

# The effects of seasonal climate variability on dengue annual incidence in Hong Kong: A modelling study

Hsiang-Yu Yuan<sup>1,\*</sup>, Jingbo Liang<sup>1</sup>, Pei-Sheng Lin<sup>2,3</sup>, Kathleen Sucipto<sup>1</sup>, Mesfin Mengesha Tsegaye<sup>1</sup>, Tzai-Hung Wen<sup>4</sup>, Susanne Pfeiffer<sup>5</sup>, and Dirk Pfeiffer<sup>6</sup>

<sup>1</sup>Department of Biomedical Sciences, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong

<sup>2</sup>Institute of Population Health Sciences, National Health Research Institutes, Taiwan

<sup>3</sup>National Mosquito-borne Disease Control Research Center, National Health Research Institutes, Taiwan

<sup>4</sup>Department of Geography, National Taiwan University, Taiwan

<sup>5</sup>Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong

<sup>6</sup>Centre for Applied One Health Research and Policy Advice, City University of Hong Kong, Hong Kong

\*sean.yuan@cityu.edu.hk

## ABSTRACT

In recent years, dengue has been rapidly spreading and growing in the tropics and subtropics. Located in southern China, Hong Kong's subtropical monsoon climate may favour dengue vector populations and increase the chance of disease transmissions during the rainy summer season. An increase in local dengue incidence has been observed in Hong Kong ever since the first case in 2002, with an outbreak reaching historically high case numbers in 2018. However, the effects of seasonal climate variability on recent outbreaks are unknown. As the local cases were found to be spatially clustered, we developed a Poisson generalized linear mixed model using pre-summer monthly total rainfall and mean temperature to predict annual dengue incidence (the majority of local cases occur during or after the summer months), over the period 2002-2018 in three pre-defined areas of Hong Kong. Using leave-one-out cross-validation, 5 out of 6 observations of area-specific outbreaks during the major outbreak years 2002 and 2018 were able to be predicted. 42 out of a total of 51 observations (82.4%) were within the 95% confidence interval of the annual incidence predicted by our model. Our study found that the rainfall before and during the East Asian monsoon (pre-summer) rainy season is negatively correlated with the annual incidence in Hong Kong while the temperature is positively correlated. Hence, as mosquito control measures in Hong Kong are intensified mainly when heavy rainfalls occur during or close to summer, our study suggests that a lower-than-average intensity of pre-summer rainfall should also be taken into account as an indicator of increased dengue risk.

## Introduction

Dengue is one of the most prevalent vector-borne arboviruses, transmitted between humans by *Aedes* mosquitoes. The number of dengue cases reported to the World Health Organization (WHO) has increased steadily from an average of less than a thousand cases per year globally in the 1950s to over 3.34 million in 2016<sup>1,2</sup>. Moreover, a recent study estimated that 390 million dengue infections occur each year<sup>3</sup> while another indicated that 3.9 billion people in 128 countries are at risk of infection<sup>4</sup>. The number of cases is still increasing as the disease spreads to new areas and explosive outbreaks occur. This global expansion of dengue virus (DENV) can be due to several causes, including environmental and climate factors. Climate is an important determinant of vector-borne disease epidemics as it directly influences the abundance and distribution of the vector. Together with environmental factors such as the degree of urbanisation, climate variations may result in the geographic expansion of mosquitoes and dengue into new areas<sup>5</sup>.

Studies have shown that a positive association exists between temperature and dengue transmission<sup>6-9</sup>. One possible reason is that warmer temperatures lead to a higher abundance of *Aedes* mosquitoes by increasing their survival and development rates<sup>10,11</sup>. Furthermore, the extrinsic incubation period (EIP; the time required for viruses to become transmissible after the initial infection of a mosquito) of DENV may be shortened under warmer conditions<sup>12,13</sup>, resulting in increased efficiency of viral transmission by allowing more mosquitoes to transmit the virus after exposure. The effect of rainfall amount on dengue

incidence, on the other hand, is more controversial. Rainfall has been generally thought to be a positive predictor of dengue vector abundance as it provides essential mosquito habitats<sup>11,14</sup>. Recent studies, however, have demonstrated that excess rain may also flush away eggs, larvae or pupae of mosquitoes and remove mosquito breeding sites<sup>15,16</sup>, hence negatively effecting dengue incidence. Another study has found that dry seasons followed by a period of excess rainfall can lead to a high risk of dengue outbreaks, indicating that either positive or negative effects can occur, however, with different time lags<sup>17</sup>. Therefore, a better understanding of how rainfall affects dengue infection rates in areas with large outbreaks or increasing incidence could provide better projections of climate effects on dengue expansion.

The South-East Asia and Western Pacific regions, two of the six WHO regions, together have accounted for nearly 75% of the current global burden of dengue infections<sup>1</sup>. A rapid increase of dengue incidence has been recently observed near the centre part of the Western Pacific region, such as in Guangzhou (the capital city of Guangdong province in southern China) and southern Taiwan<sup>18,19</sup>. Hong Kong, located geographically close to both Taiwan and Guangzhou, has a typical East Asian subtropical monsoon climate, characterised by higher temperatures and copious amounts of rain in the spring and summer months<sup>20</sup>, which may favour dengue vector abundance and cause outbreaks later during or after summer. Despite the successful dengue control measures introduced after the first local case was identified in Hong Kong in 2002<sup>21</sup>, there has been a gradual increase in dengue incidence after 2014 with the highest number of dengue so far recorded in 2018. Identifying the role of climate variables, particularly the effects of rainfall, on recent distribution trends in Hong Kong is urgently needed in order to understand the recent large outbreak and geographical expansion of dengue in the Western Pacific region or globally. However, the effects of pre-summer rainfall on dengue incidence around southern China are often ignored since the rainfall season normally has a lead time up to three to five months before the outbreaks. These lag effects were often not considered in many of the previous studies in this region<sup>22–25</sup>.

This study aims to enable prediction the annual number of dengue cases using seasonal monthly temperature and rainfall data before summer and assessing the effects of climate variability on the recent increase of dengue incidence in Hong Kong. Additionally, given that all dengue outbreaks/infections in Hong Kong occurred as either local clusters or sporadic cases in certain areas instead of large regional scale epidemics, the study utilised a generalized linear mixed model that contains area-specific random effects to account for environmental factors which are difficult to measure and fixed effects to account for climate effects on dengue transmission. Understanding how the climate may affect annual dengue incidence is important in order to forecast future possible dengue outbreaks, thereby allowing for more effective mosquito control measures in each region, leading to early preparedness and prevention of dengue.

## Results

### Area-Specific Dengue Cases in Hong Kong

After the first local case of dengue fever was discovered in 2002, a large dengue outbreak occurred in the same year and another in 2018 while sporadic cases mainly occurred between 2014 and 2017. We divided Hong Kong into 3 areas: i) New Territories-South (NTS), ii) New Territories-North (NTN), and iii) Hong Kong Island & Kowloon (HKL), based on the administrative districts, the population density, and the usage of the Mass Transit Railways (see Methods). The existence of spatial heterogeneity of dengue infections was observed between the defined areas (Fig. 1 and its inset). For example, HKL and NTS together consisted of all the infections during 2018 outbreak, while HKL and NTN contained all the sporadic cases from 2014 to 2017. Note the special case of the year 2002 when 17 out of 20 locally acquired dengue cases, which included one case transmitted via blood transfusion, were all located around a construction site on an island geographically at the centre of the three areas (shown in red in Fig. 1). After the blood transfusion case was removed, the remaining cases were equally assigned to the three areas with one extra case assigned to NTN.

### Climate Predictors Selection

In order to select suitable climate parameters for area-specific annual dengue forecasting, we compared 4 models with different combinations of temperature (either monthly mean or minimum; indicated by  $T_{mean}$  or  $T_{min}$ ) and rainfall (either monthly total or maximum; indicated by  $R_{tot}$  or  $R_{max}$ ). These climate parameters were suggested by the prior studies conducted in nearby countries<sup>8,26,27</sup>, but their association with recent dengue outbreaks in Hong Kong is still unknown and the effects of rainfall on dengue spreading are mixed. Furthermore, spatial differences in the climate data were not considered when the effects of monthly climate predictors were analysed<sup>8</sup>. Thus, all the climate data within the studied years from 2002 to 2018 were retrieved from 11 weather stations and 2 automatic rainfall stations across the defined areas (white and yellow circles Fig. 1), with the

averages from January to August shown in Fig. 2. In general, the highest temperatures and the highest rainfalls were recorded between June and August and after April, respectively.

Since most of the local (indigenous) dengue cases occurred during the summer seasons starting in or after August (Fig. 3) and in Hong Kong mosquitoes are not commonly observed during the winter until March<sup>28,29</sup>, for each combination of climate parameters we used monthly climate predictors (Fig. 4) from March to August in the 3 defined areas to forecast area-specific annual dengue cases using a Poisson mixed effects model. The predictors were first normalised in order to compare the effects between each predictor. The stepwise algorithm based on the Akaike Information Criterion with a correction (AICc) was used to select the monthly predictors for each model (Supplementary Table S1). The best-fitting model was determined according to both the AICc and Bayesian Information Criterion (BIC).

## Area-specific Annual Dengue Forecasting

The results showed that the models with the  $T_{mean}$  climate parameters set performed better than the models with the  $T_{min}$  parameters (Table 1). The best-fitting model, Model:  $T_{mean} + R_{tot}$  was used as our predictive model for dengue cases while Model:  $T_{mean} + R_{max}$  was used as an alternative model as both the  $\Delta AICc$  and  $\Delta BIC$  results were equal to only 1 for Model:  $T_{mean} + R_{max}$ . The best-fitting model estimated using the Laplace approximation is given by:

$$\log(\mu^i) = -6.2 + 3.594 \times T_3^i + 2.542 \times T_4^i + 1.21 \times T_5^i + 10.072 \times T_7^i - 4.002 \times R_4^i - 4.029 \times R_5^i - 8.047 \times R_6^i + \alpha^i \quad (1)$$

where the standard deviation of the random intercepts  $\sigma = 0.789$ . A significance test was performed for each fixed effect predictor using the likelihood ratio test. P-values were less than 0.05 for 8 out of 9 regression coefficients (Table 2). The random intercepts represent different levels of effects resulting from unobserved area-specific factors, which may include varying land usages or population densities between different areas.

The model performance was evaluated using leave-one-out cross-validation (LOOCV) and leave-one-year-out (also called leave-three-out) cross-validation (LOYOCV). For LOOCV, the observation in one area in a given year was removed before refitting the model. The best-fitting model successfully predicted the major outbreaks by year and by area using LOOCV (Fig. 5a-c, supplementary Table S2 and supplementary dataset S1 online). 42 out of a total of 51 (82.4%) total observations by year and by area (annual incidence rates in 17 years and 3 areas) were within the 95% confidence interval of the annual incidence predicted by our model. Note that the upper and the lower bounds of the confidence intervals were rounded to the nearest integer when assessing the prediction accuracy of observations in Table S2. 5 out of 6 observations of area-specific outbreaks in 2002 and 2018 were able to be predicted. Although the 2018 outbreak in NTS was not predicted, the model still identified an outbreak (defined as the annual number of dengue cases  $> 2$ ) in that area. When LOYOCV was used, the model predicted outbreaks (defined as the annual number of dengue cases  $> 2$ ) in 5 years (i.e., the years 2002, 2003, 2007, 2015 and 2018), which include the two major outbreak years (2002 and 2018) and a year with sporadic cases (2015) (Fig. 5d).

The best-fitting model produced an  $MSE_{tr}$  of 0.592, an  $MSE_{va}$  of 3.538 and an  $MSE_{ratio}$  ( $MSE_{va}/MSE_{tr}$ ) of 5.976 using LOOCV (Table 3). Both the  $MSE_{va}$  and the  $MSE_{ratio}$  of the best-fitting model were the lowest compared with the alternative (Model:  $T_{mean} + R_{max}$ ), fixed effects (same predictors as the best-fitting model but without random effects), and full models, indicating the best prediction performance among all the tested models. The alternative model using  $T_{mean}$  and  $R_{max}$  produced an  $MSE_{tr}$  of 0.672 and an  $MSE_{va}$  of 5.143, which confirmed that the chosen best model fitted the data better than the alternative. The fixed effects model performed similarly to but slightly worse than the best-fitting model. The full model, including all predictors, had the lowest  $MSE_{tr}$  value but the highest  $MSE_{va}$  and the highest  $MSE_{ratio}$ , indicating an overfitting phenomenon. The best model also performed better than the other models when the Normalised Mean Squared Error (NMSE) was used. When LOYOCV was used, the  $MSE_{ratio}$  (12.629) of the best-fitting model was slightly higher than that obtained from LOOCV results. Although the fixed effects model can perform similarly or slightly better according to the  $MSE_{va}$  (or  $NMSE_{va}$ ), the best-fitting model performed better than the other models according to the  $MSE_{ratio}$  (or  $NMSE_{ratio}$ ) (Supplementary Table S3). These results demonstrate that while working with a small number of dengue observations in Hong Kong, a mixed effects model with AICc-selected variables can reduce model overfitting.

## Effects of Climate Variables

The predictive model showed that all monthly mean temperature predictors ( $T_3$ ,  $T_4$ ,  $T_5$ , and  $T_7$ ) were positively correlated with the number of annual cases, while monthly total rainfall predictors ( $R_4$ ,  $R_5$ , and  $R_6$ ) were all negatively correlated. These behaviors were also found in the stepAICc results of the other 3 models with different climate parameter sets (Supplementary

Table S1), confirming the positive correlations of the temperature and the negative correlations of the early rainfall predictors with the number of annual cases. Among all the temperature predictors,  $T_7$  (the mean temperature in July) was the most significant predictor ( $p < 10^{-8}$ ) and its regression coefficient had the highest magnitude, indicating the strongest effect. The coefficients of the 3 rainfall predictors (from April to June) were of similar magnitude, indicating a much longer delayed effect of rainfall compared to temperature. In the NTS and HKL areas, relatively high values for  $T_3$ ,  $T_5$ , and  $T_7$  and relatively low values for  $R_4$  and  $R_5$  were found in 2018, which may explain the outbreak in 2018 in those two areas.

We further examined the marginal effects of the 4 predictors ( $T_7$ ,  $R_4$ ,  $R_5$ , and  $R_6$ ) with the greatest magnitude of regression coefficients on the relative risk  $RR$  (Supplementary Fig. S1).  $RR$  was defined as the number of annual cases divided by the average number of annual cases. Marginal effects of the collected temperature (without normalization)  $T'_m$  and rainfall  $R'_m$  (without normalization) were also obtained after refitting the same model (Supplementary Fig. S2). If the rainfall is lower than 200mm (per month) between May and June, or lower than 100mm in April, a higher relative risk of dengue incidence is observed.

## Discussion

This study is the first to explore the association between seasonal climate variables and dengue incidence while taking into account the spatial heterogeneity of the dengue cases in Hong Kong. After dividing Hong Kong into 3 areas, a Poisson mixed effects model was fitted using the monthly temperature and rainfall data from the early months of the year to forecast annual dengue incidence in the defined areas. Given that most of the dengue infections occurred in August and September, the model successfully predicts annual dengue incidence using seasonal weather data from the preceding months. Instead of using a distributed lag model, our approach allows us to capture the impact of the weather in each month on the annual number of dengue cases, which is more suitable for regions affected by seasonal epidemics<sup>8</sup>.

Because most of the outbreaks occurred in August and September, the period of the delay of climate effects can be estimated. Our results showed that the temperature is positively associated with the relative risk within time lags of at least 1 or 2 months (from  $T_7$  to the start of the outbreak), which is consistent with previous studies<sup>17,30-34</sup>. These lagged effects could be caused by the incubation periods of dengue viruses and the mosquito's reproductive life cycle. The delayed effect of rainfall from between 3 and 5 months (from  $R_4$ ,  $R_5$  and  $R_6$  to the start of the outbreak) in our study is also consistent with recent findings that rainfall can be negatively associated with the risk of dengue outbreaks with a lead time up to 5 months<sup>17</sup>.

The three defined areas NTS, NTN, and HKL in our study were selected to reflect the spatial heterogeneity of the dengue cases in Hong Kong. This spatial heterogeneity may be caused by area-specific factors other than climate variables; for instance, urbanisation, transportation, population density, etc., could all contribute to the expansion of dengue. For the purpose of our study, as we mainly focus on the impact of climate, those area-specific factors were considered as random effects in our model. Introducing random effects into the model can lead to higher accuracy of parameter estimation especially when the dataset contains non-independent observational units that are hierarchical in nature<sup>35</sup>.

The significance of understanding the climate-dengue association lies in three aspects. First, being a highly populated region with more than 50 million people travelling to the city annually<sup>36</sup>, infectious diseases can spread rapidly not only within Hong Kong but also to other countries due to its high level of contact mixing within- and between-countries<sup>37-39</sup>. Therefore, whether an even larger dengue outbreak will occur in Hong Kong in the future has become a primary global health concern.

Second, we demonstrated that in addition to global warming, the variation in rainfall intensity associated with the East Asian monsoon may play a significant role in the recent dengue expansions in some East Asian subtropical countries/cities. An increasing risk of seasonal dengue epidemics has been observed in both southern Taiwan and Hong Kong in the recent years, and similar to our findings, a negative association between early rainfall and dengue infections has previously been identified in southern Taiwan<sup>8</sup>. As Hong Kong and Taiwan are both located in the East Asian subtropics, they are both affected by the East Asian rainy season occurring annually between February and June<sup>40</sup>. Our results suggest that the variability during or even before the East Asian monsoon (pre-summer) rainy season may pose a high risk for dengue expansion in the region and complex non-linear effects should be studied further. As for Guangzhou, we noticed that before the 2014 dengue outbreak period (the largest one in Guangzhou), heavy rainfall mixed with drought (or low initial water level) were observed in that particular year<sup>41,42</sup>.

Third, our findings on rainfall effects can be used to inform mosquito control strategies in Hong Kong and possibly in other nearby countries similarly affected by the East Asian monsoon. The authorities of Hong Kong often enhance mosquito control measures while increasing public awareness especially during periods of heavy rainfall<sup>43,44</sup>. Our results suggest that in this particular region, the relative risk of incidence is higher when the intensity of rainfall before summer is lower. This implies that

stepping up mosquito control measures to reduce mosquito activities is the most critical when the early rainfall amount is lower than the expected level.

In conclusion, we have developed a generalized linear mixed model that is able to successfully forecast annual dengue incidence across Hong Kong by taking into account both the climate data and the unobserved area-specific factors. Our findings may provide a valuable assessment on how to plan dengue prevention and control measures more effectively in Hong Kong.

## Methods

### Study Area

Hong Kong is officially divided into 18 districts, which are grouped into three main areas: Hong Kong Island, Kowloon, and the New Territories (Supplementary Fig. S3). Based on the geography, population density, and the usage of the Mass Transit Railways (MTR), which is the most heavily used public transportation in Hong Kong, we redivided the city into three main areas. We first divided the New Territories into two regions: New Territories-South (NTS) and New Territories-North (NTN) such that NTS consists of the smaller and outlying islands while NTN consists of the region connected to Mainland China. We then merged Hong Kong Island and Kowloon into Hong Kong Island & Kowloon (HKL) because of their high similarities in terms of population density (Supplementary Fig. S4) and public transportation network. Kwai Tsing District was reassigned from the New Territories into this area as it also has a high population density, is located close to Kowloon, and is linked to the area by the busiest MTR line<sup>45</sup>.

### Dengue Data

The confirmed numbers of local dengue fever cases from 2002 to 2018 were collected through press releases issued by the Centre for Health Protection (CHP) of the Department of Health of the Hong Kong Government<sup>46</sup>. The official statements indicate the location and the date of every local case confirmed through laboratory tests such as NS1 antigen and IgM serological assays, where the possible infection location is determined by either the residence or the travel history of the patient. The annual number of dengue cases was defined as the total number of confirmed local (indigenous) cases in each year (Fig. 1 inset); the onset of dengue fever symptoms in all years between 2002 and 2018 was observed between June and December (Fig. 3). Imported cases, confirmed by CHP, were excluded in order to rule out travel-related factors influencing infection rates.

### Meteorological Data

Both the daily mean temperature and the daily total rainfall from 2002 to 2018 inclusive were retrieved from the Hong Kong Observatory<sup>47</sup>. A total of 11 out of the available 49 automatic weather stations were selected to represent all districts in Hong Kong (Fig. 1). The stations were chosen as they provided the most complete temperature and rainfall data in those years; any weather stations with more than 30% missing data were excluded. The selected station with the highest number of missing data is in Tuen Mun District since it only started recording from 2007 onwards. Two additional automatic rainfall stations on the island were selected due to the incomplete rainfall data from the weather stations in Hong Kong Island. Daily weather data were then averaged across different stations in each defined area to obtain average area-specific weather data. Monthly mean (minimum) temperature is defined as the average (minimum) of the daily mean temperature while the monthly total (maximum) rainfall is defined as the total (maximum) of the daily total rainfall. In general, a strong seasonal pattern was observed among the climate parameters from 2002 to 2018 (Fig. 2). An increasing trend in monthly temperature and rainfall was observed from January until July and June, respectively.

### Model Framework

Given the similar seasonality of dengue epidemics in Southern Taiwan and Hong Kong where infections mostly appeared in the latter half of the year starting in summer, we extended a previous study of forecasting dengue annual cases in Southern Taiwan by fitting a model using the temperature and rainfall data from the earlier part of the year to predict the subsequent area-specific outbreaks<sup>8</sup>. In order to account for any unobserved factors relating to the studied areas, we developed a Poisson generalized linear mixed model (GLMM) to fit the confirmed cases in the different defined areas. Let  $Y_t^i$  denote the annual number of dengue cases and  $\mu_t^i$  denote the expected number of dengue cases, which is the corresponding distribution mean in area  $i$  and year  $t$ , then:

$$Y_t^i \sim \text{Poiss}(\mu_t^i) \quad (2)$$

and the model is specified by:

$$\log(\mu_i^i) = \alpha + \alpha^i + \sum_{m=m_1}^{m_n} (\beta_{T_m} \times T_{mt}^i + \beta_{R_m} \times R_{mt}^i) \quad (3)$$

where  $T_{mt}^i$  and  $R_{mt}^i$  denote the  $m^{\text{th}}$  month's (ranges from  $m_1$  to  $m_n$ ) normalised temperature and normalised rainfall predictors in the year  $t$  and area  $i$ ,  $\beta_{T_m}$  and  $\beta_{R_m}$  denote the regression coefficient of  $T_{mt}^i$  and  $R_{mt}^i$  respectively,  $\alpha$  represents the intercept, and  $\alpha^i$  represents the area-specific random intercept.

The monthly predictors were obtained using monthly climate data that were normalised into the range [0,1] from the collected temperature  $T'_m$  and rainfall  $R'_m$  data, so that the effects of the predictors on the responses are easier to compare when fitting the model<sup>33,48</sup>:

$$\begin{aligned} T_m &= (T'_m - \min(T'_m)) / (\max(T'_m) - \min(T'_m)) \\ R_m &= (R'_m - \min(R'_m)) / (\max(R'_m) - \min(R'_m)) \end{aligned} \quad (4)$$

Despite a number of zero counts in the dengue case data for certain years, we did not use zero-inflated models as they assume that excess zeros are generated by a separate process from the count values (e.g. underreporting of infections). We assumed that the zero values represent the true case numbers as CHP conducts strict dengue prevention and control measures in Hong Kong.

## Model Selection

A two-step procedure was used for the model selection. First, a candidate model was chosen to represent each of the four climate parameter sets ( $T_{mean}$  or  $T_{min}$  with  $R_{tot}$  or  $R_{max}$ ). The stepwise algorithm based on the Akaike Information Criterion with correction (AICc), also known as the stepAICc, was used to select the fixed effect variables (monthly climate predictors) for each of the candidate models in a forward selection approach (Supplementary Fig. S5). The stepwise algorithm has been commonly used in selecting relevant environmental or climatic determinants to predict disease occurrences<sup>8,30</sup>. This approach iteratively added a new variable to the model, starting with zero variables, to fulfil the selection criteria (such as smaller AIC or AICc) until the most optimal set of variables was reached. For our case, the AICc was used instead of AIC due to the small number of observations. For the purpose of model selection, all the models were estimated using Maximum Likelihood by Laplace approximation provided by the R package, `glmmTMB`<sup>49</sup>, in order to compare models with different fixed effects. Laplace approximation is a suggested method for fitting Poisson mixed effects models<sup>50</sup>. Secondly, the candidate models with the most optimal set of variables were then compared with each other based on their AICc and Bayesian Information Criterion (BIC). The model with the lowest AICc and BIC (or only AICc) was chosen as the best-fitting model.

## Model Evaluation

Leave-one-out cross-validation (LOOCV) was used in order to evaluate the best-fitting model's overall performance and to ensure that the model does not overfit. During each iteration, the observation of the annual number of dengue cases in a particular year and area was excluded while fitting the model using the restricted maximum likelihood method. The number of annual dengue cases of this particular observation was then estimated (predicted) given the coefficients produced by the fitted model in that iteration. In order to measure the performance of the model against the observation data, the Mean Squared Error (MSE) was calculated for both training and validation sets using the following general formulae:

$$MSE_{tr} = \sum_{j=1}^{17} \sum_{i \neq j} (\hat{\mu}_i^{(-j)} - \mu_i)^2 / (17 \cdot (17 - 1)) \quad (5)$$

$$MSE_{va} = \sum_{j=1}^{17} (\hat{\mu}_j^{(-j)} - \mu_j)^2 / 17 \quad (6)$$

where  $j$  indicates the index of a year to be tested,  $\hat{\mu}_i^{(-j)}$  indicates the estimated number of cases in year  $i$  when the number in year  $j$  was removed, and  $\mu_i$  is the observed number of cases in year  $i$ . When pre-defined areas were considered, the highest

value of index  $j$  became 51 to represent the total number of observations of annual dengue incidence by year and by area. In addition, leave-one-year-out (leave-three-out) cross-validation (LOYOCV) was also used to evaluate the best-fitting model when the observations of the annual number of dengue cases in all three areas in a particular year were removed and used as the validation set while fitting the model.

Bootstrapped confidence intervals were computed. For each predicted  $\hat{\mu}$  of a model, we generated 1000 samples of outcomes from a Poisson distribution with the rate parameter equal to the estimated parameter ( $\hat{\mu}$ ) of that fitted model. Based on these 1000 simulated outcomes, we refitted the model 1000 times to obtain the distribution of the parameter's values. The 95% confidence interval defines a range of values between the 97.5% and 2.5% quantiles. We considered the model to be satisfactory if it managed to successfully predict the two major outbreaks (within confidence intervals) in 2002 and 2018 in the LOOCV process and if it produced relatively low values for  $MSE_{va}$  or  $MSE_{ratio}$  (defined as  $MSE_{va}/MSE_{tr}$ ). The Normalised Mean Squared Error (NMSE) was also calculated by dividing the MSE of the validation or training set by the range of the observed variable.

## References

1. World Health Organization. Dengue guidelines for diagnosis, treatment, prevention, and control: new edition. Available at <https://apps.who.int/iris/handle/10665/44188> (2009).
2. World Health Organization. Dengue and severe dengue [Fact sheet]. Available at <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue> (2019).
3. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nat.* **496**, 504–507 (2013).
4. Brady, O. J. *et al.* Refining the Global Spatial Limits of Dengue Virus Transmission by Evidence-Based Consensus. *PLoS Neglected Trop. Dis.* **6**, e1760 (2012). DOI 10.1371/journal.pntd.0001760.
5. Ebi, K. L. & Nealon, J. Dengue in a changing climate. *Environ. Res.* **151**, 115–123 (2016).
6. Nagao, Y. *et al.* Climatic and social risk factors for Aedes infestation in rural Thailand. *Trop. Medicine Int. Heal.* **8**, 650–659 (2003).
7. Wu, P. C. *et al.* Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. *Sci. Total. Environ.* **407**, 2224–2233 (2009).
8. Yuan, H. Y. *et al.* Prediction of annual dengue incidence by hydro-climatic extremes for southern Taiwan. *Int. J. Biometeorol.* **63**, 259–268 (2019).
9. Lowe, R., Chirombo, J. & Tompkins, A. M. Relative importance of climatic, geographic and socio-economic determinants of malaria in Malawi. *Malar. J.* **12**, 416 (2013).
10. Tun-Lin, W., Burkot, T. R. & Kay, B. H. Effects of temperature and larval diet on development rates and survival of the dengue vector Aedes aegypti in north Queensland, Australia. *Med. Vet. Entomol.* **14**, 31–37 (2000).
11. Morin, C. W., Comrie, A. C. & Ernst, K. Climate and dengue transmission: Evidence and implications. *Environ. Heal. Perspectives* **121**, 1264–1272 (2013).
12. Rohani, A., Wong, Y. C., Zamre, I., Lee, H. L. & Zurainee, M. N. The effect of extrinsic incubation temperature on development of dengue serotype 2 and 4 viruses in Aedes aegypti (L.). *Southeast Asian J. Trop. Medicine Public Heal.* **40**, 942–950 (2009).
13. Mutheneni, S. R., Morse, A. P., Caminade, C. & Upadhyayula, S. M. Dengue burden in India: Recent trends and importance of climatic parameters. *Emerg. Microbes Infect.* **6**, e70 (2017).
14. Barrera, R., Amador, M. & MacKay, A. J. Population dynamics of aedes aegypti and dengue as influenced by weather and human behavior in san juan, puerto rico. *PLoS Neglected Trop. Dis.* **5**, e1378 (2011). DOI 10.1371/journal.pntd.0001378.
15. Lai, Y. H. The climatic factors affecting dengue fever outbreaks in southern Taiwan: An application of symbolic data analysis. *BioMedical Eng. Online* **17**, 148 (2018). DOI 10.1186/s12938-018-0575-4.

16. Benedum, C. M., Seidahmed, O. M. E., Eltahir, E. A. B. & Markuzon, N. Statistical modeling of the effect of rainfall flushing on dengue transmission in Singapore. *PLoS neglected tropical diseases* **12**, e0006935 (2018). DOI 10.1371/journal.pntd.0006935.
17. Lowe, R. *et al.* Nonlinear and delayed impacts of climate on dengue risk in Barbados: A modelling study. *PLoS Medicine* **15**, e1002613 (2018). DOI 10.1371/journal.pmed.1002613.
18. Wang, S. F. *et al.* Severe dengue fever outbreak in Taiwan. *Am. J. Trop. Medicine Hyg.* **94**, 193–197 (2016).
19. Luo, L. *et al.* The dengue preface to endemic in mainland China: The historical largest outbreak by *Aedes albopictus* in Guangzhou, 2014. *Infect. Dis. Poverty* **6**, 148 (2017).
20. Yihui, D. & Chan, J. C. The East Asian summer monsoon: An overview. *Meteorol. Atmospheric Phys.* **89**, 117–142 (2005).
21. Chuang, V. W. *et al.* Review of dengue fever cases in Hong Kong during 1998 to 2005. *Hong Kong Med. J.* **14**, 170–177 (2008).
22. Xu, L. *et al.* Climate variation drives dengue dynamics. *Proc. Natl. Acad. Sci. United States Am.* **114**, 113–118 (2017).
23. Li, R. *et al.* Climate-driven variation in mosquito density predicts the spatiotemporal dynamics of dengue. *Proc. Natl. Acad. Sci.* **116**, 3624–3629 (2019).
24. Li, C., Lu, Y., Liu, J. & Wu, X. Climate change and dengue fever transmission in China: Evidences and challenges. *Sci. Total. Environ.* **622-623**, 493–501 (2018).
25. Wu, X. *et al.* Non-linear effects of mean temperature and relative humidity on dengue incidence in Guangzhou, China. *Sci. Total. Environ.* **628-629**, 766–771 (2018).
26. Chien, L. C. & Yu, H. L. Impact of meteorological factors on the spatiotemporal patterns of dengue fever incidence. *Environ. Int.* **73**, 46–56 (2014).
27. Hii, Y. L., Zhu, H., Ng, N., Ng, L. C. & Rocklöv, J. Forecast of Dengue Incidence Using Temperature and Rainfall. *PLoS Neglected Trop. Dis.* **6**, e1908 (2012). DOI 10.1371/journal.pntd.0001908.
28. Food and Environmental Hygiene Department, the Government of the Hong Kong Special Administrative Region. Dengue Fever Ovitrap Index Update. Available at [https://www.fehd.gov.hk/english/pestcontrol/dengue\\_fever/ovitrap\\_index.html](https://www.fehd.gov.hk/english/pestcontrol/dengue_fever/ovitrap_index.html) (2019).
29. Wong, M. C., Mok, H. Y., Ma, H. M., Lee, M. W. & Fok, M. Y. A climate model for predicting the abundance of *Aedes* mosquitoes in Hong Kong. *Meteorol. Appl.* **18**, 105–110 (2011).
30. Lowe, R. *et al.* The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Stat. Medicine* **32**, 864–883 (2013).
31. Gharbi, M. *et al.* Time series analysis of dengue incidence in guadeloupe, french west indies: Forecasting models using climate variables as predictors. *BMC Infect. Dis.* **11**, 166 (2011).
32. Cheong, Y. L., Burkart, K., Leitão, P. J. & Lakes, T. Assessing weather effects on dengue disease in Malaysia. *Int. J. Environ. Res. Public Heal.* **10**, 6319–6334 (2013).
33. Mordecai, E. A. *et al.* Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. *PLoS Neglected Trop. Dis.* **11**, e0005568 (2017). DOI 10.1371/journal.pntd.0005568.
34. Sharmin, S., Glass, K., Viennet, E. & Harley, D. Geostatistical mapping of the seasonal spread of under-reported dengue cases in Bangladesh. *PLoS Neglected Trop. Dis.* **12**, e0006947 (2018). DOI 10.1371/journal.pntd.0006947.
35. Harrison, X. A. *et al.* A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* **6**, e4794 (2018). DOI 10.7717/peerj.4794.
36. Hong Kong Tourism Board. Visitor Arrival Statistics. Available at [https://partnernet.hktb.com/en/research\\_statistics/latest\\_statistics/index.html](https://partnernet.hktb.com/en/research_statistics/latest_statistics/index.html) (2019).
37. Hung, L. S. The SARS Epidemic in Hong Kong: What Lessons have we Learned? *J. Royal Soc. Medicine* **96**, 374–378 (2003).



38. Wong, J. Y. *et al.* Infection fatality risk of the pandemic A(H1N1)2009 virus in Hong Kong. *Am. J. Epidemiol.* **177**, 834–840 (2013).
39. Kwok, K. O., Cowling, B., Wei, V., Riley, S. & Read, J. M. Temporal variation of human encounters and the number of locations in which they occur: A longitudinal study of Hong Kong residents. *J. Royal Soc. Interface* **15**, 20170838 (2018).
40. Chen, T. C., Wang, S. Y., Huang, W. R. & Yen, M. C. Variation of the East Asian summer monsoon rainfall. *J. Clim.* **17**, 744–762 (2004).
41. Guangzhou Meteorology. Guangzhou city weather bulletin in 2014. Available at <http://www.tqyb.com.cn/gz/climaticprediction/bulletin> (2019).
42. Cheng, Q. *et al.* The interplay of climate, intervention and imported cases as determinants of the 2014 dengue outbreak in Guangzhou. *PLoS Neglected Trop. Dis.* **11**, e0005701 (2017). DOI 10.1371/journal.pntd.0005701.
43. News.gov.hk. Anti-mosquito operation starts. Available at [https://www.news.gov.hk/eng/2019/04/20190415/20190415\\_161333\\_051.html](https://www.news.gov.hk/eng/2019/04/20190415/20190415_161333_051.html) (2019).
44. The Government of the Hong Kong Special Administrative Region. Co-ordinating meeting on dengue fever held to follow up on Government's anti-mosquito work (with photo). Available at <https://www.info.gov.hk/gia/general/201809/18/P2018091801039.htm> (2019).
45. Legislative Council Secretariat (Information Services Division), the Government of the Hong Kong Special Administrative Region. MTR train service performance. Available at <https://www.legco.gov.hk/research-publications/english/1718issh07-mtr-train-service-performance-20171220-e.pdf> (2017).
46. Department of Health, The Government of the Hong Kong Special Administrative Region. Press Release: Update on dengue fever. Available at <https://www.dh.gov.hk/english/press/press.html>.
47. Hong Kong Observatory. Climatological Information Services. Available at <https://www.hko.gov.hk/en/cis/climat.htm>.
48. Lee, H., Kim, J. E., Lee, S. & Lee, C. H. Potential effects of climate change on dengue transmission dynamics in Korea. *PLoS ONE* **13**, e0199205 (2018). DOI 10.1371/journal.pone.0199205.
49. Magnusso, A. *et al.* glmmTMB: Generalized Linear Mixed Models using Template Model Builder. Available at <https://cran.r-project.org/package=glmmTMB> (2019).
50. Bolker, B. M. *et al.* Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* **24**, 127–135 (2009).

## Acknowledgements

We would like to thank Drs. Wen Zhou (City University of Hong Kong), Xiaolin Zhu (Hong Kong Polytechnic University) and James Hay (Imperial College London) for their valuable comments on model development, weather patterns and geographical information. We also thank the financial supports from startup grant and the Applied Strategic Development Centres at City University of Hong Kong.

## Author contributions statement

K.S. and M.M.T. conducted data collection. K.S., J.L., M.M.T., P.L. and H.Y. conducted the statistical model development and wrote the main manuscript text, K.S. and H.Y. analysed the results. All authors reviewed the manuscript.

## Additional information

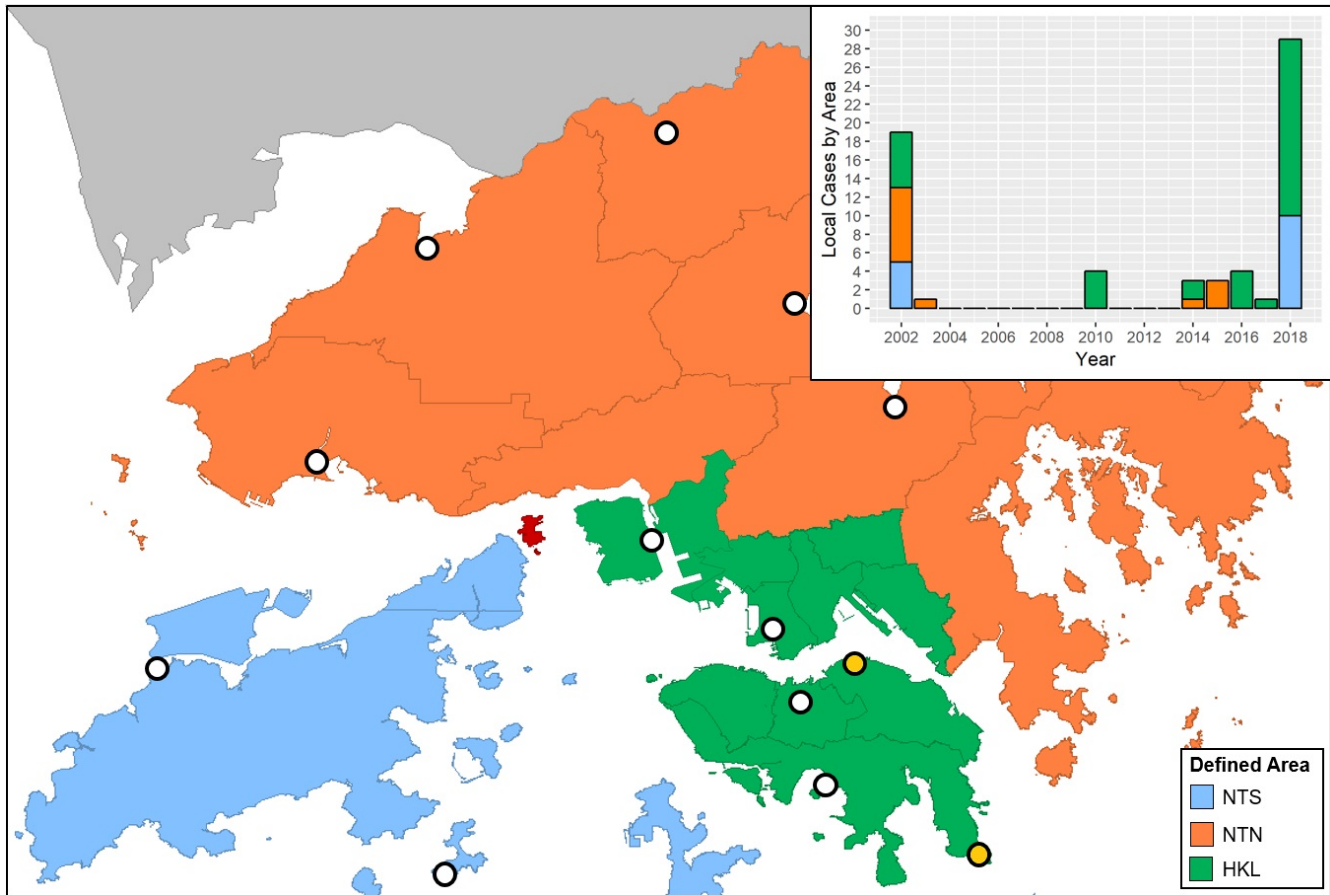
## **Data Availability**

The climate data are available on the Hong Kong Observatory website while the dengue cases data are available on the online Press Release published by the Centre for Health Protection, Hong Kong. The overall dataset including reformatted climate data and the R code files produced and applied in this study are available from the corresponding author on request.

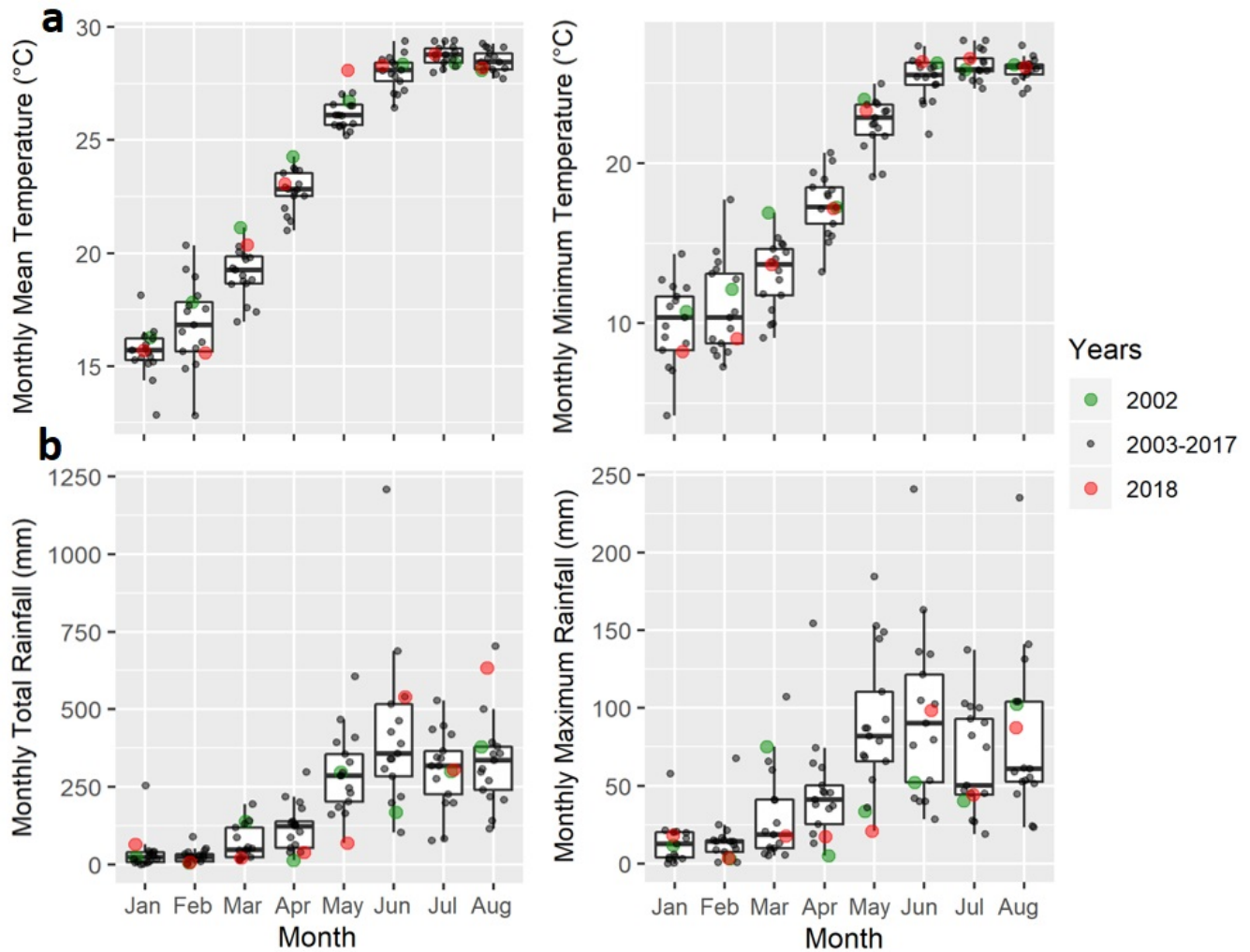
## **Competing interests**

The authors declare no competing interests.

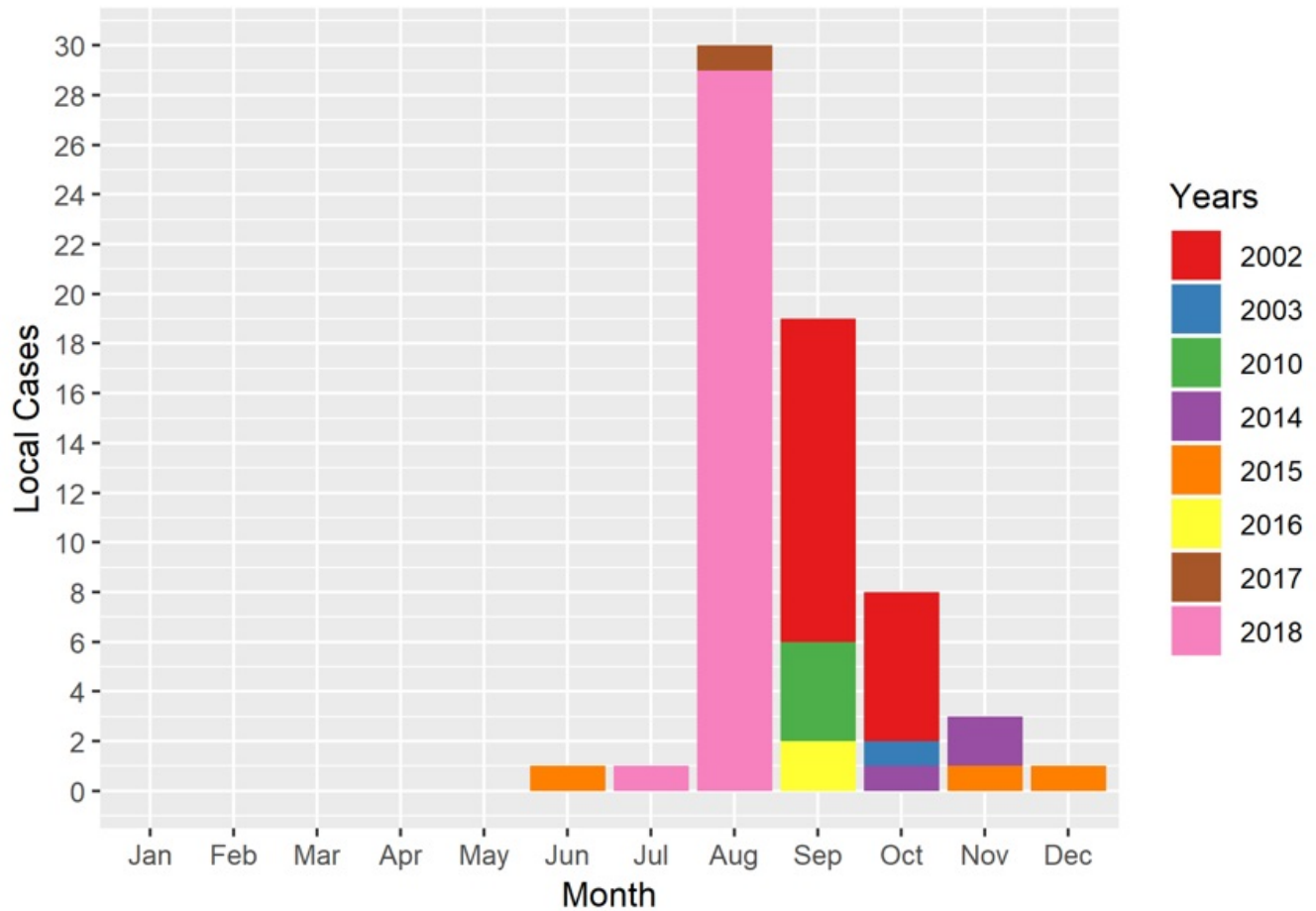
## Figures and Tables



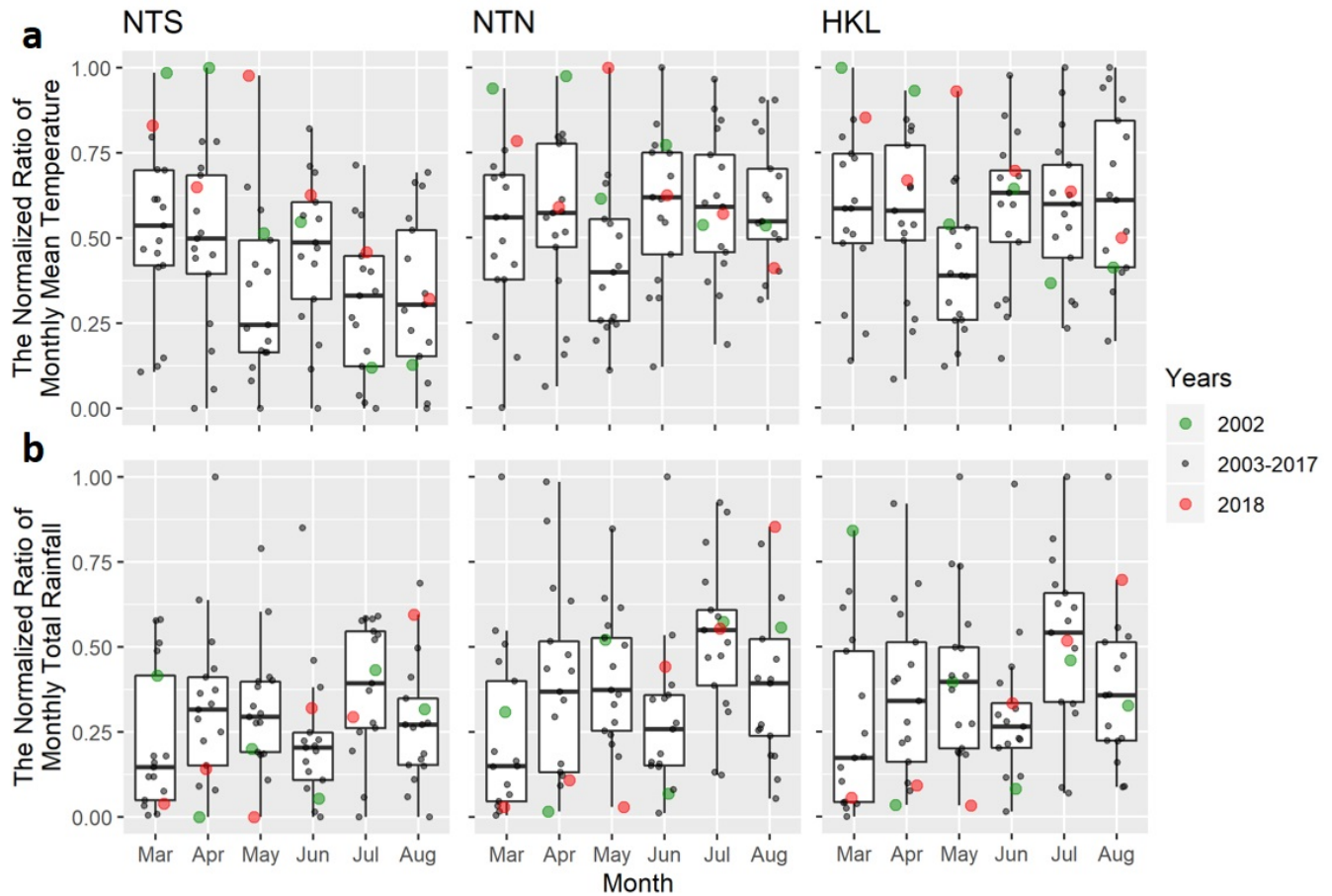
**Figure 1.** The area division: New Territories - South (NTS; light blue), New Territories - North (NTN; orange), and Hong Kong & Kowloon (HKL; green) along with the selected automatic weather stations (white circles) and rainfall stations (light yellow circles). The grey lines indicate the official district borders. As Ma Wan Island (red) is located at the center, all the 17 non-blood transfusion cases on that island are assigned into the 3 areas in our study: 5 cases in NTS, 6 cases in NTN, and 5 cases in HKL. The inset provides annual local case numbers from 2002 to 2018



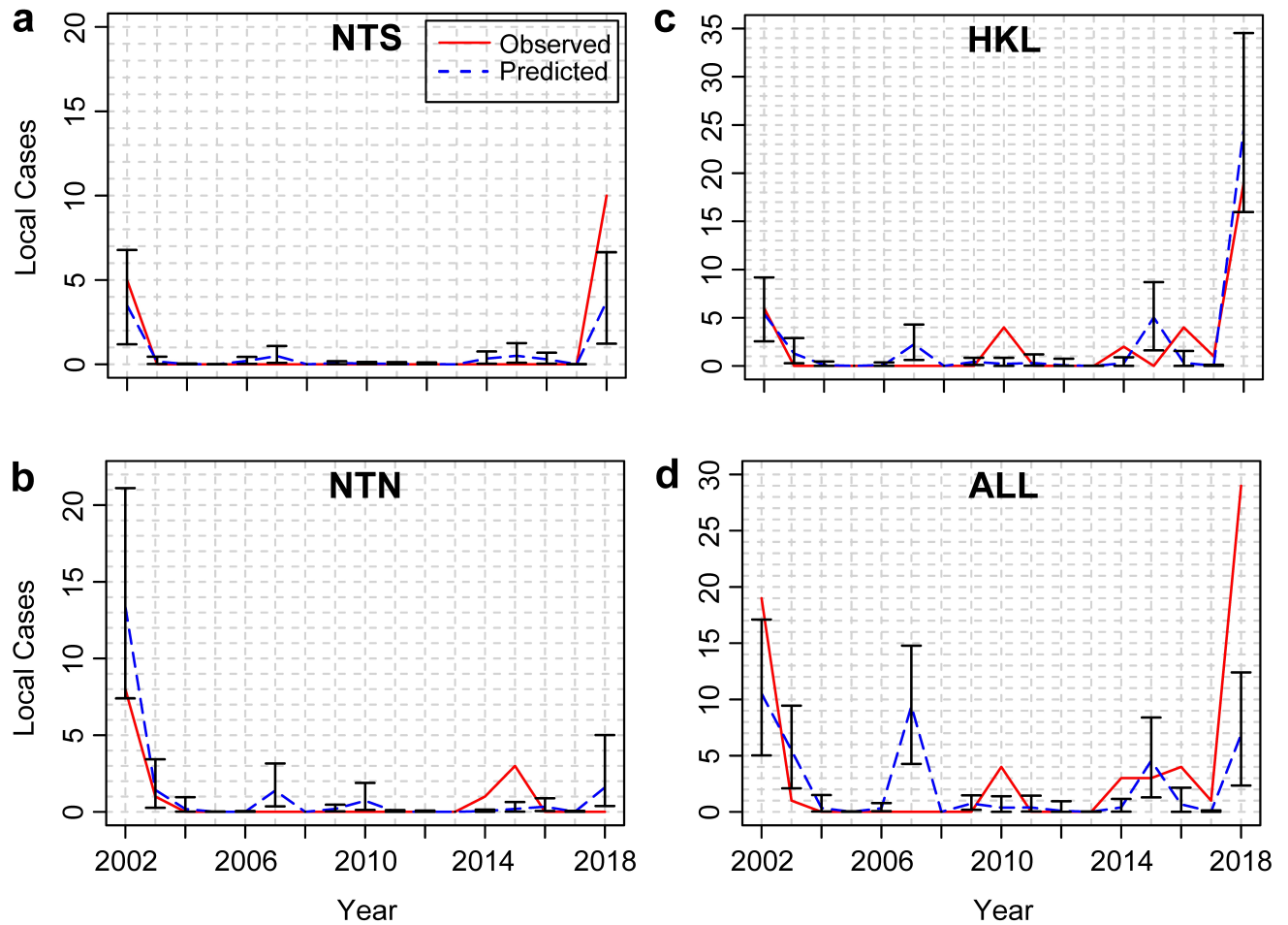
**Figure 2.** The monthly climate data for the whole of Hong Kong from 2002 to 2018. (a) Monthly mean and monthly minimum temperature, and (b) monthly total and monthly maximum rainfall from January to August for the years 2002, 2003-2017 combined, and 2018 are shown. Climate data were retrieved from 11 weather stations and 2 automatic rainfall stations across the three pre-defined areas



**Figure 3.** The number of monthly reported local dengue cases from 2002 to 2018. Different colours represent the number of dengue cases in different years.



**Figure 4.** The normalised climate predictors for each area using (a) the monthly mean temperature and (b) the monthly total rainfall. The areas NTS, NTN, and HKL are defined as in Figure 1.



**Figure 5.** Comparison between observed and predicted number of annual dengue cases. (a-c) Observed and predicted number of annual dengue cases in each pre-defined area (NTS, NTN, and HKL) with 95% confidence intervals using leave-one-out cross-validation. The predicted values represent the mean of the Poisson distribution of the annual dengue cases in each area, estimated using a generalized linear mixed model with a restricted maximum likelihood method. (d) Observed and predicted number of annual dengue cases in the whole Hong Kong area (ALL) with 95% confidence intervals using leave-one-year-out cross-validation.

**Table 1.** The AICc and BIC of each model with the selected predictors. The most optimal set of predictors was determined using the StepAICc for each climate parameters set. Model:  $T_{mean} + R_{tot}$  was selected as the best-fitting model and Model:  $T_{mean} + R_{max}$  was used as the alternative model.

Set	$T_{mean}$		$T_{min}$	
	$R_{tot}$	$R_{max}$	$R_{tot}$	$R_{max}$
<i>AICc</i>	107.84	108.84	134.34	117.91
$\Delta AICc$	0.00	1.00	26.50	10.07
<i>BIC</i>	120.83	121.83	147.33	132.39
$\Delta BIC$	0.00	1.00	26.50	11.56



**Table 2.** The Likelihood Ratio Test (LRT) results for each selected climate predictor in the best-fitting model. Each predictor was removed and compared to the original model using ANOVA. The p-value was then calculated given that the test statistic of the LRT is asymptotically approximating a  $\chi^2$  distribution.  $T_m$  and  $R_m$  denote the temperature and rainfall predictors in the  $m^{th}$  month of the year.

Climate Predictors	$\beta$	$\chi^2$	<i>p-value</i>
$T_3$	3.594	8.337	0.004
$T_4$	2.542	7.217	0.007
$T_5$	1.210	1.256	0.262
$T_7$	10.072	35.347	2e-09
$R_4$	-4.002	5.430	0.020
$R_5$	-4.029	15.494	8e-05
$R_6$	-8.047	10.067	0.002

**Table 3.** The mean squared errors (MSE) and normalised mean squared errors (NMSE) of each model tested using leave-one-out cross-validation. The MSE and NMSE for the training and the validation sets, and the ratios of the MSE and NMSE of the validation set to that of the training set, respectively, are listed. The best-fitting model (Model:  $T_{mean} + R_{tot}$ ) was compared with the alternative model (Model:  $T_{mean} + R_{max}$ ), the fixed effects model (same predictors as the best-fitting model but without random effects) and the full model (mixed effects with all predictors).

	MSE			NMSE		
	<i>Training</i>	<i>Validation</i>	<i>Ratio</i>	<i>Training</i>	<i>Validation</i>	<i>Ratio</i>
Best-fitting	0.592	3.538	5.976	0.032	0.186	5.813
Alternative	0.672	5.143	7.653	0.036	0.271	7.528
Fixed effects	0.551	3.543	6.430	0.029	0.186	6.414
Full	0.408	189.393	464.199	0.022	9.968	453.091