PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Time varying methods to infer extremes in dengue transmission dynamics

Jue Tao Lim[1]ᵒ*, **Yiting Han**[2]ᵒ, **Borame Sue Lee Dickens**[1], **Lee Ching Ng**[3], **Alex R. Cook**[1]

1 Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, **2** School of Pharmacy, Fudan University, Shanghai, China, **3** Environmental Health Institute, National Environmental Agency, Singapore

ᵒ These authors contributed equally to this work.
* ephljt@nus.edu.sg

Abstract

Dengue is an arbovirus affecting global populations. Frequent outbreaks occur, especially in equatorial cities such as Singapore, where year-round tropical climate, large daily influx of travelers and population density provide the ideal conditions for dengue to transmit. Little work has, however, quantified the peaks of dengue outbreaks, when health systems are likely to be most stretched. Nor have methods been developed to infer differences in exogenous factors which lead to the rise and fall of dengue case counts across extreme and non-extreme periods. In this paper, we developed time varying extreme mixture (tvEM) methods to account for the temporal dependence of dengue case counts across extreme and non-extreme periods. This approach permits inference of differences in climatic forcing across non-extreme and extreme periods of dengue case counts, quantification of their temporal dependence as well as estimation of thresholds with associated uncertainties to determine dengue case count extremities. Using tvEM, we found no evidence that weather affects dengue case counts in the near term for non-extreme periods, but that it has non-linear and mixed signals in influencing dengue through tvEM parameters in the extreme periods. Using the most appropriate tvEM specification, we found that a threshold at the 70th (95% credible interval 41.1, 83.8) quantile is optimal, with extreme events of 526.6, 1052.2 and 1183.6 weekly case counts expected at return periods of 5, 50 and 75 years. Weather parameters at a 1% scaled increase was found to decrease the long-run expected case counts, but larger increases would lead to a drastic expected rise from the baseline correspondingly. The tvEM approach can provide valuable inference on the extremes of time series, which in the case of infectious disease notifications, allows public health officials to understand the likely scale of outbreaks in the long run.

Author summary

Dengue is an arbovirus affecting populations across much of the globe. Frequent outbreaks occur, especially in equatorial cities such as Singapore, where the year-round tropical climate, large daily influx of travelers and population density provide the ideal

conditions for dengue transmission. Little work has however quantified the peaks of dengue outbreaks, when health systems are likely to be most stretched. Nor have methods been developed to infer differences in exogenous factors which lead to the rise and fall of dengue case counts across extreme and non extreme periods. In this paper, we developed time varying extreme mixture (tvEM) methods to account for the temporal dependence of dengue case counts across extreme and non-extreme periods. tvEM is able to infer differences in climatic forcing across non-extreme and extreme periods of dengue case counts, their temporal dependence as well as estimate suitable thresholds with associated uncertainties to determine dengue case count extremities. Using tvEM, we found no evidence that weather affects dengue case counts in the near term for non extreme periods, but has non-linear and mixed signals in influencing dengue through tvEM parameters in the extreme periods. Using the most appropriate tvEM specification, we found that a high percentile threshold is estimated, with dengue outbreak events far larger than currently observed to be expected in 5, 50 and 75 years. Weather parameters at a 1% scaled increase was found to decrease the long-run expected case counts, but larger increases would lead to a drastic expected rise from the baseline correspondingly. tvEM can provide valuable inference on the extremes of time series, which in the case of infectious disease data, allows public health officials to understand factors and the likely scale of infectious disease outbreaks in the long run.

This is a *PLOS Computational Biology* Methods paper.

Introduction

An estimated 390 million dengue infections occur annually, imposing major economic and health burdens globally [1]. It is widespread in South-east Asia, with outbreaks occurring annually, sometimes exhibiting synchronous behaviour [2]. Furthermore, it is hyper-endemic within the region due to all four serotypes being in active circulation. Increased urbanization and elevated human movement rates via both domestic and international travel have increased the transmission potential of dengue, particularly across highly connected cities such as Singapore. With favourable vector-breeding conditions due to the year-round tropical climate, a large daily influx of travelers and population density, Singapore has well-suited conditions for dengue infections to occur, with an average of over 100 cases being reported every week from 2000 to 2017.

Primarily, vector control is used to mitigate dengue transmissions in Singapore and its success is evidenced in the decreasing seroprevalence nationally for the past two decades [3–5]. However, this low seroprevalence complicates the implementation of vaccination using the tetravalent Dengvaxia (CYD-TDV) [6, 7] vaccine on the national scale [8] due to potentially longer-term risks of severe dengue in vaccinated but seronegative individuals [9]. Fogging and breeding site reduction, in conjunction with novel biocontrol techniques such as *Wolbachia* are utilized instead to prevent and control dengue epidemics [10]. However, major outbreaks still occur sporadically, with some attributed to population level phenomena such as serotype switching [11–13], in which a large change in the proportion of circulating serotypes lead to an increase in the number of reported infections. While sufficient healthcare capacity is usually

available to deal with muted levels of dengue infections, a large and prolonged rise in the number of cases may lead significant impact on public health resources.

Being able to infer the properties of extreme values of dengue transmissions allows public health planning at the national level to be equipped with the information necessary to handle outbreaks. Typically, modelling or forecasting case counts is conducted using mathematical, statistical and machine learning tools [14, 15]. These tools allow understanding of dengue transmission dynamics, aid short to medium term resource planning for disease surveillance and inform decisions about strengthening vector control for outbreak control [16]. However, these tools are used for predicting and inferring case counts to minimize overall error for the entire dataset and are not optimized to characterize extreme conditions such as the peak of dengue outbreaks [17]. Peaks of dengue outbreaks may also not occur in the near term, which may further degrade the quality of predictions for standard models [14, 16]. As a result, it may be difficult to implicitly quantify the long-term risk and scale of these events.

Similar problems exist in other fields such as climatology, oceanography and geography, where the risk and scale of potential calamitous events such as floods, earthquakes and surge storms need to be quantified [18]. Characterization of these rare but serious events are conducted through tools developed from Extreme Value Theory (EVT), with statistical estimation and inference conducted on the extremes of observed events across time and/or space [18]. Some public health work has explored the use of standard EVT tools such as distributional inference on extreme health events such as infectious disease outbreaks; there have been particular applications of EVT to Pneumonia and Influenza (PI) death rates [19] and food-borne disease outbreaks [20], for instance. These works signal its promise for application in a public health setting. Yet, while relatively large leaps have occurred in the development of extreme value theory—such as inference on both extreme and non-extreme portions of data [21], hierarchical modelling on extreme distributions, formulating EVT parameter estimation as a regression problem through Bayesian data augmentation and quantifying space and/or time dependence of extreme values [22, 23]—to the authors' knowledge, no work in biology nor public health has developed the use of these potentially highly informative extensions of EVT tools.

This paper therefore develops and explores the utility of tools derived from extended extreme value theory to investigate and compare the dynamic signature of extreme and non-extreme periods of dengue transmissions in Singapore. Briefly, we aim to determine thresholds and thus classify extreme and non-extreme periods of dengue transmission, quantify the temporal dependence of extreme and non-extreme periods of dengue transmissions as well as dengue's potentially non-linear relationship with weather. To do so, we first developed four separate extreme-bulk mixture models which are able to characterize the different dynamic signatures of dengue transmissions, in both extreme and non-extreme periods. These methods were compared against several diagnostic checks, such as residual autocorrelation, quantile-quantile plots, Bayes factor and deviance information criterion, to assess the most suitable model given our data generating process. Next, we explored the potential for climatic fluctuations to affect both extreme and non-extreme periods of dengue transmissions and the differences in transmission behaviour between separate periods. Lastly, we projected the expected scale of dengue events in the long run using the most suitable EVT model and provide discussions on these results.

Results

Dengue in Singapore from 2000 to 2017

An average of over 100 dengue cases are reported every week from 2000 to 2017 in Singapore, with large rises in periods between 2004 to 2005, the middle of 2007, 2013 to 2014 and the end

of 2015 and beginning of 2016. On visual inspection of the time series, no clear seasonal pattern is observed, with large rises in dengue observed in 2007 and 2013 to 2016 (Fig 1). We fitted four competing models (M1–M4) to infer the driving patterns of biologically relevant covariates on both elevated (extreme) and baseline (non-extreme) periods of dengue in Singapore from 2000 to 2017.

Model assessment

Convergence of MCMC chains for parameters across all four competing models was indicated by Gewecke convergence tests, signalling that each marginal parameter posterior space is well explored. Visual inspection of trace plots was also conducted (S1 Fig). Residual autocorrelation for the bulk distribution also indicate that an autoregressive (AR) order 3 model is sufficient to explain dengue transmission dynamics as each lag's autocorrelation is within the 95% confidence interval bound. Comparing each model's deviance information criterion (DIC) and log Bayes factor (logBF) showed increasing DIC across models and increasing logBF as models become more complex, indicating that more complex models are penalized as expected from the DIC but overall more favoured in explaining the data according to logBF. In general, a higher logBF and lower DIC indicates a more suitable model for the data. In particular, we obtained the highest DIC and logBF for the model with constant beta term in the tail distribution (Table 1 M4 DIC: 14 268, logBF: 663), followed by a trend of decreasing DIC and increasing logBF in the case where no regression structure was imposed on the tail distribution (Table 1 M3 DIC: 13 111, logBF: 545), no weather variables being added to the bulk distribution (Table 1 M2 DIC: 12 781, logBF: 510), as well as no regression structure being imposed on either bulk nor tail distribution (Table 1 M1 DIC: −1563).

The quantile-quantile (QQ) plots compare the similarity of fitted distributions to data. Although the DIC for the first model (M1) is low, QQ plots indicate that fitting the gamma distribution to the bulk data leads to significant deviations from the line of unity as compared to regression structures (M2–4): this shows that M1 does not provide distributional characteristics which are similar to the data. For the extreme distribution, QQ plots for the regression structure (M4) generalized Pareto distribution (GPD) has points lying closest to the line of unity, indicating that it is the most suitable structure to account for extreme value data (S1 Fig). There is a high logBF and favourable QQ plots but also higher DIC for M4 compared to other models. However, as regression structure was imposed on both the bulk and tail distribution, the increased penalty on model complexity under DIC is expected. M4 is also further able to compare the effects of biologically relevant covariates across non-extreme and extreme states of dengue transmission dynamics, compared to all other models. Further exposition on results will thus refer to M4 unless specifically mentioned otherwise.

Inference on extremes of dengue transmissions

Across 2000 to 2017, when the time varying extreme mixture model (tvEM) was estimated, an observation of around 150 weekly case counts was estimated to be threshold for data to be classified as extreme, with 95% credible interval (CrI) 72 to 244 cases. This corresponds to dengue case counts being above the 70th percentile (95% CrI 41.1–83.8) to be taken as extreme (Fig 1a). In periods where weekly dengue case counts are around or above 150, extreme value parameters correspond to larger fluctuations compared to nuisance parameters estimated in non-extreme periods (Fig 1b and 1c). Scaled return level results are consistent across M1–M4 (Fig 2a–2c), with return levels computed by integrating time varying parameters in the tvEM indicate that at baseline levels, an average scaled level of 0.70, 1.30 and 1.45 will be reached in 5, 50 and 75 years, corresponding to an event where 526.6, 1052.2 and 1183.6 weekly dengue
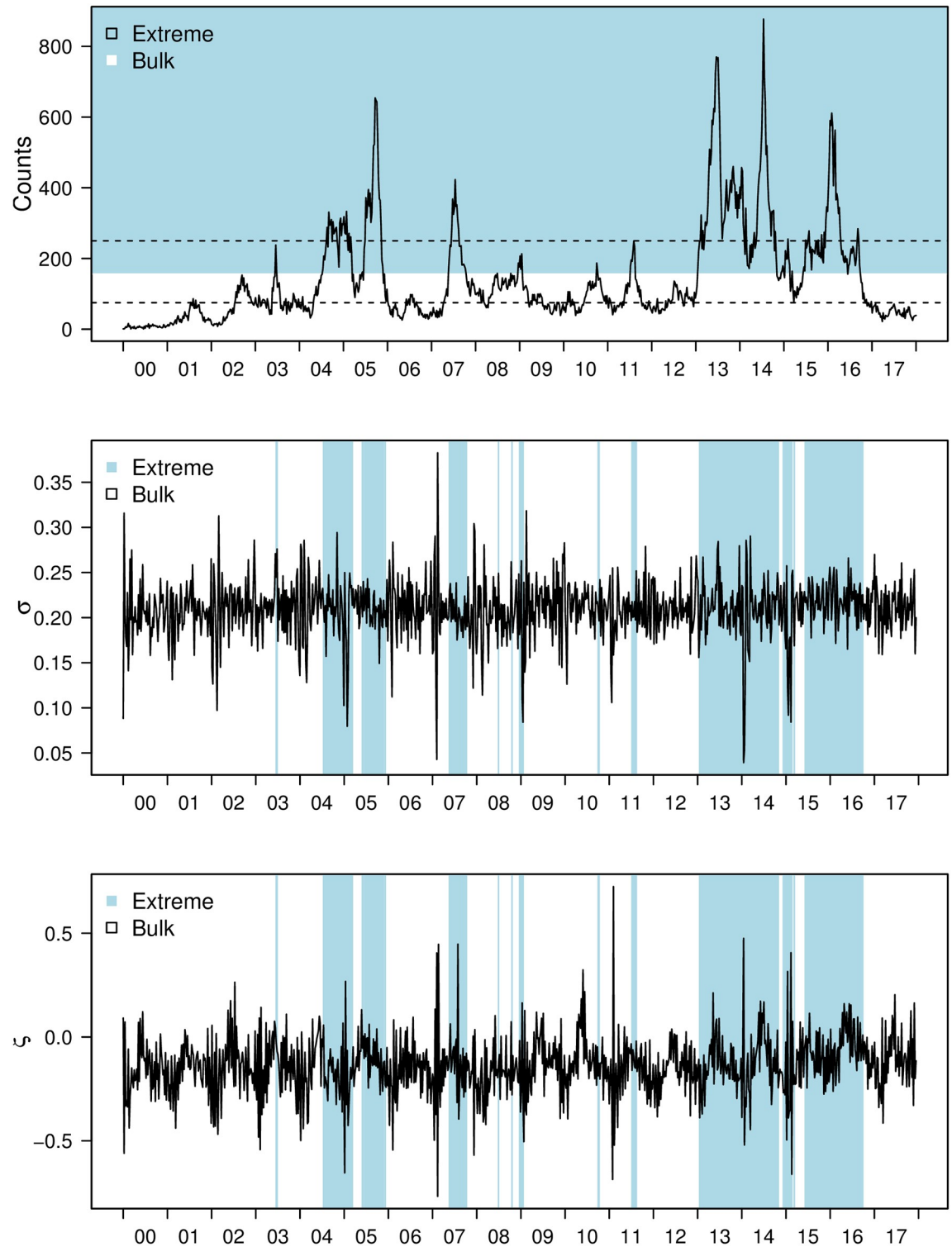
Fig 1. From top to bottom: 1) Dengue case counts from 2000 to 2017 with data above the posterior mean threshold shaded. 95% credible intervals for the threshold are given in dotted lines 2) Posterior mean time-varying scale parameter for constant beta model, with highlighted areas representing timepoints when data are above the threshold from 2000 to 2017 3) Posterior mean time-varying shape parameter for constant beta model, with highlighted areas representing timepoints when data are above the threshold from 2000 to 2017.

https://doi.org/10.1371/journal.pcbi.1008279.g001

Table 1. Model fit for 4 competing mixture models.

Model	DIC	logBF
M1: Gamma + Constant GPD	−1 563	
M2: Normal(AR) + Time-Varying GPD	12 781	510
M3: Normal(AR+weather) + Time-Varying GPD	13 111	545
M4: Normal(AR+weather) + Regression GPD	14 268	663

https://doi.org/10.1371/journal.pcbi.1008279.t001

case counts are expected to be observed (Fig 2b). These estimated return levels flatten off after 100 years after to a scaled level of around 1.60 (Fig 2b).

Duality of climatic effects below and above thresholds of dengue transmission

For the tvEM bulk distribution where the AR model of order 3 was estimated along with weather variables of up to 3 week lags, we found that 95% credible intervals for climatic coefficients all overlap with 0, with posterior mean estimates not lesser or greater than −0.1 and 0.1 respectively. The directions of these coefficients are also mixed, with all 95% credible intervals not crossing more than −0.3 and 0.4. These credible intervals are also consistent in their spread, at around 0.5 to 0.6 in magnitude, asides from precipitation, where smaller intervals are evident and range from −0.01 to 0.011 (Table 2, Fig 3c).

However, we found that 95% credible intervals for relative humidity, absolute humidity and temperature coefficients exclude 0 across certain lags in the tvEM extreme distribution. Specifically for the weather coefficients governing the extreme distribution scale parameter β_σ, we found negative effects coming from precipitation (Table 2: −0.069, 95% CrI: −0.119, −0.016) and absolute humidity (Table 2: −3.448, 95% CrI: −5.759, −0.703) at a 1 week lag. Correspondingly, positive effects from temperature (Table 2: 4.724, 95% CrI: 1.005, 7.826) and relative humidity (Table 2: 3.597, 95% CrI: 0.765, 5.969) at a 1 week lag also. The magnitude of the precipitation coefficient and the spread of the 95% credible intervals are similarly smaller than that of other weather coefficients. For weather coefficients governing the extreme distribution
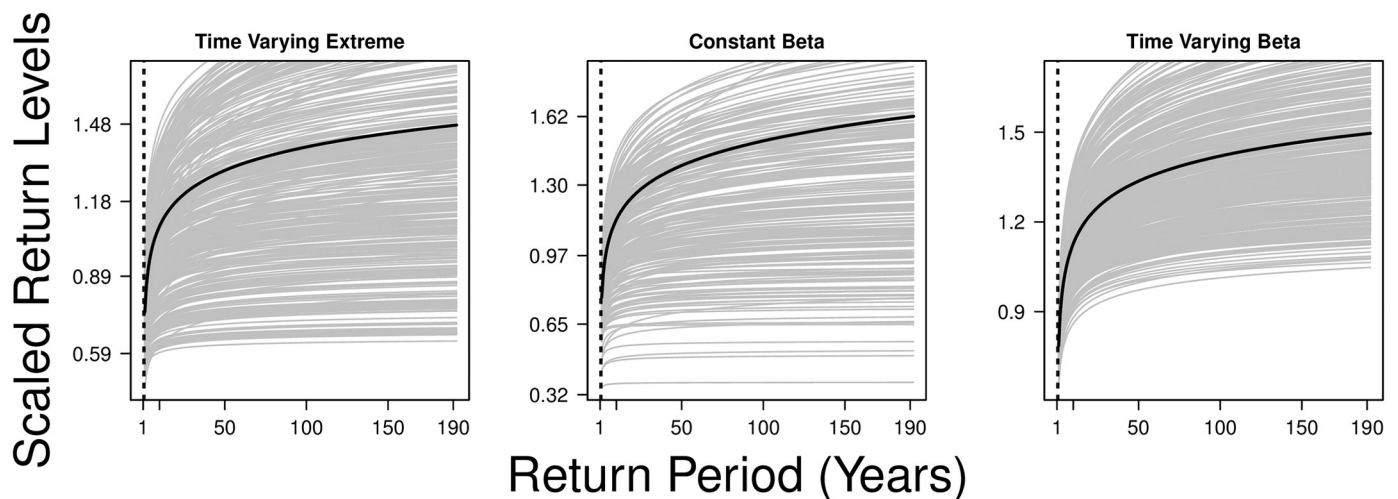
Fig 2. Return levels over a 1 to 190 year return period, with posterior mean estimate given in solid black lines, estimates from the following models are expressed from left to right: 1) Time-varying generalized Pareto distribution 2) Time-varying generalized Pareto distribution with constant beta regression structure 3) Time-varying generalized Pareto distribution with time-varying beta regression structure.

https://doi.org/10.1371/journal.pcbi.1008279.g002

Table 2. Coefficient posterior mean and 95% credible intervals for the constant beta regression with the generalized Pareto regression parameters and bulk regression. Columns from left to right represent the regression coefficients with the dependent variable being [1]Generalized Pareto distribution shape parameter [2]Generalized Pareto distribution scale parameter [3]Dengue case counts below the threshold. *Coefficients whose 95% credible intervals are away from 0.

	β_ξ		β_σ		β_{bulk}	
Coefficient	Estimate[1]	95% CrI[1]	Estimate[2]	95% CrI[2]	Estimate[3]	95% CrI[3]
Precipitation						
Lag 1	−0.032	(−0.067, 0.003)	-0.069*	(−0.119, −0.016)	0	(−0.01, 0.009)
Lag 2	−0.011	(−0.052, 0.043)	0.017	(−0.05, 0.066)	0.002	(−0.008, 0.011)
Lag 3	0.022	(−0.028, 0.077)	0.01	(−0.053, 0.067)	−0.001	(−0.01, 0.008)
Temperature						
Lag 1	−2.514*	(−4.059, −0.43)	4.724*	(1.005, 7.826)	0.036	(−0.238, 0.316)
Lag.2	3.279*	(0.123, 5.267)	−1.5	(−4.276, 1.572)	−0.011	(−0.281, 0.286)
Lag 3	−2.136*	(−3.513, −0.665)	−0.852	(−2.377, 0.763)	−0.01	(−0.277, 0.242)
Absolute Humidity						
Lag 1	1.912*	(0.32, 3.083)	−3.448*	(−5.759, −0.703)	−0.026	(−0.241, 0.18)
Lag 2	−2.479*	(−4.01, −0.051)	1.014	(−1.253, 3.071)	0.011	(−0.215, 0.215)
Lag 3	1.693*	(0.531, 2.764)	0.717	(−0.518, 1.881)	0.007	(−0.184, 0.21)
Relative Humidity						
Lag 1	−1.924*	(−3.136, −0.275)	3.597*	(0.765, 5.969)	0.026	(−0.178, 0.238)
Lag 2	2.587*	(0.133, 4.135)	−1.077	(−3.143, 1.249)	−0.01	(−0.212, 0.214)
Lag 3	−1.729*	(−2.761, −0.604)	−0.746	(−1.933, 0.51)	−0.004	(-0.206, 0.186)

scale parameter, all 2 to 3 week lag variables have 95% credible intervals containing 0 (Table 2).

For the weather coefficients governing the extreme distribution shape parameter β_ξ, the precipitation coefficients had all 95% credible intervals containing 0 from 1 to 3 weeks lag. Conversely, all other weather covariates had 95% credible intervals for their coefficients that excluded 0 for 1 to 3 weeks' lag term. These coefficients also vary in direction across lags for the same variable. For example, the temperature coefficient had 1 and 3 weeks lag being negative in sign, but the 2 weeks lag was positive (Table 2). For the 1 week lag climate coefficients, the coefficients governing the extreme distribution shape parameter β_ξ and the extreme distribution scale parameter β_σ are in opposite directions always (Table 2, Fig 3). The absolute humidity coefficient for the extreme distribution shape parameter is positive but the same coefficient for the extreme distribution scale parameter is negative. Conversely, the relative humidity coefficient for the extreme distribution shape parameter is negative but the same coefficient for the extreme distribution scale parameter is positive (Table 2).

Nonlinearities in climatic effects on extremes of dengue transmission

We have lastly examined how weather coefficients governing tvEM extreme distributions affect expected return levels. We have only included coefficients whose 95% credible intervals excluded 0, namely, temperature, absolute humidity and relative humidity. In general, we found the effects of weather on return levels to be highly nonlinear and dependent on the specific value of shocks applied to each coefficients.

For a 1% positive shock across each parameter at each lag, we found in general a decrease in the expected return levels across time for temperature lag 1 and 3, absolute humidity lag 1 and 2 and relative humidity at lag 1 and 3 (Fig 4). However, the changes in return levels are almost negligible when shocks are applied to temperature lag 2, absolute humidity lag 3 and relative humidity lag 2, corresponding roughly to points when the effects of weather on the scale and
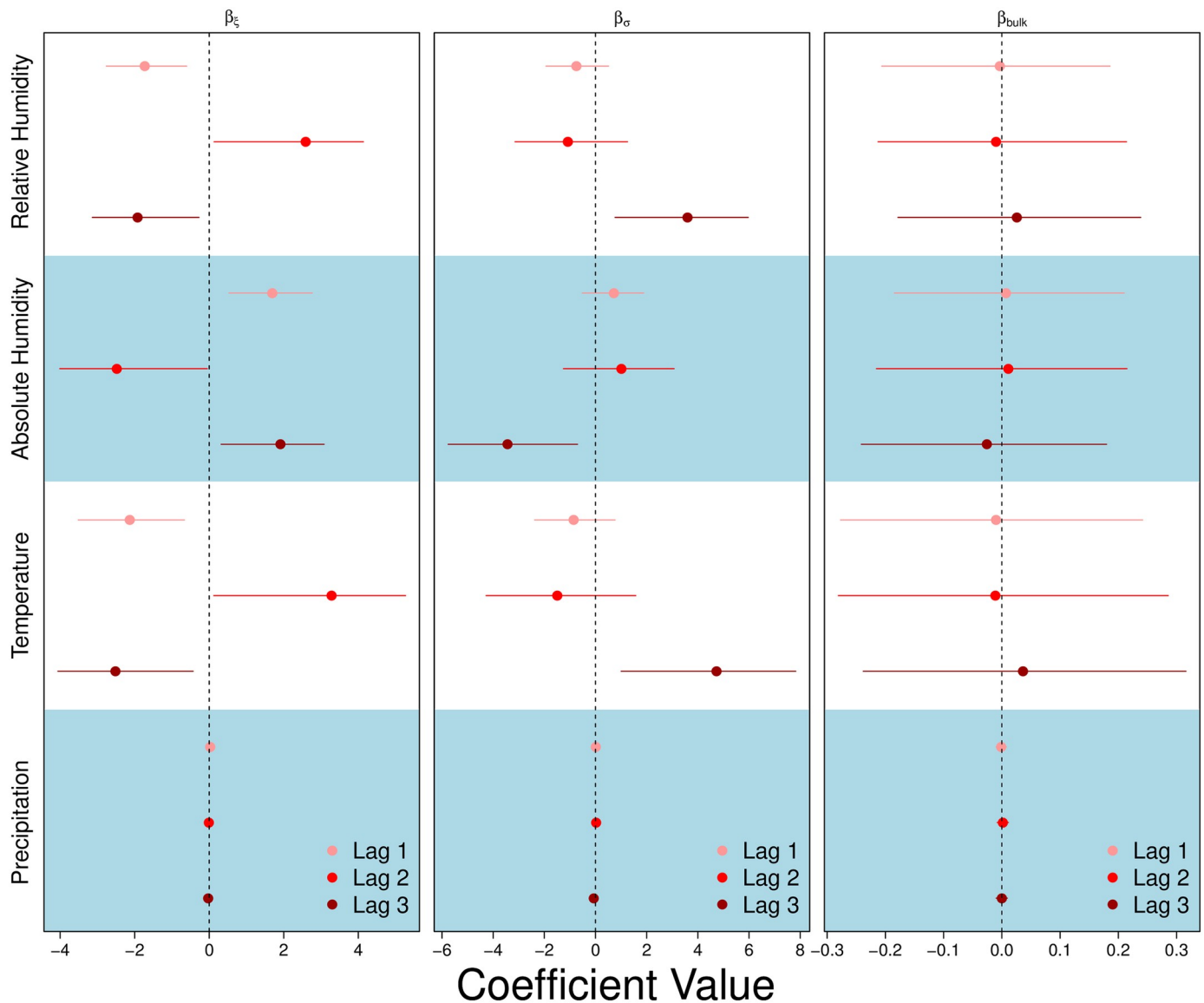
Fig 3. Coefficient plots for the constant beta regression for the generalized Pareto regression parameters and bulk regression. Panels from left to right represent the regression coefficients with the dependent variable being 1) Generalized Pareto distribution shape parameter 2) Generalized Pareto distribution scale parameter 3) Dengue case counts below the threshold.

https://doi.org/10.1371/journal.pcbi.1008279.g003

shape parameter are positive and negative respectively for temperature lag 2 and absolute humidity lag 3. They also correspond to instances when the effects of weather on the scale and shape parameter are negative and positive, respectively, for relative humidity lag 3 (Fig 3a–3c).

For a 5% independent, positive shock across each parameter at each lag, we found markedly larger magnitudes of change in the expected return levels across time (Fig 5). The same decreases were found in the expected return levels across time for temperature lag 1 and 3 and relative humidity lag 1 and 3, but these are markedly larger compared to the 1% shock (Figs 4 and 5). However, with a 5% shock in absolute humidity lag 1 and 2, their effects on return levels are now positive. The formerly negligible change when shocks are applied to temperature lag 2, absolute humidity lag 3 and relative humidity lag 2 are now positive and rise much more steeply after extending the return period beyond 10 years (Fig 5).
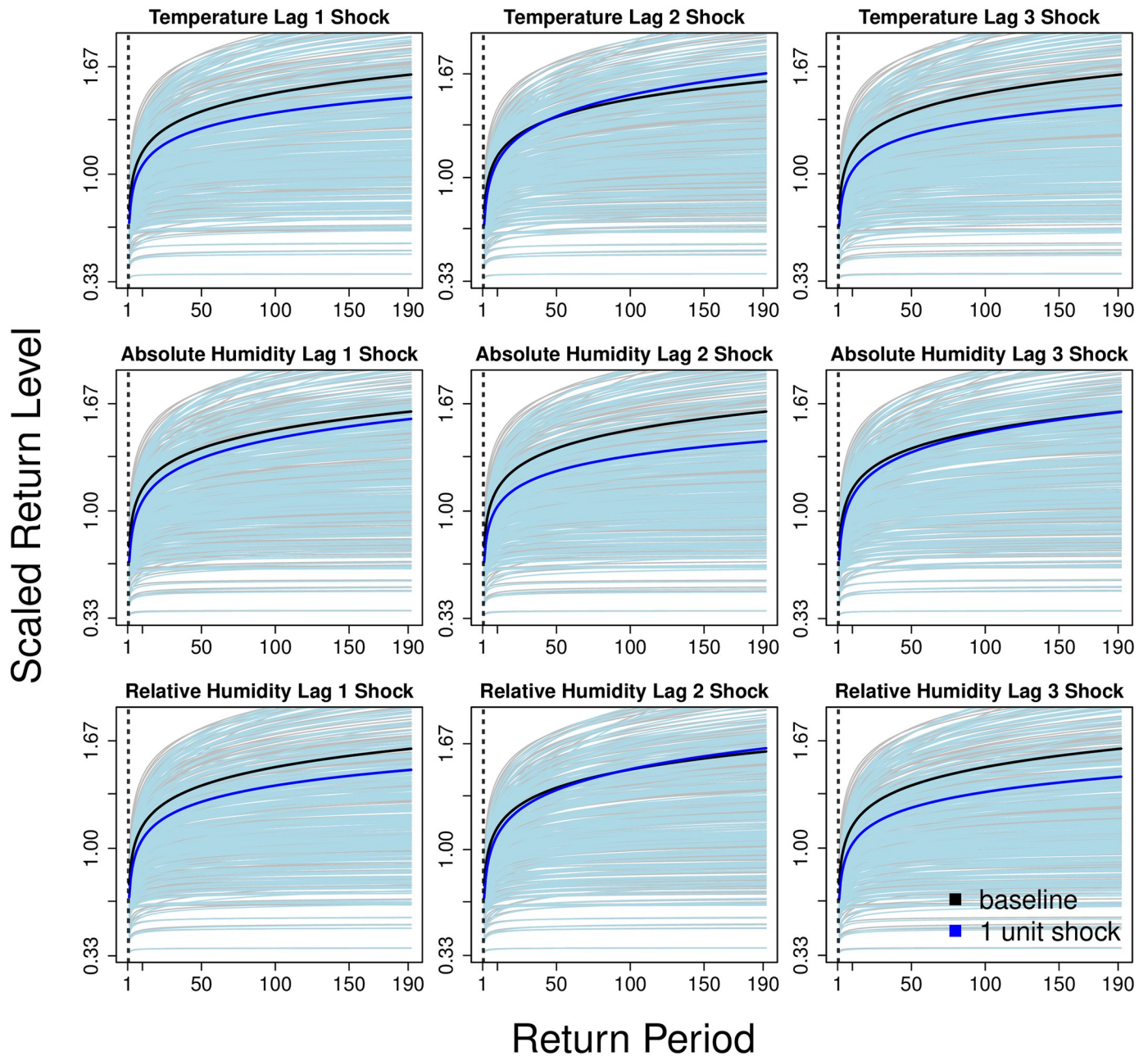
Fig 4. Posterior mean return levels at baseline (black solid line) and posterior mean return levels (blue solid line) given a 1 unit shock on each respective climatic parameter for the constant beta regression over 1 to 190 years.

Discussion

Results indicate that the tvEM model with regression structure on the extreme and non-extreme distributions of dengue case counts (M4) could characterize the data well under model assessment checks (Table 1, S1 Fig). Using M4 to infer weather effects below and above thresholds of dengue transmission showed a duality of weather effects below and above thresholds of dengue transmission, where it is found to affect the GPD distribution parameters but not case counts below the threshold (Table 2). The effects of weather on extremes in dengue
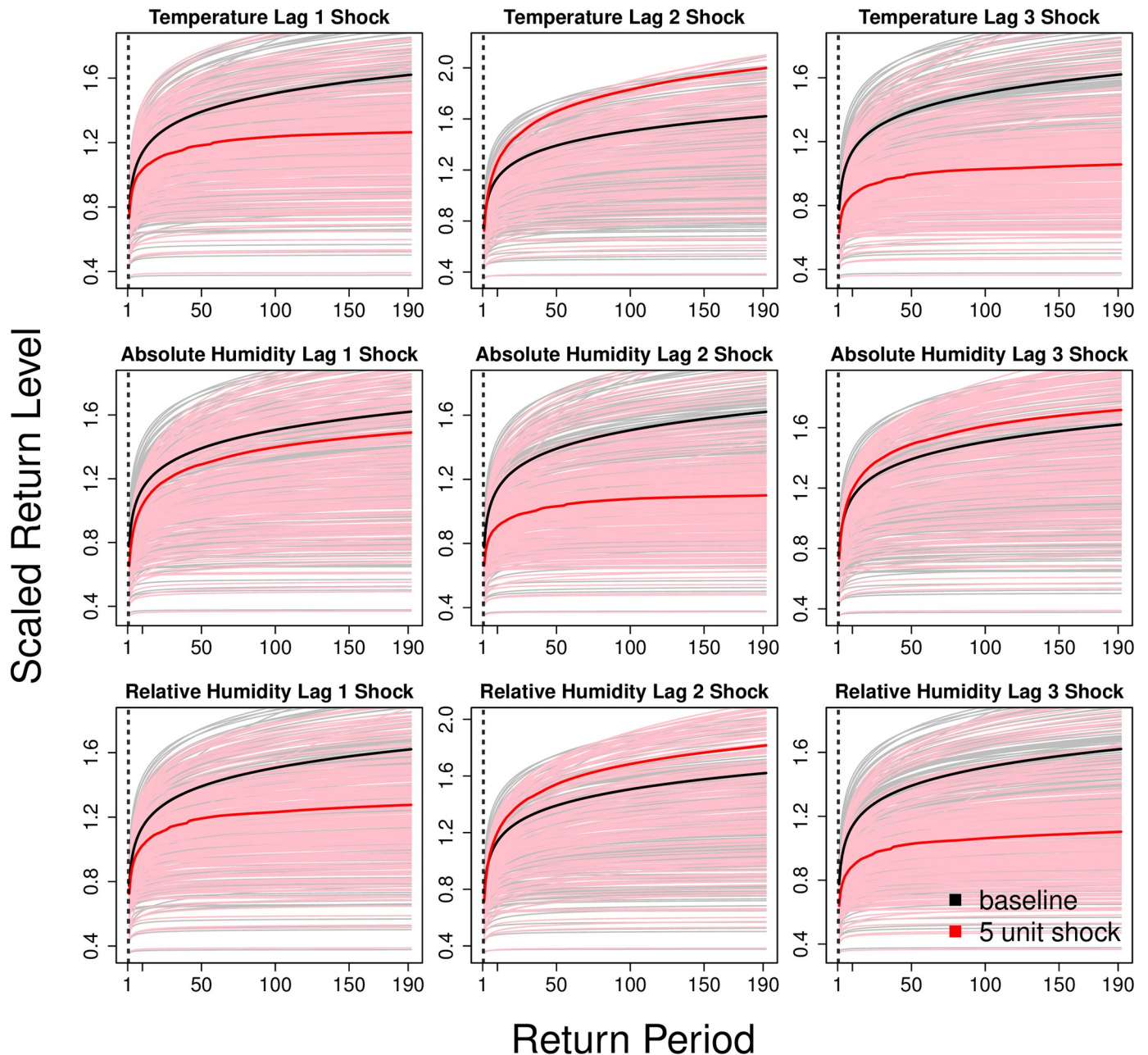
Fig 5. Posterior mean return levels at baseline (black solid line) and posterior mean return levels (red solid line) given a 5 unit shock on each climatic parameter for the constant beta regression over 1 to 190 years.

transmission are also highly nonlinear and severe, with a mild to moderate increase in weather observations leading to long run changes in expected return levels over time (Figs 4 and 5).

Inference of extreme and non-extreme periods of dengue transmissions across time can be conducted using tvEM, with the model being able to characterize the temporal dependence of dengue case counts across time. In general, inference through tvEM indicates that weather measurements such as absolute humidity, relative humidity and temperature are able to influence projected return levels over time, but these same variables do not have a near term

influence on dengue case counts directly. These results are in line with previous work [24] which found that dengue case counts in Singapore can be sufficiently described autoregressively. While said weather variables may exert influence on dengue case data over longer time-lags, spurious correlations may result from such an analysis due to temporal dynamics already adequately accounted for in the bulk distribution.

Our results found that a mild to moderate shock in weather variables would lead to considerable changes in the expected return levels over a long return period. As indicated from this analysis, independent 1% positive shocks on weather variables on 1 to 3 week lags would lead to either little change or decreases in projected return levels, while larger 5% shocks would more often lead to increases in projected return levels in dengue case counts across time (Figs 4 and 5). Other work has also revealed non-linear, interactive effects of weather on dengue transmission dynamics, wherein weather measurements may often exhibit mixed, concave and/or second order interaction effects on reported cases. For example, large rises in precipitation may reduce viable breeding grounds immediately, but may also lead to the formation of future breeding grounds through rainwater accumulation [25–29]. Evidence also suggests that mosquito ovipositing, development from mosquito larva to adult, biting rate and viral incubation time in the dengue vector are enhanced at raised ambient temperatures, but these effects have marginal gains at much higher temperatures [30, 31]. Similarly, optimally humid conditions can boost vector egg development and subsequent adult population size that may itself be correlated with the transmission potential of dengue [32–34].

Using tvEM, events of at least 526.6, 1052.2 and 1183.6 weekly dengue case counts are expected to occur in 5, 50 and 75 years respectively in Singapore. This notion of return levels is helpful for public health resource planning purposes. In the case of dengue, the expected surge in vector control, primary and secondary care required to deal with dengue fever (DF) and dengue hemorrhagic fever (DHF) can easily be computed using prior data on DF/DHF occurrence [35] and the expected number of visits and medication required for DF/DHF patients [36], both as a factor of the return levels. This method is also easily translatable to other diseases in different localities, regions or countries where disease case counts are collected over a long period of time, where the notion of exceeding a certain number of cases in any disease can be obtained for risk management purposes.

The methods described in this paper are also able to resolve certain issues in empirical EVT applications. Thresholds are estimated automatically using the MCMC approach employed in this paper, with threshold estimation uncertainty also quantified in its posterior distribution. This allows individuals to ignore the often subjective methods to select thresholds using graphical diagnostic tests and various rules of thumb approaches while working with extreme values empirically [17, 18, 37, 38]. Next, imposing an hierarchical temporal structure in tvEM also resolves the temporal dependence of observed data for the extreme value distribution, allowing stronger inference on time dependent data such as dengue case counts. This could easily be translated for use in other endemic infectious disease data such as influenza, for which case counts are also temporally dependent. Lastly, by inducing a regression structure on the extreme value distribution, statistical learning on relevant covariates could be conducted, which then allows meaningful interpretations of the covariates on projected return levels. These covariate effects could then also be compared to that of non-extreme value data within a single model structure, as evidenced by tvEM in this paper.

Several limitations are recognised. First, explicitly quantifying the effect of covariates is difficult in the extreme portion of data, as their effects are on the time-varying extreme value distribution parameters rather than the dependent variable itself. However, while direct comparison between extreme and non-extreme portions of the data is challenging, we used the shock notion to explore the effects on return levels. Time varying parameters estimated

over the extreme value distribution are also difficult to interpret and return levels have to be computed after integrating out the parameters sampled over the range of values over the threshold. The large number of parameters estimated due to time varying extreme value parameters also makes the MCMC estimation approach slow and future work should explore the use of approximate but faster estimation methodologies such as particle MCMC or approximate Bayesian computation [39, 40]. The suitability for each EVT tool on data also cannot be evaluated directly using quantities such as the mean-square error or R^2 as estimation above the threshold provides a time varying regression on parameters but not observed case counts unlike the bulk distribution formulation, leading us to use other diagnostic measures such as DIC or logBF in model comparison. The results of this paper are also only based on passive surveillance data on dengue case counts in Singapore, and may not apply to other health systems with different case definitions or surveillance systems. Lastly, while dengue case counts are plausibly spatio-temporally dependent and may have a large number of past case counts, different multicollinear weather lags affecting contemporaneously reported dengue cases, only the most parsimonious model adequately accounting for temporal dynamics is used due to the inherent difficulty in estimating a time varying distribution with regression structure.

Future work can explore nesting sparse regression priors or spatio-temporal priors on the extreme value regression to infer and incorporate spatio-temporal behaviours or the effect of many covariates on dengue extremes. Incorporation of sparse regression priors may also resolve the issue of multicollinearity between many variables in both bulk and extreme portions of data [41–43]. When data become available, the same assessment checks described in this paper can also be conducted on separate dengue time series to examine model validity, climate-extreme value exposure-response relationships and disparities in return levels.

The methods developed in this paper are easily applicable to any other infectious disease where time series are recorded over a sufficient period of time. These methods are able to conduct statistical learning of covariates on extreme and non-extreme periods of time series data. Lastly, tvEM is able to provide meaningful notions of return levels on a pre-specified return period, which in the case of infectious disease data, allows public health officials to prepare for the likely scale of an infectious disease outbreak in the long run.

Methods

Data

Dengue incidence data in Singapore are collected by the Ministry of Health with mandatory notification of virologically confirmed or laboratory-confirmed cases. Laboratory confirmation of dengue cases is conducted through (1) nonstructural protein 1 (NS1) antigen detection, viral RNA detection by polymerase chain reaction (PCR), or (2) immunoglobulin M detection [44, 45]. Data are publicly available from the Infectious Disease Bulletin, published weekly by the Ministry of Health, Singapore. Data is available from 2010 to 2017. No ethical approval is required for this study.

Weather data were obtained from ERA5, published by the European Centre for Medium-Range Weather Forecasts. ERA5 provides hourly estimates across a 30km grid [46], which we aggregate nationally over a weekly timescale and spatially averaged for Singapore. Mean, minimum and maximum air temperature at 2m is calculated to represent thermal forcing and stress on vector population growth, and weekly total rainfall is obtained for breeding site availability. Air temperature and dewpoint temperature is utilized to calculate saturation vapor pressure and actual vapor pressure using Teten's formula, with which relative and absolute humidity could then be estimated using standard formula [47]. Data are available from 2010 to 2017.

Extreme value mixtures to infer bulk and extreme periods of dengue

We have used mixtures of bulk and extreme distributions as our key framework to derive statistical inference on both non-extreme and extreme periods of dengue transmission dynamics. The general structure of our model follows (1), where y_t denotes our data at timepoint t, $F_t(y_{1:t}|\Theta)$ the cumulative distribution function of our model, $H(y_t; u, \Theta_{-u})$ the bulk cumulative distribution with parameters Θ_{-u} below the threshold u and $G(y_t|u, \Theta_u)$ the tail cumulative distribution with parameters Θ_{-u} above the threshold u.

$$F_t(y_{1:t}|\Theta) = \begin{cases} H(y_t; u, \Theta_{-u}) & \text{if } y_t < u \\ H(u; \Theta_{-u}) + (1 - H(u; \Theta_{-u}))G(y_t; u, \Theta_u) & \text{if } y_t \geq u \end{cases} \qquad (1)$$

Bulk distribution. Four competing models were estimated, first, the canonical case where no regression structure was imposed on y_t, with $h(y_t; u, \Theta_{-u}) \sim \text{Gamma}(a, b)$ and $g(y_t|u, \Theta_u) \sim \text{GPD}(u, \xi, \sigma)$, where GPD refers to the generalized Pareto distribution [23, 48]. Next, the autoregressive (AR) structure without (M2) or with (M3) weather variables (2) was imposed to the bulk distribution with normally distributed errors. Namely, P denotes the AR order, with J being the number of exogenous variables and K the maximal number of lags estimated for exogenous variables. $X_{t-k,j}$ refers to the exogenous variable at lag k of type j. AR and lagged parameters to be estimated are $\Theta_{-u} = \{\beta, \alpha, \sigma_b\}$. ϵ is some noise term with 0 mean and second moment σ_b^2. The AR structure was used for easily interpretable parameters and being able to capture the time-dependent behaviour of dengue case data. It is also supported by a large body of work using AR as a baseline to both understand and forecast dengue transmissions [34, 49–51].

$$y_t = \beta_0 + \sum_{i=1}^{P} y_{t-i}\beta_i + \sum_{j=1}^{J}\sum_{k=1}^{K} X_{t-k,j}\alpha_{k,j} + \epsilon \qquad (2)$$

$$\epsilon \sim N(0, \sigma_b^2)$$

Tail distribution. For values above some threshold u, it is standard to fit data using the generalized Pareto distribution (GPD) with $\Theta_u = \{\xi, \sigma, u\}$ shape, scale and location parameters to be estimated respectively. Our first model (M1) considers the canonical, static distribution with no regression structure imposed on the GPD (3).

$$G(y_t; u, \Theta_u) = \begin{cases} 1 - \left(1 + \frac{\xi(y_t - u)}{\sigma}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp(-(y_t - u)/\sigma) & \text{if } \xi = 0 \end{cases} \qquad (3)$$

We then consider imposing time varying parameters on the GPD, with a fixed threshold u, but having the scale and shape parameter vary across time $\Theta_{u,t} = \{\xi_t, \sigma_t, u\}$ (4). The additional flexibility of using a time varying structure was imposed on ξ_t, σ_t as observed dengue case counts may fluctuate greatly over time, especially in dengue hyper-endemic Singapore, which can skew parameter estimation if we only consider static parameters for the GPD distribution. Furthermore, imposing a suitable structure for time variation allows incorporation of dengue case count autocorrelations from successive timepoints. Lastly, past information on model parameters can be incorporated contemporaneously, strengthening inference procedures for the

GPD.

$$\Theta_{u,t} = G(\Theta_{u,t-1}, \Theta_{u,t-2}, \ldots, \Theta_{u,t-g}, w_t) \tag{4}$$

Temporal dependence. The time evolution of our GPD parameters follow some function G with input variables being the lagged GPD parameters and w_t some noise term (4). The time varying GPD parameters were considered using two functional forms. Firstly, we consider the simplest case of a first order dynamic linear model [52], where parameters follow random walk state equations, with some white noise terms $w_{\xi,t}$ and $w_{\sigma,t}$ for the shape and scale equations (M2-3) given some initial information drawn from $\theta_{\xi,0} \sim N(m_{\xi,0}, C_{\xi,0})$ and $\theta_{\sigma,0} \sim N(m_{\sigma,0}, C_{\sigma,0})$. We log-transform ξ_t, σ_t to allow parameters to be within the allowable bounds for the GPD ($\xi > -1$) with $l\xi_t = \log(\xi_t + 1)$ and $l\sigma_t = \log(\sigma_t)$:

$$l\xi_t = \theta_{\xi,t} + v_{\xi,t} \qquad v_{\xi,t} \sim N(0, 1/V_\xi) \qquad \theta_{\xi,t} = \theta_{\xi,t-1} + w_{\xi,t} \qquad w_{\xi,t} \sim N(0, 1/W_\xi)$$

$$l\sigma_t = \theta_{\sigma,t} + v_{\sigma,t} \qquad v_{\sigma,t} \sim N(0, 1/V_\sigma) \qquad \theta_{\sigma,t} = \theta_{\sigma,t-1} + w_{\sigma,t} \qquad w_{\sigma,t} \sim N(0, 1/W_\sigma)$$

Secondly, to allow inference and comparison between the possible effects that weather has on non-extreme and extreme periods of dengue transmissions, we imposed lagged regression structures on the time-varying GPD parameter (M4). To belabour, we allowed inference on $l\xi_t$, $l\sigma_t$ by additionally imposing state-determining Eqs (5) and (6). In M3, $l\xi_t$, $l\sigma_t$ is influenced by the effect of $X_{t-k,j}$ weather variables with a maximum of J weather terms and K lags. The parameters $\beta_{\xi,j,k}$, $\beta_{\sigma,j,k}$ determine the degree of influence that weather variables have on the GPD parameters.

$$l\xi_t = \beta_{\xi,0} + \sum_{j=1}^{J}\sum_{k=1}^{K}\beta_{\xi,j,k}X_{t-k,j} + v_{\xi,t} \qquad v_{\xi,t} \sim N(0, 1/V_\xi) \tag{5}$$

$$l\sigma_t = \beta_{\sigma,0} + \sum_{j=1}^{J}\sum_{k=1}^{K}\beta_{\sigma,j,k}X_{t-k,j} + v_{\sigma,t} \qquad v_{\sigma,t} \sim N(0, 1/V_\sigma) \tag{6}$$

Model estimation

We estimated all models (M1–M4) using Markov chain Monte Carlo (MCMC) approaches. In general, Gibbs sampling was conducted when suitable conditional conjugate distributions can be derived, such as the case of regression parameters in the bulk distribution or their respective noise terms. For the constant and time varying GPD parameters, there are no known conditionally conjugate priors. Marginal sampling of their parameters was conducted using Metropolis or Metropolis-Hastings steps instead. For all models, a total of 10 000 MCMC steps were taken with a burnin of 1000.

Briefly, for the first model where parameters of interest are the gamma bulk shape and rate parameters a, b and static GPD tail distribution ξ, σ, u, we conducted Metropolis-within-Gibbs steps for each of their respective marginal distributions with proposal distributions tuned to allow efficient exploration of the posterior space. This was also conducted due to the GPD distribution not having any known conjugate prior distributions, which precludes the use of more efficient Gibbs sampling. We sampled a, b, ξ, u using truncated normal distributions and σ using either a wide Gamma or truncated normal distribution depending on the sign of ξ. For

time varying tail distribution, random walk Metropolis steps were taken for $\{l\xi_t\}_{t=1}^{T}$ and $\{l\sigma_t\}_{t=1}^{T}$ with the proposal distribution a symmetric normal distributions with mean given by the previous iteration. With appropriate prior choices on the regression terms and second moments $\{V_\xi, W_\xi, \theta_{\xi,t}, V_\sigma, W_\sigma, \theta_{\sigma,t}, \beta, \sigma_b^2, u\}$, where diffuse normal priors are placed for regression parameters and inverse-gamma priors for their respective second moment, conjugate conditional posterior distributions of these terms allow for Gibbs sampling. Full technical details are explicitly provided in the supplementary information.

Model assessment

Convergence of MCMC chains was first assessed by visual inspection of trace plots and Gewecke convergence diagnostic checks. Residual autocorrelation is computed for up to 20 week lags to ensure that the transmission dynamics of dengue are properly accounted for and to determine the maximum autoregressive lag order for each specification. Quantile-Quantile plots are used to see whether the bulk and tail specifications adequately account for the data structure. The log Bayes factor was computed using naïve Monte Carlo simulation and deviance information criterion as detailed in S1 Text to provide summary measures of model fit and model appropriateness to data.

Model inference

As we have found the constant beta model (M4) to be favourable on model assessment, we define only return levels for that particular model for clear exposition. Estimation of the GPD provides us with return levels $\hat{z}_r$ for the r-year. This is the level expected to be exceeded once every r years, with $\hat{\lambda}_u$ the empirical threshold exceedance rate observed in the data set. We generalize return levels from the static distribution case to account for the presence of time varying GPD parameters which are also affected by other covariates in M3 (7).

$$\hat{z}_{r,t}\left(\hat{u}, \hat{\sigma}_t, \hat{\bar{\xi}}_t, \hat{\beta}_\xi, \hat{\beta}_\sigma, \mathbf{X}\right) = \begin{cases} \hat{u} + \frac{\hat{\sigma}_t(\hat{\beta}_\sigma, \mathbf{X})}{\hat{\xi}_t(\hat{\beta}_\xi, \mathbf{X})}\left((r\hat{\lambda}_u)^{\hat{\xi}_t(\hat{\beta}_\xi, \mathbf{X})} - 1\right), & \text{if } \hat{\bar{\xi}}_t(\hat{\beta}_\xi, , \mathbf{X}) \neq 0 \\ \hat{u} + \hat{\sigma}_t(\hat{\beta}_\sigma, \mathbf{X})\log(r\hat{\lambda}_u), & \text{if } \hat{\bar{\xi}}_t(\hat{\beta}_\xi, , \mathbf{X}) = 0 \end{cases} \tag{7}$$

The expression (7) above provides some return level for the r-year at every time-point t where GPD parameters were estimated, given some specified return period r. Integrating out the expression across time provides us with the mean return level for dengue transmissions over the tail dataset (8).

$$\mathbb{E}_t[\hat{z}_{r,t}] = \int \hat{z}_{r,t}(\hat{u}, \hat{\sigma}_t, \hat{\bar{\xi}}_t, \hat{\beta}_\xi, \hat{\beta}_\sigma, \mathbf{X}) \, dt \tag{8}$$

The subsequent effect of some weather variable on expected return levels can be inferred by comparing the return levels at baseline versus some unit shock $\omega_{l,p}$ for some prespecified variable and lag across the dataset (9) and (10). It was then back transformed to allow computation of return levels in (7) and (8) given the new scenario. In this paper, we have looked at the effect of 1% and 5% increases in the observed range of each of our weather variables on our return

levels.

$$\hat{l\xi}_{t,shock} = \sum_{j\neq l}^{J}\sum_{k\neq p}^{K}\hat{\beta}_{\xi,j,k}X_{t-j,k} + \hat{\beta}_{\xi,l,p}(X_{t-l,p} + \omega_{l,p}) \tag{9}$$

$$\hat{l\sigma}_{t,shock} = \sum_{j\neq l}^{J}\sum_{k\neq p}^{K}\hat{\beta}_{\sigma,j,k}X_{t-j,k} + \hat{\beta}_{\sigma,l,p}(X_{t-l,p} + \omega_{l,p}) \tag{10}$$

Supporting information

S1 Text. Details of MCMC procedures and model assessment checks.
(PDF)

S1 Fig. Results of model asssessment checks.
(PDF)

Author Contributions

Conceptualization: Jue Tao Lim.

Data curation: Jue Tao Lim.

Formal analysis: Jue Tao Lim, Yiting Han.

Funding acquisition: Alex R. Cook.

Investigation: Jue Tao Lim, Yiting Han.

Methodology: Jue Tao Lim, Yiting Han.

Project administration: Jue Tao Lim.

Resources: Jue Tao Lim.

Software: Jue Tao Lim, Yiting Han.

Supervision: Alex R. Cook.

Validation: Jue Tao Lim, Yiting Han.

Visualization: Yiting Han.

Writing – original draft: Jue Tao Lim, Yiting Han.

Writing – review & editing: Jue Tao Lim, Borame Sue Lee Dickens, Lee Ching Ng, Alex R. Cook.

References

1. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013; 496(7446):504. https://doi.org/10.1038/nature12060 PMID: 23563266

2. Cummings DA, Irizarry RA, Huang NE, Endy TP, Nisalak A, Ungchusak K, et al. Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. Nature. 2004; 427(6972):344–347. https://doi.org/10.1038/nature02225 PMID: 14737166

3. Low SL, Lam S, Wong WY, Teo D, Ng LC, Tan LK. Dengue seroprevalence of healthy adults in Singapore: serosurvey among blood donors, 2009. The American journal of tropical medicine and hygiene. 2015; 93(1):40–45. https://doi.org/10.4269/ajtmh.14-0671 PMID: 26013376

4. Wilder-Smith A, Foo W, Earnest A, Sremulanathan S, Paton NI. Seroepidemiology of dengue in the adult population of Singapore. Tropical Medicine & International Health. 2004; 9(2):305–308. https://doi.org/10.1046/j.1365-3156.2003.01177.x

5. Tan LK, Low SL, Sun H, Shi Y, Liu L, Lam S, et al. Force of Infection and True Infection Rate of Dengue in Singapore: Implications for Dengue Control and Management. American journal of epidemiology. 2019; 188(8):1529–1538. https://doi.org/10.1093/aje/kwz110 PMID: 31062837

6. Tricou V, Low JG, Oh HM, Leo YS, Kalimuddin S, Wijaya L, et al. Safety and immunogenicity of a single dose of a tetravalent dengue vaccine with two different serotype-2 potencies in adults in Singapore: a phase 2, double-blind, randomised, controlled trial. Vaccine. 2020; 38(6):1513–1519. https://doi.org/10.1016/j.vaccine.2019.11.061 PMID: 31843269

7. Biswal S, Reynales H, Saez-Llorens X, Lopez P, Borja-Tabora C, Kosalaraksa P, et al. Efficacy of a tetravalent dengue vaccine in healthy children and adolescents. New England Journal of Medicine. 2019; 381(21):2009–2019. https://doi.org/10.1056/NEJMoa1903869 PMID: 31693803

8. Lee JS, Lourenco J, Gupta S, Farlow A. A multi-country study of dengue vaccination strategies with Dengvaxia and a future vaccine candidate in three dengue-endemic countries: Vietnam, Thailand, and Colombia. Vaccine. 2018; 36(17):2346–2355. https://doi.org/10.1016/j.vaccine.2018.03.002 PMID: 29573874

9. Organization WH, et al. Dengue vaccine: WHO position paper, July 2016–recommendations. Vaccine. 2017; 35(9):1200–1201. https://doi.org/10.1016/j.vaccine.2016.10.070

10. Achee NL, Grieco JP, Vatandoost H, Seixas G, Pinto J, Ching-Ng L, et al. Alternative strategies for mosquito-borne arbovirus control. PLoS neglected tropical diseases. 2019; 13(1). https://doi.org/10.1371/journal.pntd.0006822

11. Hapuarachchi HC, Koo C, Rajarethinam J, Chong CS, Lin C, Yap G, et al. Epidemic resurgence of dengue fever in Singapore in 2013-2014: a virological and entomological perspective. BMC infectious diseases. 2016; 16(1):300. https://doi.org/10.1186/s12879-016-1606-z PMID: 27316694

12. Rajarethinam J, Ang LW, Ong J, Ycasas J, Hapuarachchi HC, Yap G, et al. Dengue in Singapore from 2004 to 2016: cyclical epidemic patterns dominated by serotypes 1 and 2. The American journal of tropical medicine and hygiene. 2018; 99(1):204–210. https://doi.org/10.4269/ajtmh.17-0819 PMID: 29848407

13. Lim JT, Dickens BS, Haoyang S, Ching NL, Cook AR. Inference on dengue epidemics with Bayesian regime switching models. PLoS computational biology. 2020; 16(5):e1007839. https://doi.org/10.1371/journal.pcbi.1007839 PMID: 32357146

14. Yamana TK, Kandula S, Shaman J. Superensemble forecasts of dengue outbreaks. Journal of The Royal Society Interface. 2016; 13(123):20160410. https://doi.org/10.1098/rsif.2016.0410

15. Chen Y, Ong JHY, Rajarethinam J, Yap G, Ng LC, Cook AR. Neighbourhood level real-time forecasting of dengue cases in tropical urban Singapore. BMC medicine. 2018; 16(1):1–13. https://doi.org/10.1186/s12916-018-1108-5

16. Shi Y, Liu X, Kok SY, Rajarethinam J, Liang S, Yap G, et al. Three-month real-time dengue forecast models: an early warning system for outbreak alerts and policy decision support in Singapore. Environmental health perspectives. 2016; 124(9):1369–1375. https://doi.org/10.1289/ehp.1509981 PMID: 26662617

17. Coles S, Bawa J, Trenner L, Dorazio P. An introduction to statistical modeling of extreme values. vol. 208. Springer; 2001.

18. De Haan L, Ferreira A. Extreme value theory: an introduction. Springer Science & Business Media; 2007.

19. Thomas M, Lemaitre M, Wilson ML, Viboud C, Yordanov Y, Wackernagel H, et al. Applications of extreme value theory in public health. PloS one. 2016; 11(7). https://doi.org/10.1371/journal.pone.0159312

20. Ebel ED, Williams MS, Ward-Gokhale LA, Kisselburgh HM. Assessing the maximum size of annual foodborne outbreaks in the United States: An analysis of 1973–2016 outbreaks. Microbial Risk Analysis. 2019; 12:20–26. https://doi.org/10.1016/j.mran.2019.02.002

21. Hu Y. Extreme Value Mixture Modelling with Simulation Study and Applications in Finance and Insurance. 2013.

22. Sang H, Gelfand AE. Hierarchical modeling for extreme values observed over space and time. Environmental and ecological statistics. 2009; 16(3):407–426. https://doi.org/10.1007/s10651-007-0078-0

23. Huerta G, Sansó B. Time-varying models for extreme values. Environmental and Ecological Statistics. 2007; 14(3):285–299. https://doi.org/10.1007/s10651-007-0014-3

24. Chen Y, Cook AR, Lim AX. Randomness of dengue outbreaks on the equator. Emerging infectious diseases. 2015; 21(9):1651. https://doi.org/10.3201/eid2109.141030 PMID: 26291469

25. Vezzani D, Rubio A, Velazquez S, Schweigmann N, Wiegand T. Detailed assessment of microhabitat suitability for Aedes aegypti (Diptera: Culicidae) in Buenos Aires, Argentina. Acta Tropica. 2005; 95 (2):123–131. https://doi.org/10.1016/j.actatropica.2005.03.010 PMID: 15993832

26. Valença M, Marteis L, Steffler L, Silva A, Santos R. Dynamics and characterization of Aedes aegypti (L.)(Diptera: Culicidae) key breeding sites. Neotropical entomology. 2013; 42(3):311–316. https://doi.org/10.1007/s13744-013-0118-4 PMID: 23949815

27. Weeraratne TC, Perera MDB, Mansoor MM, Karunaratne SP. Prevalence and breeding habitats of the dengue vectors Aedes aegypti and Aedes albopictus (Diptera: Culicidae) in the semi-urban areas of two different climatic zones in Sri Lanka. International Journal of Tropical Insect Science. 2013; 33(4):216–226. https://doi.org/10.1017/S174275841300026X

28. Xu HY, Fu X, Lee LKH, Ma S, Goh KT, Wong J, et al. Statistical modeling reveals the effect of absolute humidity on dengue in Singapore. PLoS neglected tropical diseases. 2014; 8(5). https://doi.org/10.1371/journal.pntd.0002805

29. Do TTT, Martens P, Luu NH, Wright P, Choisy M. Climatic-driven seasonality of emerging dengue fever in Hanoi, Vietnam. BMC public health. 2014; 14(1):1078. https://doi.org/10.1186/1471-2458-14-1078 PMID: 25323458

30. Watts DM, Burke DS, Harrison BA, Whitmire RE, Nisalak A. Effect of temperature on the vector efficiency of Aedes aegypti for dengue 2 virus. The American journal of tropical medicine and hygiene. 1987; 36(1):143–152. https://doi.org/10.4269/ajtmh.1987.36.143 PMID: 3812879

31. Dickerson CZ. The effects of temperature and humidity on the eggs of Aedes aegypti (L.) and Aedes albopictus (Skuse) in Texas. Texas A&M University; 2007.

32. Vargas REM, Ya-umphan P, Phumala-Morales N, Komalamisra N, Dujardin JP. Climate associated size and shape changes in Aedes aegypti (Diptera: Culicidae) populations from Thailand. Infection, Genetics and Evolution. 2010; 10(4):580–585. https://doi.org/10.1016/j.meegid.2010.01.004

33. Descloux E, Mangeas M, Menkes CE, Lengaigne M, Leroy A, Tehei T, et al. Climate-based models for understanding and forecasting dengue epidemics. PLoS Negl Trop Dis. 2012; 6(2):e1470. https://doi.org/10.1371/journal.pntd.0001470 PMID: 22348154

34. Wu PC, Guo HR, Lung SC, Lin CY, Su HJ. Weather as an effective predictor for occurrence of dengue fever in Taiwan. Acta tropica. 2007; 103(1):50–57. https://doi.org/10.1016/j.actatropica.2007.05.014 PMID: 17612499

35. Shepard DS, Undurraga EA, Halasa YA, Stanaway JD. The global economic burden of dengue: a systematic analysis. The Lancet infectious diseases. 2016; 16(8):935–941. https://doi.org/10.1016/S1473-3099(16)00146-8 PMID: 27091092

36. Rigau-Pérez JG, Clark GG, Gubler DJ, Reiter P, Sanders EJ, Vorndam AV. Dengue and dengue haemorrhagic fever. The Lancet. 1998; 352(9132):971–977. https://doi.org/10.1016/S0140-6736(97)12483-7

37. Hüsler J, Li D. Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields. 2007.

38. Ferreira A, de Haan L, Peng L. On optimising the estimation of high quantiles of a probability distribution. Statistics. 2003; 37(5):401–434. https://doi.org/10.1080/0233188021000055345

39. Andrieu C, De Freitas N, Doucet A. Sequential MCMC for Bayesian model selection. In: Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics. SPW-HOS'99. IEEE; 1999. p. 130–134.

40. Andrieu C, Doucet A, Holenstein R. Particle markov chain monte carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2010; 72(3):269–342. https://doi.org/10.1111/j.1467-9868.2009.00736.x

41. Nobre AA, Schmidt AM, Lopes HF. Spatio-temporal models for mapping the incidence of malaria in Pará. Environmetrics: The official journal of the International Environmetrics Society. 2005; 16(3):291–304. https://doi.org/10.1002/env.704

42. Alegana VA, Atkinson PM, Wright JA, Kamwi R, Uusiku P, Katokele S, et al. Estimation of malaria incidence in northern Namibia in 2009 using Bayesian conditional-autoregressive spatial–temporal models. Spatial and spatio-temporal epidemiology. 2013; 7:25–36. https://doi.org/10.1016/j.sste.2013.09.001 PMID: 24238079

43. Park T, Casella G. The bayesian lasso. Journal of the American Statistical Association. 2008; 103 (482):681–686. https://doi.org/10.1198/016214508000000337

44. Lai YL, Chung YK, Tan HC, Yap HF, Yap G, Ooi EE, et al. Cost-effective real-time reverse transcriptase PCR (RT-PCR) to screen for Dengue virus followed by rapid single-tube multiplex RT-PCR for serotyping of the virus. Journal of clinical microbiology. 2007; 45(3):935–941. https://doi.org/10.1128/JCM.01258-06 PMID: 17215345

45. Pok KY, Lai YL, Sng J, Ng LC. Evaluation of nonstructural 1 antigen assays for the diagnosis and surveillance of dengue in Singapore. Vector-Borne and Zoonotic Diseases. 2010; 10(10):1009–1016. https://doi.org/10.1089/vbz.2008.0176 PMID: 20426686

46. ERA5. Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS) Copernicus Climate Change Service (C3S); 2017. Available from: https://www.ecmwf.int/en/about/what-we-do/environmental-services/copernicus-climate-change-service.

47. Snyder R. Humidity conversion. Davis, University of California, Biometeorology Program (http://biomet.ucdavis.edu/conversions/HumCon.htm). 2005.

48. Do Nascimento FF, Gamerman D, Lopes HF. Time-varying extreme pattern with dynamic models. Test. 2016; 25(1):131–149. https://doi.org/10.1007/s11749-015-0444-4

49. Eastin MD, Delmelle E, Casas I, Wexler J, Self C. Intra-and interseasonal autoregressive prediction of dengue outbreaks using local weather and regional climate for a tropical environment in Colombia. The American journal of tropical medicine and hygiene. 2014; 91(3):598–610. https://doi.org/10.4269/ajtmh.13-0303 PMID: 24957546

50. Hii YL, Zhu H, Ng N, Ng LC, Rocklöv J. Forecast of dengue incidence using temperature and rainfall. PLoS neglected tropical diseases. 2012; 6(11). https://doi.org/10.1371/journal.pntd.0001908

51. Abdul Rahim H, Ibrahim F, Taib M. System identification of nonlinear autoregressive models in monitoring dengue infection. International Journal on Smart Sensing and Intelligent Systems. 2010; 3(4):783–806. https://doi.org/10.21307/ijssis-2017-421

52. West M, Harrison J. Bayesian forecasting and dynamic models. Springer Science & Business Media; 2006.