Nota Lepi. 49 2026: 197–209 | DOI 10.3897/nl.49.204400 Research Article

Advanced *Dendrolimus punctatus* forecasting through Grey-Markov integration

Guangjing Qian[1]

1 *Department of Basic Sciences, Yancheng Institute of Technology, Yancheng 224003, China; 1451106787@qq.com*

https://zoobank.org/353B6B40-4663-470F-A30C-C356948E4E6F

Received 17 June 2026; accepted 17 August 2026; published: 12 October 2026
Subject Editor: Marcin Wiorek.

Abstract. *Dendrolimus punctatus* Walker (Lepidoptera: Lasiocampidae) is a major defoliator that causes significant ecological and economic damage to forests in China. Accurate forecasting of *D. punctatus* outbreaks is essential for effective pest management. This study develops hybrid models that integrate BP neural networks or stepwise regression with Grey and Markov chain models to improve the prediction accuracy of the peak period of *D. punctatus* for the first generation of eggs. Model performance was evaluated using the Mean Relative Error (MRE) and the Mean Square Error (MSE). The results indicate that the hybrid models, particularly the BP-Markov model, outperformed the single models except for the Regression-Markov model. The BP-Markov model yielded the lowest MRE (0.2422%) and MSE (0.0197), indicating high predictive accuracy. This study demonstrates that integrating different mathematical models can improve forecasting precision, suggesting potential utility for supporting proactive forest pest management. The proposed model can provide timely early warnings of potential *D. punctatus* outbreaks, thereby helping forest managers optimize monitoring schedules, determine appropriate intervention timing, and allocate pest-control resources more efficiently.

Introduction

Dendrolimus punctatus Walker, 1855 has been one of the most pervasive forest defoliators in China since the 20th century, causing severe ecological and economic damage from the Qinling Mountains to the south of the Huaihe River (Xiao 1992; Zhang 2006; Fu 2020). Its high adaptability and fecundity lead to frequent large-scale outbreaks that significantly impair tree growth and increase forest susceptibility to secondary pests and diseases (Zhang and Li 2008; Wang et al. 2024). Furthermore, these outbreaks pose considerable challenges to forestry production and public health, particularly for individuals engaged in forest management (Zou and Wang 1989; Zou et al. 1990). Within its life cycle, the egg peak period represents the onset of a pest population outbreak and consequently offers a critical window for control. Focusing on this stage is more advantageous than targeting larvae or adults, as it facilitates easier monitoring and enables the deployment of highly effective, eco-friendly control tactics (Yu et al. 2019). Given that the population dynamics of *D. punctatus* are highly dependent on environmental conditions, the development of accurate forecasting systems is crucial for implementing timely and effective pest control measures (Zhang and Li 2008; Zhao 2018).

Copyright *Qian.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The challenge of predicting pest outbreaks is analogous to that of other major forest pests worldwide. For instance, the Gypsy moth (*Lymantria dispar* Linnaeus, 1758) in North America (Tobin et al. 2007) and the Spruce Budworm (*Choristoneura fumiferana* Clemens, 1865) in boreal forests exhibit similar outbreak patterns driven by climatic variables, host availability, and population feedback mechanisms (Elkinton and Liebhold 1990; Royama et al. 2005). Numerous modeling approaches, including spatiotemporal models, population dynamics' frameworks, and machine learning methods, have been employed to forecast these outbreaks, demonstrating the value of integrating ecological knowledge with advanced computational techniques (Royama et al. 2005; Régnière et al. 2012). These studies provide a valuable context for improving prediction methods for *D. punctatus*.

A variety of mathematical models have been applied to forecast *D. punctatus* dynamics. For example, Xu et al. (2014) compared Fisher discriminant analysis and random forest models to identify pest damage patterns, while Cheng et al. (2019) used fuzzy comprehensive evaluation to predict outbreak peak periods. Qian et al. (2021) focused on BP neural network, Contingency table analysis and Fuzzy comprehensive evaluation method to predict the peak period of *D. punctatus* for the first generation egg. Although these approaches have improved outbreak prediction, existing studies have generally evaluated individual models separately or relied primarily on the output of a single forecasting method. This is an important limitation because *D. punctatus* population dynamics are simultaneously characterized by nonlinear responses to environmental variables, short and uncertain monitoring records, temporal fluctuations, and transitions among different population states. A single model may therefore capture only part of the information contained in the data, resulting in reduced predictive accuracy.

To overcome these limitations, this study develops a hybrid forecasting framework that combines the complementary strengths of BP neural networks or stepwise regression with Grey and Markov chain models. Consequently, the hybrid framework is expected to represent nonlinear environmental effects, general temporal trends, and state-dependent fluctuations more comprehensively than any individual method. The objectives of this study are: (1) to develop and validate BP neural network and stepwise regression models for predicting the peak period of *D. punctatus* for the first generation egg; (2) to integrate these with Grey and Markov chain models to construct hybrid forecasting systems; (3) to evaluate the performance of all models using the Mean Relative Error (MRE) and the Mean Square Error (MSE). The analyses presented in this study provide a more reliable forecasting tool for *D. punctatus* management.

Methods

The life cycle of *Dendrolimus punctatus*

The life cycle of *Dendrolimus punctatus* comprises four stages: egg, larva, pupa, and adult. The oval, pinkish eggs are typically laid in clusters on pine needles and hatch within 10–14 days. The larvae, measuring 60–80 mm in length with dark grey bodies and orange-red patterns, exhibit gregarious behavior and tend to descend on silk threads when disturbed. This larval stage typically lasts 20–30 days, during which they molt 5–6 times. The pupae, dark brown and covered in yellowish hairs, are encased in grayish cocoons and undergo metamorphosis over 11–22 days. Adults have a body length of 20–30 mm and variable coloration. Females have short, comb-like antennae, while males have feathery antennae. Active at night, they mate and lay eggs on pine needles, with a

lifespan of about 5–9 days. The entire life cycle is influenced by climate and host plant conditions, with varying generations per year depending on the region-two generations in Henan, three to four in Guangdong, and two to three in other provinces.

Data source

The Forest Pest and Disease Management Station (FPDMS) of the National Forestry and Grassland Administration (NFGA) and the Forest Pest Control Station of Qianshan County, Anhui Province, provided the data on the incidence of *D. punctatus*. The China Meteorological Administration (CMA) provided the meteorological information. The data were collected between 1983 and 2018.

Field investigations were conducted along highways, railroads, and other lines as well as forest roads. Twenty standard plants were also selected using a parallel sampling method for detailed investigation. The damage severity and insect occurrence were then visually identified and recorded within demarcated standard-sized areas. Plant samples were collected for laboratory examination. After the peak emergence of the female moths, the egg stage was examined, revealing an average of one to two eggs per quadrat. The number of withered and curled branches was counted for the first- and second-instar larvae in order to study the larval stage. Large trees were examined using the "Insect dung particle estimation algorithm", whereas larvae above the third instar or small trees under 3 meters in height were physically examined to count the larvae on the crown. After the overwintering period of larvae, the overall population size was calculated based on the quantity of larvae found in the bark slits at the trunk's base. A black light trap was used to attract the adults two to three days before and throughout the emergence phase. This method leverages the insects' phototaxis, particularly their attraction to ultraviolet light emitted by the black light. A count of the entire population was conducted, taking into account the male-to-female ratio. Data on the cumulative period of *D. punctatus* provided insights into its life cycle, including the start, peak, and end of each key stage.

No wild plants were collected or used in this study. All research was focused on the modeling and forecasting of *D. punctatus*.

Grading standard

To facilitate data analysis, calendar dates in April were converted to a numerical sequence, where April 1st = 1, April 2nd = 2, ..., April 30th = 30, May 1st = 31, ..., and so on. This method enabled a continuous numerical scale for statistical calculations.

The peak period y for the first generation egg of *D. punctatus* was classified into four levels to provide actionable thresholds for timely pest management decisions according to the following criteria: level 1, y<56 d; level 2, 56 d≤y<61 d; level 3, 61 d≤y<66 d; level 4, y≥66 d. This transformation provides a continuous response variable for regression and neural-network analyses. These values were subsequently classified into four ordinal levels for Markov-state analysis and operational interpretation.

The peak period y for the first generation egg of *D. punctatus* was classified into four levels based on historical outbreak data and forestry pest control practices in Qianshan County: Level 1 (y < 56 d, corresponding to before May 26th); Level 2 (56 d ≤ y < 61 d, May 26th–May 30th); Level 3 (61 d ≤ y < 66 d, May 31st - June 4th); Level 4 (y ≥ 66 d, June 5th or later). These thresholds were not selected arbitrarily; rather, they were determined from historical monitoring and outbreak records of *D. punctatus* in Qianshan County and were aligned with the timing windows commonly used in

local forest-pest monitoring and control practices. The classification distinguishes relatively early, moderately early, intermediate, and relatively late egg-peak periods, thereby allowing model outputs to be translated into practical monitoring and intervention schedules.

Mathematical analysis method

Stepwise regression model

Regression analysis is a statistical method for examining relationships among quantitative variables, plays a vital role in modern applied statistics (You and Yan 2017). It is commonly used to develop models for predicting pest dynamics and is one of the most prevalent methods in current forecasting practices. The essence of the stepwise regression analysis method is to establish the optimal multiple linear regression equation, introduce variables into the model one by one, and test the explanatory variables that have been selected respectively. When variables are not significant due to the introduction of new explanatory variables, they should be deleted to ensure that only significant variables are included in the regression equation. In this paper, a forecast model is proposed based on the ecological factors affecting the growth and development of *D. punctatus*. The prediction formula is $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$. y is the predicted variable, x_i are the significant independent variables, and the regression coefficient b_i is estimated by the least-squares method.

BP neural network model

Neural networks, inspired by the human brain, are powerful tools in artificial intelligence for learning complex patterns and making predictions. Among various neural network models, the Backpropagation (BP) neural network is particularly noteworthy for its ability to adjust weights based on error rates, making it highly effective for training multilayer feedforward networks (Wang 2020). This method is crucial for tasks such as pest management, where accurate predictions can significantly enhance decision-making processes. The BP neural network consists of three main layers: the input layer, hidden layer(s), and output layer. The input layer receives raw data (e.g., environmental variables), and passes it to the hidden layer(s) (e.g., pest population levels). The hidden layer(s) perform complex computations on these data, with the number of layers and nodes depending on task complexity.

The BP neural network operates through two distinct computational phases: (1) forward propagation, where input signals are sequentially processed through the hidden layer to the output layer; (2) backward propagation, where output errors are recursively transmitted backward through the network to adjust connection weights. The BP neural network has an M-H-N structure (Fig. 1), comprising M input nodes in the input layer (x), H nodes in the hidden layer (y), and N output nodes in the output layer (z). The weight between the input layer and the hidden layer is $w_{jk}^{(1)}$. The weight between the hidden layer and the output layer is $w_{jk}^{(2)}$. The number of nodes in the input layer of the network is the number of system characteristic factors (independent variables), and that in the output layer is the number of target forecasted factors. The hidden layer contains a number of nodes equal to roughly 75% of the input nodes (Hornik et al. 1989).

x_i: the i-th node in the input layer;
w_{ij}: the weight of the j-th node in the hidden layer to the i-th node in the input layer;
z_j: the threshold of the j-th node in the hidden layer;
o_k: the activation function of the output layer;
z_k: the k-th node in the output layer.

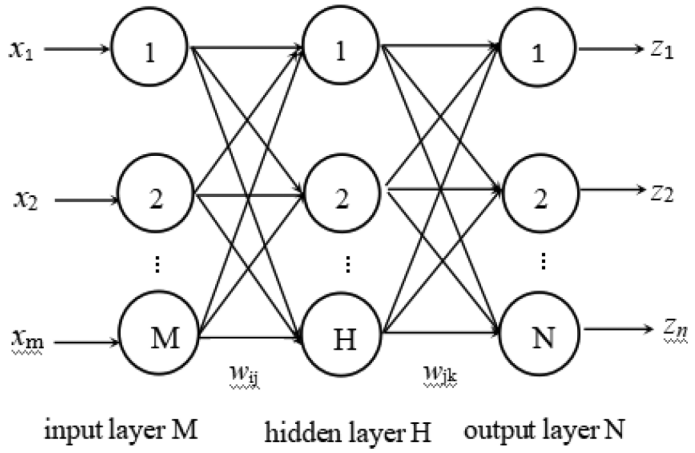

Figure 1. Structure of BP neural network.

Markov chain model

Markov chains are simple yet powerful tools for modeling state transitions in probabilistic systems (Jia 2006; Zou 2020). They rely on state transition probabilities and exhibit a "memoryless" property, where future states depend only on the current state (Cui 2016; Bai et al. 2017). This makes them suitable for modeling population dynamics in ecology and disease spread in epidemiology, where long-term behavior can be predicted based on current conditions (Anderson and May 1991).

In this study, Markov chains were used to model transitions between occurrence levels of *D. punctatus*. The occurrence state of *D. punctatus* has undergone *m* transitions, and the *m*-order transition probability matrix is:

$$p(m) = \begin{vmatrix} P_{11} & P_{12} & \cdots & P_{1k} \\ P_{21} & P_{22} & \cdots & P_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ P_{\mathrm{n1}} & P_{\mathrm{n2}} & \cdots & P_{nk} \end{vmatrix}$$

where $P^{(1)}$ is the one-step transition probability matrix of all levels between consecutive years. Similarly, $P^{(2)}$ is 2 years apart, and $P^{(3)}$ is 3 years apart. P_{l} is the probability of transition from level "1" to level "1" in the next year. The occurrence level for the forecasted year is determined by identifying the level associated with the maximum sum of transition probabilities.

Grey model

The Grey model is specifically designed to work with limited and incomplete data, making it highly adaptable for scenarios where information is scarce. By constructing a grey differential prediction model, it provides a long-term, albeit somewhat fuzzy, description of how systems evolve over time (Jiao et al. 2020; Wang et al. 2021). This approach begins with an original data sequence, referred to as sequence (0): $x^{(0)} = \left(x_{(1)}^{(0)}, x_{(2)}^{(0)}, \ldots, x_{(\mathrm{n})}^{(0)}\right)$, which is then transformed into sequence (1): $x^{(1)} = \left(x_{(1)}^{(1)}, x_{(2)}^{(1)}, \ldots, x_{(n)}^{(1)}\right)$ through a cumulative generation process. This transfor-

mation helps to reduce the inherent randomness in the original data, thereby revealing more pronounced patterns and trends. Based on this transformed sequence (1), a differential equation model $x_{(k)}^{(0)} + a\left(x_{(k)}^{(1)} + x_{(k-1)}^{(1)}\right)/2 = b$, specifically a GM model, is established. The Grey model, which incorporates new data while discarding old data, is particularly effective for handling dynamic systems with random disturbances (Forestry 2020; Xu et al. 2020).

Grey models have been successfully applied in a variety of contexts where data are limited or incomplete. For example, in environmental science, they have been used to predict the impacts of climate change on ecosystems, offering valuable insights despite the limited availability of historical data (Wang et al. 2021). In economics, these models have been used to forecast market trends and economic indicators, aiding policymakers in making informed decisions (Li et al. 2022). Their adaptability and robustness make them well-suited for predicting pest population trends.

Hybrid models

To address the limitations of single models and harness the complementary strengths of multiple models (Zhang et al. 2021; Cao and Qin 2022; Li et al. 2022), we developed a hybrid predictive framework that integrates the BP neural network model, stepwise regression model, Grey model, and Markov chain model. For clarity, hybrid-model names are expressed as "base model-correction model." Thus, BP-Grey, BP-Regression, and BP-Markov denote BP neural network models corrected using the Grey, stepwise regression, and Markov chain models, respectively. Similarly, Regression-Grey, Regression-BP, and Regression-Markov denote stepwise regression models corrected using the Grey, BP neural network, and Markov chain models, respectively.

This framework aims to enhance predictive accuracy and robustness by leveraging the unique capabilities of each model while mitigating their individual weaknesses. Our hybrid framework involves the following steps:

Initial prediction-model construction:

- Data Input: Utilize the actual data of *D. punctatus* to construct the initial prediction model.
- Model Selection: Choose the BP neural network model as the initial prediction tool due to its ability to capture complex nonlinear relationships in the data.
- Prediction Generation: Derive the predicted values using the BP neural network model.

Relative-residual sequence generation:

- Residual Calculation: Compare the predicted values from the initial model with the original observed values to generate a sequence of relative residuals. This sequence quantifies the discrepancy between the model predictions and the actual data.

Residual-level division:

- Level Categorization: Divide the relative residuals into distinct levels (seven levels in this study) based on their magnitude. Higher division values indicate more precise adjustments needed in the model.
- Level Assignment: Assign each residual to a specific level based on its magnitude, facilitating targeted refinement of the initial predictions.

Prediction-model refinement:

- Secondary Model Development: Develop a secondary prediction model using the stepwise regression model, Grey model, and Markov chain model.
- Model Refinement: Integrate the predictions from the secondary models to refine the initial predictions based on the categorized levels of residuals.
- Final Prediction: Combine the refined predictions from the secondary models to produce the final prediction values, which are expected to be more accurate and robust.

Results

Based on the four models (BP neural network, stepwise regression, Grey model, and Markov chain) described above, we predicted the peak period of *D. punctatus* for the first generation egg in Qianshan, Anhui Province, China. The data used for this study span from 1983 to 2016, and are sourced from the State Forestry Administration Forest Pest Control Station (Suppl. material 1: fig. S1). For the year 1998, an extreme drought event led to the near-extinction of *D. punctatus* population in Qianshan County, resulting in missing data for y and x_1–x_4. The final training sample consisted of 33 valid years (1983–1997, 1999–2016), which was verified for model training (Suppl. material 1: table S1).

In Qianshan County, *D. punctatus* typically undergoes three generations annually: the overwintering generation (January to June), the first generation (mid-May to mid-August), and the second generation (mid-August to December). We computed the Pearson correlation coefficients r and conducted statistical significance testing (p<0.01). The correlation coefficients between the peak period for the first generation egg (y) and four independent variables were calculated: the beginning period for the overwintering pupa (x_1, r=0.9897), the peak period for the overwintering pupa (x_2, r=0.9611), and the beginning period for the overwintering adults (x_3, r=0.9765), the peak period for the overwintering adults (x_4, r = 0.9809) (Fig. 2). When df = 31, r values exceeded the critical value ($r_{0.01}$(0.442)), which indicated that each x_i is highly correlated with y.

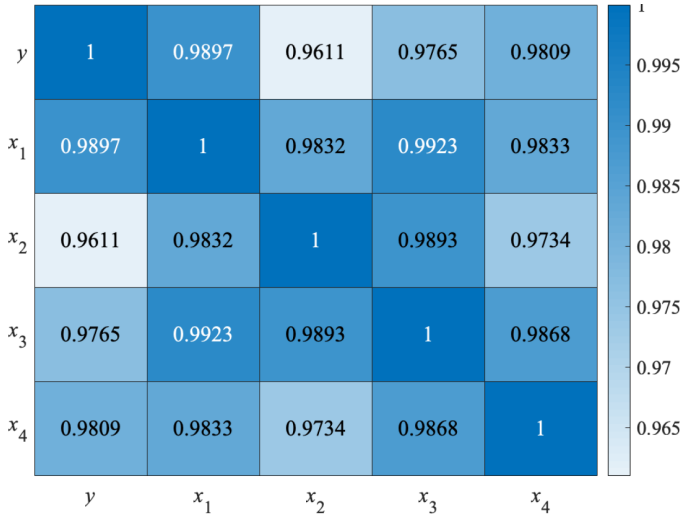

Figure 2. Heatmap of Pearson Correlation Coefficients.

Prediction for BP neural network model

We constructed a BP neural network with a 4-4-1 architecture (input-hidden-output nodes) using variables x_1 to x_4 as inputs to predict the peak period (y) (Suppl. material 1: table S2). The model was trained using observations from 1983–2016 and independently evaluated using data from 2017 and 2018. A prediction was considered acceptable when its absolute error was less than 4 days, based on the operational pest-control window used in Qianshan County. Both validation predictions satisfied this criterion.

Prediction for stepwise regression model

Based on the training data (1983–2016) and independent validation data (2017–2018), the stepwise regression model was formulated as follows: $y = 21.6075 + 1.6891x_1 - 0.6306x_3 + 0.3919x_4$. The correlation coefficient for this model is 0.9872. After substituting the values of x_1, x_3 and x_4, the difference between the predicted and actual values was calculated. All training predictions fell within the predefined 4-day tolerance. For the predictions in 2017 and 2018, the predicted values are 60.4058 and 55.4421, respectively. The difference from the actual value is less than 4 days.

Model performance comparison

We evaluated the predictive performance of hybrid models using the Mean Relative Error (MRE) and the Mean Square Error (MSE) as criteria. The BP model displays a lower Mean Relative Error of 1.5323%, while the regression analysis model has a higher Mean Relative Error of 5.3686%. In terms of Mean Square Error, the BP model has a value of 0.7848, whereas the regression analysis model has a significantly higher Mean Square Error of 9.2890. This superior baseline performance of the BP model underscores its capability to capture the nonlinear relationships in the ecological data.

When the BP neural network model is refined using the Grey, Regression, and Markov chain models (Table 1), the BP-Markov model demonstrated the lowest error rates (MRE = 0.2422%, MSE = 0.0197), indicating substantially lower prediction errors than those of the standalone BP model within the validation dataset. When the stepwise regression model was used as the base predictor, its integration with the Grey model (Regression-Grey) yielded the best results within

Table 1. The prediction results for the combination of BP neural network model, regression analysis model, Grey model, Markov chain model.

Year	Peak period	BP neural network model		BP-Grey model	
		Predicted values	Relative errors(%)	Predicted values	Relative errors(%)
2017	63.0000	62.4052	–0.9441	62.1500	–1.3492
2018	52.0000	53.1026	2.1204	52.8855	1.7028
	MRE(%)	1.5323		1.5259	
	MSE	0.7848		0.7533	
Year	**Peak period**	**BP-Regression model**		**BP-Markov model**	
		Predicted values	**Relative errors(%)**	**Predicted values**	**Relative errors(%)**
2017	63.0000	62.4455	–0.880	62.8439	–0.2478
2018	52.0000	52.9904	1.9045	51.8770	–0.2365
	MRE(%)	1.3924		0.2422	
	MSE	0.6442		0.0197	

that category (MRE = 5.3551%, MSE = 9.2244). These results suggest that most hybrid models improved predictive performance.

A similar trend was observed when the Stepwise Regression served as the base model (Table 2). The results show that the Regression-Grey model (MRE = 5.3551%, MSE = 9.2244) and the Regression-BP neural network model (MRE = 5.3679%, MSE = 9.2848) exhibit lower errors than the Regression-Markov model, indicating their superior prediction accuracy. This underscores the critical importance of selecting a robust base model, with the BP network providing a stronger foundation due to its nonlinear modeling capability.

In summary, the key findings from the hybrid model analysis are twofold: First, all hybrid models outperformed the single models with the exception of the Regression–Markov model. Second, among all tested combinations, the BP-Markov model achieves the optimal balance of precision and robustness, making it the recommended approach for forecasting the peak period of *D. punctatus* (Fig. 3).

Table 2. The prediction results for the combination of regression model, Grey model, BP neural network model and Markov chain model.

Year	Peak period	Regression model		Regression-Grey model	
		Predicted values	Relative errors(%)	Predicted values	Relative errors(%)
2017	63.0000	60.4058	–4.1178	60.2547	–4.3575
2018	52.0000	55.4421	6.6194	55.3034	6.3727
	MRE(%)	5.3686		5.3551	
	MSE	9.2890		9.2244	
Year	**Peak period**	**Regression-BP model**		**Regression-Markov model**	
		Predicted values	**Relative errors(%)**	**Predicted values**	**Relative errors(%)**
2017	63.0000	60.3988	–4.1289	60.8192	–3.4616
2018	52.0000	55.4356	6.6069	55.8214	7.3488
	MRE(%)	5.3679		5.4052	
	MSE	9.2848		9.6795	

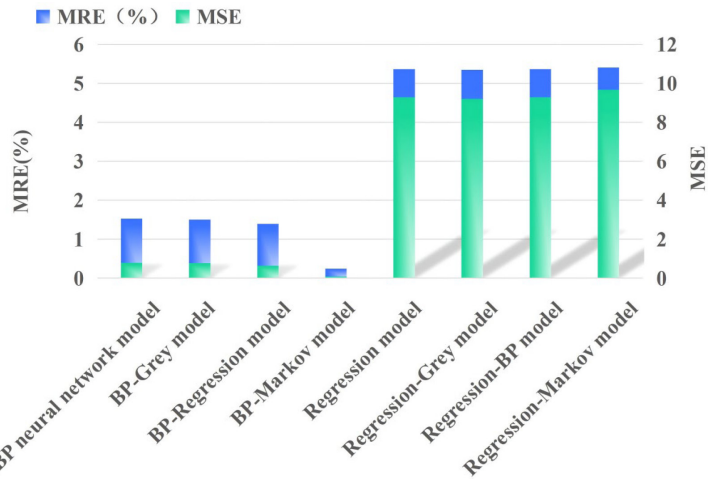

Figure 3. Comparison of the Mean Relative Error (MRE, left y-axis) and Mean Square Error (MSE, right y-axis) among the individual and hybrid forecasting models.

Discussion

Our study demonstrates that a hybrid modeling framework integrating BP neural networks with Markov models improves the prediction accuracy of the peak period for the first generation egg of *D. punctatus*. Comparative analysis showed that the BP-Markov hybrid model produced the lowest Mean Relative Error and Mean Square Error, suggesting the optimal predictive performance within our dataset. This performance may reflect the complementary functions of the BP neural network and Markov components.

The BP neural network excels at capturing complex, nonlinear relationships from multi-dimensional data (overwintering pupa and adult beginning/peak periods) (Xu 2002). However, standalone BP models may be sensitive to annual fluctuations in the input data, as illustrated by the 2016 prediction where the BP neural network model's error (–1.0553 days) produced a larger absolute prediction error than most other training years (Suppl. material 1: table S2). This limitation can be mitigated by the Markov model, which classifies the BP residuals into discrete states and uses historical state-transition probabilities to estimate the most likely residual state in the forecast year. Rather than smoothing annual fluctuations, the Markov model provides state-dependent corrections to the BP predictions and may therefore better account for temporal shifts in prediction errors. In the present validation dataset, integrating the Markov model reduced the MSE from 0.7848 to 0.0197 and the MRE from 1.5323% to 0.2422% (Table 1).

Our findings align with and extend the work of Fei et al. (2014) and Wang et al. (2016), but mark a distinct advance. Their studies primarily focused on optimizing single models like random forest or fuzzy evaluation. In contrast, the BP-Markov model combines nonlinear relationship capture with state-transition modeling, making it better suited to account for temporal dependence and discrete changes in *D. punctatus* population dynamics. Hybrid models have also been applied to other forest pests, such as *Hyphantria cunea* Drury, 1773 (Shi et al. 2024) and *Sphaerolecanium prunastri* Fonscolombe, 1834 (Guo et al. 2024). The success of hybrid models across different pest systems underscores a broader principle: ecological complexity often necessitates methodological pluralism. However, similarities in model performance among different pest systems should not be regarded as direct evidence that the present model is readily generalizable. Pest species differ in their life histories, environmental sensitivities, monitoring systems, and outbreak mechanisms; consequently, species- and region-specific validation remains necessary.

Although the BP-Markov model performed best, the comparative analysis of the BP-Grey model provided valuable insights. Its inferior performance is likely attributable to the smoothing characteristics of Grey modeling. Although this smoothing can reduce random fluctuations in the BP residual sequence, it may also weaken short-term changes that are important for accurately correcting annual prediction errors. In contrast, the Markov model explicitly accounted for transitions among residual states and generated more appropriate state-dependent corrections, resulting in predicted values of 62.8439 and 51.8770 days for 2017 and 2018, respectively. These results suggest that, within the present dataset, the temporal state-transition structure of the BP residuals contained more useful predictive information than the general trend captured by the Grey model, thereby leading to the substantially lower MSE of the BP–Markov model (Table 1).

Several limitations of this study should be acknowledged. Model development and validation were based on a limited number of independent observations. With a small validation dataset, performance estimates may be strongly influenced by individual observations. Furthermore, all

observations were collected exclusively in Qianshan County; consequently, the model may primarily reflect local climatic conditions, forest characteristics, pest population dynamics, thereby limiting its transferability to other regions. Several potentially important ecological drivers, including soil moisture and natural-enemy abundance, were also not incorporated into the current model. Therefore, the BP-Markov model should be regarded as a promising forecasting tool whose robustness requires further evaluation using longer time series, independent temporal and regional datasets, and a broader range of biological and environmental predictors before its widespread application in the early warning and management of *D. punctatus*.

Conclusion

This study demonstrates that hybrid forecasting models integrating a BP neural network with a Markov model improve the prediction of peak period of *D. punctatus*. Among all tested models, the BP-Markov hybrid achieved the best accuracy, confirming the value of methodological integration in ecological forecasting. Although the framework shows potential for pest forecasting, its transferability to other regions and pest species remains to be established through independent validation. Future research should focus on expanding the geographical scope, incorporating additional ecological variables (e.g., soil moisture, natural enemies), and developing adaptive, real-time forecasting platforms.

References

Anderson RM, May RM (1991) Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford. https://doi.org/10.1093/oso/9780198545996.001.0001

Bai Y, Shu K, Jiang H (2017) Application of mathematical analysis model in marketing of medical institutions. Chinese Market 28: 104–106.

Cao H, Qin JT (2022) Research on freight volume forecast based on ARIMA-BP combined model. Software Guide 21: 32–36.

Cheng X, Zhang SP, Yu Y, Bi SD, Zhou XZ, Zou YD, Zhang GQ, Zhang Z, Fang GF, Song YS (2019) Prediction of the peak occurrence of *Dendrolimus punctatus* larvae by fuzzy comprehensive evaluation method. Plant Protection 45: 116–121.

Cui LP (2016) Research on Markov chain theory and its application in the field of economic management. Chinese Market 11: 48[+50].

Elkinton JS, Liebhold AM (1990) Population dynamics of gypsy moth in North America. Annual Review of Entomology 35: 571–596. https://doi.org/10.1146/annurev.en.35.010190.003035

Fei HZ, Wang HB, Kong XB, Zhang Z, Zhang SF, Song XG (2014) Selection and prediction of meteorological factors correlated with *Dendrolimus punctatus* out-break. Journal of Northeast Forestry University 41: 136–140.

Forestry (2020) Researchers from Fuzhou University report new studies and findings in the area of forestry (BP neural networks and random forest models to detect damage by *Dendrolimus punctatus* Walker). Agriculture Week.

Fu ZA (2020) Biological characteristics and prevention of *Dendrolimus punctatus* Walker. Green Technology 21: 169–170[+173].

Guo FC, Ma GL, Gao GZ (2024) Prediction of potential habitat areas for the invasive pest *Sphaerolecanium prunastri* based on Biomod2 combined model. Forest Research 37: 136–147.

Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural Networks 2: 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Jia CS (2006) Forecast of occurrence level of *Dendrolimus punctatus* by Markov Chains. Journal of Northeast Forestry University 34: 21–22.

Jiao YQ, Feng J, Yang LL, Zhao GS, Fan HJ (2020) Research on prediction for pine growth of grey model GM(1,1). Practice and Understanding of Mathematics 50: 83–88.

Li HD, Liu YF, Li FM (2022) Prediction of the number of newborns based on combined model. Journal of Shandong University of Technology 36: 45–52.

Qian GJ, Song XY, Zhang SP, Zhou XZ, Zou YD, Bi SD (2021) The three models for forecasting the peak period of *Dendrolimus punctatus* for the first generation egg. International Journal of Pest Management 68(1): 1–7. https://doi.org/10.1080/09670874.2020.1784495

Régnière J, St-Amant R, Duval P (2012) Predicting insect distributions under climate change from physiological responses: spruce budworm as an example. Biological Invasions 14: 1571–1586. https://doi.org/10.1007/s10530-010-9918-1

Royama T, MacKinnon WE, Kettela EG (2005) Analysis of spruce budworm outbreak cycles in New Brunswick, Canada, since 1952. Ecology 86: 1212–1224. https://doi.org/10.1890/03-4077

Shi CC, Zhou Y, Wang KJ, Wang C, Li HP (2024) Prediction of forest insect pests based on the WOA-BiLSTM-BA algorithm. China Agricultural Machinery Chemistry Journal 45: 221–227.

Tobin PC, Whitmire SL, Johnson DM, Bjørnstad ON, Liebhold AM (2007) Invasion speed is affected by geographical variation in the strength of Allee effects. Ecology Letters 10: 36–43. https://doi.org/10.1111/j.1461-0248.2006.00991.x

Wang Q (2020) Application of BP neural network model in data analysis. Electronic Technology and Software Engineering 12: 189–190.

Wang Q, Bi M, Du T (2016) Spatial pattern of *Dendrolimus punctatus* incidence rate based on meteorological factors. Forest Research 29: 256–260.

Wang FH, Zhou XY, Song J (2021) Research on traffic volume prediction based on grey model and BP neural network. Engineering and Construction 35: 429–432.

Wang YJ, Zhao YJ, Miao GT, Zhou XT, Yu CJ, Cao Y (2024) Predicting the potential distribution of *Dendrolimus punctatus* and its host *Pinus massoniana* in China under climate change conditions. Frontiers in Plant Science 15: 1362020. https://doi.org/10.3389/fpls.2024.1362020

Xiao GR (1992) Forest Insects of China (2nd edn.). Chinese Academy of Forestry, Beijing, 948–953.

Xu KX (2002) Biological mathematics. Science Press, Beijing, 40–50.

Xu ZH, Li CH, Liu J (2014) Fisher discriminant analysis of *Dendrolimus punctatus* Walker pest levels. Transactions of the Chinese Society for Agricultural Machinery 45: 401–406.

Xu ZH, Huang XY, Lin L, Yu KY, Liu J, Xie SJ, Li XP, Chen FH, Qi XL, Chen GR, Li ZL (2020) BP neural networks and random forest models to detect damage by *Dendrolimus punctatus* Walker. Journal of Forestry Research 31: 107–121. https://doi.org/10.1007/s11676-018-0832-1

You SB, Yan Y (2017) Stepwise regression analysis and its application. Statistics and Decision 14: 31–35.

Yu Y, Zhang SP, Bi SD, Zhou XZ, Zou YD, Zhang GQ, Zhang Z, Fang GF, Song YS (2019) Analysis of variance periodic extrapolation prediction of the damage index of *Dendrolimus punctatus* larvae. Plant Protection 45: 125–133.

Zhang XX (2006) Insect Ecology and Forecast (3rd edn.). China Agriculture Press, Beijing, 2015–221.

Zhang Z, Li DM (2008) Approach to outbreak mechanism of *Dendrolimus punctatus* (Lepidoptera: Lasiocampidae). Scientia Silvae Sinicae 44: 140–150.

Zhang XH, Ma JJ, Xu L, Zhang GD, Tian X (2021) Research on forecasting GDP of Zhejiang Province based on a combined model. Technology and Industry 21: 370–376.

Zhao YJ (2018) Analysis on the integrated prevention of *Dendrolimus punctatus* Walker. Flower 14: 275–276.

Zou LQ (2020) Markov chain model and its applications. Science and Technology Innovation Herald 17: 97–98.

Zou YD, Wang HF (1989) Insect Ecology of Agriculture and Forestry. Anhui Science and Technology Press, Hefei, 462–469.

Zou YD, Cheng FJ, Zha GJ (1990) Effect of the chemical components of pine needle on the existence and growth of pine caterpillars *Dendrolimus punctatus*. Scientia Silvae Sinicae 26: 142–148.

Additional information

Conflict of interest

The author has declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Artificial Intelligence (AI) use

The author accept full responsibility for the content of the manuscript, including the disclosure of any use of AI.
No AI tools were used in the preparation of this manuscript.

Funding

No funding was reported.

Author contributions

Guangjing Qian conceived this study, and further performed numerical simulations and analysis. The author contributed to the discussion of the results and the final text.

Data availability

All the required data are uploaded as supplementary material.

Supplementary material 1

Supporting information

Author: Guangjing Qian

Data type: docx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/nl.49.204400.suppl1