Application study of IFAS and LSTM models on runoff simulation and flood prediction in the Tokachi River basin

Yue-Chao Chen\(^a\), Jia-Jia Gao\(^a\), Zhao-Hui Bin\(^a,b,\)*, Jia-Zhong Qian\(^c\), Rui-Liang Pei\(^d\) and Hua Zhu\(^b\)

\(^a\) Graduate School of Engineering, Muroran Institute of Technology, Muroran 0508585, Hokkaido, Japan
\(^b\) Henan Institute of Geological Sciences, Zhengzhou 450001, China
\(^c\) School of Resources and Environmental Engineering, Hefei University of Technology, Hefei 230009, China
\(^d\) Henan Institute of Geological Survey, Zhengzhou 450001, China

*Corresponding author. E-mail: binzhaohui@163.com

ABSTRACT

Floods are often caused by short-term heavy rainfall. An Integrated Flood Analysis System (IFAS) model is good at runoff simulation and a Long Short-Term Memory (LSTM) model is good at learning massive data and realizing rainfall forecast. In this paper, the applicability of the IFAS model to runoff simulation in the Tokachi River basin and the LSTM model to forecast hourly rainfall was studied, and the accuracy of flood prediction was also studied by inputting the optimal rainfall data forecasted by the LSTM model into the IFAS model. The research results show that the IFAS model can accurately simulate the runoff process in the Tokachi River basin. In the calibration period and the verification period, the Nash–Sutcliffe efficiency coefficient (NSE) of all simulation results are above 0.75; the LSTM model can achieve forecast hourly rainfall with high precision, the NSE of best forecast results is 0.86; the IFAS model can achieve flood prediction with high precision by using the optimal rainfall data forecasted by the LSTM model, the NSE of simulation result is 0.81. The above conclusions show that it is of great significance to combine the hourly rainfall forecasted by the LSTM model with the IFAS model for flood prediction.

Key words: flood prediction, IFAS model, LSTM model, rainfall and runoff simulation, Tokachi River basin

HIGHLIGHTS

- The Integrated Flood Analysis System (IFAS) model can accurately simulate runoff in the Tokachi River basin.
- The Long Short-Term Memory (LSTM) model can achieve high-precision hourly rainfall forecast.
- The combination of the optimal hourly rainfall predicted by the LSTM model and the calibrated IFAS model can achieve high-precision runoff simulation, which makes a beneficial exploration for the realization of flood forecasting.

1. INTRODUCTION

The history of human development is a history of fighting against natural disasters, especially floods. Frequent floods around the world have been a serious threat to human security. Although nowadays the flood is difficult to avoid, it can be predicted in advance. The establishment of the hydrological model and simulation study of the hydrological process will help people to better understand the water cycle process, effectively utilize water resources and cope with flood disasters. At present, hydrological model is an important tool to study the evolution process of water resources. In essence, the hydrological model is a directly corresponding logical relationship between the mathematical expression and the hydrological process, that is, a mathematical description of the hydrological process (Todini 2011).

The distributed hydrological model can not only better reflect the water cycle of river basin but also describe the different distribution of rainfall in time and space, and finally get better simulation results (Gao et al. 2017; Gumindoga et al. 2017; Romagnoli et al. 2017; Hao et al. 2018; Jeziorska & Niedzielski 2018; Rujner et al. 2018; Xue et al. 2018). At the end of the last century, with the rapid development of computer technology, Geographic Information System (GIS), Remote Sensing (RS), satellite rainfall measurement technology and other high-tech, the development of the distributed hydrological model had a solid, real-time technical platform, and the Integrated Flood Analysis System (IFAS) model came into being in the period. The IFAS model is a simulation software invented and developed by a Japanese research team. Its core is Public

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).
Works Research Institute-Distributed Hydrological Model (PWRI-DHM). In the model, surface flow, groundwater flow, groundwater seepage, and downstream flow are played back by the form of tank and channel. Japanese scholars and river basin managers applied the IFAS model to various river basins in Japan to study river basin runoff. Nguyen et al. (2020) and Riaz et al. (2017) applied the IFAS model using the GSMaP_NRT satellite rainfall data for flood prediction in the Chenab River, IFAS showed the capability to generate sufficient lead time flood forecast. Kimmanny et al. (2020) applied the IFAS model to simulate the side flow into the NN1 reservoir in the Nam Ngum River basin (NNRB), and the fit is good. Umer et al. (2020) make innovative use of the Asian Precipitation–Highly-Resolved Observational Data Integration Towards Evaluation (APHRODITE) with the IFAS model for flood simulation in the River Jhelum Basin and had good results. Chow et al. (2019) proved that it is possible to estimate the IFAS parameters using regression-based techniques for flood simulation. The accuracy of runoff simulation and flood forecast in the large river basin by combining different types of rainfall data with the IFAS model needs to be further studied. Rainfall forecast is of great significance for agriculture, engineering, drought, and flood control. There are many common forecast methods and models: Markov chain, grey model, machine learning, Back Propagation Neural Network model (BP model), Recurrent Neural Network model (RNN model), etc. In terms of time-series data forecasting, new models such as machine learning and hybrid models have shown greater and greater advantages over traditional forecasting models (Goyal et al. 2014; Kumar et al. 2021). With the development of the neural network model, the Long Short-Term Memory (LSTM) model, with the function of self-learning and self-training, had been applied to many fields, such as the time-space relationship between location and earthquake (Wang et al. 2017), remaining useful life prediction of supercapacitor (Zhou et al. 2019), identifying the bearing degradation states and accurately predict the remaining useful life (Zhang et al. 2019), predicting the nutrient removal efficiency for full-scale sewage treatment plant (Yaqub et al. 2020), and has achieved good results. In hydrology, the LSTM model has been used in streamflow and rainfall forecasting (Ni et al. 2020) and prediction of water breakthrough in subsurface systems (Bai & Tahmasbei 2020) and also has achieved good results.

At present, with the great development of computer technology and the continuous improvement of big data platform, new methods of computer field should also be introduced into the traditional field of hydrology research. For example, the self-learning function of the neural network model can be expanded to rainfall forecast. So far, using the neural network model to forecast rainfall, there are still many shortcomings: the amount of data used for training is usually small, and the data are usually daily rainfall or monthly rainfall, which cannot fully reflect the rainfall process; the database and big rainfall data are not fully used in rainfall forecast, and the utilization rate of existing data is low; floods are often caused by short-term heavy rainfall, usually within a few hours, so hourly rainfall forecast is more useful for runoff simulation and flood prediction. As the improved RNN model, can the LSTM model achieve accurate forecast of hourly rainfall? If so, can the results be used in the IFAS model to improve the accuracy of flood prediction?

Our research is just to study and attempt to answer the above questions. The Tokachi River basin is located in the east of Hokkaido, Japan. The drainage area of the Tokachi River basin is larger, topography is complex and flood disasters occur frequently. There are many hydrological observation stations, which collect a large number of data of rainfall, flow, and water level. The river basin is suitable for the research of rainfall-runoff simulation. In this study, the Tokachi River basin is the experimental platform and our research aims to (1) study the applicability of the IFAS model in runoff simulation of the Tokachi River basin by comparing the runoff simulation results with the measured flow; (2) study whether the self-learning and self-training function of the LSTM model can achieve the high-precision prediction of hourly rainfall forecast based on nearly 1.5 million hourly rainfall data in the Tokachi River basin; (3) input the optimal rainfall forecasted by the LSTM model into the IFAS model to simulate the runoff and predict the flood in the Tokachi River basin, and evaluate the accuracy of simulation and prediction. Figure 1 shows the flowchart of this study.

2. MATERIALS AND METHODOLOGY

2.1. Tokachi River basin

The Tokachi River originates from Tokachi Mountain, 2,077 m above sea level in the middle of Hokkaido, with a total length of 156 km and a drainage area of 9,010 km², between 142°36′38″–144°11′48″E and 42°25′40″–43°38′47″N. It is a first-class River in Japan. The climate of Hokkaido can be divided into the West Pacific coast climate zone, the East Pacific coast climate zone, the Japan coast climate zone, and the Okhotsk coast climate zone. The climate feature is that there is no plume rains period. The rising temperature in spring and the increase of rainfall are easy to cause snow melting and flood. There
is often heavy rain in summer and autumn due to the impact of typhoon, so the maximum precipitation period of Hokkaido is from August to September. The annual average precipitation in Hokkaido is 1,135.6 mm, which is lower than the national average precipitation of 1,607.7 mm in Japan. It belongs to an area with little precipitation. The annual sunshine time in Hokkaido is 1,817 h, lower than the 1,983 h of the national average sunshine time in Japan. The average wind speed in Hokkaido is 3.6 m/s, which is higher than the national average wind speed of 2.8 m/s. Figure 2 shows the map of the Tokachi River basin.

2.2. IFAS model
The main functions of the IFAS model include rainfall-runoff simulation, regional rainfall distribution, and flood prediction. It can use various rainfall data to simulate the runoff process. It has the advantages of simple operation, abundant supporting data, convenient use, and high degree of model automation.

2.2.1. Structure of the IFAS model
The IFAS model consists of four TASKs and one calculation engine module: Overall control (TASK 1), Importing rainfall data (TASK 2), Creating runoff model (TASK 3), Calculation engine module (PWRI-DHM), and Result display (TASK 4).

2.2.2. Public Works Research Institute-Distributed Hydrological Model (PWRI-DHM)
The IFAS model uses PWRI-DHM as the runoff simulation engine with the features listed below: (1) according to the principle of tank mode, the outflow from each cell is calculated by the nonlinear relation based on Manning hyperbola; (2) parameters can roughly be estimated by using grid-based global data sets on topography, soil, geology, land use, etc.; (3) storage function runoff model enhances floods reproducibility by modifying saturation rainfall for each flood event; (4) for numerical calculation, PWRI-DHM uses approximation functions to solve the time integral equation. For this reason, the system can conduct numerical calculations smoothly and to realize real-time operation; (5) to calculate discharge in the river course tank, PWRI-DHM uses the Kinematic wave equation.
In this study, the two-layer tank model of PWRI-DHM was selected, including two tanks in the vertical direction, one was the Surface Tank, the other was the Aquifer Tank, and the right side of the model was the River Channel.

2.2.3. Parameters of PWRI-DHM in the IFAS model

In this study, the default parameters of the IFAS model and experience parameters were combined to set the parameters of PWRI-DHM. The IFAS model itself had the function of generating default parameters according to the soil data, land-use data, and elevation data of the study area. The IFAS model also could be adjusted to the parameters by historical hydrological data. In this study, the calibration period was from 20 August 2011 to 30 September 2011, the period from 1 September 2013 to 30 September 2013 was used for the validation. Tables 1–3 show the parameters and their classification of the Tokachi River basin.

Table 1 | Ground surface parameters of the Tokachi River basin

| Parameter                  | Final infiltration capacity $S_{fo}$ | Maximum storage height of surface layer $S_{r2}$ | Height to generate fast runoff $S_{r1}$ | Height to generate infiltration $S_{r0}$ | Roughness coefficient of surface $N$ | Fast intermediate flow regulation coefficient $d_{f0}$ | Initial storage height $N$ | IFAS map color |
|----------------------------|-------------------------------------|-----------------------------------------------|----------------------------------------|-----------------------------------------|------------------------------------|-----------------------------|-----------------|----------------|
| 1                          | 0.0005                              | 0.1                                           | 0.01                                   | 0.005                                   | 0.7                                | 0.8                         | 0               | Blue           |
| 2                          | 0.000002                            | 0.05                                          | 0.01                                   | 0.005                                   | 2                                  | 0.6                         | 0               | Red            |
| 3                          | 0.000001                            | 0.05                                          | 0.01                                   | 0.005                                   | 2                                  | 0.5                         | 0               | Light green    |
| 4                          | 0.0000001                           | 0.001                                         | 0.005                                  | 0.0001                                  | 0.1                                | 0.9                         | 0               | Dark green     |
| 5                          | 0.000001                            | 0.05                                          | 0.01                                   | 0.005                                   | 2                                  | 0.5                         | 0               | Pink           |
2.2.4. Performing calculation

The collected data were applied to build the model, and the river basin boundary was created according to the land-use data and digital elevation map. Then, PWRI-DHM was configured and parameters were set. After that, the simulation window was opened and ‘simulation management’ was clicked. The simulation manager could perform calculation in combining the rainfall data, parameters setup. After the calculation, the IFAS model could display the runoff simulation result of any position of the river.

2.3. LSTM model

The LSTM model can not only deal with all kinds of massive data, realize data prediction through its self-learning and self-training function but also solve the problem of ‘gradient explosion’ and ‘gradient vanishing’. ‘Gradient vanishing’ mainly refers to the fact that the weights of the hidden layer close to the input layer update slowly or even stagnate, while ‘gradient explosion’ mainly refers to the phenomenon that the weights of the hidden layer close to the input layer increase exponentially. Gradient vanishing and gradient explosion are essentially the same, both of which are caused by the gradient back propagation (back propagation refers to the back propagation of the error calculated according to the loss function to guide the updating and optimization of network parameters). In the process of practical application, the LSTM model can realize the continuous updating of the existing database, theoretically can realize the forecast of infinite data in the future, but the precision of forecast results will be lower and lower. If the real-time data can be updated manually or by the system itself, the forecast precision of follow-up data will be greatly enhanced. Specifically, for hourly rainfall forecast, the LSTM model can input hundreds of thousands or even millions of data at a time, and forecast hourly rainfall in the next few weeks or months through self-training. If the rainfall data can be updated manually or in real-time by the system itself to provide timely and sufficient database for the LSTM model, the precision of rainfall forecast will be greatly improved.

To improve the RNN model and eliminate its defects, Hochreiter & Schmidhuber (1997) introduced a new LSTM model which could be considered as the improved RNN model. Experiments showed that the LSTM model could fully solve the shortcomings of RNN model training. The most important parts of LSTM are cells, which can store information for a long time. The cells with the function of store are called logical cells, which are used to set the weight. LSTM uses gate to choose which information should enter the cell. When the Sigmoid layer is 1, information can enter, and when it is 0, information cannot enter. The only thing users need to do is setting input gate, output gate, and forget gate, then LSTM uses them to control cells (Figure 3(a)).

The LSTM model inputs the first data, and then produces a result, which will have an impact on the next calculation, and the impact will gradually disappear as this calculation is further away from the future calculation (Figure 3(b)). Therefore, it is

### Table 2 | Aquifer parameters of the Tokachi River basin

| Parameter       | Runoff coefficient of unrestricted groundwater $(1/mm/day)^{1/2}$ | Runoff coefficient of restricted groundwater $1/day$ | Storage height to produce unrestricted groundwater $m$ | Initial value for calculation $m$ |
|-----------------|-----------------------------------------------------------------|-----------------------------------------------------|---------------------------------------------------|----------------------------------|
|                 | 1                                                               | 0.1                                                 | 0.003                                             | 2                                |

### Table 3 | River tank parameters of the Tokachi River basin

| River tank parameter | Coefficient set from actual river width | Constant coefficient | Manning coefficient | Initial value for calculation | Infiltration coefficient from river to aquifer layer tank | Water level of submerged high water channel | River width of high/low water channel | Slope gradient of high water channel | Collection coefficient of river length |
|----------------------|----------------------------------------|----------------------|---------------------|-----------------------------|----------------------------------------------------------|------------------------------------------|-------------------------------------|-------------------------------------|----------------------------------------|
| RBW                  | c                                      | s                    | n                   | 2                           | 1/3                                                     | RBW                                      | RBW                                 | RBW                                 | RBW                                   |
| RBS                  | –                                      | –                    | m $^{-1}/3$ $s^{-1}$ | m                           | –                                                       | RBD                                      | RBH                                 | RBH                                 | RBH                                   |
| RNS                  | 1                                      | 7                    | 0.5                 | 0.035                       | 0                                                       | RGWD                                     | 9,999                               | 0.5                                 | 0.05                                  |
| BRID                 | 1                                      | 0                    | 2                   | 0.2                         | 0                                                       | RHW                                      | RBH                                 | 1.4                                 | 1.4                                   |
| RGWD                 | 1                                      | 0                    | 2                   | 0.2                         | 0                                                       | RHW                                      | RBH                                 | 1.4                                 | 1.4                                   |
| RHW                  | 1                                      | 0                    | 2                   | 0.2                         | 0                                                       | RHW                                      | RBH                                 | 1.4                                 | 1.4                                   |
| RBH                  | 1                                      | 0                    | 2                   | 0.2                         | 0                                                       | RHW                                      | RBH                                 | 1.4                                 | 1.4                                   |
| RBET                 | 1                                      | 0                    | 2                   | 0.2                         | 0                                                       | RHW                                      | RBH                                 | 1.4                                 | 1.4                                   |
| RLCOF                | 1                                      | 0                    | 2                   | 0.2                         | 0                                                       | RHW                                      | RBH                                 | 1.4                                 | 1.4                                   |
very important to update the existing rainfall database manually or automatically to ensure the precision of rainfall forecast in the next period. If the data are updated circularly, the rainfall forecast can be precise all the time.

The formulas of the LSTM model are as follows.

(a) The value of next cell is \( C_t^* \):

\[
C_t^* = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)
\]

where \( C_t^* \) is the value of the next unit; \( \tanh \) is hyperbolic tangent function; \( x_t \) is the current input datum; and \( W_{xc} \) is its weight (weight value has no unit, the same below); \( h_{t-1} \) is output of previous cell and \( W_{hc} \) is its weight; \( b_c \) is bias vector.

(b) The value \( i_t \) of input gate at time \( t \) is related with the current input datum \( x_t \), output of last cell \( h_{t-1} \), and the value of previous cell \( c_{t-1} \):

\[
i_t = \delta(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)
\]

where \( i_t \) is the value input gate; \( \delta \) is function of logistic sigmoid; \( W_{xi}, W_{hi}, \) and \( W_{ci} \) are weights of \( x_t, h_{t-1}, \) and \( c_{t-1} \), respectively; \( b_i \) is bias vector.

(c) The value \( f_t \) of forget gate at time \( t \):

\[
f_t = \delta(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)
\]

where \( f_t \) is the value forget gate; \( \delta \) is function of logistic sigmoid; \( W_{xf}, W_{hf}, \) and \( W_{cf} \) are weights of \( x_t, h_{t-1}, \) and \( c_{t-1} \), respectively; \( b_f \) is bias vector.

(d) \( c_t \) is the status value in current cell at time \( t \):

\[
c_t = f_t c_{t-1} + i_t \tilde{c}_t
\]

(e) \( o_t \) is the value of output gate at time \( t \):

\[
o_t = \delta(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o)
\]

where \( W_{xo}, W_{ho}, \) and \( W_{co} \) are weights of \( x_t, h_{t-1}, \) and \( c_t \), respectively; \( b_o \) is bias vector.
(f) The output $h_t$ of the LSTM model at time $t$:

$$h_t = o_t \tanh (c_t)$$ (6)

We use the way of rolling forecast, specifically using the first 24 data to forecast the 25th data. For the setting of the data set, the rainfall data in September 2013 was taken as the test set to test the final simulation fitting results. The hourly rainfall data from 1 August 2010 to 30 August 2013 was divided into training set and verification set by the ratio of 7:3.

2.4. Basic data

2.4.1. Basic data of the IFAS model

The establishment of the hydrological model of river basin mainly requires the basic data of geography, rainfall, and river channel. The IFAS model has the function of importing various relevant data according to cell size and coordinate set function, and the data can be loaded and imported completely by the IFAS model itself, which greatly facilitates the professional researchers and nonprofessional users to carry out river basin rainfall-runoff process research.

(a) Geography Data. In this study, Global Map Elevation data, Global Map Land-cover data, and Digital Soil Map of the World were downloaded by the IFAS model. Figure 4 shows the elevation, land use, and geological (lithology) conditions in the Tokachi River basin.

(b) Rainfall Data. Two kinds of rainfall data were used, which were satellite rainfall data (two sets) and ground observation rainfall data (one set). The difference is that the satellite data can be directly downloaded by IFAS, and the rainfall data of the ground observation station needs manual input. The satellite rainfall data employed the GSMAP_NRT satellite data and the

![Figure 4](Downloaded from http://iwaponline.com/jh/article-pdf/doi/10.2166/hydro.2021.035/911964/jh2021035.pdf)
3B42RT (V7) satellite data. As for the ground rainfall data, IFAS has the function of distributing the rainfall of each observation point to each unit of the calculation grid of the Thiessen Polygon of the point, only inputting the rainfall data and the coordinates of observation stations into IFAS.

(c) River Channel Data. The IFAS model can automatically generate river channel based on standard elevation data. In addition, the IFAS model can also use the imported river channel shape file to correct the automatically generated river network. In this study, the function of the IFAS model to generate river network automatically was used.

2.4.2. Basic data of the LSTM model
There are 57 rainfall ground observation stations of the Tokachi River basin, and the observation stations have continuous and long-term precipitation data. From the data over the years, the flow in the 2013 flood season belongs to the normal year, so the 2013 flood season was selected as the research period. The hourly precipitation data from 1 August 2010 to 30 September 2013 were downloaded and sorted out. Each station had 27,745 data, with a total of near 1.5 million hourly precipitation data.

2.5. Methodology

2.5.1. Applicability study of the IFAS model in the Tokachi River basin
The measured rainfall data (Measured Data, abbreviated as MD) from ground observation stations of the Tokachi River basin and two sets of satellite rainfall data of the river basin (Satellite GSMAP Data and Satellite 3B42RT Data, abbreviated as SD1 and SD2) were input into the IFAS model to simulate the rainfall-runoff process of the Tokachi River basin, and three sets of runoff simulation results were generated (abbreviated as MDR, SD1R, and SD2R), and the measured runoff result from the ground surface hydrological stations were also collected. (Ground Result, abbreviated as GR) The following comparative studies were carried out: (1) A calibration period of the IFAS model was set from 20 August to 30 September 2011. The ground observation rainfall data was used to simulate the rainfall-runoff in the basin and the simulation results accuracy was improved by adjusting the parameters. (2) A validation period of the IFAS model was set from 1 September to 30 September 2013. During the validation period, first, the runoff simulation results generated by the measured rainfall data were compared with the measured runoff result (MDR vs GR) to study the applicability of the IFAS model in large river basin runoff simulation; second, the runoff simulation results generated by the satellite rainfall data were compared with the simulation results (SD1R and SD2R vs MDR) generated by the measured rainfall data to study the feasibility of runoff simulation using the satellite rainfall data; and finally, the accuracy of the IFAS model using satellite data to simulate runoff was studied by comparing the three sets of simulation results with the measured runoff (MDR, SD1R and SD2R vs GR).

2.5.2. Precision research of the LSTM model forecasting hourly rainfall
The LSTM model can solve the problem of ‘gradient explosion’ and ‘gradient vanishing’ in the process of long sequence training, so that the LSTM model can have more performance in a longer sequence. However, too long training sequence will not necessarily increase the precision: when the training times reach a certain value, it will not increase that of precision but the workload only. In this study, the download rainfall data were used to train the LSTM model, and the forecast rainfall results by different training times were compared with the measured rainfall to study the optimal training times of the LSTM model.

2.5.3. Combination application study of IFAS and LSTM models
Combining application of the two models through inputting LSTM predicting rainfall data into IFAS, the accuracy of flood prediction was studied.

2.5.4. Error analysis
To verify the accuracy of the simulation results, the error analysis was carried out. It was based on the Nash–Sutcliffe Efficiency coefficient (NSE; Chen et al. 2021), and the wave shape error, the volume error, the peak flow error defined by the Japanese Institute of Construction Engineering. The error analysis methods and indicators are shown in Table 4.

It should be noted that the value ranges of $E_v$ and $E_p$ are between $-1$ and 1, the value ranges of $E_w$ is between 0 and 1 and when the absolute values are small, it means that the errors are small and the accuracies are high; while the value range of $E$ is between $-\infty$ and 1, the value is close to 1, indicating good simulation and high model reliability; the value of $E$ is close to 0,
indicating that the simulation result is close to the average value of the observation values, that is, the overall result is reliable, but the process simulation error is large; the value of $E$ is far less than 0, then the model is not credible.

3. RESULTS AND DISCUSSION

3.1. Applicability analysis of the IFAS model in runoff simulation in the Tokachi River basin

In this study, rainfall data from 57 ground observation stations and two satellites were input into the IFAS model to simulate the runoff process of the Tokachi River basin, and the simulation results were compared with the measured flow. The Maoiwa hydrological station was selected for data analysis and comparison. It is located in the downstream of the Tokachi River basin (Figure 1) with the largest flow and long-term runoff data, so it is the most representative station. The simulation results in the calibration period and validation period are quite close to the actual flow (Figures 5 and 6).

In the calibration period, the NSE is 0.78, the wave shape error $E_w$ is 0.1, the volume error $E_v$ is 0.22, and the peak flow error $E_p$ is −0.16. The error of the runoff simulation is small, and the simulation effect is good. In the verification period, the NSE of three simulation runoff results of the IFAS model are all above 0.75, and the simulation effects are good. Among them, the value based on ground observation station data is the smallest, which is 0.08; the error values of satellites 3B42RT (V7) and GMAP satellite data are 0.1 and 0.11, respectively. Generally speaking, the errors are all very small. As for the volume error $E_v$, the error value obtained by the GMAP satellite data is the smallest, which is −0.01, followed by the error value of 3B42RT (V7) satellite data, which is 0.02, and the error value of ground observation station data is the largest, which is 0.2. And as for the peak flow error $E_p$, the error values are similar to those of the volume error, in which the error of 3B42RT (V7) satellite data is the smallest, which is 0.01; the error of the GMAP satellite data is the second, which is 0.03; and the error of ground observation station data is the largest, which is 0.04. In brief, no matter which rainfall data is used, the simulation results are very accurate.

The simulation results and error analysis show that: first, the simulation runoff results generated by the IFAS model using the measured rainfall data are very close to the measured runoff (MDR ≈ GR), so the IFAS model is fully competent for the Tokachi River basin runoff simulation; second, the runoff simulation results generated by the IFAS using the satellite rainfall

Table 4 | Methods and indicators of error analysis

| Wave shape error | Volume error | Peak flow error | Nash–Sutcliffe Efficiency |
|------------------|--------------|----------------|--------------------------|
| $E_w = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Q_{oi} - Q_{ci}}{Q_{oi}} \right)^2$ | $E_v = \frac{\sum_{i=1}^{n} Q_{oi} - \sum_{i=1}^{n} Q_{ci}}{\sum_{i=1}^{n} Q_{oi}}$ | $E_p = \frac{Q_{op} - Q_{cp}}{Q_{op}}$ | $E = 1 - \frac{\sum_{i=1}^{n} (Q_{oi} - Q_{ci})^2}{\sum_{i=1}^{n} (Q_{oi} - Q_{av})^2}$ |

where $E_w$, $E_v$, and $E$ represent the wave shape error, the volume error, and $E$ is the calculation time (in hours); $Q_{oi}$ is the measured flow at time $i$, $Q_{ci}$ is the calculated flow at time $i$, $Q_{op}$ is the measured maximum runoff; $Q_{cp}$ is the calculated maximum runoff; and $Q_{av}$ is the measured average flow.

Figure 5 | Comparison diagram of the simulated and measured runoff processes in the calibration period (in the Maoiwa hydrological station).
data are also very close to the simulation results generated by the IFAS model using the measured rainfall data (SD1R ≈ MDR and SD2R ≈ MDR), so it is feasible for IFAS to use the satellite rainfall data in runoff simulation; and finally, the three sets of simulation results are consistent with the actual runoff and measured flow (MDR ≈ SD1R ≈ SD2R ≈ GR), so the accuracy of IFAS using the satellite data to simulate the runoff process is very high. In a word, no matter which kind of rainfall data are used, the IFAS model can accurately simulate the runoff process, accurately capture the flood peak, and the simulation accuracy is ideal, which fully shows that the applicability of the IFAS model for the Tokachi River basin is very good.

3.2. Precision analysis of hourly rainfall forecasted by the LSTM model

To test the self-learning and self-training performance of the LSTM model in forecasting hourly rainfall, 50, 100, 150, 200, 250, and 300 times trainings were conducted, respectively. Data from the Second Okawa Bridge station, Furoshiki station, and Pyrigapantine station were selected to compare the forecast results under the same observation station data and different training times (Figure 7).

The comparison results show that increasing training times in the early stage can significantly enhance the precision of the forecast results and reach the optimal effect at 200 times. At this moment, the forecast results including time distributions, accumulated rainfall values and total rainfalls are all closest to the measured values.

The volume error, peak flow error, and NSE are still used to evaluate the forecast results. With the increase of training times, the forecast effect is increasing. When the training times are 200 times, the peak rainfall error is 0.4, the volume error is 0.17, and the NSE is 0.86. At this time, the forecast precision is the highest. As the training times continue to increase, the forecast errors increase, the NSE decreases, and the forecast precision decreases (Table 6). With the increase of training times, the training time becomes longer, and neither timeliness nor precision can meet the demand of hourly rainfall forecast.

The above analysis shows that the LSTM model is suitable for hourly rainfall forecast based on a large amount of data. With sufficient early-stage rainfall data and proper training, the LSTM model can forecast future rainfall with high precision, so it is a recommendable tool to forecast hourly rainfall. The disadvantage lies in the maximum precipitation intensity period, and

---

**Table 5** | Error analysis of the IFAS model in runoff simulation and flood prediction

| Classification       | Data Source          | NSE  | $E_w$ | $E_r$ | $E_p$ |
|----------------------|----------------------|------|-------|-------|-------|
| Calibration period   | Ground Observation Station | 0.78 | 0.10  | 0.22  | −0.16 |
| Verification period  | Satellile: 3B42RT (V7)  | 0.75 | 0.10  | 0.02  | 0.03  |
|                      | Satellile: GSMAP_NRT   | 0.75 | 0.11  | −0.01 | 0.01  |
|                      | Ground Observation Station | 0.79 | 0.08  | 0.20  | −0.04 |
| Flood prediction     | The optimal forecast rainfall data | 0.81 | 0.15  | 0.05  | 0.21  |
Figure 7 | Rainfalls forecasted by the LSTM model with different training times in the Second Okawa Bridge station (a), Furoshiki station (b), and Pyrigapantine station (c).
the forecast peak rainfall is less than the measured maximum precipitation. In future research, it should be taken as a research focus, or introduce a peak adjustment coefficient to make up for the error of rainfall forecast in the period of heavy precipitation.

In previous studies, the RNN model and the grey model had been used to forecast rainfall. Although the RNN model could process a large number of data and forecast rainfall, the RNN model often brought gradient vanishing and gradient explosion in the process of forecasting, resulting in low forecast accuracy and low reliability. The grey model had high accuracy in forecasting monthly rainfall; however, it was difficult to deal with a large number of data, let alone to achieve the prediction of hourly rainfall with a large proportion of 0. The LSTM model can not only deal with a large number of data but also achieve rainfall prediction with high accuracy, which has obvious advantages.

3.3. Accuracy evaluation on runoff simulation and flood prediction by the combination application of IFAS and LSTM models

In this study, the optimal forecast rainfall data obtained from 200 times trainings of the LSTM model were inputted into the IFAS model to carry out runoff simulation and flood prediction of the Tokachi River basin. Comparing the simulated runoff process with the measured runoff process of the Maoiwa hydrological station, it can be found that except the simulated peak flow is less than the measured peak flow, the remaining parts are all well-fitted (Figure 8).

The wave shape error is 0.15, the volume error is 0.05, the peak flow error is 0.21, and the NSE is 0.81 (Table 5), so the overall simulation effect is good. It is worth noting that the simulation of flow peak value by the IFAS model is less than the measured flow, which is related to the small forecast value of LSTM for maximum rainfall in Section 3.2: the insufficient forecast of the LSTM model for maximum rainfall results in the small prediction value of maximum flood by the IFAS model.

However, despite the minor shortcoming, the combined utilization of IFAS and LSTM models still has a good application prospect (or vision). All the following work can be finished in the office without going out: (1) Using the IFAS model to download the satellite rainfall data directly and then update the rainfall database automatically; or collecting rainfall data of the ground observation station to update the rainfall database through programming language; (2) Input updated rainfall data into the LSTM model, and forecast the future rainfall through self-training functions; (3) The optimal forecast rainfall data is fed back to the IFAS model to complete runoff simulation and flood prediction.

Table 6 | Error analysis of the LSTM model in rainfall forecast

| Training Times | 50   | 100  | 150  | 200  | 250  | 300  |
|----------------|------|------|------|------|------|------|
| $E_p$          | 0.68 | 0.50 | 0.55 | 0.40 | 0.57 | 0.63 |
| $E_v$          | 0.56 | 0.30 | 0.28 | 0.17 | 0.36 | 0.47 |
| NSE            | 0.57 | 0.79 | 0.77 | 0.86 | 0.73 | 0.66 |

Figure 8 | Comparison diagram of the simulated and measured runoff processes (in the Maoiwa hydrological station).
It should be noted that since the rainfall data of the ground observation station needs to be collected and input manually, the workload is very large, and there is uncertainty about whether the ground observation station is available and whether the data can be collected. Even if there is no rainfall data of the ground observation station, the above steps can still be completed only by downloading the satellite rainfall data. It is good to have the data of the ground observation station, which can be compared and rechecked; if not, it will still not affect the realization of the goal of flood prediction. In a word, the combination of LSTM and IFAS models can accurately simulate runoff and predict flood in those areas without the ground observation station.

4. SUMMARY AND CONCLUSIONS

Taking the rainfall-runoff process of the Tokachi River basin during the autumn flood in 2011 and 2013 as the research platform, the applicability of the IFAS model in rainfall-runoff simulation in the Tokachi River basin, the precision of hourly rainfall forecast by the LSTM model, and the combination application of IFAS and LSTM models were studied. And the wave shape error, the volume error, the peak flow error, and the NSE coefficient were employed to evaluate the results.

The main conclusions are as follows: (1) the IFAS model can simulate the surface runoff process in the Tokachi River basin based on the satellite rainfall data or ground observation station rainfall data, and the simulation accuracy is quite high. (2) The LSTM model can forecast hourly rainfall with high precision after proper self-training times. The forecast rainfall results have a good consistency with the actual precipitation value in time distribution, but in the period of dense precipitation, the peak rainfall is underestimated. (3) The runoff process can be simulated with high accuracy and the flood can be accurately predicted by the IFAS model combining with the optimal forecast rainfall of the LSTM model, and the main simulation error occurs in the peak flow.

ACKNOWLEDGEMENTS

This study was supported by the National Natural Science Foundation of China (Grant Nos. 41831289, 41772250, and 41877191), the Foundation of Key Scientific Research Projects of Henan Colleges and Universities in 2019 (19A170008), and Key Laboratory of Mine Geological Hazards Mechanism and Control and Department of Land and Resources of Shaanxi Province Foundation (KF2018-06).

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Bai, T. & Tahmasebi, P. 2020 Efficient and data-driven prediction of water breakthrough in subsurface systems using deep long short-term memory machine learning. *Computational Geosciences* **25** (1), 1–13. https://doi.org/10.1007/s10596-020-10005-2.

Chen, Y., Nakatsugawa, M. & Ohashi, H. 2021 Research of impacts of the 2018 Hokkaido Eastern Iburi Earthquake on sediment transport in the Atsuma River basin using the SWAT model. *Water* **15** (3), 356. https://doi.org/10.3390/w15030356.

Chow, M. F., Jamil, M. M., Ros, F. C., Yuzir, M. A. M. & Hossain, M. S. 2019 Evaluation of parameter regionalization methods for flood simulations in Kelantan river basin. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*. Available from: http://dspace.uniten.edu.my/jspui/handle/123456789/13033.

Gao, Y. Q., Yuan, Y., Wang, H. Z., Schmidt, A. R. & Ye, L. 2017 Examining the effects of urban agglomeration polders on flood events in Qinhuai River basin, China with HEC-HMS model. *Water Science and Technology* **75** (9–10), 2130–2138. https://doi.org/10.2166/wst.2017.023.

Goyal, M. K., Bharti, B., Quilty, J., Adamowski, J. & Pandey, A. 2014 Modeling of daily pan evaporation in subtropical climates using ANN, LS-SVR, fuzzy logic, and ANFIS. *Expert Systems with Applications* **41** (11), 5267–5276. https://doi.org/10.1016/j.eswa.2014.02.047.

Gumindoga, W., Rwasoka, D. T., Nhapi, I. & Dube, T. 2017 Ungauged runoff simulation in Upper Manyame Catchment, Zimbabwe: application of the HEC-HMS model. *Physics and Chemistry of the Earth, Parts A/B/C* **100**, 371–382. https://doi.org/10.1016/j.pce.2016.05.002.

Hao, G. R., Li, J. K., Song, L. M., Li, H. E. & Li, Z. L. 2018 Comparison between the TOPMODEL and the Xin’anjiang model and their application to rainfall runoff simulation in semi-humid regions. *Environmental Earth Sciences* **77** (7), 1–13. https://doi.org/10.1007/s12665-018-7477-4.

Hochreiter, S. & Schmidhuber, J. 1997 Long short-term memory. *Neural Computation* **9** (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.
