

RESEARCH ARTICLE

Integrating dynamic models and neural networks to discover the mechanism of meteorological factors on Aedes population

Mengze Zhang¹, Xia Wang^{1*}, Sanyi Tang²

1 School of Mathematics and Statistics, Shaanxi Normal University, Xi'an, People's Republic of China,
2 School of Mathematics and Statistics, Shanxi University, Taiyuan, People's Republic of China

* xiawang@snnu.edu.cn



OPEN ACCESS

Citation: Zhang M, Wang X, Tang S (2024) Integrating dynamic models and neural networks to discover the mechanism of meteorological factors on Aedes population. PLoS Comput Biol 20(9): e1012499. <https://doi.org/10.1371/journal.pcbi.1012499>

Editor: Samuel V. Scarpino, Northeastern University, UNITED STATES OF AMERICA

Received: May 13, 2024

Accepted: September 18, 2024

Published: September 27, 2024

Copyright: © 2024 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The code and data used are available at: <https://github.com/Zhangmze/Dynamic-Network-Aedes>.

Funding: XW was funded by National Natural Science Foundation of China (12171295). ST was funded by National Natural Science Foundation of China (12031010). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare no competing interests.

Abstract

Aedes mosquitoes, known as vectors of mosquito-borne diseases, pose significant risks to public health and safety. Modeling the population dynamics of Aedes mosquitoes requires comprehensive approaches due to the complex interplay between biological mechanisms and environmental factors. This study developed a model that couples differential equations with a neural network to simulate the dynamics of mosquito population, and explore the relationships between oviposition rate, temperature, and precipitation. Data from nine cities in Guangdong Province spanning four years were used for model training and parameter estimation, while data from the remaining three cities were reserved for model validation. The trained model successfully simulated the mosquito population dynamics across all twelve cities using the same set of parameters. Correlation coefficients between simulated results and observed data exceeded 0.7 across all cities, with some cities surpassing 0.85, demonstrating high model performance. The coupled neural network in the model effectively revealed the relationships among oviposition rate, temperature, and precipitation, aligning with biological patterns. Furthermore, symbolic regression was used to identify the optimal functional expression for these relationships. By integrating the traditional dynamic model with machine learning, our model can adhere to specific biological mechanisms while extracting patterns from data, thus enhancing its interpretability in biology. Our approach provides both accurate modeling and an avenue for uncovering potential unknown biological mechanisms. Our conclusions can provide valuable insights into designing strategies for controlling mosquito-borne diseases and developing related prediction and early warning systems.

Author summary

When simulating population dynamics, scholars often rely on predefined functions to describe processes such as growth, development, and mortality. However, these functions may not be able to accurately capture the underlying biological mechanisms, posing challenges for related research. In this study, we adopted a data-driven approach integrated

with mechanistic principles to simulate mosquito dynamics. Given the unknown functional relationships between oviposition rate, temperature, and precipitation, we employed a neural network to model these relationships, resulting in a model that combines the neural network with traditional mechanistic approaches. This method can successfully simulate the changing trends of mosquitoes, infer unknown parameters, train the neural network, and capture the relationships between oviposition rate, temperature, and precipitation. Data from nine cities in Guangdong Province were used for model training, while data from the remaining three cities were used for model validation. Our model successfully fitted data from all cities using the same set of parameters, demonstrating satisfactory performance. On this basis, we successfully identified the functional form between oviposition rate, temperature, and precipitation through the symbolic regression algorithm, which was consistent with the results obtained from the neural network. These findings show that our method can accurately simulate population dynamics while exploring unknown biological mechanisms.

Introduction

Vector-borne diseases pose significant threats to human and animal populations, with mosquito-borne diseases being one of the primary causes of mortality in humans for centuries [1]. Aedes mosquitoes, as one of the major threats to severe outbreaks of mosquito-borne viruses, have been proven to be an effective vector of multiple arboviruses, such as dengue virus, chikungunya virus, yellow fever virus, and so on [2]. The behavior and life cycles of Aedes mosquitoes are highly sensitive to meteorological factors, such as temperature and precipitation [3–5]. Thus, investigating the impact of meteorological factors on the population dynamics of Aedes mosquitoes has always been a focus of attention for researchers. The existing research can be roughly divided into three main categories: biological experiment research, statistical model research, and mechanistic model research.

Biological experiments mainly refer to cultivating mosquitoes in a controlled laboratory environment to observe and analyze their survival and development under varying climatic conditions, such as temperature and humidity. For example, Delatte, Apperson, Brady, and other scholars have studied the effects of temperature on mosquito development, sexual nutrient cycle, the maximum number of eggs laid, and survival time of adult mosquitoes when other conditions remained unchanged [6–8]. However, such controlled conditions are difficult to reproduce in actual scenes, and the conclusions obtained from biological experiments may be challenging to validate in real environments. Statistical models are mainly based on controlled experiments or field observations to explore the correlations between mosquito abundance and meteorological factors [9–13]. Usually, the analysis of these correlations is linear, and the conclusions are obvious. However, it is difficult to precisely quantify the effects of meteorological factors like temperature and humidity on different life stages, such as development and mortality.

A common method in mechanistic models is to use differential equations to describe various stages of the mosquito life cycle, incorporating climate factors like temperature and precipitation through different forms of functions, such as oviposition rate, mortality rate, and development rate. Based on diverse research contents, scholars have divided the life cycle of Aedes mosquitoes into various stages and established differential equations of different dimensions [14–16]. For instance, Fukui et al. developed a four-stage model [17], Erickson et al. established a six-stage model [18], and the model proposed by Cailly et al. includes ten different stages [19]. However, when establishing these dynamics models, the functional forms for

key parameters such as development rate, mortality rate, and oviposition rate need to be predefined, which poses a significant challenge to model establishment. Scholars have proposed some functional forms based on biological experiments, including Gaussian functions, and forms based on degree-day relationships for development rate, as well as monotonic exponential functions, and linear functions for mortality rate [9,20–26]. However, it is important to note that these functional forms are typically derived under controlled laboratory conditions and may not necessarily be applicable in natural environments. In addition, some functional forms remain unknown. For example, the oviposition behavior of mosquitoes may be influenced by both temperature and precipitation [6,27], but existing proposed oviposition rate functions often focus solely on one factor, either temperature or precipitation [9,25,28]. The precise form of oviposition rate under the combined influence of temperature and precipitation has yet to be determined.

In recent years, neural networks, distinguished by their unique network structure and information processing methods, have made brilliant achievements in many fields. They have been effectively employed in many tasks including mosquito species classification and abundance prediction. For example, Lee, Kinney, and others have successfully simulated mosquito abundance by constructing neural network models such as feedforward neural networks, long short-term memory (LSTM) networks, and gated recurrent units (GRUs), achieving ideal simulation results [1,29–32]. However, the common operation of these methods is to take factors such as time and weather as inputs, with mosquito abundance as output. This end-to-end learning method cannot reveal the biological mechanism of mosquito growth and reproduction. The black-box nature of the algorithm makes the deep learning method encounter unexplained risks [33]. In order to improve the interpretability of deep learning in physics, biology, chemistry, and other issues, more and more works pay attention to coupling or embedding differential equations and deep neural networks. It has recently been shown to be advantageous to merge differential equations with machine learning [34], such as neural differential equations and universal differential equations [35–37]. Despite the remarkable progress in coupling the differential equation and deep learning, these methods haven't attracted enough attention and widespread application, with most existing research focusing on epidemiological dynamics problems [33,38–39]. In fact, there are many unknown mechanisms in biological growth and development processes. Introducing the idea of coupling neural networks with differential equations into population dynamics models can help us address many issues. Furthermore, despite the improved interpretability of neural networks achieved through this coupling model, it still faces criticism for its "black-box" nature, which has prompted researchers to seek exact function expressions. Some data-driven methods, such as symbolic regression and sparse recognition of nonlinear dynamical systems, provide methods to explore unknown functional forms between variables based on data [40–43]. Combining these methods with neural networks can further enhance the interpretability of the model.

In this study, we will propose a modeling framework that combines deep learning and population dynamic models, which allows the neural network to adhere to the biological mechanism of population development and reproduction throughout the training process. In the absence of explicit mathematical formulations for the relationship between oviposition behavior and temperature/precipitation, we use real data to drive the model, approximating the oviposition rate function through a neural network, and embedding it into the traditional mosquito population dynamics model. After the model training, symbolic regression is used to interpret the input and output of the neural network and to explore the explicit expression of the oviposition rate function. By combining data and mechanisms, this modeling framework can effectively simulate population dynamics while exploring some unknown biological mechanisms. This approach can offer new insights for traditional population dynamics problems.

Method

Data

Guangdong Province is located in the southern coastal region of China and falls within the subtropical monsoon climate zone. The monthly average temperature and monthly cumulative precipitation for each city in Guangdong Province in June 2016 are shown in Fig 1A and 1B. As depicted in the figures, notable variations in temperature and rainfall among cities are observed due to geographical differences, such as disparities between the northern and southern parts of the province and the influence of mountainous and hilly terrain. These diverse climate patterns contribute to significant differences in mosquito abundance among cities (shown in Fig 1C). Based on these observations, we selected twelve representative cities—Shenzhen, Yangjiang, Guangzhou, Shantou, Shanwei, Heyuan, Meizhou, Zhaoqing, Shaoguan, Zhanjiang, Zhuhai, and Huizhou—for further analysis and research. The daily average temperature and daily cumulative precipitation of these cities can be obtained from the National Centers for Environmental Information (NCEI) website (<https://www.ncei.noaa.gov>), which are shown in Fig 2.

The Mosquito Ovitrap Index (MOI) of these cities, providing insights into the fluctuation of Aedes mosquito populations, was obtained from the Guangdong Provincial Health Commission (<https://wsjkw.gd.gov.cn>). According to the monitoring guidelines for dengue vector published by the Guangdong Provincial Health Commission, monitoring staff establish multiple monitoring points in each city to assess the populations of Aedes mosquitoes, including both *Aedes albopictus* and *Aedes aegypti*. The combined results of these two mosquito species will be used as the overall monitoring results for Aedes mosquitoes. The Guangdong Provincial Health Commission reports monitoring results every half month. These reports detail the proportions of four categories of monitoring points in each city—meeting epidemic prevention and control standards, low-density, medium-density, and high-density areas—with corresponding MOI values for each category. Based on the distribution of monitoring points across different monitoring levels, we can calculate the data of 100 monitoring points for each city. In addition, considering the differences in data reporting methods before 2015 and the potential impact of the COVID-19 pandemic on the population dynamics of Aedes mosquitoes after 2020, the study period was selected from 2016 to 2019 to ensure reliability and accuracy.

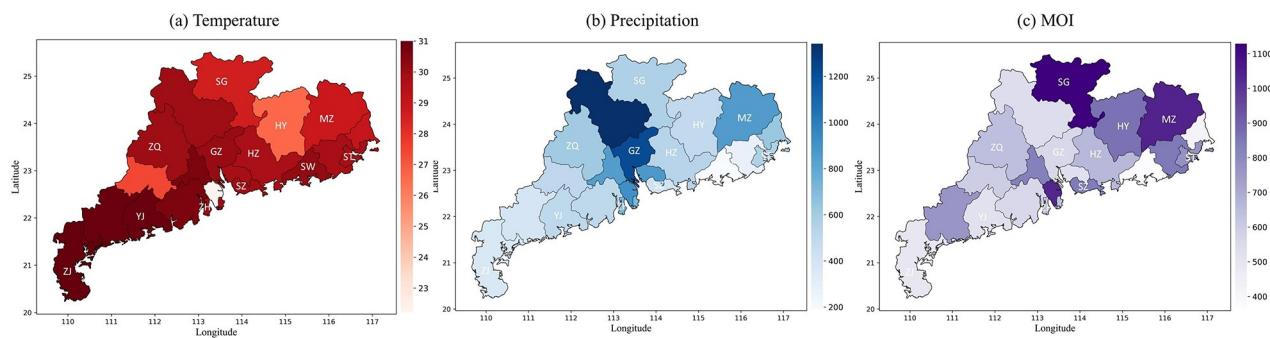


Fig 1. Temperature, precipitation, and Mosquito Ovitrap Index (MOI) data in Guangdong Province. (a)-(b) Average monthly temperature and cumulative monthly precipitation for all regions of Guangdong Province in June 2016. (c) MOI data for different regions of Guangdong Province during the latter half of June 2016. Darker colors indicate higher numerical values, with study areas marked by the initial letters of city names. The base layer of the map was created using the publicly available dataset 'China—Subnational Administrative Boundaries' contributed by the Office for the Coordination of Humanitarian Affairs (OCHA) Regional Office for Asia and the Pacific. The dataset can be accessed at: <https://data.humdata.org/dataset/china-administrative-boundaries>. The data license can be found at: <https://data.humdata.org/faqs/licenses>. We affirm that no modifications have been made to the original data.

<https://doi.org/10.1371/journal.pcbi.1012499.g001>

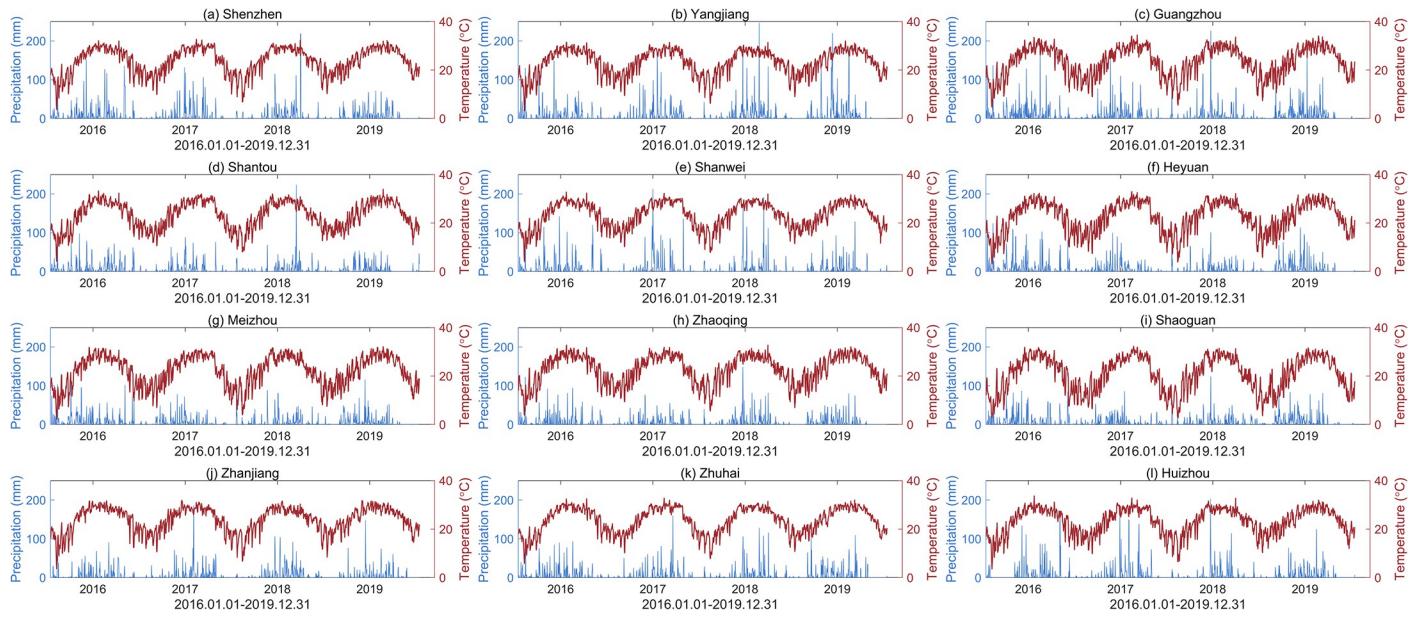


Fig 2. Meteorological data for twelve cities. Daily cumulative precipitation (blue) and daily average temperature (red) from January 1, 2016, to December 31, 2019, for twelve cities.

<https://doi.org/10.1371/journal.pcbi.1012499.g002>

Biological mechanism model

The life cycle of Aedes mosquitoes includes three water-dependent stages (egg, larva, and pupa), and one aerial stage (adult). The development and death of each stage depend on several factors, such as temperature, or water availability. Based on the work of Gong et al. and Wang et al. [21,26], in this study, we collectively refer to the first three water-dependent stages as the immature stage, and the aerial stage as the adult stage. Combining biological mechanisms such as reproduction, development, and mortality, we establish a two-stage population dynamics model comprising immature (P) and adult stages (A):

$$\begin{aligned} \frac{dP}{dt} &= ANN(T(t), R(t)) \cdot A - dp(T(t)) \cdot P - mp(T(t)) \cdot P \\ \frac{dA}{dt} &= ef(R_{norm}(t)) \cdot dp(T(t)) \cdot P - ma(T(t)) \cdot A \end{aligned} \quad (1)$$

Where $ANN(T(t), R(t))$ is the oviposition rate, which depends on the daily temperature and precipitation, $mp(T(t))$ is the mortality rate of immatures, $ma(T(t))$ is the mortality rate of adults, $dp(T(t))$ is the development rate from immatures to adults, and $ef(R_{norm}(t))$ is the success rate of emergence dependent on the weekly precipitation. Parameter definitions can be found in Table 1, and these parameters are on a daily time scale.

Many studies have revealed the significant impact of temperature on insect development, and the developmental and mortality processes of mosquitoes also exhibit temperature dependence. Therefore, the setting of mortality rates for both immatures and adults in this study is based on the temperature-dependent function proposed by Otero et al. [22], which resembles the Gaussian function. This is a common functional form for mosquitoes [21] and was used by scholars such as Li et al. in the study of Aedes [44]. The development rate from immatures to adults follows the Sharpe & DeMichele equation proposed by Sharpe et al. [45], which was widely used for describing the development process of Aedes. [20,22–23]. The success rate for

Table 1. Definitions of the parameters used in the model.

Parameter	Definitions
$ANN(T(t), R(t))$	The daily oviposition rate affected by temperature and precipitation (day^{-1})
$dp(T(t))$	The daily development rate of immatures (day^{-1})
$mp(T(t))$	The daily mortality rate of immatures (day^{-1})
$ma(T(t))$	The daily mortality rate of adults (day^{-1})
$ef(R_{norm}(t))$	The daily density-dependent success of adult emergence (day^{-1})

<https://doi.org/10.1371/journal.pcbi.1012499.t001>

emergence is constrained by environmental carrying capacity and precipitation, with the functional form inspired by the research of Fukui et al. and Tran et al. [17,25].

Furthermore, diapause is a survival strategy adopted by organisms in adverse conditions, crucial for maintaining population size and regulating population growth [46]. There are two primary methods to integrate diapause into mosquito population models. One method is to use piecewise functions to distinguish the survival rate or development rate between normal and diapause stages, while another method treats diapause as an independent dynamic process with significantly different dynamics from the normal period. In addition, it is meaningful to further consider the incubation period before returning to the normal state after the diapause period [47]. In this study, to maintain model simplicity and avoid introducing too many parameters, we opted to follow the first method and modified the mortality and development rate functions for immature stage. Based on the work of Liu et al. [48], we set that mosquitoes enter diapause when the average temperature of half a month drops below 21 degrees and the duration of daylight falls below 13 or 14 hours. Considering the sunshine data in Guangdong Province, these sunshine conditions are usually met in winter, so we focus solely on temperature. During diapause, the mortality rate of diapausing eggs is low, enabling them to survive the long winter [49–50]. Thus, we segmented the mortality function of immature stage to maintain a low rate when the temperature is continuously below 21°C. We also adjusted the developmental rate during diapause to be δ times that of the normal period.

The expressions for the development rate, the mortality rate, and the emergence rate in the model are defined as follows:

$$\left\{ \begin{array}{l} mp(T(t)) = \begin{cases} 1 - \mu_p \exp\left(-\frac{(T(t) - T_p)^2}{V_p^2}\right), & T_{month}(t) \geq 21^\circ C \\ 1 - \mu_p \exp\left(-\frac{(21 - T_p)^2}{V_p^2}\right), & T_{month}(t) < 21^\circ C \end{cases} \\ ma(T(t)) = 1 - \mu_a \exp\left(-\frac{(T(t) - T_a)^2}{V_a^2}\right) \\ dp(T(t)) = \begin{cases} AA \cdot \frac{T(t) + 273.15}{298.15} \cdot \frac{\exp(\frac{HA}{1.987} (\frac{1}{298.15} - \frac{1}{T(t) + 273.15}))}{1 + \exp(\frac{HH}{1.987} (\frac{1}{TH} - \frac{1}{T(t) + 273.15}))}, & T_{month}(t) \geq 21^\circ C \\ \delta \cdot AA \cdot \frac{T(t) + 273.15}{298.15} \cdot \frac{\exp(\frac{HA}{1.987} (\frac{1}{298.15} - \frac{1}{T(t) + 273.15}))}{1 + \exp(\frac{HH}{1.987} (\frac{1}{TH} - \frac{1}{T(t) + 273.15}))}, & T_{month}(t) < 21^\circ C \end{cases} \\ ef(R_{norm}(t)) = \exp\left(-0.1 \cdot \left(1 + \frac{dp(T(t)) \cdot P(t)}{250000 \cdot (R_{norm}(t) + 1)}\right)\right) \end{array} \right. \quad (2)$$

Table 2. Definitions of the parameters.

Parameter	Definition	Value	Source
μ_p	The survival rate at T_p	1.0000	Estimated
T_p	The optimum survival temperature of immatures	20.7896	Estimated
V_p	The variance for immatures	17.4774	Estimated
μ_a	The survival rate at T_a	0.6000	Estimated
T_a	The optimum survival temperature of adults	23.3492	Estimated
V_a	The variance for adults	19.3648	Estimated
AA	The development rate in the absence of temperature inactivation of key developmental enzymes	0.0150	Estimated
HA	The enthalpy of activation of enzyme-catalyzed reactions	28116.4131	Estimated
HH	The change in enthalpy associated with high temperature inactivation of enzymes	35378.2344	Estimated
TH	The temperature at which 50% of the enzyme is inactivated by high temperature	301.6750	Estimated
δ	The ratio of the developmental rate between diapause and non-diapause stages.	0.0023	Estimated
$T_{month}(t)$	The average temperature over half a month		Data
$R_{norm}(t)$	Standardized weekly cumulative precipitation		Data

<https://doi.org/10.1371/journal.pcbi.1012499.t002>

Definitions of the parameters are shown in Table 2.

Neural network component

In model (1), $ANN(T(t), R(t))$ is the oviposition rate, which represents the average number of eggs laid by each mosquito after finding a breeding site per unit time. Mosquito oviposition behavior is influenced by both temperature and precipitation. Precipitation increases surface humidity, providing breeding habitats and stimulating mosquito oviposition [27], while temperature affects the mosquito's gonotrophic cycle, resulting in significant differences in maximum egg production at different temperatures [6]. Most previous studies have proposed oviposition rate functions that only consider temperature or precipitation individually [9,12,18]. Furthermore, when considering precipitation factor, more scholars choose to integrate it with the processes of egg hatching and development [25,51]. To explore the combined effects of temperature and precipitation on oviposition rates and make the oviposition rate function more realistic, we utilize a neural network to replace traditional oviposition rate and establish a model that couples neural network with differential equations.

We define an L-layer deep feedforward network, $\mathcal{L}^L(U) : R^{d_{input}} \rightarrow R^{d_{output}}$, where U represents the inputs. Let N^k denotes the number of neurons of the k th layer, then the number of neurons of the input and output layers are $N_0 = d_{input}, N_L = d_{output}$. As we aim to explore the combined effects of temperature and precipitation on oviposition rate variation, the inputs to the neural network are temperature and precipitation: $U = (T, R)$, the output is oviposition rate: $ANN(T, R)$, then $d_{input} = 2, d_{output} = 1$. Define the weight matrix and bias vector corresponding to each layer as W^k and b^k , and the activation function as σ , then the architecture of the neural network can be recursively represented as follows:

- Input : $U(T, R) \in R^2$,
- Input layer : $\mathcal{L}^0(U) = U$,
- Hidden layers : $\mathcal{L}^k(U) = \sigma(W^k \mathcal{L}^{k-1}(U) + b^k) \in R^{N_k}$, for $k = 1 : L - 1$,
- Output layer : $\mathcal{L}^L(U) = \sigma(W^L \mathcal{L}^{L-1}(U) + b^L) \in R$,
- Output : $ANN(T, R) = \mathcal{L}^L(U)$

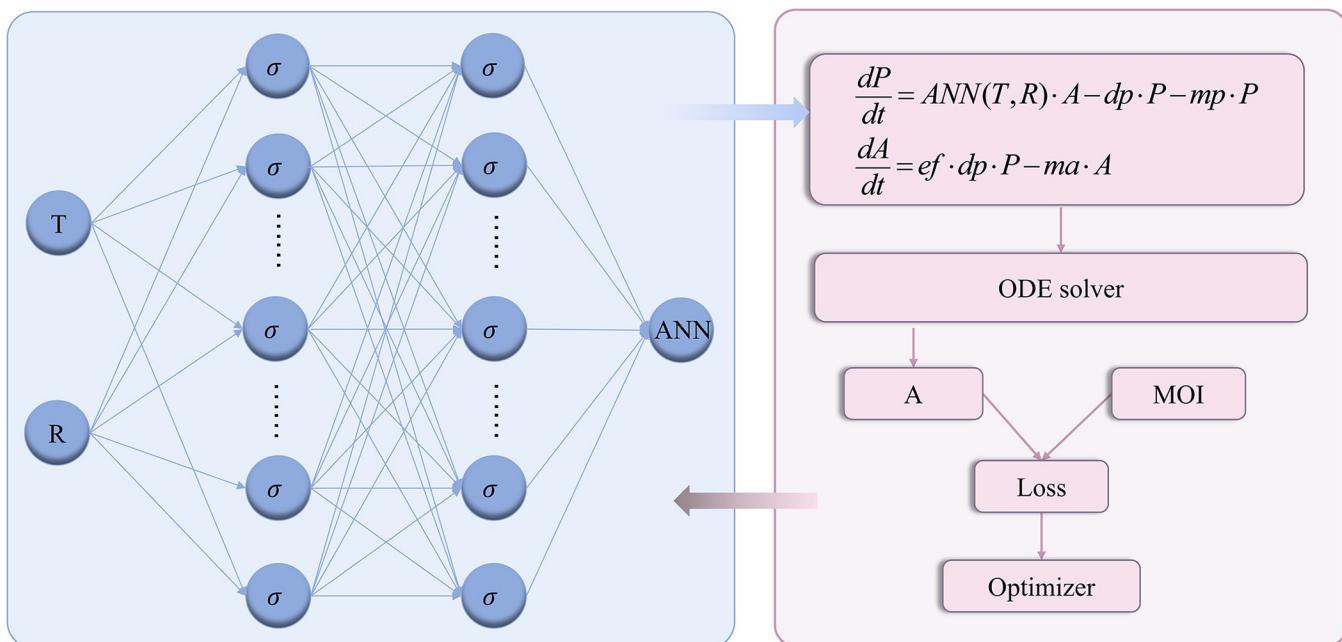


Fig 3. Schematic diagram of the coupling model of neural network and differential equations. The blue shading represents the neural network, and the red shading represents the mechanistic model. The output of the neural network serves as an internal dependency of the differential equation, and the neural network is indirectly trained through the differential equation.

<https://doi.org/10.1371/journal.pcbi.1012499.g003>

Theoretically, a two-layer neural network is sufficient to fit any bounded continuous function. In practice, after multiple iterations and adjustments, we finally constructed a three-layer feedforward network, that is $L = 3$, and the number of neurons in the hidden layer are $N_1 = 10$, $N_2 = 10$. According to the universal approximation theorem [52–53], this is enough to approach the oviposition rate function. We choose the Sigmoid function as the activation function and map the output of the network to lie between the maximum oviposition rate and the baseline rate, which is derived from the conclusion of Liu et al., Tran et al., and Metelmann et al. [25,48,51].

After constructing and initializing the network, the next key step is to embed the network into the mechanistic model. The essence of this coupled model is described in Fig 3. The equation comprises two parts: one part describes the mechanisms of the growth, development, and reproduction of Aedes mosquitoes based on traditional single-population dynamical system theory, while the other part depicts the unknown mechanism, specifically the relationship between oviposition rate, temperature, and precipitation. The neural network's output serves as an internal dependency for the numerical solution of the differential equation. By employing suitable optimization methods to minimize the error between the numerical solution of the equation and the observed values, we can indirectly constrain the output of the neural network and realize the training of the network. This method emphasizes the connection between biological mechanisms and neural networks, integrates data with mechanisms, and plays a crucial role in exploring unknown mechanisms.

Symbolic regression for interpreting the coupling term

In Model (1), we employ a neural network to explore the unknown mechanisms between oviposition rate, temperature, and precipitation. However, due to the complex internal structure and numerous parameters of neural networks, the coupling term remains as a black-box,

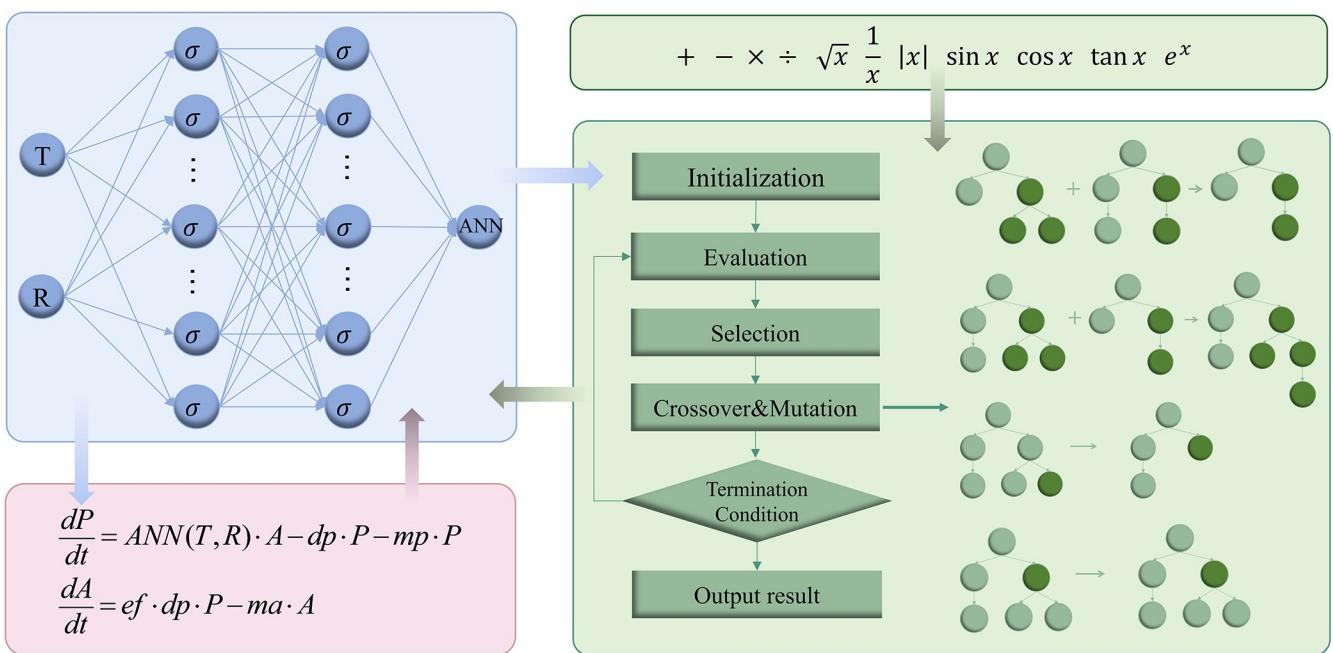


Fig 4. Symbolic regression flow diagram. The blue shading represents the neural network, the red shading represents the mechanistic model, and the green shading represents the process of symbolic regression using the idea of the genetic algorithm. The training data of symbolic regression is obtained according to the input and output of the neural network, and the optimal expression is found through a series of genetic, variation, and evolutionary operations.

<https://doi.org/10.1371/journal.pcbi.1012499.g004>

lacking specific mathematical expressions. In recent years, some data-driven methods such as symbolic regression have been proposed to address this interpretability issue. Symbolic regression is a type of regression analysis that searches the space of mathematical expressions to find the model that best fits a given dataset, both in terms of accuracy and simplicity, without requiring any prior domain expertise or predefined underlying regression structures [54].

We traverse the temperatures from 10°C to 35°C and precipitation from 0 mm to 200 mm, generating a set of two-dimensional arrays. These arrays serve as inputs to the neural network $\mathcal{L}^L(U)$ to compute the corresponding oviposition rates. Subsequently, this dataset will be utilized as training data for symbolic regression.

The purpose of symbolic regression is to find the best function expression that can approximate the results of the neural network. Note that the result of symbolic regression is the function $SR(T, R): R^2 \rightarrow R$, then the fitness function can be defined as:

$$MSE = \sqrt{\frac{1}{m \cdot n} \sum_{i=1}^m \sum_{j=1}^n (SR(T_i, R_j) - ANN(T_i, R_j))^2} \quad (3)$$

Where m and n represent the sample size.

Using the training data obtained from the above operations, with the goal of minimizing this fitness function, we employ symbolic regression to seek explicit expressions among variables, as shown in Fig 4. Genetic programming is a specific implementation of this procedure. Its core idea is to simulate natural selection, genetic variation, and other phenomena in biological systems following Darwin's theory of evolution, evolve the solution to a given problem, and find the most suitable solution after several generations. This approach utilizes a binary tree structure to describe function expression and applies string-based genetic operations to manipulate the binary tree [55–57]. The process begins with the initial binary trees generated

randomly. After generations of reproduction and evolution, we can obtain the optimal expression.

Parameter estimation

Based on the study of Tran et al. [25], we assumed that there are 1000000 diapause eggs at the initial moment to simulate the dynamics of Aedes mosquito population. To mitigate the impact of initial conditions on simulation results, simulations commenced on January 1, 2015, and spanned five years in daily units until 2019. Data from 2016 to 2019 were utilized to constrain the parameters. The dataset was split into two segments: data from nine cities—Shenzhen, Yangjiang, Guangzhou, Shantou, Shanwei, Heyuan, Meizhou, Zhaoqing, and Shaoguan—served as training set for model development and parameter estimation, while data from three cities—Zhanjiang, Zhuhai, and Huizhou—were used as testing set to evaluate model generalization.

During the model training process, we simultaneously calculated the coupling models of the nine cities in the training set using the same set of parameters, and evaluated the overall loss function. It should be noted that the MOI doesn't directly reflect the exact count of mosquitoes, rather, it is a commonly used index to represent the abundance of adults, which is proportional to the actual number. In fact, directly measuring the real number of mosquitoes is challenging, so we pay more attention to the changing trend in mosquito population dynamics. Our goal of parameter estimation is to align the fitted trends of adult mosquitoes with the changes in MOI. Since the MOI is published half-monthly, we calculated the average simulation results every half-month and adjust them using rotation factors to align their scales with MOI. Our definition of the loss function is

$$\text{Loss}(\theta_1, \theta_2) = \frac{1}{m \cdot n} \sum_{j=1}^m \sum_{t=1}^n |MOI_{j,t} - \lambda_j A_{j,t}(\theta_1, ANN_{\theta_2}(T_{j,t}, R_{j,t}))| \quad (4)$$

Where θ_1 represents the biological parameters and θ_2 represents the network parameters, m, n represent the number of cities and months, λ_j presents the rotation factor of city j .

We employed local adjoint sensitivity analysis [36] to minimize the loss function and utilized the adaptive moment estimation algorithm (ADAM) to update the parameters. The parameters to be estimated can be mainly divided into two categories. The first category comprises parameters related to the established known mechanisms (θ_1), including these unknown parameters in the development rate and mortality function, as shown in Table 2. The second category is the parameters involved in the neural network (θ_2), which are used to adjust the connection strength and activation rules between neurons. Notably, the same parameter set is shared across the nine cities in the training set, indicating a level of parameter generalization achieved in our study. Considering the numerous internal parameters in the neural network and the associated training costs, we implemented a three-stage optimization approach to search for the best parameters:

- Step 1: Set the parameters in sets θ_1 and θ_2 as trainable and adjustable parameters. Utilize the ADAM optimizer with an initial learning rate of 0.01 and employ the TRBDF2 method to solve the differential equation. Iterate the optimization process 500 times to estimate all parameters;
- Step 2: Fix the parameters of the mechanistic model (θ_1) with the optimal parameters obtained in step 1. Set the parameters in set θ_2 as trainable and adjustable parameters. Utilize the ADAM optimizer with an initial learning rate of 0.001, and iterate the optimization process 2000 times to estimate the network parameters θ_2 .

Table 3. Hyperparameter in symbolic regression process.

Hyperparameter	Definition	Value
population_size	The number of individuals retained in each generation	6000
generations	The number of iterations for the algorithm	40
stopping_criteria	The stopping criteria for the algorithm	0.01
p_crossover	The probability of crossover operation occurrence	0.1
p_subtree_mutation	The probability of subtree mutation operation occurrence	0.1
p_hoist_mutation	The probability of hoist mutation operation occurrence	0.07
p_point_mutation	The probability of point mutation operation occurrence	0.1
parsimony_coefficient	Controls the complexity of the generated expressions, balancing model fit and simplicity	0.02

<https://doi.org/10.1371/journal.pcbi.1012499.t003>

- Step 3: Fix the parameters in the neural network (θ_2) with the optimal parameters obtained in step 2. Set the parameters in the set θ_1 as trainable and adjustable parameters. Set the initial learning rate of the ADAM optimizer to 0.01, and perform iterative optimization 1000 times.

After the above training, a well-trained neural network can be constructed. Next, we employed symbolic regression to explore explicit expressions between the network inputs (temperature and precipitation) and outputs (oviposition rate). The training dataset for symbolic regression was constructed as follows: temperature values were sampled at 1°C intervals from 15°C to 35°C, while precipitation values were selected based on different ranges: from 0mm to 30mm at 1 mm intervals, from 31mm to 150mm at 10mm intervals, and from 150mm to 200mm at intervals of 5mm. This selection method was determined based on the observed curve shapes of the oviposition rates generated by the neural network. Details on the hyperparameters and their respective values used during the symbolic regression are listed in [Table 3](#). The termination condition was set such that the error of the algorithm running to the optimal model is less than 0.01% or it runs to the 40th generation. The population size was set to be 6000, and the depth of individuals in the initial population varied randomly between 2 and 6. During the evolutionary process, we controlled the occurrence of point mutations in winning individuals with a probability of 0.1, hoist mutations with a probability of 0.07, and subtree mutations with a probability of 0.1. The probability of crossover among winners was set to 0.1. To control the complexity of the formula, several attempts were made to set the parsimony coefficient from 0.001 to 0.1, and it was finally determined to be 0.02.

Result

Fitting results

We utilized the established model to simulate mosquito populations in nine cities from the training set. To facilitate comparison and presentation, we calculated the mean of the simulated results for each half-month and standardized them to the same scale as the MOI. The optimal fitting results for adults across these nine cities were presented in [Fig 5A–5I](#), where purple solid lines represent the fitting results, purple dots denote the training data, and shaded areas indicate periods of diapause. To evaluate the model's generalization ability and prevent overfitting, we further extended the trained model to predict mosquito populations in the testing set cities—Zhanjiang, Zhuhai, and Huizhou—as depicted in [Fig 5J–5L](#), where red solid lines represent the predictions and red dots denote the testing data.

It can be seen that our simulated results are closely consistent with the actual MOI in both the training and testing sets, indicating that our model effectively captures the trend of Aedes

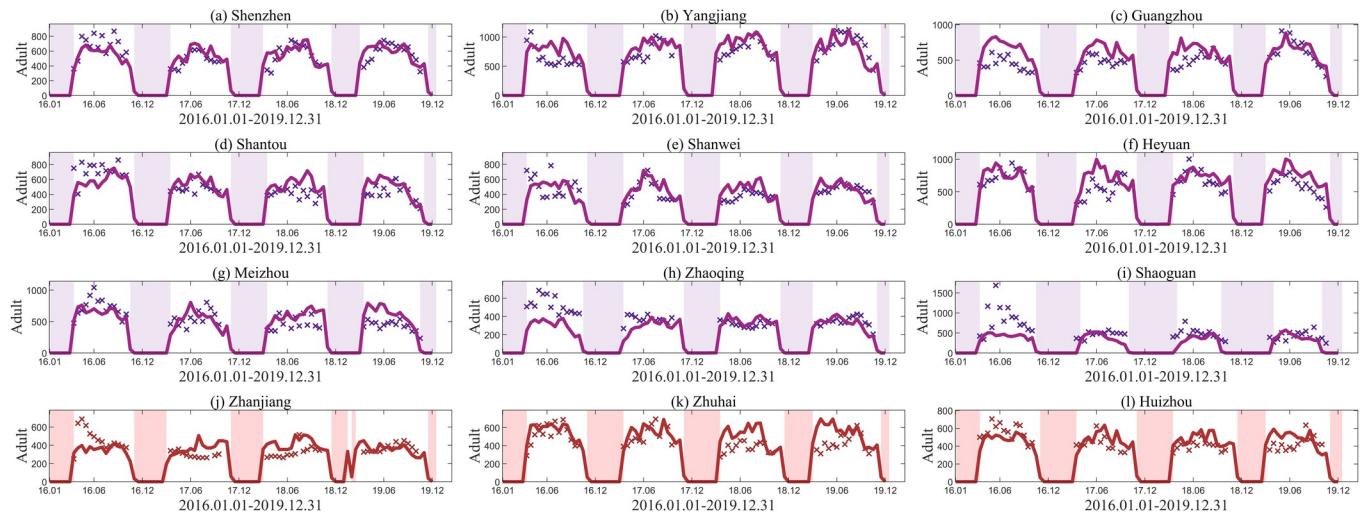


Fig 5. Data fitting and verification results. (a)-(i) Fitting results of adults from January 2016 to December 2019 in the training set. Purple solid lines represent the fitting results, dots indicate the training set data, and the shaded areas indicate the diapause periods. (j)-(l) Prediction results of adults from January 2016 to December 2019 for three cities in the testing set. Red solid lines represent the predicted results, dots denote test set data, and shaded areas indicate periods of diapause.

<https://doi.org/10.1371/journal.pcbi.1012499.g005>

mosquito population dynamics. To further verify the fitting accuracy, the correlation coefficient (R) was used to compare the fitting values with the MOI (shown in Fig 6A). The analysis revealed that the correlation coefficients for all regions were above 0.7. Specifically, in the training set, the correlation coefficient surpassed 0.85 in Shenzhen, Guangzhou, and Heyuan, exceeded 0.8 in Yangjiang, and was above 0.75 in Shantou, Meizhou, and Shanwei City. In the testing set, the correlation coefficients of Zhuhai and Huizhou exceeded 0.8. The strong correlations observed between the simulation results and the actual values for these cities, along with their high statistical significance ($P < 0.01$), provide robust evidence for the reliability of the simulation outcomes.

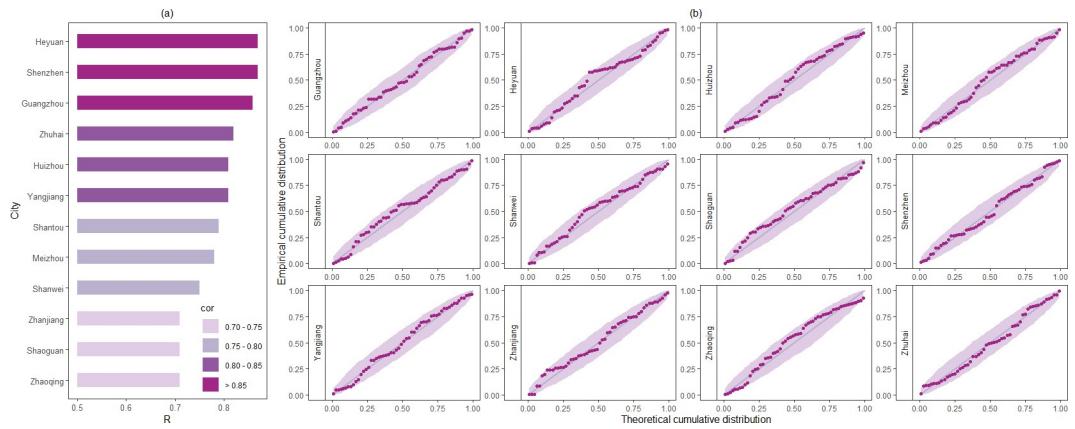


Fig 6. Correlation analysis and residual analysis. (a) The correlation coefficient between the simulation results and the actual values in twelve areas. The correlation coefficient is distinguished by the bar length and color depth. (b) The Probability-Probability (P-P) plots across twelve cities to assess the goodness-of-fit of residual distributions to the theoretical normal distribution. The x-axis in each plot denotes the theoretical cumulative probabilities under the normal distribution, while the y-axis represents the empirical cumulative probabilities derived from the residuals of each city. The shaded area around the diagonal dashed line represents the 95% confidence band.

<https://doi.org/10.1371/journal.pcbi.1012499.g006>

It is noteworthy that using the same set of parameters to fit data across all regions inevitably leads to a loss of local fitting accuracy. Additionally, variations exist in the standardization and accuracy of data monitoring procedures among different cities, compounded by the influence of mosquito control measures, collectively contributing to heightened data uncertainty. Considering the inherent uncertainty and randomness of the MOI, we also conducted a normality test on the residuals of the simulated results to account for them. Probability-Probability (P-P) plots were constructed to assess the goodness-of-fit of residual distributions across twelve cities to the theoretical normal distribution (shown in Fig 6B). Each panel in the figure represents a P-P plot for a specific city, comparing the observed cumulative probabilities of residuals with those expected under the normal distribution. The alignment of points along the diagonal dashed line in each panel indicates good agreement between observed and expected cumulative probabilities, suggesting that the normal distribution assumption is reasonable for modeling the residuals in these cities, thus providing additional evidence to support the validity of our model.

Analysis of oviposition rate via the neural network

Through the optimization method mentioned earlier, we obtained a set of optimal parameters. Notably, we did not provide any prior information regarding the relationships between oviposition rates, temperature, and precipitation in the model, meaning we did not assume a specific functional form for these interactions in advance. However, the trained network provided valuable insights into how oviposition rates fluctuate in response to changes in temperature and precipitation. Figs 7 and 8 show the results derived from our neural network.

The pattern of oviposition rates at various temperatures is shown in Fig 7. The figure illustrates that under constant precipitation levels, the oviposition rate follows an S-shaped curve in response to temperature changes. At lower temperatures, the oviposition rate remains low. As the temperature gradually rises beyond a certain threshold, the oviposition rate increases rapidly. However, when the temperature becomes too high, the rate of increase slows down

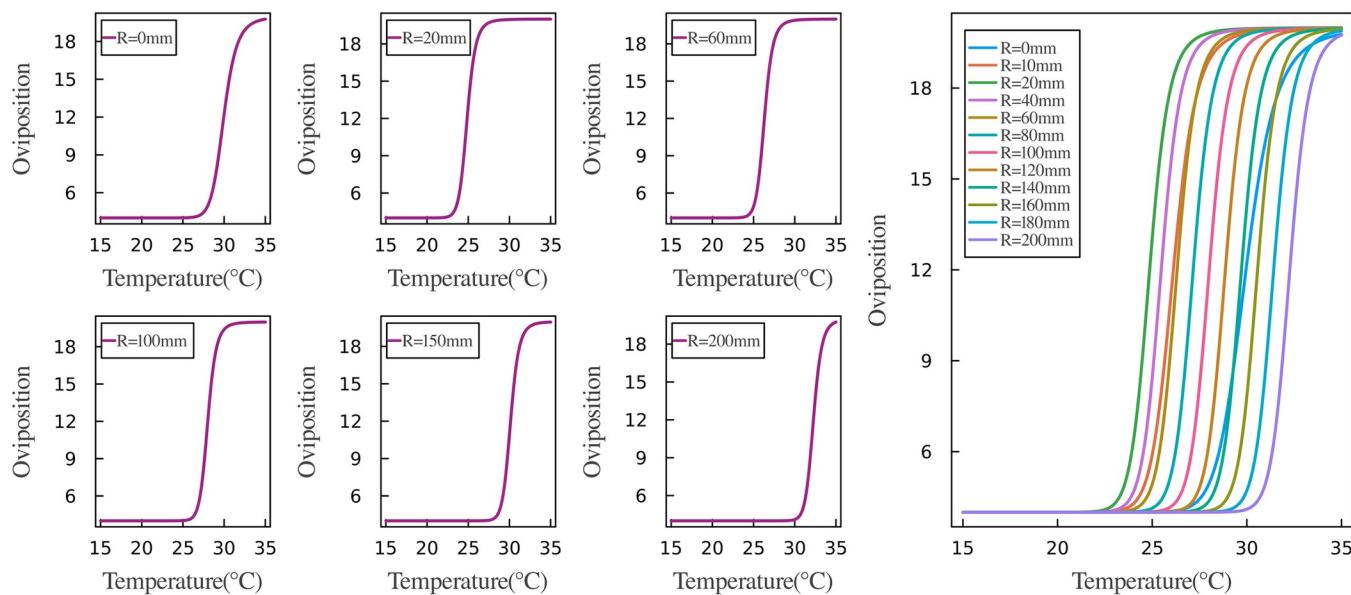


Fig 7. Oviposition rates fluctuating with temperature variations. To enhance observation and analysis, the left plot illustrates the variation in oviposition rate with temperature across different precipitation levels: 0, 20, 60, 100, 150, and 200 mm. Conversely, the right plot consolidates all findings for comparative analysis, employing distinct colors to differentiate between oviposition rates at varying precipitation levels.

<https://doi.org/10.1371/journal.pcbi.1012499.g007>

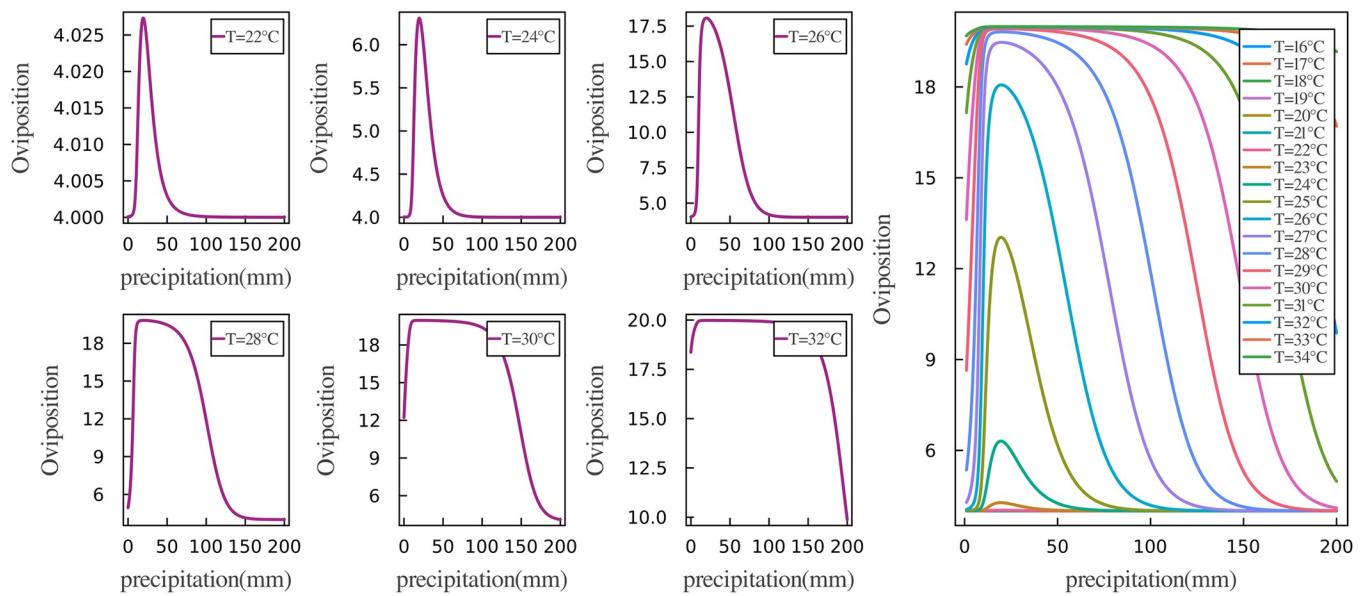


Fig 8. Oviposition rates fluctuating with precipitation variations. The left plot depicts the fluctuation of oviposition rates in response to precipitation across temperatures ranging from 22°C to 32°C. Conversely, the right plot consolidates all data points, with distinct colors to denote varying temperatures for ease of comparison.

<https://doi.org/10.1371/journal.pcbi.1012499.g008>

and eventually levels off. Comparing the curves under different levels of precipitation, we observed distinct temperature thresholds for rapid increases in oviposition rate. When precipitation is either too low or too high, a temperature exceeding 25°C is required for a significant increase in oviposition rate. Under excessive precipitation, even a temperature of 30°C results in a low oviposition rate. However, at optimal precipitation levels, the oviposition rate begins to rise quickly at temperatures as low as 23°C. This curve pattern leads to fewer mosquitoes in winter and higher populations in summer and fall, which aligns with mosquitoes' natural behavior [58]. The results demonstrate that the neural network can effectively model the relationship between oviposition rate and temperature.

The relationship between oviposition rate and precipitation variation is depicted in Fig 8. The curves representing oviposition rates at various temperatures exhibit a characteristic parabolic shape, rising initially and then falling. This pattern indicates that as precipitation levels increase, the oviposition rate also rises and maintains high values at optimal precipitation levels. However, once precipitation surpasses a certain threshold, the oviposition rate will begin to decline. Notably, the parabolic shapes vary significantly at different temperatures. The initial values of the curves differ across different temperatures, with lower temperatures associated with lower initial oviposition rates. For example, at 22°C and 24°C with low precipitation, the oviposition rate is approximately 4, but when the temperature rises to 30°C, the oviposition rate can be as high as 12. Additionally, the peak values and range of variation in oviposition rates also differ with temperature. Specifically, at lower temperatures, changes in oviposition rate in response to varying precipitation are minimal. For instance, at 22°C, the oviposition rate increases by only around 0.025 at its peak, remaining close to 4. However, at more favorable temperatures like 28°C, the range of variation in oviposition rate can exceed 12, with peak rates reaching up to 19. Furthermore, when analyzing the precipitation threshold at which the oviposition rate begins to decline significantly after reaching its peak, we found that higher temperatures extend the range of precipitation that kept the oviposition rate near its peak. For instance, at 26°C, the oviposition rate peaks at around 15mm of precipitation and

subsequently decreases as precipitation continues to increase. However, at 30°C, the oviposition rate peaks when precipitation is around 10mm and remains stable until precipitation exceeds 100mm before a noticeable decrease occurs. At 32°C, oviposition rate remains high throughout a precipitation range of 10-150mm.

While previous studies generally suggested that precipitation has a positive impact on mosquito populations [59–61], some researchers argue that excessive rainfall may have negative impacts, as it may wash away mosquito habitats, affecting oviposition, egg hatching, and other factors, which in turn affect the adult mosquito populations [12,62–63]. Our study, combining biological mechanisms and neural networks, concluded that moderate precipitation positively influences mosquito populations, while excessive precipitation can have negative effects.

In summary, our results elucidate the fluctuations in oviposition rates resulting from the combined effects of temperature and precipitation. These findings are consistent with observed biological phenomena, emphasizing the efficacy of our integrated model comprising neural networks and mechanistic principles. Our coupled model not only achieves a robust fit to observed data but also enhances our understanding of the underlying biological processes.

Oviposition rate function from symbolic regression

Through the trained neural network, we plotted the curve of oviposition rates varying with temperature and precipitation. However, the network does not provide explicit mathematical expressions. To further investigate the relationship between oviposition rate, temperature, and precipitation, and to enhance the interpretability of deep learning results, we conducted symbolic regression on the results of the neural network. The methods and parameter settings for symbolic regression are thoroughly detailed in the Methods section.

The binary tree obtained through symbolic regression is shown in Fig 9A. Traversing the binary tree using a depth-first algorithm and combining the mathematical operators at each node with the values of the leaf nodes, the optimal symbolic expression can be constructed. After further simplifying and optimizing, the final expression for oviposition rate with respect to temperature and precipitation is derived as follows:

$$\begin{aligned} ANN(T, R) &\approx \frac{16}{1 + e^{-\phi}} + 4, \\ \phi = T - \left(\frac{16 - 1.3593 \frac{R}{T}}{1 + e^{0.3427R}} + \frac{16.0729}{1 + e^{-1.0742T}} + 1.3593 \frac{R}{T} + 7.9271 \right) \end{aligned} \quad (5)$$

To validate the accuracy and assess the generalization capability of this expression, we compared the oviposition rates calculated by the neural network and Eq (5) across a broader input space (temperature ranging from 15°C to 35°C at 0.5°C intervals and precipitation ranging from 0mm to 200mm at 1mm intervals). The results shown in Fig 9B and 9C exhibit a high degree of similarity between the two surfaces in patterns and fluctuations. Furthermore, we calculated the fitness function (3) for the whole spatial domain to be 1.51, providing further evidence that the derived expression effectively captures the trends in oviposition rate variations.

To validate the reasonableness of the functional form obtained through symbolic regression, we integrated the oviposition rate function (5) into model (1) to establish and solve a traditional dynamics model. The comparison of the simulation results obtained by approximating oviposition rates using the neural network and function (5) respectively is shown in Fig 10. Although the two approaches use different underlying principles, their

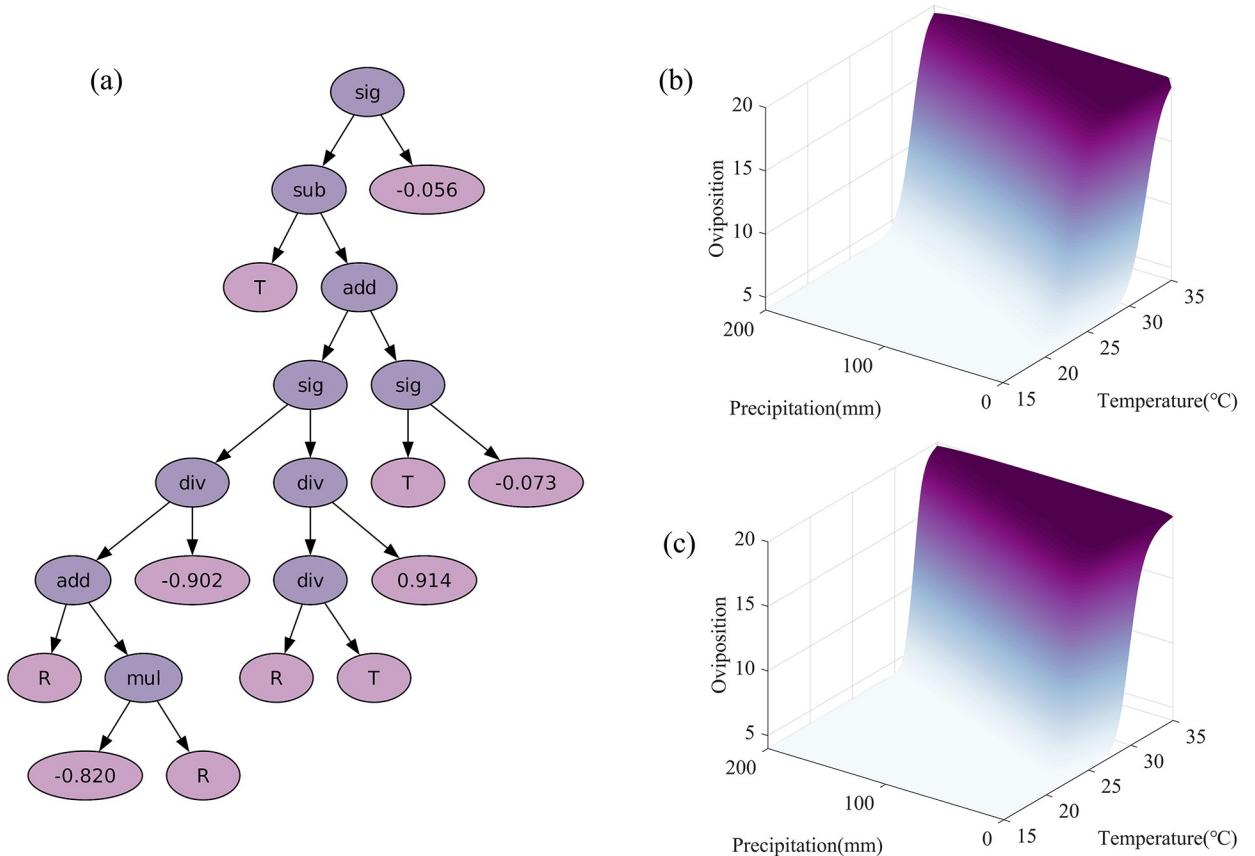


Fig 9. Results from symbolic regression. (a) The optimal binary tree obtained from symbolic regression using genetic programming, "sig" denotes a sigmoid function, "sub" denotes subtraction, "add" denotes addition, "mul" denotes multiplication, and "div" denotes division. (b)-(c) represent oviposition rates calculated using the neural network and symbolic regression under different temperatures and precipitation, respectively.

<https://doi.org/10.1371/journal.pcbi.1012499.g009>

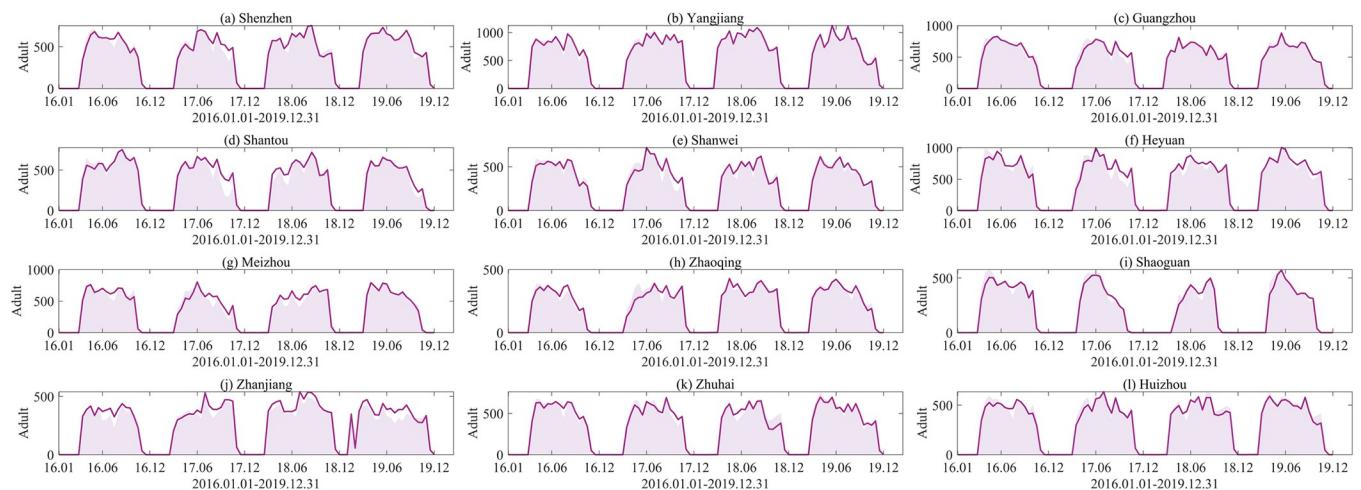


Fig 10. Comparison of simulation results. The shaded part represents the simulation result of using a neural network to approximate the oviposition rate function, and the solid line represents the simulation result of using the expression obtained by symbolic regression to approximate the oviposition rate.

<https://doi.org/10.1371/journal.pcbi.1012499.g010>

simulation results are mostly consistent. This demonstrates that function (5) closely aligns with the neural network's performance. Moreover, utilizing function (5) to construct a mechanistic model effectively captures the changes in mosquito populations. These findings validate the suitability and efficacy of the obtained functional form for oviposition rate.

Comparison and analysis of oviposition rate functions

In this section, we compared and discussed the function we obtained with various forms used in earlier research. The forms of oviposition rates in previous studies can be categorized into four main types: constant, temperature-dependent, humidity-dependent, and dependent on both temperature and precipitation, as shown in [Table 4](#).

Temperature-dependent functions primarily originate from the works of Yang et al. and Metelmann et al. [28,51], who fitted data from controlled temperature experiments in laboratories using forms such as 4th-degree polynomials and skewed Gaussian probability density functions. Experimental data indicate a quasi-linear increase in oviposition rates with rising temperature. The oviposition rates we obtained also demonstrate a similar trend when fixing precipitation. However, the experiments observed a decrease in oviposition rate at extreme temperatures, which was not captured by our model. Several factors may account for this difference. Laboratory conditions are tightly controlled, whereas natural environmental factors such as temperature and precipitation exhibit coordinated variability, highlighting significant differences between these environments. Therefore, in the actual ecological setting of Guangdong Province, this decline may not necessarily manifest. Furthermore, due to the limited sample size of extreme weather conditions in Guangdong, further research is needed to determine whether extreme temperatures lead to decreased oviposition rates under real ecological conditions.

Humidity-dependent function was proposed by Gong [21]. Although his research focused on the Culex mosquito, the oviposition rate was proposed based on the broader behavioral patterns of mosquito populations. Based on the effect of rainfall on mosquito dynamics, Gong hypothesized that there is a positive correlation between oviposition rate and humidity index and formulated it through a sigmoid function. The oviposition rate computed by this function increases exponentially with rising humidity, gradually reaches saturation, and then remains stable. However, this function's independent variable is humidity, whereas our function's independent variables are temperature and precipitation. The difference in the independent variables leads to the incomparability of the two functions. However, a simple analysis of oviposition rate trends can be conducted. Gong's findings indicate that rainfall positively impacts mosquito oviposition. But some scholars [12] argue that excessive precipitation may

Table 4. Comparison of oviposition rate functions.

Variable	Function	Source
Constant	—	Erickson [18], Tran [25]
Temperature	$B(T) = b_0 + b_1 \times T + b_2 \times T^2 + b_3 \times T^3 + b_4 \times T^4$ $B(T) = b_0 + b_1 \times T + b_2 \times T^2 + b_3 \times T^3$	Yang [28], Jia [9], Fukui [17]
Temperature	$B(T) = ae^{-b\frac{(T-T_b)^2}{v_b}(T_d-T)^{1.5}}$	Metelmann [51], Wang [64]
Humidity	$B(W) = \frac{\bar{F}_{max}}{1+e^{-\frac{W-\bar{W}_{mean}}{F_{var}}}} + b_0$	Gong [21]
Temperature, precipitation	$B(T, R) = \alpha u(T)v(R)$ $u(T) = e^{-a(T-T_b)^2}, v(R) = \frac{(1+s)e^{-r(R-R_b)^2}}{e^{-r(R-R_b)^2} + s}$	Abdelrazec [65], Liu [66]

<https://doi.org/10.1371/journal.pcbi.1012499.t004>

negatively affect oviposition, disrupting activities such as blood feeding, host seeking, and the search for oviposition sites. Our results show a decrease in oviposition rates with excessive precipitation, supporting this viewpoint. Additionally, if we regard the humidity index as a function of precipitation and temperature, then our function closely resembles Gong's function in form.

In addition, a limited number of scholars have indeed considered both temperature and precipitation in the oviposition rate function. For instance, Abdelrazec et al. [65] separately modeled the effects of temperature and precipitation and then multiplied these functions to represent their combined impact. After selecting appropriate parameters, this method shows a nearly linear increase in oviposition rate with rising temperatures. At a constant temperature, oviposition rates exhibit an initial increase followed by a decrease with increasing precipitation, which is broadly consistent with our results. However, this approach has limitations because it assumes temperature and precipitation act independently in the oviposition process. Under this formulation, the optimal precipitation level remains constant across different temperatures, and the optimal temperature remains constant across varying precipitation conditions, which contradicts biological knowledge. In contrast, our model incorporates interaction terms between temperature and precipitation, demonstrating that their effects on oviposition rates are interdependent, which aligns more closely with biological principles.

In summary, the trends of oviposition rate with changes in temperature and precipitation in our study closely align with previous research findings. Our function integrates the combined effects of temperature and precipitation on oviposition rate, revealing their complex interactions that better reflect biological behaviors. Our research can serve as a reference for future studies and offer new perspectives on the application of neural networks in biological dynamics.

Discussion and conclusion

Integrating ideas derived from machine learning with traditional engineering-related methods facilitates the incorporation of domain knowledge and known physical information into the learning process [67]. Recent studies have shown the feasibility of merging differential equations with neural networks [34]. For example, Dandekar et al. developed an SIR model for disease transmission and integrated neural networks to quantify the impact of government control measures [38]. However, this approach has not received significant attention. In fact, coupling neural networks with mechanistic models and using neural networks to explore unknown mechanisms is highly valuable. In this paper, we apply this concept to population dynamics, focusing on using a coupled model of mechanistic model and neural network to simulate changes in mosquito population abundance and explore the relationships between oviposition behavior, temperature, and precipitation. This approach allows for more accurate simulations of mosquito dynamics while also enhancing the biological interpretability of our model.

In this work, we first constructed a two-stage mosquito model based on the life cycle, including immatures and adults. This approach is consistent with traditional dynamics models, with differential equations describing intrinsic mechanisms such as growth and reproduction. Secondly, recognizing that the oviposition process is influenced by both temperature and precipitation, yet existing studies rarely offer a specific function to describe this relationship, we employed a neural network to approximate the oviposition function and embedded it into the dynamic equations. This approach resulted in a coupled model that combines the strengths of both neural networks and traditional mechanistic models. Subsequently, data from nine cities in Guangdong Province were used for model training, and data from three cities were used

as testing sets for model verification, which allowed us to obtain a set of biological and network parameters applicable to all cities, ensuring the model's generalizability.

The simulated results of our model closely aligned with the observed data. Compared with previous studies [9,19,31,51], our model simultaneously fitted data from nine cities, with correlation coefficients above 0.7 for all cities (shown in Figs 5 and 6). Particularly in Heyuan, Shenzhen, and Guangzhou, the coefficients exceed 0.85, indicating highly satisfactory simulation performance. Moreover, evaluations on the testing set reveal correlation coefficients exceeding 0.8 for Zhuhai and Huizhou, further confirming the robustness and effectiveness of the model. However, deviations between simulated and actual values are noticeable in some localized positions. In fact, from a modeling perspective, while neural networks can enhance predictive capabilities, attempting to fit data from all cities with the same set of parameters inevitably requires sacrificing some local fitting accuracy to achieve better overall fitting performance. Additionally, several factors contribute to deviations in simulation results. Since the initial number of immature mosquitoes cannot be precisely determined, we followed Tran's initial value settings [25] and advanced the simulation start time to 2015. Nevertheless, this adjustment may still impact simulation results in 2016. Furthermore, inherent uncertainties and stochastic fluctuations in observational data also affect the model's performance. Variations in the standardization and accuracy of data monitoring procedures among different regions contribute to heightened data uncertainty. Moreover, differences in the timing and intensity of mosquito control measures implemented by various cities in response to population increases could affect mosquito data, thereby influencing the effectiveness of our model.

The inputs and outputs of the neural network revealed the relationship between oviposition rate, temperature, and precipitation (shown in Figs 7 and 8). Under varying levels of precipitation, the trend of oviposition rate with temperature changes closely resembled an S-shaped curve, indicating that optimal temperatures promote mosquito egg-laying, while excessively high temperatures may have adverse effects. The impact of precipitation on the oviposition rate is also nonlinear, with the oviposition rate exhibiting a parabolic trend in response to precipitation changes. Optimal precipitation levels provide more breeding sites, stimulating mosquito egg-laying. However, excessive precipitation may negatively impact oviposition. Both low and high levels of precipitation can alter the threshold temperature required for a significant increase in mosquito oviposition rates. Additionally, different temperatures may also influence the precipitation levels needed for the oviposition rate to reach its peak and then decline from its maximum.

In addition, based on the results of the neural network, we employed symbolic regression to search for the optimal expression of the oviposition rate that depends on temperature and precipitation. The expression we obtained closely corresponded to the neural network trained previously (shown in Fig 9). The simulation results of the mechanistic model established using this oviposition rate function were highly consistent with our prior simulations and aligned well with the trends of the MOI, validating the effectiveness of the symbolic regression results. Furthermore, we compared and discussed the obtained expression with those used in previous studies. In contrast, our function integrates both temperature and precipitation factors affecting oviposition, emphasizing their coordinated influence and thereby providing a more accurate reflection of biological behaviors in practice.

Recently, deep neural networks have shown remarkable effectiveness in addressing traditionally challenging problems such as image recognition and natural language processing. In studies on population dynamics of mosquitoes, some scholars have achieved promising results using machine learning methods. However, these studies often directly employ factors such as temperature, precipitation, and sunlight as inputs, with mosquito abundance as the output, heavily relying on data for model training. The complexity and multidimensionality of these

algorithms cause models to function as black-boxes, which makes it challenging to interpret the various physiological mechanisms involved in the simulation process. Traditional population dynamics models describe a range of biological mechanisms, but the functional forms involved such as oviposition rate, development rate, and mortality rates must be predetermined, which significantly limits the performance of dynamics models. The highlight of our model lies in the coupling of the mechanistic model and the neural network, which allows it to adhere to biological mechanisms while mining trends in data, thus effectively overcoming these limitations. This method also allows us to capture information about reproduction, development, and other processes directly from data. Consequently, for less understood mechanisms, we can approximate them using neural networks, without the need for predefined functions. This coupling approach enhances the biological interpretability of the model and provides new insights into the dynamics of complex systems.

In summary, we developed a modeling method for population dynamics that integrates mechanistic models with machine learning techniques, and accurately modeled the trends of *Aedes* mosquitoes. Our study also explored how the oviposition rate is influenced by the combined effects of temperature and precipitation and provided a precise mathematical expression. However, our model also has some limitations. Firstly, the *Aedes* mosquito data we have collected actually includes information on both *Aedes albopictus* and *Aedes aegypti*, and we are unable to distinguish data for each species individually. Since these two species have different thermal responses, exploring the specific effects of temperature and rainfall on each mosquito species still requires further research. Secondly, in natural environments, excessive rainfall can cause a decrease in temperature due to the interconnection between the two factors. As a result, scenarios with both high rainfall and relatively high temperatures are less common, so the network's learning outcomes should be treated as a reference. Additionally, our set of diapause is relatively simple, and we did not consider the effects of human interventions, such as government mosquito control measures. We intend to address these aspects in future work, but we believe that this method, combining dynamics models with deep learning, holds potential for broader applications.

Author Contributions

Conceptualization: Mengze Zhang, Xia Wang, Sanyi Tang.

Data curation: Mengze Zhang, Xia Wang.

Formal analysis: Mengze Zhang.

Funding acquisition: Xia Wang, Sanyi Tang.

Investigation: Mengze Zhang, Xia Wang.

Methodology: Mengze Zhang, Xia Wang.

Project administration: Mengze Zhang, Xia Wang.

Resources: Mengze Zhang, Xia Wang.

Software: Mengze Zhang.

Supervision: Xia Wang, Sanyi Tang.

Validation: Mengze Zhang, Xia Wang, Sanyi Tang.

Visualization: Mengze Zhang.

Writing – original draft: Mengze Zhang, Xia Wang, Sanyi Tang.

Writing – review & editing: Mengze Zhang, Xia Wang, Sanyi Tang.

References

1. Suarez-Ramirez CD, Duran-Vega MA, Sanchez C HM, Gonzalez-Mendoza M, Chang L, Marshall JM. Deep Learning Architectures Applied to Mosquito Count Regressions in US Datasets. In: Batyrshin I, Gelbukh A, Sidorov G, editors. Advances in Computational Intelligence. Cham: Springer; 2021. p. 199–212. https://doi.org/10.1007/978-3-030-89817-5_15
2. Vontas J, Kioulos E, Pavlidi N, Morou E, della Torre A, Ranson H. Insecticide resistance in the major dengue vectors *Aedes albopictus* and *Aedes aegypti*. *Pestic Biochem Physiol*. 2012; 104(2):126–131. <https://doi.org/10.1016/j.pestbp.2012.05.008>
3. Pardo-Araujo M, Eritja R, Alonso D, et al. Present and future suitability of invasive and urban vectors through an environmentally-driven Mosquito Reproduction Number. *bioRxiv*. 2024: 2024.05.31.596775. <https://doi.org/10.1101/2024.05.31.596775>
4. Mordecai EA, Cohen JM, Evans MV, Gudapati P, Johnson LR, Lippi CA, et al. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. *PLoS Negl Trop Dis*. 2017; 11(4):e0005568. <https://doi.org/10.1371/journal.pntd.0005568> PMID: 28448507
5. Reinhold JM, Lazzari CR, Lahondere C. Effects of the Environmental Temperature on *Aedes aegypti* and *Aedes albopictus* Mosquitoes: A Review. *Insects*. 2018; 9(4):158. <https://doi.org/10.3390/insects9040158> PMID: 30404142
6. Delatte H, Gimmonneau G, Triboire A, Fontenille D. Influence of temperature on immature development, survival, longevity, fecundity, and gonotrophic cycles of *Aedes albopictus*, vector of chikungunya and dengue in the Indian Ocean. *J Med Entomol*. 2009; 46(1):33–41. <https://doi.org/10.1603/033.046.0105> PMID: 19198515
7. Brady OJ, Johansson MA, Guerra CA, Bhatt S, Golding N, Pigott DM, et al. Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures in laboratory and field settings. *Parasit Vectors*. 2013; 6:351. <https://doi.org/10.1186/1756-3305-6-351> PMID: 24330720
8. Teng HJ, Apperson CS. Development and survival of immature *Aedes albopictus* and *Aedes triseriatus* (Diptera: Culicidae) in the laboratory: effects of density, food, and competition on response to temperature. *J Med Entomol*. 2000; 37(1):40–52. <https://doi.org/10.1603/0022-2585-37.1.40> PMID: 15218906
9. Jia P, Lu L, Chen X, Chen J, Guo L, Yu X, et al. A climate-driven mechanistic population model of *Aedes albopictus* with diapause. *Parasit Vectors*. 2016; 9:175. <https://doi.org/10.1186/s13071-016-1448-y> PMID: 27009065
10. Nagao Y, Thavara U, Chitnumsup P, Tawatsin A, Chansang C, Campbell-Lendrum D. Climatic and social risk factors for *Aedes* infestation in rural Thailand. *Trop Med Int Health*. 2003; 8(7):650–659. <https://doi.org/10.1046/j.1365-3156.2003.01075.x> PMID: 12828549
11. Oidtmann RJ, Lai S, Huang Z, Yang J, Siraj AS, Reiner RC Jr, et al. Inter-annual variation in seasonal dengue epidemics driven by multiple interacting factors in Guangzhou, China. *Nat Commun*. 2019; 10(1):1148. <https://doi.org/10.1038/s41467-019-09035-x> PMID: 30850598
12. Roiz D, Rosa R, Arnoldi D, Rizzoli A. Effects of temperature and rainfall on the activity and dynamics of host-seeking *Aedes albopictus* females in northern Italy. *Vector Borne Zoonotic Dis*. 2010; 10(8):811–816. <https://doi.org/10.1089/vbz.2009.0098> PMID: 20059318
13. Xu L, Stige LC, Chan KS, Zhou J, Yang J, Sang S, et al. Climate variation drives dengue dynamics. *Proc Natl Acad Sci U S A*. 2017; 114(1):113–118. <https://doi.org/10.1073/pnas.1618558114> PMID: 27940911
14. Ewing DA, Cobbold CA, Purse BV, Nunn MA, White SM. Modelling the effect of temperature on the seasonal population dynamics of temperate mosquitoes. *J Theor Biol*. 2016; 400:65–79. <https://doi.org/10.1016/j.jtbi.2016.04.008> PMID: 27084359
15. Huang M, Tang M, Yu J, Zheng B. A stage structured model of delay differential equations for *Aedes* mosquito population suppression. *Discrete & Continuous Dynamical Systems*. 2020; 40(6):3467–3484. <https://doi.org/10.3934/dcds.2020042>
16. Liu Y, Jiao F, Hu L. Modeling mosquito population control by a coupled system. *J Math Anal Appl*. 2022; 506(2):125671. <https://doi.org/10.1016/j.jmaa.2021.125671>
17. Fukui S, Kuwano Y, Ueno K, Atsumi K, Ohta S. Modeling the effect of rainfall changes to predict population dynamics of the Asian tiger mosquito *Aedes albopictus* under future climate conditions. *PLoS One*. 2022; 17(5):e0268211. <https://doi.org/10.1371/journal.pone.0268211> PMID: 35613220
18. Erickson RA, Presley SM, Allen LJS, Long KR, Cox SB. A stage-structured, *Aedes albopictus* population model. *Ecol Model*. 2010; 221(9):1273–1282. <https://doi.org/10.1016/j.ecolmodel.2010.01.018>

19. Cailly P, Tran A, Balenghien T, L'Ambert G, Toty C, Ezanno P. A climate-driven abundance model to assess mosquito control strategies. *Ecol Model.* 2012; 227:7–17. <https://doi.org/10.1016/j.ecolmodel.2011.10.027>
20. Focks DA, Haile DG, Daniels E, Mount GA. Dynamic life table model for *Aedes aegypti* (Diptera: Culicidae): analysis of the literature and model development. *J Med Entomol.* 1993; 30(6):1003–1017. <https://doi.org/10.1093/jmedent/30.6.1003> PMID: 8271242
21. Gong H, DeGaetano AT, Harrington LC. Climate-based models for West Nile Culex mosquito vectors in the Northeastern US. *Int J Biometeorol.* 2011; 55(3):435–446. <https://doi.org/10.1007/s00484-010-0354-9> PMID: 20821026
22. Otero M, Solari HG, Schweigmann N. A stochastic population dynamics model for *Aedes aegypti*: formulation and application to a city with temperate climate. *Bull Math Biol.* 2006; 68(8):1945–1974. <https://doi.org/10.1007/s11538-006-9067-y> PMID: 16832731
23. Rueda LM, Patel KJ, Axtell RC, Stinner RE. Temperature-dependent development and survival rates of *Culex quinquefasciatus* and *Aedes aegypti* (Diptera: Culicidae). *J Med Entomol.* 1990; 27(5):892–898. <https://doi.org/10.1093/jmedent/27.5.892> PMID: 2231624
24. Shaman J, Spiegelman M, Cane M, Stieglitz M. A hydrologically driven model of swamp water mosquito population dynamics. *Ecol Model.* 2006; 194(4):395–404. <https://doi.org/10.1016/j.ecolmodel.2005.10.037>
25. Tran A, L'Ambert G, Lacour G, Benoit R, Demarchi M, Cros M, et al. A rainfall- and temperature-driven abundance model for *Aedes albopictus* populations. *Int J Environ Res Public Health.* 2013; 10(5):1698–1719. <https://doi.org/10.3390/ijerph10051698> PMID: 23624579
26. Wang X, Tang S, Cheke RA. A stage structured mosquito model incorporating effects of precipitation and daily temperature fluctuations. *J Theor Biol.* 2016; 411:27–36. <https://doi.org/10.1016/j.jtbi.2016.09.015> PMID: 27693525
27. Shaman J, Day JF. Reproductive phase locking of mosquito populations in response to rainfall frequency. *PLoS One.* 2007; 2(3):e331. <https://doi.org/10.1371/journal.pone.0000331> PMID: 17396162
28. Yang HM, Macoris MLG, Galvani KC, Andrigatti MTM, Wanderley DMV. Assessing the effects of temperature on the population of *Aedes aegypti*, the vector of dengue. *Epidemiol Infect.* 2009; 137(8):1188–1202. <https://doi.org/10.1017/S0950268809002040> PMID: 19192322
29. Ceia-Hasse A, Sousa CA, Gouveia BR, Capinha C. Forecasting the abundance of disease vectors with deep learning. *Ecol Inform.* 2023; 78:102272. <https://doi.org/10.1016/j.ecoinf.2023.102272>
30. Kinney AC, Current S, Lega J. Aedes-AI: Neural network models of mosquito abundance. *PLoS Comput Biol.* 2021; 17(11):e1009467. <https://doi.org/10.1371/journal.pcbi.1009467> PMID: 34797822
31. Lee KY, Chung N, Hwang S. Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas. *Ecol Inform.* 2016; 36:172–180. <https://doi.org/10.1016/j.ecoinf.2015.08.011>
32. Mudele O, Frery A, Zanandrez L, Eiras A, Gamba P. Dengue Vector Population Forecasting Using Multisource Earth Observation Products and Recurrent Neural Networks. *IEEE J-STARS.* 2021; 14:4390–4404. <https://doi.org/10.1109/jstars.2021.3073351>
33. Song P, Xiao Y, Wu J. Discovering First Principle of Behavioural Change in Disease Transmission Dynamics by Deep Learning. In: David J, Wu J, editors. *Mathematics of Public Health: Mathematical Modelling from the Next Generation.* Cham: Springer; 2023. p. 37–54. https://doi.org/10.1007/978-3-031-40805-2_2
34. Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys.* 2019; 378:686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>
35. Nogueira IBR, Santana VV, Ribeiro AM, Rodrigues AE. Using scientific machine learning to develop universal differential equation for multicomponent adsorption separation systems. *Can J Chem Eng.* 2022; 100(9):2279–2290. <https://doi.org/10.1002/cjce.24495>
36. Rackauckas C, Ma Y, Martensen J, Warner C, Zubov KV, Supekar RB, et al. Universal Differential Equations for Scientific Machine Learning. *arXiv preprint arXiv:2001.04385.* 2020.
37. Silvestri M, Baldo F, Misino E, Lombardi M. An Analysis of Universal Differential Equations for Data-Driven Discovery of Ordinary Differential Equations. In: International Conference on Computational Science. Cham: Springer Nature Switzerland; 2023. p. 353–366. https://doi.org/10.1007/978-3-031-36027-5_27
38. Dandekar R, Rackauckas C, Barbastathis G. A Machine Learning-Aided Global Diagnostic and Comparative Tool to Assess Effect of Quarantine Control in COVID-19 Spread. *Patterns.* 2020; 1(9):100145. <https://doi.org/10.1016/j.patter.2020.100145> PMID: 33225319

39. He M, Tang S, Xiao Y. Combining the dynamic model and deep neural networks to identify the intensity of interventions during COVID-19 pandemic. PLoS Comput Biol. 2023; 19(10):e1011535. <https://doi.org/10.1371/journal.pcbi.1011535> PMID: 37851640
40. Brunton SL, Proctor JL, Kutz JN. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc Natl Acad Sci U S A. 2016; 113(15):3932–3937. <https://doi.org/10.1073/pnas.1517384113> PMID: 27035946
41. Tu JH, Rowley CW, Luchtenburg DM, Brunton SL, Kutz JN. On dynamic mode decomposition: Theory and applications. J Comput Dyn. 2014; 1(2):391–421. <https://doi.org/10.3934/jcd.2014.1.391>
42. Kaheman K, Kutz JN, Brunton SL. SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics. Proc Math Phys Eng Sci. 2020; 476(2242):20200279. <https://doi.org/10.1098/rspa.2020.0279> PMID: 33214760
43. Schaeffer H. Learning partial differential equations via data discovery and sparse optimization. Proc Math Phys Eng Sci. 2017; 473(2197):20160446. <https://doi.org/10.1098/rspa.2016.0446> PMID: 28265183
44. Li M, Zhang X, Ding W, Ma J. Estimating the Mosquito Density in Guangzhou City, China. J Appl Anal Comput. 2023; 13(1):329–343. <https://doi.org/10.11948/20220129>
45. Sharpe PJ, Curry GL, DeMichele DW, Cole CL. Distribution model of organism development times. J Theor Biol 1977; 66(1):21–38. [https://doi.org/10.1016/0022-5193\(77\)90309-5](https://doi.org/10.1016/0022-5193(77)90309-5) PMID: 875419
46. Denlinger DL. Insect diapause. UK: Cambridge University Press; 2022.
47. Lou Y, Liu K, He D, Gao D, Ruan S. Modelling diapause in mosquito population growth. J Math Biol. 2019; 78(7):2259–2288. <https://doi.org/10.1007/s00285-019-01343-6> PMID: 30847501
48. Liu Y, Wang X, Tang S, Cheke RA. The relative importance of key meteorological factors affecting numbers of mosquito vectors of dengue fever. PLoS Negl Trop Dis. 2023; 17(4):e0011247. <https://doi.org/10.1371/journal.pntd.0011247> PMID: 37053307.
49. Hanson SM, Craig GB Jr. Aedes albopictus (Diptera: Culicidae) eggs: field survivorship during northern Indiana winters. J Med Entomol. 1995; 32(5):599–604. <https://doi.org/10.1093/jmedent/32.5.599> PMID: 7473614
50. Hawley WA, Pumpuni CB, Brady RH, Craig GB Jr. Overwintering survival of Aedes albopictus (Diptera: Culicidae) eggs in Indiana. J Med Entomol. 1989; 26(2):122–129. <https://doi.org/10.1093/jmedent/26.2.122> PMID: 2709388
51. Metelmann S, Caminade C, Jones A, Medlock J, Baylis M, Morse A. The UK's suitability for Aedes albopictus in current and future climates. J R Soc Interface. 2019; 16(152):20180761. <https://doi.org/10.1098/rsif.2018.0761> PMID: 30862279
52. Cybenko G. Approximation by superpositions of a sigmoidal function. MCSS. 1989; 2(4):303–314. <https://doi.org/10.1007/bf02551274>
53. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw. 1989; 2(5):359–366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
54. Wang Y, Wagner N, Rondinelli JM. Symbolic regression in materials science. MRS Communications. 2019; 9(3):793–805. <https://doi.org/10.1557/mrc.2019.85>
55. Schmidt M, Lipson H. Distilling free-form natural laws from experimental data. Science. 2009; 324(5923):81–85. <https://doi.org/10.1126/science.1165893> PMID: 19342586
56. Quade M, Abel M, Shafi K, Niven RK, Noack BR. Prediction of dynamical systems by symbolic regression. Phys Rev E. 2016; 94(1):012214. <https://doi.org/10.1103/PhysRevE.94.012214> PMID: 27575130
57. Yang G, Li X, Wang J, Lian L, Ma T. Modeling oil production based on symbolic regression. Energy Policy. 2015; 82:48–61. <https://doi.org/10.1016/j.enpol.2015.02.016>
58. Lührsen DS, Zavitsanou E, Cerecedo-Iglesias C, Pardo-Araujo M, Palmer JR, Bartumeus F, et al. Adult Aedes albopictus in winter: implications for mosquito surveillance in southern Europe. The Lancet Planetary Health, 2023, 7(9): e729–e731. [https://doi.org/10.1016/S2542-5196\(23\)00170-5](https://doi.org/10.1016/S2542-5196(23)00170-5) PMID: 37673540
59. Estrada-Franco JG, Craig GB. Biology, disease relationships, and control of Aedes albopictus. Washington, D.C: Pan American Health Organization; 1995.
60. Ho BC, Chan KL, Chan YC. Aedes aegypti (L.) and Aedes albopictus (Skuse) in Singapore City. 3. Population fluctuations. Bull World Health Organ. 1971; 44(5):635–641. PMID: 5316747
61. Richards SL, Apperson CS, Ghosh SK, Cheshire HM, Zeichner BC. Spatial Analysis of Aedes albopictus(Diptera: Culicidae) Oviposition in Suburban Neighborhoods of a Piedmont Community in North Carolina. J Med Entomol. 2006; 43(5):976–989. [https://doi.org/10.1603/0022-2585\(2006\)43\[976:saoaad\]2.0.co;2](https://doi.org/10.1603/0022-2585(2006)43[976:saoaad]2.0.co;2) PMID: 17017237

62. Hassan AB, Adanan CR, Rahman WA. Patterns in *Aedes albopictus* (Skuse) population density, host-seeking, and oviposition behavior in Penang, Malaysia. *J Vector Ecol.* 1996; 21:17–21.
63. Koenraadt CJM, Harrington LC. Flushing Effect of Rain on Container-Inhabiting Mosquitoes *Aedes aegypti* and *Culex pipiens* (Diptera: Culicidae). *J Med Entomol.* 2008; 45(1):28–35. [https://doi.org/10.1603/0022-2585\(2008\)45\[28:feoroc\]2.0.co;2](https://doi.org/10.1603/0022-2585(2008)45[28:feoroc]2.0.co;2) PMID: 18283939
64. Wang Y, Li Y, Ren X, Liu X. A periodic dengue model with diapause effect and control measures. *Appl Math Model.* 2022; 108:469–488. <https://doi.org/10.1016/j.apm.2022.03.043>
65. Abdelrazec A, Gumel AB. Mathematical assessment of the role of temperature and rainfall on mosquito population dynamics. *J Math Biol.* 2016; 74(6):1351–1395. <https://doi.org/10.1007/s00285-016-1054-9> PMID: 27647127
66. Liu K, Fang S, Li Q, Lou Y. Effectiveness evaluation of mosquito suppression strategies on dengue transmission under changing temperature and precipitation. *Acta Tropica.* 2024; 253. <https://doi.org/10.1016/j.actatropica.2024.107159> PMID: 38412904
67. Baker N, Alexander F, Bremer T, Hagberg A, Kevrekidis Y, Najm H, et al. Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence. United States: USDOE Office of Science, Washington, D.C; 2019. Sponsored by USDOE Office of Science, Advanced Scientific Computing Research. <https://doi.org/10.2172/1478744>