Infectious Disease Modelling 10 (2025) 691–701

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Prediction of monthly occurrence number of scrub typhus in Ganzhou City, China, based on SARIMA and BPNN models

Renfa Huang [a, 1], Kailun Pan [b, 1], Qingfeng Cai [a], Fen Lin [b], Hua Xue [a], Mingpeng Li [a], Yong Liao [a, *]

[a] Ganzhou Center for Disease Control and Prevention, Ganzhou, 341000, Jiangxi, China
[b] School of Public Health and Health Management, Gannan Medical University, Ganzhou, 341000, Jiangxi, China

ARTICLE INFO

Article history:
Received 21 October 2024
Received in revised form 28 December 2024
Accepted 12 February 2025
Available online 13 February 2025
Handling Editor: Dr Daihai He

Keywords:
Scrub typhus
SARIMA
BPNN
Predictive modelling

ABSTRACT

Scrub typhus poses a serious public health risk globally. Forecasting the occurrence of the disease is essential for policymakers to develop prevention and control strategies. This study investigated the application of modelling techniques to predict the occurrence of scrub typhus and establishes an early warning system aimed at providing a foundational reference for its effective prevention and control. In this study, the monthly occurrence of scrub typhus in Ganzhou City from January 2008 to December 2022 was utilized as the training set for the first part of the analysis, while the data from January 2008 to December 2019 served as the training set for the second part. Based[1] on these data, the SARIMA model, the BPNN model, and the combined SARIMA-BPNN model were developed and validated using data from January to December 2023. The most effective model was then selected to predict the number of occurrences of scrub typhus for the years 2024 and 2025, respectively. The root mean square error (RMSE) and mean absolute error (MAE) of the BPNN (3-9-1) model, developed using data from January 2008 to December 2022, were 8.472 and 6.4, respectively. In contrast, the RMSE and MAE of the combined SARIMA-BPNN (1-9-1) model, constructed using data from January 2008 to December 2019, were 19.361 and 16.178, respectively. In addition, the BPNN (3-9-1) model predicted 284 cases of scrub typhus in Ganzhou City for 2024, and 163 cases for 2025. The BPNN (3-9-1) model demonstrated strong applicability in predicting the monthly occurrence of scrub typhus. Furthermore, incorporating three years of data on the occurrence of new crown outbreaks when developing a predictive model for infectious diseases can substantially enhance prediction accuracy.

1. Introduction

Scrub typhus, also referred to as jungle typhus, is an acute vector-borne infectious disease caused by Orientia tsutsugamushi (Sun et al., 2017). The clinical manifestations of this disease primarily include fever, distinctive crusting, and

* Corresponding author.
E-mail address: gzcdc001@126.com (Y. Liao).
Peer review under the responsibility of KeAi Communications Co., Ltd.
[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.idm.2025.02.009

lymphadenopathy. In severe cases, patients may experience multiple organ failure, which can be life-threatening (Paris et al., 2013). The mortality rate of the disease can reach as high as 70% if not treated promptly, and even with treatment, the mortality rate remains at 1.4% (Bonell et al., 2017; Ma et al., 2017). Furthermore, the absence of an effective vaccine complicates efforts to control this disease. Consequently, the development of a prediction model for scrub typhus, along with the precise identification of its progression trends, was of significant practical importance for implementing effective prevention and control measures.

Numerous mathematical models have been developed to effectively predict the spread of infectious diseases (Saxena, 2021; Wang et al., 2019; Zhang et al., 2020). Among these models, the seasonal auto regressive integrated moving average (SARIMA) model demonstrated distinct advantages. Rather than depending on external factors or presuming data distributions, it generated predictions exclusively by analyzing historical time series data. Furthermore, the SARIMA model possessed the capability to identify anomalies and trend mutations within the time series data, thereby enhancing its reliability and robustness (Fang et al., 2023). However, in the investigation of the temporal distribution of infectious diseases and the prediction of their future incidence, researchers have discovered that the SARIMA model was limited to modelling linear relationships within time-series data. This model struggled to account for potential non-linear relationships that may arise in such datasets. Furthermore, SARIMA models typically necessitated that the series exhibited stationarity or achieved stationarity following differencing and logarithmic transformations. Despite its widespread application in predicting infectious diseases, the simulation and predictive accuracy of the SARIMA model often proves inadequate when confronted with sequences exhibiting complex change patterns or nonlinear characteristics. On the other hand, the back propagation neural network (BPNN) model employed a distributed information storage method, exhibiting strong environmental adaptability, learning capability, robustness, and fault tolerance, along with the ability to effectively model a variety of nonlinear data (Liu et al., 2019). Consequently, many researchers have started to integrate the SARIMA model with neural network models or other predictive frameworks, effectively leveraging the strengths of multiple models to enhance overall performance.

Since the outbreak of novel coronavirus pneumonia at the end of 2019, the epidemic has rapidly escalated into a global pandemic. To effectively control the spread of the novel coronavirus, China promptly initiated a level 1 response for public health emergencies and implemented a series of public health interventions (Shen et al., 2022). The 'zero outbreak' policy implemented by the Chinese government may have significant and far-reaching effects on the spread of other infectious diseases (Chen et al., 2021). However, it remains unclear how the occurrence numbers during the novel coronavirus pneumonia period influence the effects of predictive models. Consequently, it was essential to explore the necessity of incorporating data from this period into the model to enhance the prediction of infectious diseases following an outbreak.

Since the initial case was reported in Ganzhou City in 2006, the prevalence of the disease has been expanding, making it a high-incidence area for scrub typhus in Jiangxi Province. From 2006 to 2017, Ganzhou City recorded the highest number of cases and the highest incidence rate of scrub typhus in Jiangxi (Liao et al., 2019). Although researchers have analyzed the epidemiological characteristics and influencing factors of scrub typhus, they have yet to predict its incidence (Pan et al., 2024; Peng et al., 2022). Consequently, this study employed the SARIMA model, the BPNN model, and the combined SARIMA-BPNN model to predict the occurrence of scrub typhus in Ganzhou City. Additionally, it aimed to investigate whether the number of cases during the epidemic influences the predictive accuracy of the models, thereby providing a reference for assessing the applicability of predictive models in forecasting infectious diseases following an epidemic.

2. Methods

2.1. Data collection

The data on scrub typhus cases in Ganzhou City, reported by the National Disease Surveillance Information Management System (NDSIMS) from January 2008 to December 2023, were selected for analysis. All cases adhered strictly to the diagnostic criteria established by the Chinese Centre for Disease Control and Prevention (https://www.chinacdc.cn/).

2.2. Constructing predictive models

2.2.1. The SARIMA model

The SARIMA model integrates seasonal effects, long-term trend effects, and cyclical variations. The general structure of the SARIMA model is expressed as SARIMA(p,d,q)(P,D,Q)s, where p, d, and q represent the non-seasonal components, while P, D, and Q represent the seasonal components. Specifically, p and P indicate the degree of autoregression for the non-seasonal and seasonal components, respectively. Similarly, d and D denote the degree of differencing for the non-seasonal and seasonal components, respectively. Furthermore, q and Q represent the degree of moving average for the non-seasonal and seasonal components, respectively, with s denoting the sampling period.

Te expression of the SARIMA model is given as follows:

$$\phi(B^s)\varphi(B)(x_t-\mu)=\Theta(B^s)\theta(B)_{\varepsilon_t} \tag{1}$$

Te nonseasonal components are as follows:

$$\text{AR}: \varphi(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p \tag{2}$$

$$\text{MA}: \theta(B) = 1 - \theta_1 B + \cdots + \theta_q B^q \tag{3}$$

Te seasonal components are as follows:

$$\text{Seasonal AR}: \phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^{PS} \tag{4}$$

$$\text{Seasonal MR}: \Theta(B) = 1 - \Theta_1 B + \cdots + \Theta_Q B^{QS} \tag{5}$$

whereas, B indicates a reverse shift, ε_t represents a projected residual error at time t, and x_t denotes an observed value at time t (where t ranges from 1 to k). Additionally, φ refers to the route of the AR coefficients, θ represents the route of the MA coefficients, and Θ signifies the route of the seasonal MA coefficients.

The fitting steps of the SARIMA model were as follows: (1) data stationarity: utilize the augmented dickey-fuller (ADF) unit root test to assess the stationarity of the series. Convert non-stationary series into stationary series through differencing, simultaneously determining the parameters d and D. (2) Model identification and parameter estimation: generate autocorrelation plots (ACF) and partial autocorrelation plots (PACF) to select the values for the model parameters p and q. For the seasonal parameters P and Q, based on insights from previous studies, their values are typically less than 3. Therefore, values of 0, 1, and 2 were considered, leading to the establishment of nine alternative models. (3) Model testing: the optimal prediction model was selected based on the minimum Akaike Information Criterion (AIC). The validity of this optimal model was assessed using the Ljung-Box Q test. A *P*-value greater than 0.05 indicates that the residuals are white noise, thereby confirming the model's validity. (4) Model fitting: the selected optimal model was employed to predict the monthly occurrence numbers from January to December 2023.

2.2.2. The BPNN model

The BPNN model is a multilayer feed-forward network that can adjust the weights of nonlinear functions. It primarily consists of three components: the input layer, the hidden layer, and the output layer (Zhao, 2023). The steps for constructing the BPNN model were as follows: (1) constructing the training set and the sample data set: typically, the BPNN model utilizes the first M values of the samples as the input layer data and the last N values as the output layer data. (2) Data normalization: since the number of events per month varies greatly, this variation may affect the sensitivity and training speed of the model. Therefore, the data needs to be normalized to the range [0,1]. (3) Parameter setting: firstly, the number of neuron nodes in both the input and output layers is established. The number of nodes in the hidden layer is then calculated using the formula $n = \sqrt{M+N} + k$, where n represents the number of nodes in the hidden layer, M is the number of input nodes, N is the number of output nodes, and k is a natural number ranging from 1 to 10. Subsequently, the error value, learning rate, and number of training sessions are defined, with values set at 0.00001, 0.01, and 1000, respectively. Additionally, the training function for this model is designated as trainlm, the activation function for the input layer is set to tansig, and the activation function for the hidden layer is selected as purelin. Finally, the minimum root mean square error (RMSE) of the training set is utilized to determine the optimal model. (4) Model fitting: the optimal model will be utilized to predict the number of monthly occurrences from January to December 2023.

2.2.3. The combined SARIMA-BPNN model

Leveraging the advantageous characteristics of the BPNN model, the predictive performance of the SARIMA model was enhanced and optimized, thereby improving the accuracy of the overall prediction model. The fitted and predicted values generated by the best-fitting SARIMA model serve as inputs to the BPNN model, while the actual values are utilized as outputs, resulting in the creation of a combined SARIMA-BPNN model. Validation was conducted using data from January to December 2023 to assess the fitting performance of the model.

2.2.4. Assessment of the effects of predictive modelling

The assessment of the effects of prediction models was based on the error between the predicted and actual values; a smaller error indicated greater accuracy in model predictions. In this study, the root mean square error (RMSE) and the mean absolute error (MAE) were chosen to evaluate the predictive performance of various models.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{7}$$

Where n is the number of observations, yi is the actual value, $\hat{y}$ is the predicted value.

2.3. Statistical analyses

In this study, MATLAB version 2023b was utilized to construct both the BPNN model and the combined SARIMA-BPNN model, while R version 4.3.3 was employed for the development of the SARIMA fitted model, with a significance level established at $P < 0.05$. To account for the potential impact of the novel coronavirus outbreaks on the predictive model, the study divided the data into two segments for analysis. Specifically, this study utilized the monthly occurrence data of scrub typhus in Ganzhou City from January 2008 to December 2022 as the first part of the training set, while the data from January 2008 to December 2019 served as the second part of the training set. Based on these two data sets, the SARIMA model, the BPNN model, and the combined SARIMA-BPNN model were constructed to fit the monthly occurrence numbers from January to December 2023. Secondly, the study compared the model predictions with the actual values to assess the predictive effects of the various models. Finally, the optimal model was selected to predict the occurrence of scrub typhus in Ganzhou City for the years 2024–2025.

3. Results

3.1. Descriptive analyses

From 2008 to 2023, a total of 6,892 cases of scrub typhus were reported in Ganzhou City. The trend in the occurrence of scrub typhus exhibited an approximate 'N' shape, peaking in 2018 (Fig. 1A). The occurrence of scrub typhus was predominantly observed during the summer months, indicating a clear seasonal pattern (Fig. 1B).

3.2. Constructing modelling using data from January 2008 to December 2022

3.2.1. Constructing the SARIMA model

3.2.1.1. Data stationarity test. A stationarity test conducted on the time series of scrub typhus cases from January 2008 to December 2022, yielded a t-value of −8.604 and a p-value of 0.01, indicating that the series was stationary. Furthermore, plotting the time series against the original series revealed significant seasonality within the data (Fig. 2). The 'diff' function was utilized to conduct first-order seasonal differencing on the original time series, effectively eliminating the seasonal effects present in the series. The resulting of time series was then plotted (Supplementary Fig. 1). An ADF test was performed on the differentiated time series, yielding a test statistic of $t = -4.739$ and a p-value of 0.01. This indicated that there was no unit root in the original data series, suggesting that it can be regarded as stationary and suitable for modelling.

3.2.1.2. Model identification and parameter estimation. The pure white noise test was conducted on the stationary time series, revealing that the p-value of the Ljung-Box Q-test statistic was consistently low ($P < 0.05$) across all delay orders. This indicated that the series can be classified as non-white noise. Following the stationarity test after differencing, it can be preliminarily concluded that the model parameters were $d = 1$, $D = 1$, and $s = 12$. Based on the ACF and PACF diagrams (Supplementary Fig. 1), the values of p and q were identified as 1 and 0, respectively. For the seasonal parameters P and Q, previous studies suggested that their values typically did not exceed 3. Consequently, we considered P and Q to be 0, 1, and 2, leading to the establishment of nine alternative models through individual testing. As shown in Supplementary Table 1, the SARIMA$(1,1,0)(2,1,0)_{12}$ model emerged as the optimal choice. The Ljung-Box Q-test applied to the model residuals indicates that $\chi^2 = 4.251$, with a p-value of 0.528. Therefore, the null hypothesis cannot be rejected, suggesting that the residual series can be regarded as a white noise series. This implied that most of the data characteristics in the time series have been effectively captured by the optimal model.

3.2.2. Constructing the BPNN model

In this study, the BPNN model was developed using the occurrences data of scrub typhus from January 2008 to December 2022 to forecast the occurrences of scrub typhus for the period from January to December 2023. Taking into account the periodicity and autocorrelation characteristics of scrub typhus occurrences in Ganzhou City, three distinct BPNN models were established in this research. Net1: utilizes the number of occurrences from the previous six months to forecast the occurrences in the subsequent month. Net2: employs the number of occurrences from the preceding twelve months to predict the

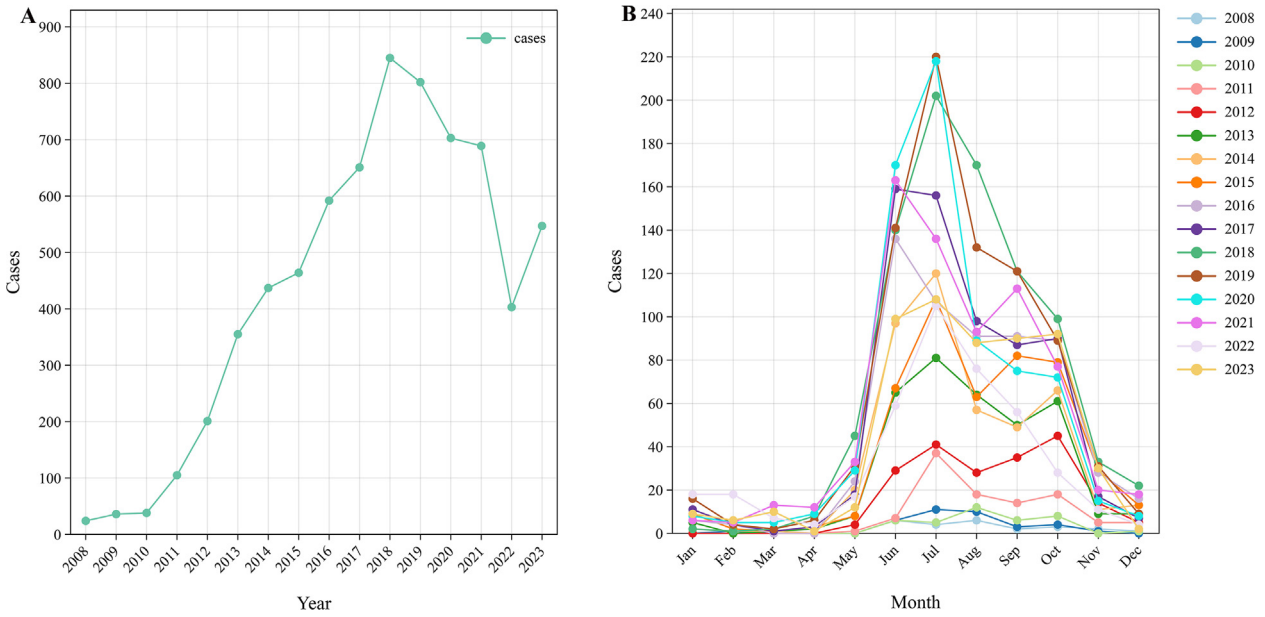

Fig. 1. Occurrence numbers of scrub typhus in Ganzhou City, January 2008–December 2023. (A: annual occurrence numbers; B: monthly occurrence numbers).

occurrences in the following month. Net3: analyzes the number of occurrences during the same period over the previous three years to estimate the occurrences in the corresponding period of the upcoming year.

3.2.2.1. Sample processing. The training and validation sets were constructed using Net1 as an example. Net1 comprised 186 samples, with the first 174 designated for the training set and the remaining 12 allocated to the validation set. The specific structure of the sample sets was illustrated in Supplementary Tables 2 and 3

3.2.2.2. Training network. The range of hidden layer nodes for the three network models was derived according to the specified equation, facilitating the construction of BPNN models with varying network architectures. The optimal network structures for these models were determined based on the principle of minimizing the RMSE of the test set. As shown in Supplementary Table 4, the optimal hidden layers for Net1, Net2, and Net3 were 11, 14, and 9, respectively. Consequently, the optimal network structures for the three models were 6-11-1, 12-14-1, and 3-9-1. Among the three BPNN models, the structure 3-9-1 emerged as the optimal configuration.

3.2.3. Constructing the combined SARIMA-BPNN model

From the preceding analysis, it was established that the SARIMA$(1,1,0)(2,1,0)_{12}$ model was optimal. The fitted and predicted values generated by this SARIMA model served as inputs for the BPNN model, while the actual values were utilized as the outputs of the BPNN model. A total of ten alternative models were constructed based on the number of neurons in the hidden layer. When the number of neurons in the hidden layer was set to eight, the *combined* SARIMA-BPNN model achieved the minimum RMSE value, indicating that the network structure of SARIMA-BPNN was 1-8-1 (Supplementary Table 4).

3.2.4. Model predictions and comparisons

The actual and predicted cases of scrub typhus in 2023 were presented in Table 1 and Fig. 3. All three constructed models demonstrated a strong ability to predict scrub typhus cases effectively. The BPNN (3-9-1) model provided predictions that were closest to the actual case numbers in January, April, May and August. In contrast, the combined SARIMA-BPNN (1-8-1) model exhibited the best performance in February, May, and June. To further compare the three models, various rating metrics, including MAE and RMSE, were employed in this study. As shown in Table 2, the BPNN (3-9-1) model demonstrated the best predictive performance, exhibiting the lowest values of MAE and RMSE. This was closely followed by the combined SARIMA-BPNN and SARIMA models, respectively.

3.3. Constructing modelling using data from January 2008 to December 2019

3.3.1. Constructing the SARIMA model

The time series of scrub typhus cases from January 2008 to December 2019 was assessed for stationarity. The results indicated that the time series was stationary ($t = -5.212$, $P = 0.01$). Furthermore, plotting the time series against the original

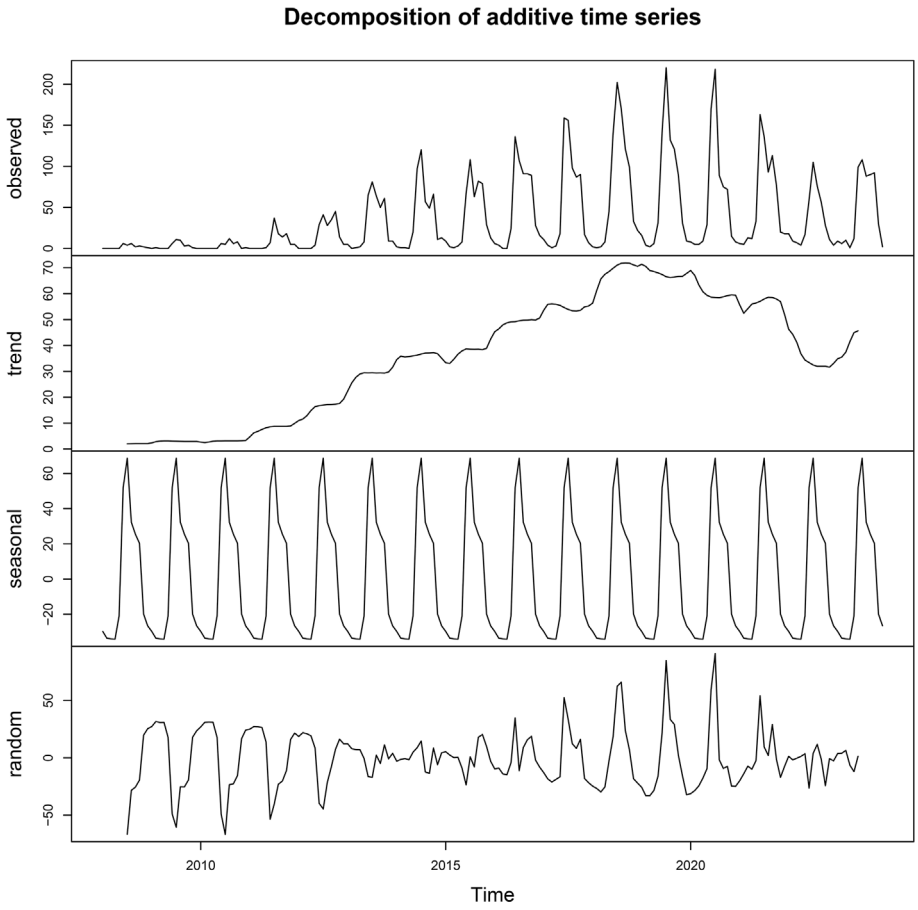

Fig. 2. Decomposition of the original sequence.

series revealed significant seasonality within the data (Fig. 2). The original time series underwent first-order seasonal differencing to remove the seasonal effects present in the series. The p and q parameters of the SARIMA model were initially determined using the ACF and PACF plots (Supplementary Fig. 2). Subsequently, the optimal SARIMA$(1,1,0)(0,1,1)_{12}$ model was identified based on the principle of AIC minimization (Supplementary Table 5). A white noise test was conducted on the model's residuals, yielding a result of $\chi^2 = 9.952$, $P = 0.077$, which confirmed the model's adequacy.

3.3.2. Construction the BPNN model and combined SARIMA-BPNN model

A total of 30 alternative models were constructed based on variations in the number of neurons in the hidden layer, as well as different configurations for the input and output layers. When the hidden layer was 13, the BPNN model achieved the lowest RMSE value, resulting in a network structure of 6-13-1(Supplementary Table 6).

From the preceding analysis, it was determined that the SARIMA$(1,1,0)(0,1,1)_{12}$ model was optimal. The fitted and predicted values generated by this SARIMA model served as inputs for the BPNN model, while the actual values were utilized as the outputs of the BPNN model. A total of ten alternative models were constructed based on the number of neurons in the hidden layer. When the hidden layer was 9, the combined SARIMA-BPNN model exhibited the minimum RMSE value, with a network structure of 1-9-1 (Supplementary Table 6).

3.3.3. Model predictions and comparisons

The BPNN (6-13-1) model predicted the number of cases most accurately in March, while the combined SARIMA-BPNN model demonstrated superior performance in December (Table 3). To further evaluate these three models, this study employed MAE and RMSE as assessment metrics. As indicated in Table 4, the combined SARIMA-BPNN (1-9-1) model exhibited the best predictive performance, achieving the lowest values for both MAE and RMSE.

Table 1
Predicted monthly number of scrub typhus cases in Ganzhou City, January 2023–December 2023.

Time	Actual value	Predicted value		
		BPNN	SARIMA- BPNN	SARIMA
January	9	10	6	−2
February	6	14	6	−3
March	10	6	6	−10
April	1	0	6	−9
May	12	11	11	6
June	99	71	99	81
July	108	101	130	135
August	88	88	90	70
September	90	82	46	51
October	92	65	33	35
Novembe	30	36	8	3
December	2	18	5	−8

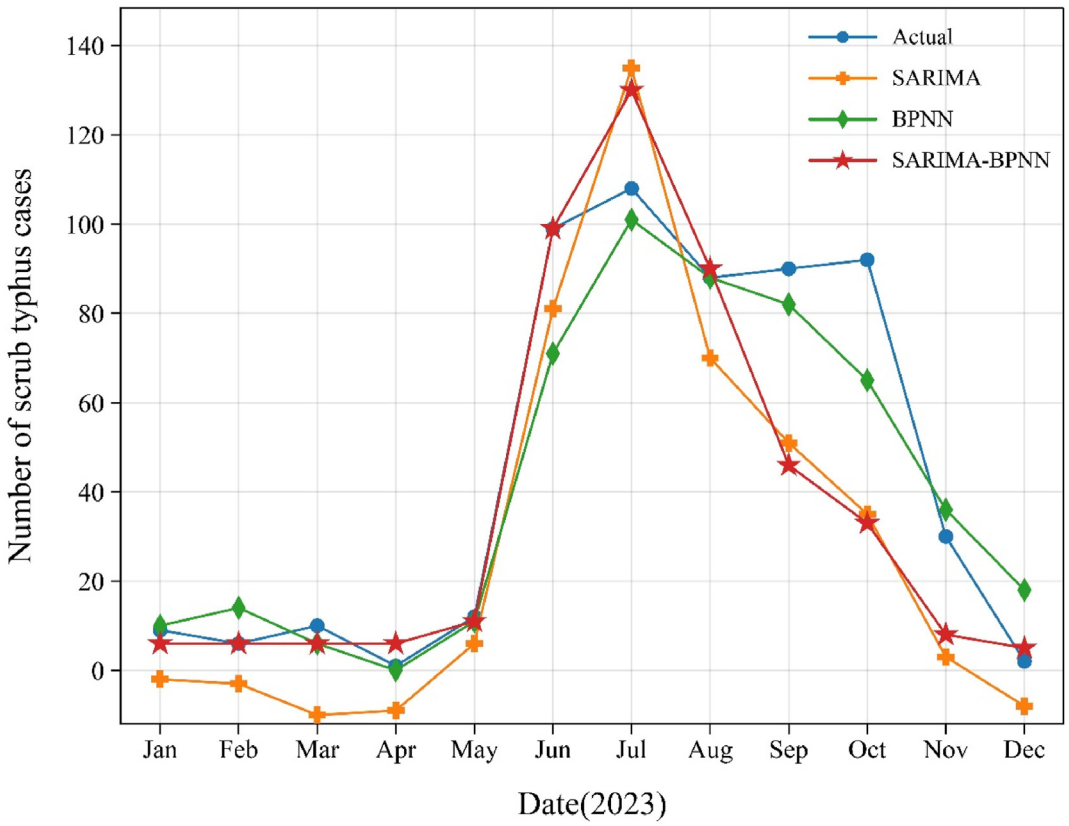

Fig. 3. Comparison between the predicted occurrences of scrub typhus and the actual number of occurrences from January 2023 to December 2023.

3.4. Final model selection and application

This study compared the MAE and RMSE values of the BPNN (3-9-1) model, which was constructed using data from January 2008 to December 2022, with those of the SARIMA-BPNN (1-9-1) model, which was developed using data from

Table 2
Comparison of three models using mean absolute error (MAE) and root mean square error (RMSE).

Evaluation index	SARIMA$(1,1,0)(2,1,0)_{12}$	BPNN (3-9-1)	SARIMA- BPNN (1-8-1)
MAE	21.969	6.4	13.873
RMSE	27.950	8.472	19.881

Table 3
Predicted monthly number of scrub typhus cases in Ganzhou City, January 2023–December 2023.

Time	Actual value	Predicted value		
		BPNN	SARIMA- BPNN	SARIMA
January	9	−19	6	7
February	6	−1	−16	−4
March	10	9	−18	−5
April	1	44	−9	−1
May	12	21	20	25
June	99	32	129	133
July	108	111	89	206
August	88	79	122	129
September	90	78	68	111
October	92	80	87	83
Novembe	30	22	18	23
December	2	13	0	3

Table 4
Comparison of three models using mean absolute error (MAE) and root mean square error (RMSE).

Evaluation index	SARIMA$(1,1,0)(0,1,1)_{12}$	BPNN (6-13-1)	SARIMA- BPNN (1-9-1)
MAE	21.413	17.558	16.178
RMSE	33.740	25.604	19.361

January 2008 to December 2019. The model exhibiting smaller MAE and RMSE values, specifically the BPNN (3-9-1), was identified as the superior forecasting model, as detailed in Tables 2 and 4

The model BPNN (3-9-1) was utilized to predict the occurrence of scrub typhus in Ganzhou City from January 2024 to December 2025, as presented in Table 5. The occurrence of scrub typhus in Ganzhou City during this period remained low, with the number of cases recorded at 284 in 2024 and 163 in 2025.

4. Discussion

Scrub typhus is one of the major public health challenges globally, posing a significant risk to human health. Each year, approximately one million cases of scrub typhus were reported, with around one billion individuals at risk of infection (Premaratna et al., 2017). The prediction of infectious diseases facilitated the timely detection of trends in disease progression. Consequently, regions that were continuously monitored for scrub typhus can be utilized to forecast the occurrence of this disease by developing a mathematical model. The findings from this model can enhance scrub typhus surveillance and provide a scientific foundation for prevention strategies in the area.

The occurrence of scrub typhus in Ganzhou City was primarily concentrated between June and October, exhibiting a notable seasonality. This phenomenon may be attributed to the alignment of this period with the local agricultural season, during which farmers engaged in fieldwork, thereby increasing their likelihood of contact with chiggers (Liao et al., 2014). Furthermore, the warm climate and extended hours of sunlight during these months created favorable conditions for the growth and development of chiggers (Wei et al., 2022). In this study, the trend of occurrence decreased since 2018, but exhibited an increasing trend once again in 2023. Alongside scrub typhus, other infectious diseases, such as gonorrhea and hepatitis, displayed similar patterns (Wang et al., 2024; Zheng et al., 2023). This may be attributed to the Chinese government's embargo strategy in response to the novel coronavirus outbreak, which involved strict control of population movement. Consequently, there was a virtual standstill of most social activities in China, leading to a reduction in the number of individuals infected with scrub typhus.

In this study, we constructed the BPNN model, the SARIMA model, and the combined SARIMA-BPNN model using the occurrence data of scrub typhus in Ganzhou City from January 2008 to December 2022. These models were compared against actual values to evaluate their fitting performance. The results indicated that the BPNN model exhibited a significantly better fitting effect than the other two models, highlighting the nonlinear characteristics of scrub typhus occurrence data in Ganzhou City. In this study, three distinct BPNN models were developed based on various methods of sample division. All three models were capable of predicting the occurrence of scrub typhus; however, the third model (Net3) demonstrated the smallest error between the predicted values and the actual data, achieving the lowest evaluation metrics, namely RMSE. In the prediction process of BPNN, various methods of sample division can result in differences in prediction accuracy, potentially leading to significant discrepancies (Fang et al., 2023). Consequently, when conducting BPNN modelling, it was advisable to explore multiple sample division techniques to identify the optimal network model. This study employed the BPNN (3-9-1) model to predict the occurrence of scrub typhus in Ganzhou City from January 2024 to December 2025. The

Table 5
Predicted monthly number of scrub typhus cases in Ganzhou City, January 2024–December 2025.

2024	Predicted value	2025	Predicted value
January	12	January	10
February	14	February	11
March	12	March	14
April	17	April	13
May	26	May	14
June	23	June	15
July	22	July	14
August	30	August	14
September	25	September	16
October	38	October	13
Novembe	56	Novembe	14
December	9	December	15

results indicated that the occurrence of scrub typhus in Ganzhou City is expected to remain low during this period, demonstrating a decreasing trend. However, it is essential to strengthen prevention and control measures for scrub typhus.

As a naturally occurring epidemic disease, the incidence and spread of scrub typhus are influenced by a combination of factors, including meteorological conditions, the health environment, and social and public health measures (Pan et al., 2024; Ranjan, 2018). Researchers Geng et al. found that non-pharmacological interventions aimed at preventing and controlling neococcal pneumonia were effective in reducing the incidence of common infectious diseases transmitted through the respiratory and digestive tracts, such as pertussis, influenza, and scarlet fever (Geng et al, 2022). The impact of the lower number of scrub typhus occurrences during the new coronavirus epidemic on the predictive efficacy of the model remained unclear, necessitating further exploration. In this study, a scrub typhus prediction model was developed using data collected from January 2008 to December 2019. The findings indicated that the SARIMA-BPNN model exhibited the highest predictive performance. This model effectively integrated the strengths of both the SARIMA and BPNN methodologies, with SARIMA focusing on extracting linear information from the data and BPNN adeptly mining its nonlinear features. This synergistic approach enabled the model to leverage the available information more comprehensively, thereby enhancing its applicability for predicting the incidence of infectious diseases (Orang et al., 2024). However, the MAE and RMSE of the SARIMA-BPNN model constructed from January 2008 to December 2019 were higher than those of the BPNN model developed from January 2008 to December 2022. This discrepancy may be attributed to the characteristics inherent in the time series data and the nature of infectious diseases. The coherence characteristics of time series demonstrated an interdependence among serial values that exhibited fluctuations over time, highlighting a significant continuity in their relationships. This coherence facilitated the identification of patterns of change between series values, thereby enabling the forecasting of future trends. Considering three years of data from the new coronavirus epidemic when constructing the prediction model can ensure the consistency of occurrence numbers in the time series, thereby enhancing the model's ability to identify data characteristics. Conversely, excluding these three years of data from the modelling process would lead to gaps in the dataset for that period, ultimately diminishing the efficacy of the developed model in predicting infectious diseases. Despite the significant reduction in the occurrence of scrub typhus due to the impact of the new coronavirus outbreak, it was essential to incorporate the occurrence data from these three years into the infectious disease prediction model for short-term forecasts to enhance the accuracy of the predictions.

This study presents several strengths. Firstly, it comprehensively analyzed the linear and non-linear characteristics of the serial data using the SARIMA model, the BPNN model, and the combined SARIMA-BPNN model. Secondly, it incorporated three years of occurrence data related to the epidemic and examined the potential impact of the new coronavirus epidemic on the prediction models.

This study also has several limitations. Firstly, the data on scrub typhus cases were obtained passively, making it susceptible to misreporting and underreporting. Secondly, the incidence of scrub typhus was influenced by meteorological factors and socio-economic conditions, which were not considered in this study. Furthermore, this study did not consider the transmission dynamics of infectious diseases, which may not adequately address the issue of time delays.

5. Conclusions

The occurrence of scrub typhus in Ganzhou City has demonstrated a decreasing trend in recent years, with a clear seasonal pattern in its occurrence. The model that exhibited the best predictive performance was the BPNN (3-9-1) model. Notably, during the development of the prediction model for scrub typhus, the three-year occurrence data of the new coronavirus epidemic was incorporated, which contributed to enhancing the prediction accuracy. By developing this prediction model, this study aimed to provide more scientific and rational recommendations for the prevention and control of scrub typhus in Ganzhou City. Additionally, it will assist decision-makers in the effective allocation of public health resources. Furthermore, timely and effective countermeasures can be implemented in response to potential epidemic peaks.

CRediT authorship contribution statement

Renfa Huang: Writing – original draft, Software, Resources, Methodology, Data curation, Conceptualization. **Kailun Pan:** Writing – original draft, Visualization, Software, Methodology, Conceptualization. **Qingfeng Cai:** Methodology, Investigation, Data curation. **Fen Lin:** Software, Methodology, Formal analysis. **Hua Xue:** Methodology, Investigation, Data curation. **Mingpeng Li:** Software, Methodology, Investigation, Data curation. **Yong Liao:** Writing – review & editing, Validation, Project administration, Methodology, Conceptualization.

Availability of data and materials

All relevant data are within the manuscript and its Supporting Information files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

Funding for this research was provided by the Science and Technology Program of Jiangxi Provincial Health and Wellness Commission (202312001).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idm.2025.02.009.

References

Bonell, A., Lubell, Y., Newton, P. N., Crump, J. A., & Paris, D. H. (2017). Estimating the burden of scrub typhus: A systematic review. *PLoS Neglected Tropical Diseases, 119*, Article e0005838. https://doi.org/10.1371/journal.pntd.0005838

Chen, B., Wang, M., Huang, X., Xie, M., Pan, L., Liu, H., Liu, Z., & Zhou, P. (2021). Changes in incidence of notifiable infectious diseases in China under the prevention and control measures of COVID-19. *Frontiers in Public Health*. , Article 9728768. https://doi.org/10.3389/fpubh.2021.728768

Fang, K., Cao, L., Fu, Z., & Li, W. (2023). Prediction of reported monthly incidence of hepatitis B in Hainan Province of China based on SARIMA-BPNN model. *Medicine (Baltimore), 10241*, Article e35054. https://doi.org/10.1097/md.0000000000035054

Geng, Y., & Zhang, L. (2022). Impact of non-pharmaceutical interventions during COVID-19 pandemic on pertussis, scarlet fever and hand-foot-mouth disease in China. *Journal of Infection, 842*, e13–e15. https://doi.org/10.1016/j.jinf.2021.12.023

Liao, Y., Huang, R. F., Hu, X. J., Guo, J., Huang, H. S., Li, J. H., Liu, X. H., & Xu, J. H. (2019). Epidemiological characteristics of scrub typhus in Jiangxi, 2006-2017. *Modern Preventive Medicine, 46*(7), 1167–1170.

Liao, Y., Li, R., Yang, J. P., & Huang, R. F. (2014). The epidemiological analysis of scrub typhus in Ganzhou from 2008 to 2012. *Chinese Journal of Disease Control & Prevention, 18*, 86–88.

Liu, Q., Li, Z., Ji, Y., Martinez, L., Zia, U. H., Javaid, A., Lu, W., & Wang, J. (2019). Forecasting the seasonality and trend of pulmonary tuberculosis in Jiangsu Province of China using advanced statistical time-series analyses. *Infection and Drug Resistance*, 122311–122322. https://doi.org/10.2147/idr.S207809

Ma, C. J., Oh, G. J., Kang, G. U., Lee, J. M., Lee, D. U., Nam, H. S., Ryu, S. Y., & Lee, Y. H. (2017). Differences in agricultural activities related to incidence of scrub typhus between Korea and Japan. *Epidemiology and health, 39*, Article e2017051. https://doi.org/10.4178/epih.e2017051

Orang, A., Berke, O., Poljak, Z., Greer, A. L., Rees, E. E., & Ng, V. (2024). Forecasting seasonal influenza activity in Canada-Comparing seasonal Auto-Regressive integrated moving average and artificial neural network approaches for public health preparedness. *Zoonoses and public health, 713*, 304–313. https://doi.org/10.1111/zph.13114

Pan, K., Huang, R., Xu, L., & Lin, F. (2024). Exploring the effects and interactions of meteorological factors on the incidence of scrub typhus in Ganzhou City, 2008-2021. *BMC Public Health, 241*, 36. https://doi.org/10.1186/s12889-023-17423-8

Paris, D. H., Shelite, T. R., Day, N. P., & Walker, D. H. (2013). Unresolved problems related to scrub typhus: A seriously neglected life-threatening disease. *The American Journal of Tropical Medicine and Hygiene, 892*, 301–307. https://doi.org/10.4269/ajtmh.13-0064

Peng, P. Y., Xu, L., Wang, G. X., He, W. Y., Yan, T. L., & Guo, X. G. (2022). Epidemiological characteristics and spatiotemporal patterns of scrub typhus in Yunnan Province from 2006 to 2017. *Scientific Reports, 121*, 2985. https://doi.org/10.1038/s41598-022-07082-x

Premaratna, R., Blanton, L. S., Samaraweera, D. N., de Silva, G. N., Chandrasena, N. T., Walker, D. H., & de Silva, H. J. (2017). Genotypic characterization of Orientia tsutsugamushi from patients in two geographical locations in Sri Lanka. *BMC Infectious Diseases, 171*, 67. https://doi.org/10.1186/s12879-016-2165-z

Ranjan, J., & Prakash, J. A. J. (2018). Scrub typhus re-emergence in India: Contributing factors and way forward. *Medical Hypotheses*, 11561–11564. https://doi.org/10.1016/j.mehy.2018.03.019

Saxena, A. (2021). Grey forecasting models based on internal optimization for Novel Corona virus (COVID-19). *Applied Soft Computing.* , Article 111107735. https://doi.org/10.1016/j.asoc.2021.107735

Shen, L., Sun, M., Song, S., Hu, Q., Wang, N., Ou, G., Guo, Z., Du, J., Shao, Z., Bai, Y., & Liu, K. (2022). The impact of anti-COVID-19 nonpharmaceutical interventions on hand, foot, and mouth disease-A spatiotemporal perspective in Xi'an, northwestern China. *Journal of Medical Virology, 947*, 3121–3132. https://doi.org/10.1002/jmv.27715

Sun, Y., Wei, Y. H., Yang, Y., Ma, Y., de Vlas, S. J., Yao, H. W., Huang, Y., Ma, M. J., Liu, K., Li, X. N., Li, X. L., Zhang, W. H., Fang, L. Q., Yang, Z. C., & Cao, W. C. (2017). Rapid increase of scrub typhus incidence in Guangzhou, southern China, 2006-2014. *BMC Infectious Diseases, 171*, 13. https://doi.org/10.1186/s12879-016-2153-3

Wang, Z., Wang, Y., Zhang, S., Wang, S., Xu, Z., & Feng, Z. (2024). Trend analysis and prediction of gonorrhea in mainland China based on a hybrid time series model. *BMC Infectious Diseases, 241*, 113. https://doi.org/10.1186/s12879-023-08969-4

Wang, Y., Xu, C., Wang, Z., & Yuan, J. (2019). Seasonality and trend prediction of scarlet fever incidence in mainland China from 2004 to 2018 using a hybrid SARIMA-NARX model. *PeerJ, 7*, Article e6165. https://doi.org/10.7717/peerj.6165

Wei, X. Y., Ou, L. L., Zhang, W. Y., & Sun, H. L. (2022). Progress of on epidemic characteristics, risk factors and prediction of tsutsugamushi disease in mainland China. *Acta Parasetologica et Medica Entomologica Sinica, 29*, 60–66.

Zhang, Y. Q., Li, X. X., Li, W. B., Jiang, J. G., Zhang, G. L., Zhuang, Y., Xu, J. Y., Shi, J., & Sun, D. Y. (2020). Analysis and predication of tuberculosis registration rates in Henan Province, China: An exponential smoothing model study. *Infectious diseases of poverty, 91*, 123. https://doi.org/10.1186/s40249-020-00742-y

Zhao, D. (2023). Research of combined ES-BP model in predicting Syphilis incidence 1982-2020 in mainland China. *Iranian Journal of Public Health, 5210*, 2063–2072. https://doi.org/10.18502/ijph.v52i10.13844

Zheng, W., Li, H., Yang, X., Wang, L., Shi, Y., Shan, H., He, L., Liu, J., Chen, H., Wang, G., Zhao, Y., & Han, C. (2023). Trends and prediction in the incidence rate of hepatitis C in Shandong Province in China from 2004 to 2030. *Preventive Medicine.* , Article 177107749. https://doi.org/10.1016/j.ypmed.2023.107749