Fig. 6a, b, and d are all earlier than the observed values, especially when flood depth decreases. Among them, in the context of the 60-min model, when flood depth increases or decreases, the predicted value is changed one hour earlier than that of the observed value. Also, as seen in Fig. 6c, the 30-min model shows better synchronization performance; more accurate prediction when it comes to the time of flooding, and the observed time of flooding was also highly consistent. With that said, there are multiple sharp spikes and jumps in the time series, and the model's predictive ability needs to be further improved.

Fig. 6 Comparison of predicted and observed flood depth value. **a** 1-min model; **b** 15-min model; **c** 30-min model; **d** 60-min model

Conclusions

The causes of floods in coastal urban are complex, it is usually the result of typhoons, heavy rain, tides and rivers entering the sea at the same time, and the data relationship between flood and flood hazard factors is complex. The performance of data-driven models based on artificial neural networks is closely related to data quality, which requires proper selection of model input factors and training data volume. In the coastal city of Macau, the BPNN-based prediction method of flood depth in the next hour was implemented and tested. We use linear correlation analysis to select closely related factors with flood depth from the numerous flood hazard factors, take typhoon optimum track, urban weather, tides, geographic height and water depth increase of submerged areas as input, and take flood depth as output. Four models trained using different sample numbers were constructed. The division of the number of samples is based on the time scales that are 1-min, 15-min, 30-min, and 60-min. Finally, MAE, RMSE, and R^2 are used to evaluate the performance of the model. As a result, the flood event data of six typhoons are collected. The model was trained using five typhoon flood events, and tested using one flood event during the MANGKHUT typhoon. This was also tested and compared against the recorded data. The test showed that the 30-min model has achieved an advance prediction accuracy that can be considered as flood depth. We should further study the insufficient input of the model as the structure is too complex, the forecast value is falsely high, and the forecast time series is not smooth to improve the prediction quality. Four models are trained with different sample numbers.

Acknowledgements We thank City University of Macau and the Meteorological Bureau of Macau for supporting this study to collect necessary data easily and help the authors to collect the necessary data without payment, Weijun Dai and Zhiming Cai for their helpful contributions to collect the data and all other sources of funding for the research collected from authors.

Author contributions WD and ZC designed this research, wrote this paper, collected the necessary data and did analysis of the data.

Funding This study was supported by National Natural Science Foundation of China and Macau Science and Technology Development Joint Fund (0066/2019/AFJ) and Research and Application of Cooperative Multi-Agent Platform for Zhuhai-Macao Manufacturing Service (MOST-FDCT Projects 0058/2019/AMJ).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agnieszka IO, Joanne C, Stephen N, Michael H (2017) High-resolution multi-scale modelling of coastal flooding due to tides, storm surges and rivers inflows. A cork city example. *Coast Eng* 121:278–296. <https://doi.org/10.1016/j.coastaleng.2016.12.006>
- Al-Abadi AM (2016) Modeling of stage–discharge relationship for Gharraf River, southern Iraq using backpropagation artificial neural networks, M5 decision trees, and Takagi-Sugeno inference system technique: a comparative study. *Appl Water Sci* 6:407–420. <https://doi.org/10.1007/s13201-014-0258-7>
- Bates PD, Dawson RJ, Hall JW, Horritt MS, Nicholls RJ, Wicks J, Hassan MA (2005) Simplified two-dimensional numerical modelling of coastal flooding and example applications. *Coast Eng* 52(9):793–810. <https://doi.org/10.1016/j.coastaleng.2005.06.001>
- Berkhahn S, Fuchs L, Neuweiler I (2019) An ensemble neural network model for real-time prediction of urban floods. *J Hydrol* 575:743–754. <https://doi.org/10.1016/j.jhydrol.2019.05.066>
- Bruen M, Yang J (2006) Combined hydraulic and black-box models for flood forecasting in urban drainage systems. *J Hydrol Eng* 11(6):589–596. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2006\)11:6\(589\)](https://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(589))
- Chang FJ, Chen PA, Lu YR, Huang E, Chang KY (2014) Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. *J Hydrol* 517:836–846. <https://doi.org/10.1016/j.jhydrol.2014.06.013>
- Christopher S, Robert JN (2003) A global analysis of human settlement in coastal zones. *J Coastal Res* 19:584–599
- CMA (2020) CMA best path dataset. <https://tcdata.typhoon.org.cn/index.html> (accessed 22 July 2020)
- Derakhshannia et al (2020) Corrosion and deposition in Karoon river Iran based on hydrometric stations. *Int J Hydrol Sci Technol* 10(4):334–345. <https://doi.org/10.1504/IJHST.2020.108264>
- Dhiman R, VishnuRadhan R, Eldho TI et al (2019) Flood risk and adaptation in Indian coastal cities: recent scenarios. *Appl Water Sci* 9:5. <https://doi.org/10.1007/s13201-018-0881-9>
- Easterling DR, Meehl GA, Parmesan C (2000) Climate Extremes: Observations, Modeling, and Impacts. *Science* 289(5487):2068–2074. <https://doi.org/10.1126/science.289.5487.2068>
- EI-Din AG, Smith DW (2002) A neural network model to predict the wastewater inflow incorporating rainfall events. *Water Res* 36(5):1115–1126. [https://doi.org/10.1016/S0043-1354\(01\)00287-1](https://doi.org/10.1016/S0043-1354(01)00287-1)
- Golian et al (2020) Prediction of tunnelling impact on flow rates of adjacent extraction water wells. *Q J Eng Geol Hydrogeol*. <https://doi.org/10.1144/qjegh2019-055>
- Hallegatte S, Green C, Nicholls RJ, Corfee MJ (2013) Future flood losses in major coastal cities. *Nat Clim Chang* 3:802–806. <https://doi.org/10.1038/nclimate1979>
- Hongtao N (2020) Smart safety early warning model of landslide geological hazard based on BP neural network. *Saf Sci* 123:104572. <https://doi.org/10.1016/j.ssci.2019.104572>

- Javadinejad et al (2021) The analysis of the most important climatic parameters affecting performance of crop variability in a changing climate. *Int J Hydrol Sci Technol (IJHST)* 11(1):1–25. <https://doi.org/10.1504/IJHST.2021.112651>
- Mondal SK, Jana S, Majumder M et al (2012) A comparative study for prediction of direct runoff for a river basin using geomorphological approach and artificial neural networks. *Appl Water Sci* 2:1–13. <https://doi.org/10.1007/s13201-011-0020-3>
- Niu HT (2020) Smart safety early warning model of landslide geological hazard based on BP neural network. *Saf Sci*. <https://doi.org/10.1016/j.ssci.2019.104572>
- Ostad-Ali-Askari et al (2020) Effect of climate change on precipitation patterns in an arid region using gcm models: case study of Isfahan-Borkhar Plain. *Nat Hazards Rev*. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000367](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000367)
- Ostad-Ali-Askari K, Shayannejad M (2021) computation of subsurface drain spacing in the unsteady conditions using artificial neural networks (ANN). *Appl Water Sci* 11:21. <https://doi.org/10.1007/s13201-020-01356-3>
- Ostad-Ali-Askari et al (2016) Artificial neural network for modeling nitrate pollution of groundwater in marginal area of zayandeh-rood river, Isfahan Iran. *KSCE J Civil Eng Korean Soc Civil Eng* 21(1):134–140. <https://doi.org/10.1007/s12205-016-0572-8>
- Ostad-Ali-Askari et al (2018a) Comparison of solutions of Saint-Venant equations by characteristics and finite difference methods for unsteady flow analysis in open channel. *Int J Hydrol Sci Technol* 8(3):229–243. <https://doi.org/10.1504/IJHST.2018.093569>
- Ostad-Ali-Askari K, Su R, Liu L (2018) Water resources and climate change. *J Water Clim Change* 9(2):239. <https://doi.org/10.2166/wcc.2018.999>
- Simon B, Lothar F, Insa N (2019) An ensemble neural network model for real-time prediction of urban floods. *J Hydrol* 575:743–54. <https://doi.org/10.1016/j.jhydrol.2019.05.066>
- SMG (2020), Description of storm surge warning level. <https://www.smg.gov.mo/zh/subpage/28/page/26> (accessed 22 July 2020)
- Solomantine DP, Ostfeld A (2008) Data-driven modelling: some past experiences and new approaches. *J Hydroinform* 10(1):3–22. <https://doi.org/10.2166/hydro.2008.015>
- Sovzil D, Kvasnicka V, Pospichal J (1997) Introduction to multi-layer feed-forward neural networks. *Chemom Intell Lab Syst* 39(1):43–62. [https://doi.org/10.1016/S0169-7439\(97\)00061-0](https://doi.org/10.1016/S0169-7439(97)00061-0)
- Tung TM, Yaseen ZM (2020) A survey on river water quality modeling using artificial intelligence models: 2000–2020. *J Hydrol*. <https://doi.org/10.1016/j.jhydrol.2020.124670>
- Usman TK, He JX, Caterina V (2017) River flood prediction using fuzzy neural networks: an investigation on automated network architecture. *Water Sci Technol* 1:238–247. <https://doi.org/10.2166/wst.2018.107>
- Wang HY, Zhang ZX, Liu LM (2020) Prediction and fitting of weld morphology of Al alloy–CFRP welding–rivet hybrid bonding joint based on GA-BP neural network. *J Manuf Process*. <https://doi.org/10.1016/j.jmapro.2020.04.010>
- Ying M, Zhang W, Yu H, Lu X, Feng J, Fan Y, Zhu Y, Chen D (2014) An overview of the China meteorological administration tropical cyclone database. *J Atmos Oceanic Tech* 31:287–301. <https://doi.org/10.1175/JTECH-D-12-00119.1>
- Yousef A, Seyed HK, Morteza O (2019) Prediction of blast induced ground vibration (BIGV) of quarry mining using hybrid genetic algorithm optimized artificial neural network. *Measurement*. <https://doi.org/10.1016/j.measurement.2019.106874>
- Yu F, Xu XZ (2014) A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. *Appl Energy* 134:102–113. <https://doi.org/10.1016/j.apenergy.2014.07.104>
- Yves AR, Oras A, Marwan S, Fadi HC (2017) Flood forecasting within urban drainage systems using NARX neural network. *Water Sci Technol* 76(9):2401–2412. <https://doi.org/10.2166/wst.2017.409>
- Zhang QQ, Li Z, Spencer SW, Siam A, El-Dakhakhni W (2019) Predictive models for wastewater flow forecasting based on time series analysis and artificial neural network. *Water Sci Technol* 80(2):243–253. <https://doi.org/10.2166/wst.2019.263>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.