Interdisciplinary modelling and forecasting of dengue

Cathal Mills*[1,2], Moritz U. G. Kraemer[2,3] and Christl A. Donnelly[1,2]

[1]Department of Statistics, University of Oxford, Oxford, United Kingdom.
[2]Pandemic Sciences Institute, University of Oxford, Oxford, United Kingdom.
[3]Department of Biology, University of Oxford, Oxford, United Kingdom.

* Corresponding Author

E-mail: cathal.mills@linacre.ox.ac.uk (C.M.)

Abstract

Understanding the past, current, and future dynamics of dengue epidemics is challenging yet increasingly important. To date, many techniques across statistics, mathematics, and machine learning have provided us with quantitative tools for studying dengue epidemics. Here, using data from provinces in northern Peru across 2010 to 2021, we provide a new interdisciplinary pipeline that draws on a new and existing techniques to provide comprehensive understanding and robust prediction of dengue epidemic dynamics.

Wavelet analyses can unveil spatiotemporal patterns in epidemic dynamics across annual and multi-annual time periods. Here, these included climatic forcing and greater spatial similarity in large outbreak years. Space-varying epidemic drivers included climatic influences and shorter pairwise distances driving greater epidemic similarity in more northerly coastal provinces. Then, using a Bayesian model, we can probabilistically quantify the timing, structure, and intensity of such climatic influences on Dengue Incidence Rates (DIRs), while simultaneously considering other influences. Recognising that a single model is generally sub-optimal for any forecasting task, we demonstrate how to form trained and untrained probabilistic ensembles for forecasting dengue cases in settings reflective of real-world conditions. We introduce a suite of climate-informed and covariate-free deep learning approaches that leverage big data and foundational time series, temporal convolutional networks, and conformal inference. We complement these modern techniques with statistically principled training and assessment of ensemble frameworks, while explicitly considering strong benchmark models, computational costs, public health priorities, and data availability limitations. In doing so, we show how ensemble frameworks consistently outperform individual models across space and time, and produce sharp and accurate forecasts with robust, reliable descriptions of uncertainty. We report interpretable classification metrics for detection of outbreaks to communicate our outputs with the wider public and public health authorities.

Looking forward, whether the objective is to understand and/or to predict epidemic dynamics, our modelling pipeline can be used in any dengue setting to robustly inform the decision-making and planning of public health authorities.

1. Introduction

Dengue is a mosquito-borne disease which poses an increasingly global public health threat. Endemic in tropical and subtropical regions, reported annual incidence has risen from 500,000 cases in 2000 to

over 5 million cases in 2023. 80% of the reported cases in 2023 occurred in the WHO Region of the Americas, yet dengue is endemic in over 100 nations [1,2]. Primarily spread by *Aedes* mosquito vectors (female *Ae. aegypti* and *Ae albopictus*), the viral infection is caused by the dengue virus (DENV), a flavivirus with four genetically distinct serotypes. Infection with an individual serotype offers long-term immunity to the specific serotype, but only short-term immunity to others [3]. Symptoms of dengue vary in severity, and up to 80% of all infections are asymptomatic [1].

The increasing number of dengue outbreaks has been linked to human, viral, and environmental factors. Climatic conditions such as temperature, precipitation, and humidity influence the mosquito abundance and DENV replication, as *Aedes* mosquitoes thrive in warm, humid conditions and precipitation is necessary for a mosquito's juvenile stages [4,5,6,7]. Extreme climatic conditions, including drought and heavy precipitation, instigated by events such as the El Niño phenomena, have been identified as epidemic drivers [8,9,10]. Human factors like rapid urbanisation have contributed to the creation of mosquito breeding sites and addition of susceptible hosts in close proximity. Similarly, increased travel, human mobility, and globalisation have contributed to greater incidence and potential for rapid geographic spread [11,12].

Faced with a difficult-to-detect, climate-sensitive disease, public health authorities require a quantitative understanding of epidemic dynamics and drivers to inform their decision-making. In the absence of complete spatial or temporal coverage of dengue surveillance, authorities may also seek alternative ways to fill gaps in surveillance. Prospectively, public health authorities aim to have early warning systems which probabilistically predict future epidemic trajectories and the likelihood of outbreaks, thus allowing sufficient time for design of evidence-based public health policies. Therefore, in Brazil, Southeast Asia, and Barbados, modellers have used wavelets, Bayesian hierarchical models, deep learning (DL)-based methods, and ensemble frameworks as effective modelling techniques. These have provided model-based understanding of spatial patterns (e.g. highly correlated outbreaks and amplifying effects), temporal patterns (e.g. strong seasonality), demographic patterns (e.g. mean ages of infection), and key epidemic drivers (e.g. climatic and human influences) [9,13,14,15,16,17]. In our studied region of northern Peru, dengue is a substantial, year-round public health burden where climate-based modelling analyses have also separately involved wavelets, ensemble forecasting, and Bayesian hierarchical models to quantify relationships between climatic conditions and reported dengue cases, and forecast short-term future epidemic trajectories [10,18,19].

Here, we build upon these previous works as we introduce an interdisciplinary pipeline to address many requirements of public health authorities. To date, modelling workflows have been presented for a single modelling class, yet are not dengue-specific nor focused on learning from and integrating approaches from multiple modelling disciplines [20,21,22,23]. Similarly, approaches rarely focus on a pipeline to assess past, current, and future dynamics. As a model is an imperfect description of reality, there is no reason to assume (potentially restrictively) that a single technique for modelling is correct and appropriate. Here, we explicitly acknowledge this fact to avoid drawing results and conclusions from a single modelling domain. Instead, our approach leverages strengths from statistical time series, wavelet analysis, Bayesian statistics, DL, and probabilistic ensemble forecasting. In doing so, we demonstrate a new pipeline to more robustly and comprehensively analyse past, current, and future dengue epidemic dynamics. Across fourteen provinces in Peru, we describe how our interdisciplinary approach can be used to make inferences from seemingly distinct approaches. These inferences should provide insights that aid policy-makers with monitoring and controlling dengue transmission. So, whether our objective is to understand (modelling climatic drivers) or predict (forecasting dengue cases), we show how superior, robust results and conclusions can be obtained. This bridges the gaps in the literature regarding modelling and forecasting beyond single domains.

2. Materials and methods

2.1. Data

We used province-level reported dengue cases per month from May 2010 to December 2021 across the 14 provinces in Piura, Tumbes, and Lambayeque (on the northern coast of Peru) which had registered dengue cases, as per the National Centre for Epidemiology, Disease Prevention and Control (Peru CDC) in Peru's Ministry of Health[24]. The reported cases included both confirmed and probable cases across all serotypes. Probable dengue cases were defined as an individual with i) febrile illness for a maximum of seven days, ii) two or more specific symptoms, and iii) resides in or has recently visited areas with dengue transmission or known *Ae. aegypti* populations. Confirmed dengue cases met the same exposure and symptoms criteria, and had a positive dengue test[25]. Monthly data were used to align with other data sources (see below).

We interpolated population data from the 2007 and 2017 national censuses, and the 2022 mid-year population estimates[26,27,28]. Then, we defined Dengue Incidence Rates (DIRs) per 100,000 by converting reported new dengue cases to rates per 100,000. Data on the proportion of provinces' populations living in urban areas were also sourced from the 2007 and 2017 censuses.

Climatological data were as previously described in[19] but we now calculated the average values across the individual province areas. Briefly, we sourced total monthly precipitation and monthly averages of daily maximum (and minimum) temperature from the WorldClim 2·1 dataset[29,30], a drought indicator called the Standardized Precipitation Index (SPI-6)[31,32] from the European Drought Observatory[33], and El Niño Southern Oscillation (ENSO) indicators, namely the El Niño Coastal Index (ICEN) and the Oceanic Niño Index (ONI) from the Geophysical Institute of Peru (IGP)[34,35] and the National Oceanic and Atmospheric Administration[36,37] respectively. Each of these climatological variables have been used in previous modelling studies in Peru and/or in other nations[10,13,17,19,38].

2.2. Wavelet and exploratory analyses

We employed wavelet methods to analyse epidemic dynamics and drivers. These methods have been used for dengue analyses in neighbouring Brazil and across Southeast Asia[9,14,15], and can analyse signals (such as DIR time series) with sharp discontinuities and time-varying periodicity (i.e. non-stationarity).

Using the R package WaveletComp, we employed a Morlet wavelet with non-dimensional frequency $\omega_0 = 6$, as used previously for wavelet analysis of epidemic dynamics[9,14,39,40]. The Morlet wavelet allowed us to decompose DIRs into reconstructed annual cycles (maximum period of two years) and multiannual cycles (period of two to twelve years). The wavelet transform produced a wavelet power spectrum, and we averaged across the provinces' reconstructed annual and multiannual cycles to obtain the average wavelet power of the annual and multiannual cycles per province per month.

For both annual and multiannual cycles, for each province pair, we computed the cross-wavelet power (the wavelet analogue of covariance) and wavelet coherence (the wavelet analogue of correlation) to compare the relative timing of epidemics across province. To measure epidemic synchrony (how the amplitude of the incidence time series covary), for each province pair, we computed the Pearson correlation coefficient between their raw DIR time series and between their reconstructed (annual and multiannual) cycles. We also quantified potentially time- and space-varying geographic, human, and climatic drivers via the cross-wavelet power and coherence between climatic conditions and reconstructed dengue cycles. This was complemented by developing generalised additive models (GAMs) for epidemic synchrony using smooth functions (thin-plate splines) of temporal random effects (month and year), spatial random effects (province), climatic variables (all variables above), and human mobility approximations (the pairwise product of provinces' populations)[41,42]. We fitted the GAMs using the mgcv package in R version 4·2·1[39,42], and we evaluated model performance using the minimised generalised cross-validation score (GCV), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)[43,44,45].

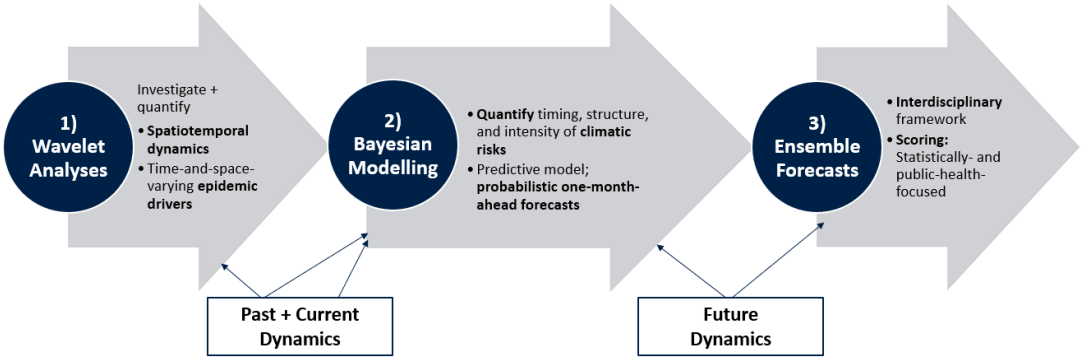

Figure 1: **Modelling and forecasting pipeline.** Our proposed pipeline which takes the researcher from studying past to future dengue epidemic dynamics. Retrospective analyses: Wavelets and Bayesian modelling are used to provide a model-based understanding of past epidemic dynamics and drivers. Future dynamics: Our interdisciplinary framework is used to probabilistically estimate dengue cases with a forecast horizon of one month, and forecast evaluation takes place with both statistically rigorous and public-health focused criteria.

Other exploratory analyses involved computing the cross-correlation between past climatic conditions and current DIR at different time lags. These cross-correlations were computed for each province.

2.3. Bayesian climate-based modelling

We developed a Bayesian climate-based modelling framework which was similar to existing models from other nations [13,17] and our previous work in Peru [19]. Here, we focus on recent modifications.

We modelled province-level dengue incidence using a zero-inflated Poisson distribution, thus accounting for anticipated excess zero counts. The final model, informed by exploratory analyses, was developed using data across 2010 to 2017, and contained: i) fixed effects (to estimate effects of recent incidence trends relative to the preceding year, shared seasonality across provinces, urbanised populations, and epidemic momentum), ii) province-level temporal effects (to estimate monthly patterns and year-to-year heterogeneities), iii) spatiotemporal random effects via an adapted Bayesian version of the Besag-York-Mollié model[46,47] (to account for similarities and uniqueness across provinces), and iv) non-linear, delayed climatic influences via distributed lag non-linear models (DLNMs)[48]. We provide additional model details in Supplementary Material 4.

Models were assessed using information criteria (e.g. Widely Applicable Information Criteria, WAIC and cross-validated logarithmic score), in-sample predictive accuracy, and leave-one-time-point-out cross-validation (with assessments of prediction interval (PI) coverage and probabilistic calibration). We fitted all Bayesian models using Integrated Nested Laplace Approximation (INLA) in R version 4·2·1[39,49].

2.4. Probabilistic ensemble forecasting

By sampling from the posterior predictive distribution, the climate-based model above readily produces probabilistic forecasts of future dengue incidence with a forecast horizon of up to one month. However, any model is just an abstraction of reality, and individual models can have contrasting strengths yet can be wrong in different ways (e.g. predictive vs understanding transmission mechanisms).

So, we consider ensemble frameworks which combine the probabilistic forecasts of several individual models. Ensemble frameworks have demonstrated strong predictive performance for many dynamical systems (from forecasting of COVID-19, seasonal influenza, and dengue to weather forecasting[50,51,52,53]). Ensemble forecasts are often represented in one of two ways; using samples or quantiles (our primary focus) of the predictive distribution. We focus on quantile-based forecasts as i) it provides

access to a wider range of models which directly target quantiles (i.e. quantile regression for which no there is no assumed likelihood), and ii) it is reflective of a computationally cheaper environment (versus processing thousands of samples) that has been regularly employed in recent years[54,55]. The combination of the individual model components can be achieved by stacking, weighted density ensembles, or model averaging[56,57,58]. Some existing ensemble approaches for dengue have not always assessed probabilistic forecasts of (e.g.[59,60]), instead focusing on pointwise accuracy metrics (which do not measure uncertainty in our predictions). These approaches also have not always separated the model development period from the testing period nor have they benchmarked performance using iteratively updating models, all of which is necessary to reflect a realistic application without bias in our conclusions. In recent years, probabilistic ensembles have been increasingly used with equal weighting of quantiles or trained weighting of sample distributions (e.g. Bayesian stacking)[51,52,53,61].

For the forecasts of dengue incidence at a one-month horizon, we used 23 quantile levels (identical to past forecasting studies[55]). Untrained approaches used the median or mean average of individual modelling components' predictive quantiles. We also briefly explored trained approaches via linear stacking with the quantgen package in R[62], as we weighted individual models' quantile forecasts for a given month based on weights that would have optimised the sum of the quantile losses for forecasts (of the logarithm of cases – see below) over the past twelve months (not including the forecasting date). We performed this weighting for each province independently (to derive province-dependent weights for each time point) and for all provinces jointly (to derive province-independent weights for each time point). We iteratively updated the model weights as each month's data became available (i.e. an online procedure)

While we performed our study retrospectively, to reflect a realistic application, we forecasted dengue cases with a one-month horizon and left the current and all future data out of model training. The four-year period of January 2018 to December 2021 was our testing period. This allowed sufficient data for the model development period (2010 to 2017 inclusive), where we also performed leave-one-time-point-out cross-validation and forecasting with leave-future-out cross-validation. This allowed for learning of spatial, temporal, and climatic relationships. The covariates and hyperparameters of each model (Section 2.5 were selected solely from performance of candidate models in the model development period. Conclusions and results were then based on subsequent performance in the testing period.

To evaluate forecasts, we used the weighted interval score (WIS) – a proper scoring rule that is a discrete approximation to the continuous ranked probability score (CRPS)[54,63]. The WIS measures the absolute distance between our predictive distribution and the observed data, and simultaneously assesses model calibration (compatibility of forecasts with observations) and sharpness (concentration of predictive distributions). Lower scores indicate better performance. As the WIS rigorously accounts for distributional uncertainty, we avoided focusing excessively on pointwise accuracy metrics. We also used the decomposition of the WIS into over-prediction, under-prediction and dispersion. Recently, it has been shown that it matters to our forecast assessment and model rankings how we apply a scoring rule. Applying these scores at the level of case counts is likely inappropriate, due to vastly different epidemic properties across space and time[64]. So, in the current work, we i) follow the recommended procedure to score forecasts on a natural-logarithm-transformed scale (i.e. log(Cases + 1)), and ii) also score forecasts on a DIR scale. While we primarily focused on the recommended logarithmic scale, the latter assessment is motivated by a desire to be comprehensive and to provide interpretable and comparable information for public health authorities across space and over time. Alongside the WIS, we assessed bias, quantile coverage, and PI coverage.

While all such measures provided aggregate and statistically rigorous perspectives of performance, we also included a public-health-oriented measure. This measure focused our attention on times of the year and areas of the predictive distribution of public health urgency. Here, we carried out a setting-specific evaluation of the ability to correctly classify (i.e. binary forecast) the following month's DIR reaching thresholds of 50 or 150 per 100,000. These thresholds were pre-specified by us to indicate different magnitudes of dengue outbreaks, and we report classifications as these can be easily interpreted by public health authorities and the wider public. We used our model-based quantile forecasts to

Table 1: **Quantile-based forecasting models and abbreviations.** The six individual forecasting models are described first (above the double lines), followed by ensemble models that are highlighted with an asterisk. The equal weighting (EW) scheme applies equal weights to the quantile of that ensemble model's individual predictions for each quantile. The median of individual models means that we take the median of the individual models' predictions for each quantile.

Model Abbreviation	Description
Baseline	Epidemiologically naive random walk model that uses the most recent month's observed cases as a forecast for the next month.
Bayes-Climate	Bayesian spatiotemporal model with non-linear, delayed climatic effects, spatial random effects, a momentum indicator, and temporal effects
SARIMA	Seasonal Auto-Regressive Integrated Moving Average model specified for each province independently at each time
TCN	Temporal Convolutional Network with climatic covariates specified for each province independently
TimeGPT	Transformer-based foundational time series forecasting model with climatic covariates specified for each province independently
TimeGPT-NoCov	Transformer-based foundational time series forecasting model without any covariates specified for each province independently
EW-Mean *	Equally-weighted mean average of all individual forecasting models
EW-Mean-NoBase *	Equally-weighted mean average of all individual forecasting models, excluding the baseline model
EW-Mean-NoBayes *	Equally-weighted mean average of all individual forecasting models, excluding the Bayesian climate-based model
EW-Mean-NoCov *	Equally-weighted mean average of all covariate-free individual forecasting models. This excludes covariate-based models – the Bayesian climate-based, TimeGPT, and TCN models.
Median *	Median of all individual forecasting models, including the baseline model
Median-NoBase	Median of all individual forecasting models, excluding the baseline model
Median-NoBayes *	Median of individual forecasting models, excluding the Bayesian climate-based model
Median-NoCov *	Median of individual forecasting models, excluding covariate-based models (the Bayesian climate-based, TimeGPT, and TCN models)
Prov-Trained*	Trained ensemble with province-dependent weights determined to minimise the sum of the quantile losses
Trained*	Trained ensemble with spatially homogeneous weights determined to minimise the sum of the quantile losses

determine a cut-off quantile level (i.e. forming a decision rule based on historical performance) which determined whether we classify an outbreak (of DIR $\geq$ 50 or 150 per 100,000) for the next month. An outbreak is forecasted if the prediction at the (historically calibrated) cut-off quantile level exceeded the corresponding outbreak threshold (e.g. 50 per 100,000). We allowed for decision rules to be iteratively updated as each month's data became available – in each month, the rule was determined to maximise the historical AUC (the area under the receiver operating curve) for classifying outbreaks. In doing so, we separated the inference stage (modelling) from the decision-making stage, and ensured that our classifications were motivated by maximising public health utility. This is because simultaneously maximising true positives and minimising false alarms helps to avoid wasted resources and lack of public confidence. We repeated this outbreak detection evaluation in the setting of detecting first onsets of outbreaks, where we assessed performance for forecasts of initial onset of outbreak with DIR $\geq$ 50 (or 150) per 100,000. This smaller dataset includes classifications for each province for months of a year corresponding to the outbreak onset, months before the onset, or in a year with no eventual onset. This classification can be important as it focuses on a public health priority of dengue incidence first reaching pre-defined thresholds.

2.5. Our ensemble framework components

Our ensemble frameworks contained two or more forecasting model components whose predictive distributions (for one-month-ahead cases), were represented by predictive quantiles (23 levels).

Our baseline model was an epidemiologically naive model which forecasted next month's dengue cases as the current month's dengue cases. This was a direct application of an existing strong benchmarking model from the U.S. COVID-19 Forecast Hub, and is a random walk model with innovations (i.e. uncertainty) based on past observed differences in monthly dengue cases [55].

Our Bayesian climate-based model (described in Section 2.3) used information up to one month before the target month to generate samples from a posterior predictive distribution – from which we computed quantile forecasts. We also developed Seasonal Auto-Regressive Integrated Moving Average (SARIMA) models. These statistical time series models were fitted independently to the logarithm of monthly cases for each province independently using the Auto-ARIMA (Auto-Regressive Integrated Moving Average) implementation of the Statsforecasts package and Darts library in Python[65,66]. This implementation selects the order of the ARIMA model based on the Akaike Information Criterion[43], and quantile forecasts are generated by assuming normality in forecast errors. We refitted the model at each individual time point to reflect real-world conditions as new information becomes available. For the baseline, Bayesian, and SARIMA models, we generated quantile forecasts by taking quantiles of the 5,000 samples of the predictive distributions.

The models above are all reflective of traditional model frameworks, yet in recent years, DL approaches have demonstrated strong performance for forecasting of infectious diseases[67]. We introduce here the first known application of a foundational time series model for forecasting dengue cases. We used TimeGPT, a transformer-based foundational time series model with self-attention that has been pre-trained on over 100 billion data points from different domains[68,69]. We implemented this model (for the logarithm of cases) with and without climatic covariates (two-month rolling averages of precipitation and minimum temperature). The model readily produces quantile forecasts by using conformal inference — a distribution-free way to produce statistically rigorous PIs[70]. We also developed a forecasting model that used a temporal convolutional network (TCN) from the Darts library in Python[65]. This DL architecture has dilated causal convolutional nets which preserve the local and temporal structures of the data. To avoid making distributional assumptions, we used quantile regression to directly estimate the quantiles of the predictive distribution. We again used climatic covariates of two-month rolling averages of precipitation and minimum temperature. These covariates were based on exploratory analyses from the model development period.

We outline model abbreviations in Table 1 and additional details on ensemble forecasting models in Supplementary Material 5. We followed EPIFORGE 2020 guidelines[71] throughout the forecasting analysis (Table SI 3).

3. Results

3.1. Investigation of epidemic dynamics using wavelets and climate-based analyses

We used wavelet analysis to analyse simultaneously signals (of our incidence and climatic time series) in both the time and frequency domains. Our wavelet-transform-based reconstructions of annual and multiannual (two to five years) cycles of DIRs enabled a range of findings. First, similar to wavelet analyses in Southeast Asia[9,15], large dengue epidemic years (such as 2015 and 2017) were characterised by statistically significant peaks in the average wavelet power and amplitude of reconstructed cycles (Figures 2, SI 3). There was a spatial trend of common large peaks across reconstructed cycles in provinces of the north and west of the region, with slightly more widespread incidence peaks in the second large epidemic year of 2017. The mean period of multiannual cycles reduced during such years, whilst across the study period, we observed similar overall trajectories of the mean period (Figure SI 7) across the provinces (median pairwise Pearson correlation: 0.66, Interquartile Range – IQR: 0.25, 0.84). Complementary visualisations of monthly DIRs suggested strong similarity in the province-level dengue incidence, in terms of both timing and magnitude. Trends in dengue incidence included waves in the south preceding increases in northerly provinces and waves in the west preceding surges in the east. The trends were observed within individual years and when analysed using monthly averages over the study period (Figures SI 4, SI 5, SI 6).

Employing wavelet coherence (measuring the relative timing of epidemics), we found generally widespread coherence across time and space (Table SI 5, figures SI 11, SI 10). Coherence generally

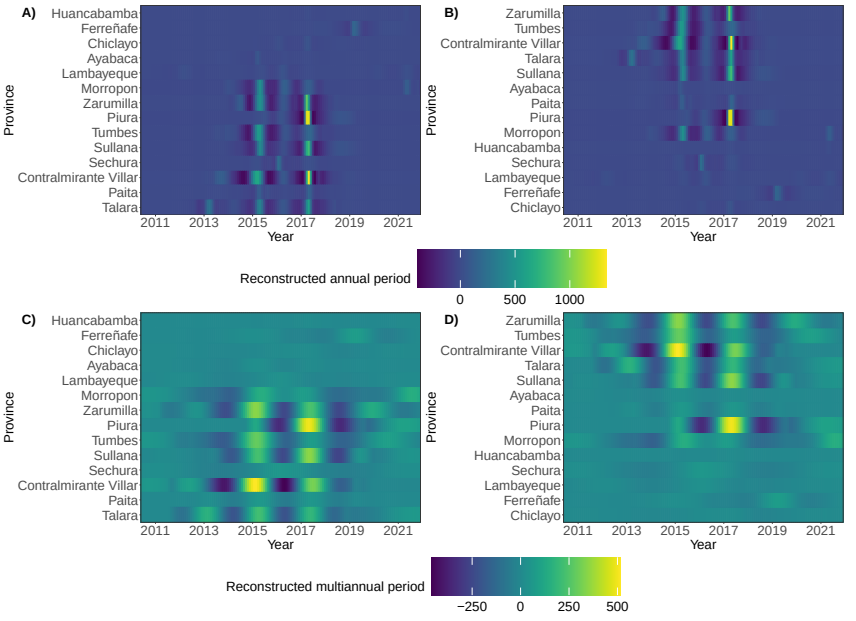

Figure 2: **Wavelet reconstructions of Dengue Incidence Rate (DIR) time series.** Reconstructed annual cycles of province-level DIRs, where provinces are sorted in ascending order by **(A)** longitude (from west to east) and **(B)** latitude (from south to north). Reconstructed multiannual cycles of DIRs are shown in **(C)** and **(D)**, where provinces are also sorted in ascending order by longitude and latitude respectively. Figure SI 2 depicts the reconstructed cycles with provinces sorted by percentage of population living in urban areas.

amplified (i.e. more widespread) during epidemic years when more provinces shared statistically significant coherence with the (annual and multiannual) dengue cycles of other provinces (Figure 3). Spatially, more northerly and more westerly provinces tended to possess a greater proportion of statistically significant coherent relationships with annual and multiannual cycles of other provinces, whilst more southerly and easterly provinces had more out-of-sync (or equivalently, self-contained) cycles. The spatial trends were observed at individual time points within years (Figure 3) and across the entire study period of 140 months (Figure SI 11), yet come with the caveat that Chiclayo (the most southerly province) had more coherent cycles, reflective of a general trend of more urbanised provinces possessing more significantly coherent annual and multiannual cycles (Table SI 6, Figure SI 2).

We found moderate-to-strong epidemic synchrony between annual raw DIR time series. Similar to phase coherence, there was strong synchrony between reconstructed cycles (both annual and multiannual, Table SI 7), and synchrony was high during large epidemic years (Figure SI 13). Spatially, particularly in the reconstructed multiannual cycles, synchrony tended to be stronger in more northerly and westerly provinces (Figure SI 12), thus capturing that correlations between epidemic curves shared similar patterns as was previously observed for seasonality of epidemics (i.e. coherence).

We quantified geographic, climatic, and human drivers. Coherence and cross-wavelet power measured the agreement between the seasonality of climatic variables and the reconstructed cycles. Here, we found widespread statistically significant coherence, and consistently greater average cross-wavelet power between climatic variables (maximum temperature, precipitation, and El Niño Coastal Index) and reconstructed dengue cycles (both annual and multiannual) during epidemic years of elevated incidence (Figure SI 14). Peaks in the average cross-wavelet power were always observed during periods of elevated incidence across each of the climatic variables and across almost all of the provinces (Figures SI 15, SI 16, SI 17). Spatially, more northerly provinces shared greater levels of statistically significant coherence with climatic variables. The El Niño Coastal Index had a greater coherence (with reconstructed cycles) in more westerly (and hence, coastal) provinces, and had a spatial gradient over

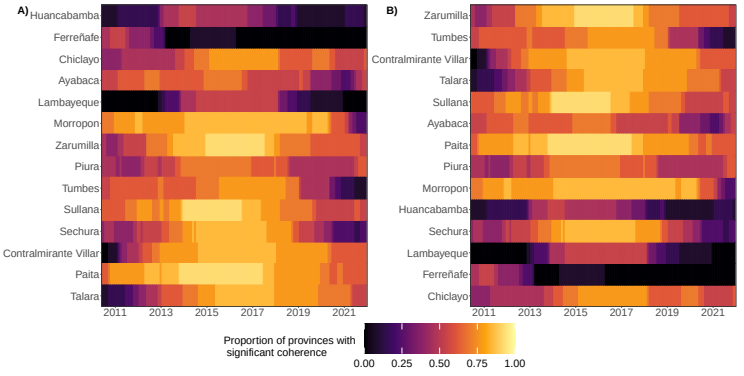

Figure 3: **Coherence of reconstructed annual dengue cycles.** Proportion of a focal province's annual dengue cycles which share statistically significant coherence with the cycles of other provinces at each time point. Provinces are sorted in ascending order by **(A)** longitude (from west to east) and **(B)** latitude (from south to north).

time in the cross-wavelet power with annual cycles, whereby greater agreement in seasonality of cycles was observed over time as one progresses from west to east and from north to south (Figure SI 17).

To assess the influence of lagged climatic conditions on human disease incidence and to inform model development, we computed cross-correlations between climatic variables and the raw monthly DIR time series (during the model development period; 2010 to 2017). There were generally moderate-to-strong, and spatially consistent, relationships between climatic conditions and DIRs across the provinces, particularly across temperature, drought, and precipitation. From a forecasting perspective, the strongest relationships were inferred with climatic lead times of one to three months and cyclical patterns in the relationship were visibly pronounced (Figure SI 1).

We investigated the impact of pairwise distance between provinces on their corresponding average correlation and proportion of significant coherent cycles. Within some provinces, there were clear relationships between shorter distances and greater average correlation or coherence (which was most widespread for coherence between multiannual cycles). There was a spatial trend of shorter pairwise province distances being associated with similar annual epidemic dynamics (timing and amplitudes, across annual and multiannual cycles) in more northerly provinces (Figures SI 18, SI 19).

We extended previously-employed modelling methodology[15] (to further incorporate the pairwise product of population sizes) as we developed GAMs to identify predictors of epidemic synchrony (see Materials and methods). Our best-fitting GAM, based on generalised cross-validation score and corrected AIC[72,73], included climatic influences (minimum temperature and precipitation), year-to-year heterogeneity, and the (logarithmic) pairwise product of population sizes. We estimated negative coefficients in 12 of the 14 provinces for the pairwise distance-province interaction effect on epidemic synchrony, whilst 6 of the 14 provinces had statistically significant (all negative) effects (Figure SI 20). Significant positive (partial) effects were also estimated for temporal (yearly) random effects during epidemic years, whilst lower (and higher) values of the pairwise product of populations were associated with statistically significant negative (and positive) effects on synchrony respectively. Climatic partial effects were subject to greater uncertainty, but greater minimum temperatures and greater precipitation were estimated to have increasingly negative and increasingly positive effects respectively (Figure SI 20 and see Discussion).

3.2. Bayesian climate-based modelling

We now describe results from our Bayesian spatiotemporal model, whose inferences complement and enhance the details provided by our wavelet- and climate-based analyses above. During the model development period (2010 to 2017), we inferred strong generalisability of our climate-based Bayesian spatiotemporal model due to accurate estimated out-of-sample trends. The leave-one-time-point-out

posterior predictive distributions' 95% credible intervals contained 82.3% of the observations (indicative of a reasonably calibrated model), the corresponding posterior median explained 66% of the variation in the DIR (i.e. $R^2 = 0.66$).

Briefly, with respect to the climatic inferences from the model fitted to the entire study period (of 2010 to 2021), we analysed risk profiles (Figure SI 21), where relative risk of greater DIR was measured with respect to the risk induced by the mean average of each climatic variable. Higher (lower) values of the two-month-rolling-averaged, monthly average of daily minimum temperature were associated with elevated (diminished) relative risk of human disease incidence. Similarly, there was a consistent estimated risk profile for the cumulative drought indicator with elevated DIR risk for extreme values (large surpluses or deficits of accumulated precipitation). Intermediate values of the indicator were associated with lower relative risk. In terms of the two-month rolling average of monthly precipitation, we estimated greater risk of disease incidence for values between 150mm and 200mm at lags of approximately one month, whilst extreme large precipitation values, except at very small time lags, were associated with diminished risk of DIR. Finally, a complex risk profile was estimated for the sea surface temperature indicator (ICEN) whereby extreme large values (El Niño events) and extreme small values (La Niña events) were each associated with both elevated and diminished risk at different time lags across four months. It was difficult to infer the reasoning behind some of the complex behaviour in the risk profile for the ICEN (see Discussion).

3.3. Ensemble forecasting

We first deployed our Bayesian climate-based model in a forecasting setting (excluding the current and all future months' data from model training) for each month in the model development period (2010 to 2017). We found inferior results versus the leave-one-time-point-out analysis. Irrespective of assessment on a logarithmic or DIR level (see Materials and methods), we identified model calibration issues due to frequently underpredicting posterior predictions and hence, poor interval and quantile coverage values (Figures SI 28-SI 29). This suggested that the Bayesian climate-based model alone would likely be inappropriate and unreliable for probabilistic forecasting, and reinforced our motivation for extending to an ensemble framework. While it would be unwise (due to potential bias from model tweaking after seeing the training data) to focus heavily on results from the model development period, we found stronger performance for each of our other ensemble candidate members (see Section 2.5). This was reflected in assessment of the whole predictive distribution (via the WIS), pointwise accuracy metrics (e.g. R^2), PI and quantile interval coverage, and bias (Table SI 8). For example, the simple yet flexible SARIMA models (which were refitted for each province at each time point) improved the calibration (e.g. quantile and interval coverage in Figures SI 28 – SI 28), accurate pointwise predictions (e.g. $R^2 = 0.79$, see Table SI 8 and Figure SI 23), and overall distributional accuracy (as summarised by the WIS). Our DL models, particularly TimeGPT (the foundational time series model), which incorporated climatic covariates also yielded strong forecasting performance relative to the baseline forecaster, models without climatic covariates, and our Bayesian climate-based model (Figures SI 22-SI 27, Table SI 8). Simple, untrained ensemble forecasting frameworks also consistently produced accurate pointwise estimates and strong forecasting skill (in terms of calibration and sharpness).

Our four-year testing window of 2018 to 2021 was an environment to assess performance in a setting reflective of a real-world application, as all models (including hyperparameters, covariates, and ensemble weights) were pre-defined from only historical data. In this testing period, we found strong predictive performance in forecasting cases for each of our ensemble models (Tables 2, SI 9). This performance was captured by values for our WIS (measuring the whole distribution), PI and quantile coverage (measuring calibration), and pointwise accuracy metrics (such as R^2 and RMSE – Root Mean Square Error). We found similarly strong performance of the ensemble models for reliably forecasting the first onset of DIR reaching thresholds of 50 (or 150) per 100,000 (Figure 4 a) and more generally for forecast classification of months with DIR above/below these thresholds (Figure 4 b). The individual models handle different types of uncertainty in different ways (e.g. uncertainty in data generating processes, random error terms,

Table 2: **Results for forecasting log(Cases + 1) in the testing period of 2018 to 2021.** We summarise proper scores, predictive performance, calibration, and outbreak detection capabilities of forecasting models. Models are arranged by mean weighted interval score (WIS), from best (lowest) to worst (highest). Further summary statistics for WIS are presented in Table SI 14, and implied similar model rankings. Bias measures the bias of each models' predictive quantiles, PI coverage measures the coverage of the 95% prediction intervals (PIs), R^2 denotes the proportion of variation in log(Cases + 1) explained by the models' posterior median estimates, and RMSE is the corresponding root mean square error (RMSE) for log(Cases + 1) by using the posterior median. For detection of outbreak onset in each province in each year, *Sensitivity*, *Specificity*, and *AUC* represent the True Positive rate, True Negative rate, and Area Under the Curve in terms of forecasting skill for correctly classifying the first possible outbreak of DIR $\geq$ 50 per 100,000. These metrics are visualised in Figure 5 a for model comparison purposes. Table SI 9 is the analogous table which depicts outbreak detection for DIR first surpassing a threshold of 150 per 100,000, while Table SI 12 presents the models scored on a DIR level. Details of the model abbreviations are provided in Table 1 (where similar model rankings were obtained). All ensemble models are highlighted with an asterisk.

	Model	WIS	Bias	PI Coverage	R^2	RMSE	Sensitivity	Specificity	AUC
1	Median *	0.34	0.03	95.2%	0.74	0.81	81.8%	94.3%	0.88
2	Median-NoBase *	0.35	0.02	94.8%	0.73	0.82	81.8%	95.7%	0.89
3	Median-NoBayes *	0.35	0.06	94.8%	0.74	0.82	81.8%	87.2%	0.85
4	EW-Mean *	0.35	0.17	87.8%	0.75	0.79	72.7%	95.9%	0.84
5	Median-NoCov *	0.36	0.04	95.2%	0.73	0.83	81.8%	91.0%	0.86
6	Ew-Mean-NoBayes *	0.36	0.21	87.5%	0.74	0.81	63.6%	97.3%	0.80
7	EW-Mean-NoCov *	0.36	0.14	91.2%	0.74	0.81	81.8%	86.5%	0.84
8	Trained *	0.36	0.07	92.4%	0.72	0.84	81.8%	95.4%	0.89
9	EW-Mean-NoBase *	0.36	0.16	87.1%	0.74	0.80	81.8%	94.7%	0.88
10	SARIMA	0.36	0.03	93.5%	0.73	0.82	63.6%	96.1%	0.80
11	Prov-Trained *	0.37	0.05	91.1%	0.72	0.85	72.7%	89.9%	0.81
12	Baseline	0.38	0.04	97.3%	0.74	0.83	90.9%	45.0%	0.68
13	TCN	0.39	0.11	91.2%	0.70	0.87	72.7%	96.6%	0.85
14	TimeGPT	0.41	0.08	88.1%	0.69	0.91	90.9%	91.7%	0.91
15	TimeGPT-NoCov	0.42	0.12	85.6%	0.67	0.92	72.7%	91.5%	0.82
16	Bayes-Climate	0.52	-0.20	87.6%	0.58	1.09	54.5%	90.1%	0.72

parameter estimates etc.), yet individually do not account for all possible sources of uncertainty and so, the coverage of their 95% PIs is less than the claimed coverage (95%). In contrast, relative to other ensembles and individual models, ensembles that were based on the median of individual models' quantiles produced statistically valid PIs and overall consistently strong forecasting ability (Figure 4). This was captured by consistently sharp and well-calibrated (albeit relatively conservative) PIs (Figure SI 41, low bias, and better WIS values (e.g. pairwise comparisons of WIS values in Figure SI 37, Table 2). These trends were generally observed across provinces (Figure 5 e) and over time (Figure 4 c), and for different magnitudes of epidemics (Figure SI 39 and Table SI 15). The superiority of the forecasts from ensemble models, particularly median-based ensembles, was also observed when we ranked models on their forecasts of DIRs. Indeed, across assessments on logarithmic and DIR scale, median-based ensembles consistently dominated the model rankings based on score ratios, outbreak detection, and pointwise accuracy metrics (Tables 2, SI 9, SI 12, SI 13, Figures 4c, Figure 5 a and SI 37).

While these assessments of WIS values (and related statistical measures) provide aggregated measures of the whole predictive distribution, we also evaluate the utility of the models' forecasts for public health authorities. For instance, the epidemiologically naive forecasting model's predictive distributions could be relatively close in distribution to the observed cases (as the true distributions do not regularly change drastically in short periods), yet this model fails to offer any valuable information about oncoming outbreaks (as it forecasts the past observed cases). So, looking at detecting the first onset of DIR reaching thresholds of 50 (or 150) per 100,000 (Figures 4 b, 5 a and 5 c), the median-based ensemble models produced AUC values of between 0.85 and 0.89, which are i) far superior to the naive baseline forecasting model (AUC: 0.68) and ii) more generally, superior to individual forecasting models. For example, if authorities used the median-based ensemble of all individual models, 81.8% of first onsets (of reaching DIR of 50 per 100,000), would have been correctly detected one month ahead of time with a corresponding false alarm rate of 5.7%. We found similarly strong performance of ensemble models for more generally classifying whether DIR in a province in any given month would be at or above the thresholds of 50 (or 150) per 100,000. Here, nearly all forecasting models demonstrated strong classification skill (e.g. AUC of between 0.79 for DIR $\geq$ 50 per 100,000, see Figure 5b and Table SI 10), likely in part due to the easier nature of predicting the next month being at or above a threshold given that we know the most recently observed high DIR level.

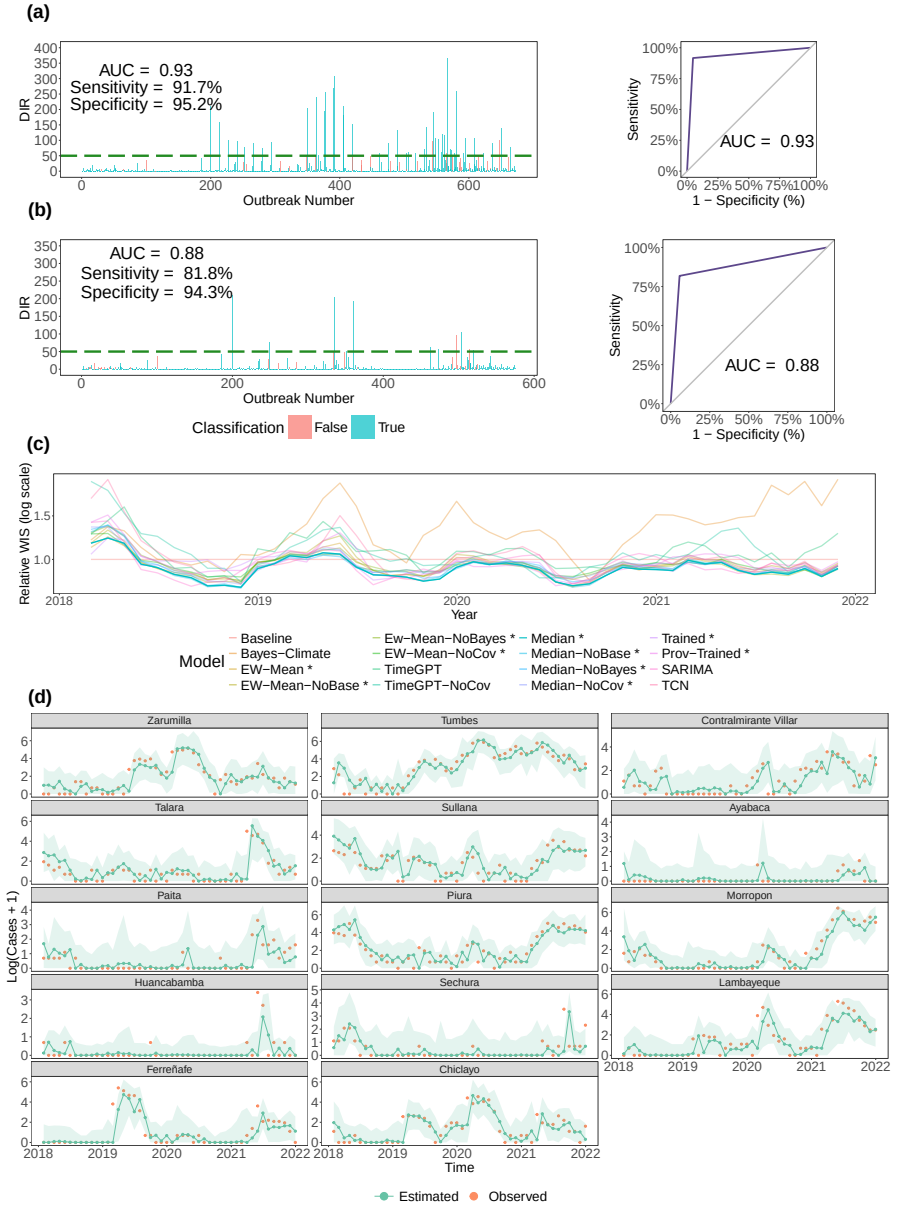

Figure 4: **Performance of median ensemble forecasting model:** This ensemble forecasting models' probabilistic forecasts were generated by taking the median of all six ensemble forecasting models' prediction at the 23 quantile levels (see Materials and methods). (a): Visualisations of forecasting ability for classifying whether the coming month will experience a DIR of $\geq$ 50 per 100,000. AUC denotes the area under the receiver operating characteristic (ROC) curve, and higher values indicate greater classification skill. (b): Analogous visualisation for classifying whether the coming month will be the first month of the year in a province with DIR of $\geq$ 50 per 100,000. (c): Relative Weighted Interval Score (WIS for log(Cases + 1)) is shown for all forecasting models in centred, rolling windows of three months, with the median model's values highlighted. Relative WIS is an individual forecasting models' WIS divided by the baseline forecasting's WIS, where values below one indicate greater forecasting skill relative to the baseline forecaster. The corresponding summary statistics for Log(Cases + 1) over time are visualised in Figure SI 39. (d): For each province (sorted in descending order by latitude from north to south), we visualise the median model's predictions (for log(cases + 1)) where green points denote the posterior median and orange points denote the observed values. The green shaded intervals denote the 95% prediction intervals (PIs). 95.2% of all observations lay within these 95% PIs. The posterior median explained 74% of the variation in log(cases + 1) i.e. $R^2 = 0.74$. Y-axes are province-dependent due to the varying sizes of the dengue epidemics observed.

In terms of further relative comparisons, the forecasts of ensemble models which included climatic covariates outperformed the analogous models without covariates (denoted NoCov in e.g. Table 2), while trained ensembles were outperformed by simple median-based ensembles (see Discussion). This was shown by values for the mean scoring rule (WIS across both logarithmic and DIR scales), pointwise accuracy (R^2), and outbreak detection capabilities (Figure 5 a). Ensemble models also outperformed almost all individual forecasting models in terms of metrics such as 95% PI coverage and accuracy of the entire distribution (WIS). These model rankings were similar across both DIR and logarithmic scales (Tables 2, SI 9.

4. Discussion

We have introduced a new interdisciplinary framework which allows a researcher to comprehensively quantify the dynamics and drivers of past epidemics, and to robustly estimate the likely future short-term epidemic trajectories.

4.1. Wavelet analysis and climate-based modelling

Using wavelet methods allows us to identify epidemic dynamics and drivers that vary in different temporal resolutions, times of the year, and/or in different geographies. Our analysis revealed strong similarities in both the seasonality and epidemic curves across the provinces. Such similarities were pronounced during epidemic years with major outbreaks. So, from a spatial perspective, it is unsurprising that northerly provinces, which experienced higher DIRs, possessed more synchronous and more coherent (annual and multiannual) dengue incidence cycles. Also, the strongest wavelet coherence and cross-wavelet power between climatic conditions and dengue cycles occurred in more westerly, and more northerly provinces (all of which were coastal provinces). Consistent with wavelet analyses in Southeast Asia[9], such significant peaks in wavelet coherence and cross-wavelet power were estimated during years with larger epidemics, and were widespread, which is indicative of an amplifying role (greater similarity in epidemic dynamics) played by climatic forcing in large epidemic years. The spatially shifting impact of the El Niño Coastal Index during such years (from west to east and north to south) may reflect initial climatic risks imposed on coastal provinces when an El Niño (or La Niña) event initially occurs, dengue incidence climbs, and subsequent climatic conditions and human mobility induce transition of outbreaks further inland and further south. From a public health policy perspective, the roles played by both climatic forcing and spatially shifting influences represent information that should be combined with seasonal climate forecasts (such as GloSea5[74] which has been employed for dengue forecasting in Vietnam[51]) to enhance early warning detection. For the causes of epidemic synchrony, we identified large outbreak years, greater pairwise products of populations and shorter pairwise distances as having enhancing effects on synchrony between provinces. This may reflect greater mobility and greater connectivity driving enhanced epidemic synchrony. Spatially, shorter pairwise distances had more significant effects on both synchrony and coherence in more northerly provinces, which may be indicative of greater connectivity (although this may be confounded by similarity in other environmental conditions). The more uncertain estimated effects of climatic conditions on synchrony may be due to our limited spatial focus (only fourteen provinces) and/or potential confounding in our models of synchrony.

The estimated consistent elevated risk for anomalous precipitation deficits and accumulations is aligned with the well-documented impacts of water shortages and heavy rainfall[13,75,76]. Drought-induced precipitation shortages, particularly in areas with poor water supply, can induce greater risk of dengue incidence due to changes to water storage practices (including more containers around the home), creation of potential egg-laying sites (for the female *Ae. aegypti* who prefer artificial water storage containers), greater survival capacity of *Ae. aegypti* eggs in dry conditions, and/or more concentrated breeding in the few remaining suitable habitats[77,78,79]. Similar elevated drought-induced risk has been inferred in both Vietnam and Brazil[13,38]. Meanwhile, the effects of heavy rainfall are mainly due to large-scale accumulation of stagnant water, which yields creation of abundant breeding sites, and increases

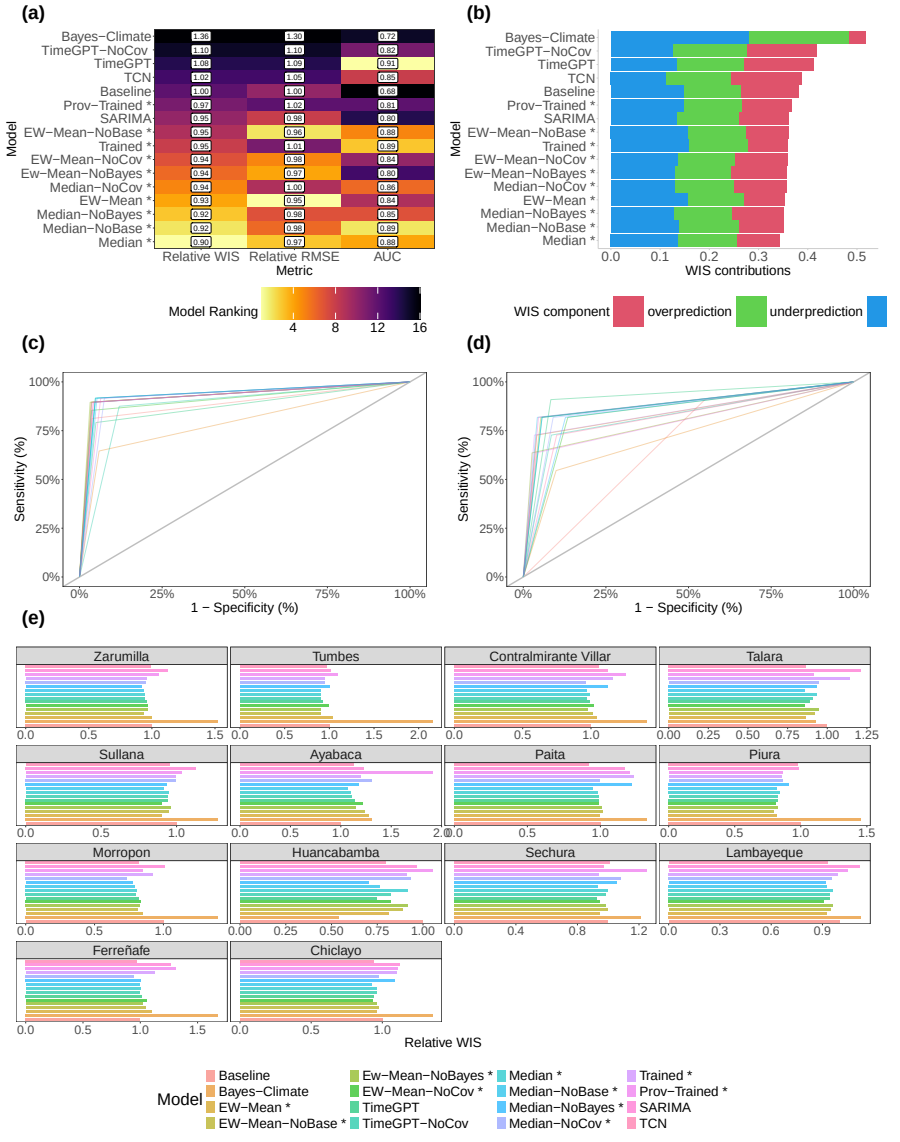

Figure 5: **Relative performance of forecasting models in the testing period of 2018 to 2021.** We compare performance of our 16 ensemble frameworks (see Table 1 for abbreviations) in terms of their predictive performance for forecasting log(Cases + 1). (a) Model rankings by different metrics, where relative root mean squared error (RMSE) and relative Weighted Interval Score (WIS) are defined relative to the baseline model. Lower values indicate greater predictive performance for these relative metrics, but higher AUC values indicate greater forecasting skill for correctly classifying the first months with DIR $\geq$ 50 per 100,000 one month ahead of time. (b): WIS is decomposed into overprediciton, overprediction, and dispersion for each of the forecasting models. (c): Receiver operating characteristic (ROC) curves by forecasting model for correctly classifying months with DIR $\geq$ 50 per 100,000 one month ahead of time. (d): Analogous ROC plot to (c) for detection of first month with DIR $\geq$ 50 per 100,000 one month ahead of time. (e): Relative WIS is shown for all forecasting models per province where provinces are sorted in descending order by latitude from north to south. Relative WIS is an individual forecastin models' WIS divided by the baseline forecaster's WIS for that province. X-axes are province-dependent due to the varying sizes of the dengue epidemics observed.

to vector populations. The risk of disease incidence was estimated to peak after a one-month lag for the two-month average of monthly precipitation, whilst we estimated diminished risk for the heaviest precipitation (above a two-month average of 200m). Diminished risk has been uncovered in Asian nations including Singapore, Malaysia, Sri Lanka, and Thailand[80,81]. Our final climatic variable (the El Niño Coastal Index) was estimated to have a complex risk profile. Whilst the estimated diminished risk for extreme indicator values at short time lags make intuitive sense (due to time taken for onshore climatic conditions of provinces to eventually change), protection (from risk) for greater absolute values

at longer time lags seem counter-intuitive. The latter issue may due to the limited size of our data and/or using our predictive model to disentangle real-world climatic inferences (which may be subject to potential confounding). The issue of potential confounding (between variables) is applicable to each climatic variable, whilst external, unobserved confounders may also have affected our conclusions and thus, we indirectly estimated such unobserved confounding effects (via unstructured spatial effects and yearly random effects).

There are several limitations to both our retrospective and forecasting analyses. These include concerns about the quality, volume, and type of data available, ranging from a lack of information on reporting rates to a lack of serotype data (See Supplementary Material 8 for further discussion). Consistent with the WHO's cited lack of data on circulating serotypes[82], we advocate for future collation of age-structured dengue cases and viral sequencing data, alongside future seroprevalence surveys across at-risk areas, all of which would provide insights about immunology and virology issues (such as multitypic immunity, genomic diversity, and repeat infections) across space and time.

4.2. Forecasting analysis

Moving from past to future dynamics, in a four-year testing window, we demonstrate the utility of a diverse, probabilistic ensemble framework for forecasting of monthly dengue incidence up to one month ahead of time.

The reliable forecasts from ensemble frameworks demonstrate the benefits of combining the strengths of multiple models from different disciplines. Each of these models (and their predictive distributions) are wrong in different ways, and yet each model can handle different aspects of forecasting dengue cases. The foundational time series model (TimeGPT) included in the current work leverages strengths from a transformer architecture, conformal inference, and training on over 100 billion data time points. The TCN model uses representation learning and dilated causal layers in its deep neural network architectures. In doing so, these DL models can explicitly account for temporal structure in the data, handle non-linearities, and detect temporal traits (from seasonality to heteroscedasticity). However, the fact that these DL models do not always outperform simple yet flexible and robust SARIMA models is interesting and perhaps unsurprising. Dengue epidemics are somewhat predictable in terms of longer-term trajectories with highly seasonal patterns (e.g. dengue cycles based on climatic conditions) and in terms of very short-term trajectories with highly autoregressive patterns (e.g. we expect the next month's DIR to somewhat similar to the current month's DIR). So, unlike some previous studies which either benchmark against models only within the DL field or do not allow the SARIMA model structure to be updated as new information becomes available, our rigorous benchmarking reveals similar performance of these statistical time series models with advanced DL methods. The similarity in performance is inferred by rigorous evaluation as we find comparable (or superior) values for PI coverage, bias, WIS, R^2, and outbreak detection metrics. Similar results have been observed for forecasting of COVID-19 cases and deaths[83] and in other unrelated forecasting tasks with strong, seasonal patterns (such as retail demand and wholesale food prices)[84,85]. Of course, this observation comes with the caveat that future work should assess how this predictive performance varies at longer forecast horizons, which are of great importance to public health authorities.

Importantly, however, the DL models do always improve our ensemble's forecasts when included alongside other individual models. The trend is reflective of a more general observation of the outperformance of ensemble models, relative to individual models (similar to previous forecasting studies for COVID-19, influenza, and dengue[52,55,86]). Intuitively, similar to our retrospective analysis, it is unreasonable to expect a single model (which has natural strengths and weaknesses) to capture all aspects of current or future epidemic dynamics. Such learned patterns may vary substantially across space and/or evolve over time. These variations may be why our DL forecasting models (TCN and TimeGPT) with climatic covariates and statistical time series model (SARIMA) forecasted more reliably than our semi-mechanistic Bayesian climate-based model which estimates parametric (climate-to-incidence) relationships for all provinces and all time points jointly. These trends are similar to previous poorer

predictive performance of models with mechanistic structure (and stronger performance of ensembles) in the dengue forecasting challenge[52]. Here, we deduce that a single model alone is likely insufficient for answering all of our research questions, yet we can answer different research questions robustly by using and potentially unifying several models. Our aim here is not to simply include every modelling approach in a single framework, but rather to identify which models are appropriate for different tasks (e.g. probabilistically explaining past epidemic drivers vs predicting future trajectories).

Of course, our forecasting results are not an upper bound on what is possible, as i) we have retained a preference for realism in our independent testing window, ii) our pipeline can readily incorporate many more forecasting models, and ii) future work could investigate further feature selection, optimal training windows for individual models and for weighting of models, as well as model weights per individual quantile. Such future works may help to improve the performance of the ensemble in provinces which experience rare, sporadic dengue cases (e.g. Ayabaca). Our focus here has instead been to showcase that we can, for any dengue epidemic, blend, rather than overemphasising, the benefits of forecasting from single modelling domains.

Our work also highlights the importance of not focusing on a single, potentially misleading performance metric. We must consider what information the forecasting metric provides for public health authorities. While used for forecast evaluation in many dengue studies to date (e.g.[59,60]), pointwise metrics obscure the natural uncertainty when we try to predict the unknown future of a dengue epidemic. We could, for example, obtain a high R^2 by just predicting the most recent value, while providing little useful information for public health authorities. Of course, similar to other recent forecasting studies (e.g.[50,51,54,55]), we therefore propose proper scoring rules. These measures of distributional accuracy (WIS) account for many deficiencies of pointwise metrics by simultaneously evaluating calibration and sharpness. This is why we focus on this metric, yet not exclusively, as i) we could again obtain reasonably good WIS values by simply forecasting the past value(s) (as drastic predictive distributional shifts may be rare in short time horizons) without conveying much information, and ii) we aim for our outputs to be easily interpreted and used by public health authorities. This is an important contribution of our work, as we complement proper scoring rules (which measure the whole distribution) with both relative WIS values (to reflect forecasting skill above the baseline) and measures of how well our forecasting models can detect specific aspects of epidemics, such as season onset. The WIS (and/or relative WIS) may be a statistically principled measure for our evaluations, yet when communicating our outputs, such a metric may matter so little to so many. Metrics must be both interpretable and public-health-oriented, and so our outbreak threshold metrics refocus evaluation strongly on parts of predictive distributions that are vital for decision-making. These new outbreak thresholds are simple yet generalisable across geographies of different sizes and across different spatial resolutions (e.g. coarser or finer resolution analyses).

In terms of further public health implications, it is interesting that we can still often obtain reliable probabilistic forecasts without climatic covariates. This may be again related to our limited one-month forecast horizon as given that we know the current level of the epidemic, the next month's cases may not require the climatic information to make reliable one-step-ahead forecasts. On the other hand, when making seasonal forecasts (and inferring epidemic drivers as in our retrospective analyses), climatic covariates are likely crucial. Another benefit of our approach is the computationally cheap nature of our forecasting pipeline. Working with just 23 predictive quantiles per observation to summarise the predictive distribution saves a modeller from processing potentially thousands of predictive samples. Finally, ensembles with untrained weighting outperform those which use trained weighting, likely in part due to weight uncertainty and/or the restricted spatial and temporal focus of our study.

4.3. Conclusions

The key advance of our research is to introduce a new framework and a new way of thinking about how we can robustly analyse dengue epidemic dynamics across space and over time. Our work introduces several state-of-the-art methods across different disciplines, yet their usage alone is not our objective. We complement these methods with traditional, reliable modelling techniques to produce a computationally

cheap and generalisable interdisciplinary pipeline that emphasises public health utility throughout. The outputs of our retrospective and forecasting techniques are designed to be interpretable, robust, and comprehensive, and can be used to inform long-term public health planning and real-time decision-making for dengue epidemics.

Acknowledgements

C.M. is supported by a studentship from the UK's Engineering and Physical Sciences Research Council. M.U.G.K. acknowledges funding from The Rockefeller Foundation (PC-2022-POP-005), Google.org, the Oxford Martin School Programmes in Pandemic Genomics & Digital Pandemic Preparedness, European Union's Horizon Europe programme projects MOOD (No. 874850) and E4Warning (No. 101086640), the John Fell Fund, a Branco Weiss Fellowship and Wellcome Trust grants 225288/Z/22/Z, 226052/Z/22/Z & 228186/Z/23/Z, United Kingdom Research and Innovation (No. APP8583) and the Medical Research Foundation (MRF-RG-ICCH-2022-100069). The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission or the other funders. C.A.D. is supported by the UK National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Emerging and Zoonotic Infections in partnership with Public Health England (PHE), Grant Number: HPRU200907. The funders had no role in the study design or analysis.

Declaration of Interests

The authors declare no competing interests.

Data Availability Statement

The dengue incidence surveillance data are publicly available from the National Centre for Epidemiology, Disease Prevention and Control (Peru CDC) in Peru's Ministry of Health. We sourced these from https://www.dge.gob.pe/salasituacional. Mid-year population estimates are made available by the National Institute of Statistics and Information of Peru at https://www.inei.gob.pe/media/MenuRecursivo/indices_tematicos/proy_04.xls. The WorldClim monthly historical climate data are available at https://www.worldclim.org/. SPI-6 data from the European Drought Observatory are available at https://jeodpp.jrc.ec.europa.eu/. The El Niño indices of the ONI and ICEN are available respectively from the NOAA (https://origin.cpc.ncep.noaa.gov/) and the Geophysical Institute of Peru (http://met.igp.gob.pe). All code and data used in our analysis is available at https://github.com/cathalmills/peru_dengue_province/.

References

[1] World Health Organization. Disease Outbreak News; Dengue – Global situation. Technical report, December 2023. URL https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON498.

[2] European Centre for Disease Prevention and Control. Dengue worldwide overview; Situation update, December 2023, December 2023. URL https://www.ecdc.europa.eu/en/dengue-monthly.

[3] Leah C. Katzelnick, Lionel Gresh, M. Elizabeth Halloran, Juan Carlos Mercado, Guillermina Kuan, Aubree Gordon, Angel Balmaseda, and Eva Harris. Antibody-dependent enhancement of severe dengue disease in humans. *Science*, 358(6365):929–932, November 2017. ISSN 0036-8075, 1095-9203. http://dx.doi.org/10.1126/science.aan6836. URL https://www.science.org/doi/10.1126/science.aan6836.

[4] Jing Liu-Helmersson, Hans Stenlund, Annelies Wilder-Smith, and Joacim Rocklöv. Vectorial Capacity of Aedes aegypti: Effects of Temperature and Implications for Global Dengue Epidemic

Potential. *PLoS ONE*, 9(3):e89783, March 2014. ISSN 1932-6203. http://dx.doi.org/10.1371/journal.pone.0089783. URL https://dx.plos.org/10.1371/journal.pone.0089783.

[5] Azael Che-Mendoza, Abdiel Martin-Park, Juan Manuel Chávez-Trava, Yamili Contreras-Perera, Hugo Delfín-González, Gabriela González-Olvera, Jorge Leirana-Alcocer, Guillermo Guillermo-May, Daniel Chan-Espinoza, Norma Pavia-Ruz, Rosa Eugenia Méndez-Vales, Alberto Alcocer-Gamboa, Fabian Correa-Morales, Jorge Palacio-Vargas, Dongjing Zhang, Gonzalo Vazquez-Prokopec, Zhiyong Xi, and Pablo Manrique-Saide. Abundance and Seasonality of *Aedes aegypti* (Diptera: Culicidae) in Two Suburban Localities of South Mexico, With Implications for *Wolbachia* (Rickettsiales: Rickettsiaceae)-Carrying Male Releases for Population Suppression. *Journal of Medical Entomology*, 58(4):1817–1825, July 2021. ISSN 0022-2585, 1938-2928. http://dx.doi.org/10.1093/jme/tjab052. URL https://academic.oup.com/jme/article/58/4/1817/6208904.

[6] Oliver J Brady, Michael A Johansson, Carlos A Guerra, Samir Bhatt, Nick Golding, David M Pigott, Hélène Delatte, Marta G Grech, Paul T Leisnham, Rafael Maciel-de Freitas, Linda M Styer, David L Smith, Thomas W Scott, Peter W Gething, and Simon I Hay. Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures in laboratory and field settings. *Parasites & Vectors*, 6(1):351, December 2013. ISSN 1756-3305. http://dx.doi.org/10.1186/1756-3305-6-351. URL https://parasitesandvectors.biomedcentral.com/articles/10.1186/1756-3305-6-351.

[7] Nur Athen Mohd Hardy Abdullah, Nazri Che Dom, Siti Aekball Salleh, Hasber Salim, and Nopadol Precha. The association between dengue case and climate: A systematic review and meta-analysis. *One Health*, 15:100452, December 2022. ISSN 23527714. http://dx.doi.org/10.1016/j.onehlt.2022.100452. URL https://linkinghub.elsevier.com/retrieve/pii/S2352771422000842.

[8] Manisha A. Kulkarni, Claudia Duguay, and Katarina Ost. Charting the evidence for climate change impacts on the global spread of malaria and dengue and adaptive responses: a scoping review of reviews. *Globalization and Health*, 18(1):1, December 2022. ISSN 1744-8603. http://dx.doi.org/10.1186/s12992-021-00793-2. URL https://globalizationandhealth.biomedcentral.com/articles/10.1186/s12992-021-00793-2.

[9] Willem G. Van Panhuis, Marc Choisy, Xin Xiong, Nian Shong Chok, Pasakorn Akarasewi, Sopon Iamsirithaworn, Sai K. Lam, Chee K. Chong, Fook C. Lam, Bounlay Phommasak, Phengta Vongphrachanh, Khamphaphongphane Bouaphanh, Huy Rekol, Nguyen Tran Hien, Pham Quang Thai, Tran Nhu Duong, Jen-Hsiang Chuang, Yu-Lun Liu, Lee-Ching Ng, Yuan Shi, Enrique A. Tayag, Vito G. Roque, Lyndon L. Lee Suy, Richard G. Jarman, Robert V. Gibbons, John Mark S. Velasco, In-Kyu Yoon, Donald S. Burke, and Derek A. T. Cummings. Region-wide synchrony and traveling waves of dengue across eight countries in Southeast Asia. *Proceedings of the National Academy of Sciences*, 112(42):13069–13074, October 2015. ISSN 0027-8424, 1091-6490. http://dx.doi.org/10.1073/pnas.1501375112. URL https://pnas.org/doi/full/10.1073/pnas.1501375112.

[10] Tia Dostal, Julianne Meisner, César Munayco, Patricia J. García, César Cárcamo, Jose Enrique Pérez Lu, Cory Morin, Lauren Frisbie, and Peter M. Rabinowitz. The effect of weather and climate on dengue outbreak risk in Peru, 2000-2018: A time-series analysis. *PLOS Neglected Tropical Diseases*, 16(6):e0010479, June 2022. ISSN 1935-2735. http://dx.doi.org/10.1371/journal.pntd.0010479. URL https://dx.plos.org/10.1371/journal.pntd.0010479.

[11] Duane J. Gubler. Dengue, Urbanization and Globalization: The Unholy Trinity of the 21st Century. *Tropical Medicine and Health*, 39(4SUPPLEMENT):S3–S11, 2011. ISSN 1349-4147, 1348-8945. http://dx.doi.org/10.2149/tmh.2011-S05. URL https://www.jstage.jst.go.jp/article/tmh/39/4SUPPLEMENT/39_2011-S05/_article.

[12] Amy Wesolowski, Taimur Qureshi, Maciej F. Boni, Pål Roe Sundsøy, Michael A. Johansson, Syed Basit Rasheed, Kenth Engø-Monsen, and Caroline O. Buckee. Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proceedings of the National Academy of Sciences*, 112(38):11887–11892, September 2015. ISSN 0027-8424, 1091-6490. http://dx.doi.org/10.1073/pnas.1504964112. URL https://pnas.org/doi/full/10.1073/pnas.1504964112.

[13] Rachel Lowe, Sophie A Lee, Kathleen M O'Reilly, Oliver J Brady, Leonardo Bastos, Gabriel Carrasco-Escobar, Rafael De Castro Catão, Felipe J Colón-González, Christovam Barcellos, Marilia Sá Carvalho, Marta Blangiardo, Håvard Rue, and Antonio Gasparrini. Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study. *The Lancet Planetary Health*, 5(4):e209–e219, April 2021. ISSN 25425196. http://dx.doi.org/10.1016/S2542-5196(20)30292-8. URL https://linkinghub.elsevier.com/retrieve/pii/S2542519620302928.

[14] Mikhail Churakov, Christian J. Villabona-Arenas, Moritz U. G. Kraemer, Henrik Salje, and Simon Cauchemez. Spatio-temporal dynamics of dengue in Brazil: Seasonal travelling waves and determinants of regional synchrony. *PLOS Neglected Tropical Diseases*, 13(4):e0007012, April 2019. ISSN 1935-2735. http://dx.doi.org/10.1371/journal.pntd.0007012. URL https://dx.plos.org/10.1371/journal.pntd.0007012.

[15] Cara E. Brook, Carly Rozins, Jennifer A. Bohl, Vida Ahyong, Sophana Chea, Liz Fahsbender, Rekol Huy, Sreyngim Lay, Rithea Leang, Yimei Li, Chanthap Lon, Somnang Man, Mengheng Oum, Graham R. Northrup, Fabiano Oliveira, Andrea R. Pacheco, Daniel M. Parker, Katherine Young, Michael Boots, Cristina M. Tato, Joseph L. DeRisi, Christina Yek, and Jessica E. Manning. Climate, demography, immunology, and virology combine to drive two decades of dengue virus dynamics in Cambodia, June 2022. URL http://medrxiv.org/lookup/doi/10.1101/2022.06.08.22276171.

[16] Rachel Lowe, Trevor C. Bailey, David B. Stephenson, Tim E. Jupp, Richard J. Graham, Christovam Barcellos, and Marilia Sá Carvalho. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Statistics in Medicine*, 32(5):864–883, February 2013. ISSN 02776715. http://dx.doi.org/10.1002/sim.5549. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.5549.

[17] Rachel Lowe, Antonio Gasparrini, Cédric J. Van Meerbeeck, Catherine A. Lippi, Roché Mahon, Adrian R. Trotman, Leslie Rollock, Avery Q. J. Hinds, Sadie J. Ryan, and Anna M. Stewart-Ibarra. Nonlinear and delayed impacts of climate on dengue risk in Barbados: A modelling study. *PLOS Medicine*, 15(7):e1002613, July 2018. ISSN 1549-1676. http://dx.doi.org/10.1371/journal.pmed.1002613. URL https://dx.plos.org/10.1371/journal.pmed.1002613.

[18] G. Chowell, M. A. Miller, and C. Viboud. Seasonal influenza in the United States, France, and Australia: transmission and prospects for control. *Epidemiology and Infection*, 136(6):852–864, June 2008. ISSN 0950-2688, 1469-4409. http://dx.doi.org/10.1017/S0950268807009144. URL https://www.cambridge.org/core/product/identifier/S0950268807009144/type/journal_article.

[19] Cathal Mills and Christl A Donnelly. Climate-based modelling and forecasting of dengue fever in three endemic departments of Peru. *PLOS Neglected Tropical Diseases (In Press)*, 2024.

[20] Léo Grinsztajn, Elizaveta Semenova, Charles C. Margossian, and Julien Riou. Bayesian workflow for disease transmission modeling in Stan. *Statistics in Medicine*, 40(27):6209–6234, November 2021. ISSN 0277-6715, 1097-0258. http://dx.doi.org/10.1002/sim.9164. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.9164.

[21] Judith A. Bouman, Anthony Hauser, Simon L. Grimm, Martin Wohlfender, Samir Bhatt, Elizaveta Semenova, Andrew Gelman, Christian L. Althaus, and Julien Riou. Bayesian workflow for time-varying transmission in stratified compartmental infectious disease transmission models. *PLOS Computational Biology*, 20(4):e1011575, April 2024. ISSN 1553-7358. http://dx.doi.org/10.1371/journal.pcbi.1011575. URL https://dx.plos.org/10.1371/journal.pcbi.1011575.

[22] Z. F. Dembek, T. Chekol, and A. Wu. Best practice assessment of disease modelling for infectious disease outbreaks. *Epidemiology and Infection*, 146(10):1207–1215, 2018. http://dx.doi.org/10.1017/S095026881800119X.

[23] Amanda Minter and Renata Retkute. Approximate Bayesian Computation for infectious disease modelling. *Epidemics*, 29:100368, December 2019. ISSN 1878-0067. http://dx.doi.org/10.1016/j.epidem.2019.100368.

[24] CDC Peru. Reporte de Tabla de casos notificados por causas, May 2023. URL https://www.dge.gob.pe/salasituacional/.

[25] César V. Munayco, Betsabet Yadira Valderrama Rosales, Susan Yanett Mateo Lizarbe, Carmen Rosa Yon Fabian, Ricardo Peña Sánchez, César Henry Vásquez Sánchez, Maria Paquita García, Carlos Padilla-Rojas, Victor Suárez, Liliana Sánchez-González, Forrest K. Jones, Luciana Kohatsu, Laura E. Adams, Juliette Morgan, and Gabriela Paz-Bailey. *Notes from the Field:* Dengue Outbreak — Peru, 2023. *MMWR. Morbidity and Mortality Weekly Report*, 73(4):86–88, February 2024. ISSN 0149-2195, 1545-861X. http://dx.doi.org/10.15585/mmwr.mm7304a4. URL http://www.cdc.gov/mmwr/volumes/73/wr/mm7304a4.htm?s_cid=mm7304a4_w.

[26] INEI. Censos Nacionales 2007: XI de Población y VI de Vivienda, 2007. URL https://www.gob.pe/36495-consultar-informacion-de-los-censos-realizados-por-el-inei-censos-nacionales-de-poblacion-y-vivien

[27] INEI. Resultados Definitivos de los Censos Nacionales 2017, 2017. URL https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1544/.

[28] INEI. Population and Housing, June 2022. URL https://www.inei.gob.pe/estadisticas/indice-tematico/poblacion-y-vivienda/.

[29] Stephen E. Fick and Robert J. Hijmans. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12):4302–4315, October 2017. ISSN 0899-8418, 1097-0088. http://dx.doi.org/10.1002/joc.5086. URL https://onlinelibrary.wiley.com/doi/10.1002/joc.5086.

[30] Ian Harris, Timothy J. Osborn, Phil Jones, and David Lister. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7(1):109, April 2020. ISSN 2052-4463. http://dx.doi.org/10.1038/s41597-020-0453-3. URL https://www.nature.com/articles/s41597-020-0453-3.

[31] B. Gruvberger, M. Bruze, and M. Tammela. Preservatives in moisturizers on the Swedish market. *Acta Dermato-Venereologica*, 78(1):52–56, January 1998. ISSN 0001-5555. http://dx.doi.org/10.1080/00015559850135850.

[32] Thomas B McKee, Nolan J Doesken, John Kleist, and others. The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, volume 17, pages 179–183. California, 1993. Issue: 22.

[33] European Drought Observatory. Standardized Precipitation Index (SPI), 2023.

[34] K. Takahashi, A. Montecinos, K. Goubanova, and B. Dewitte. ENSO regimes: Reinterpreting the canonical and Modoki El Niño. *Geophysical Research Letters*, 38(10):n/a–n/a, May 2011. ISSN 00948276. http://dx.doi.org/10.1029/2011GL047364. URL http://doi.wiley.com/10.1029/2011GL047364.

[35] Ken Takahashi, Kobi Mosquera, and Jorge Reupo. The El Niño Coastal Index (ICEN): history and update. Technical bulletin, Geophysical Institute of Peru, February 2014.

[36] NOAA. Cold & Warm Episodes by Season, July 2023. URL https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php.

[37] NOAA. Nino regions., July 2023.

[38] Rory Gibb, Felipe J. Colón-González, Phan Trong Lan, Phan Thi Huong, Vu Sinh Nam, Vu Trong Duoc, Do Thai Hung, Nguyn Thanh Dong, Vien Chinh Chien, Ly Thi Thuy Trang, Do Kien Quoc, Tran Minh Hoa, Nguyen Hu Tai, Tran Thi Hang, Gina Tsarouchi, Eleanor Ainscoe, Quillon Harpham, Barbara Hofmann, Darren Lumbroso, Oliver J. Brady, and Rachel Lowe. Interactions between climate change, urban infrastructure and mobility are driving dengue emergence in Vietnam. *Nature Communications*, 14(1):8179, December 2023. ISSN 2041-1723. http://dx.doi.org/10.1038/s41467-023-43954-0. URL https://www.nature.com/articles/s41467-023-43954-0.

[39] R Core Team. R: A Language and Environment for Statistical Computing, 2022. URL https://www.R-project.org/.

[40] Angi Roesch and Harald Schmidbauer. *WaveletComp: Computational Wavelet Analysis*. 2018. URL https://CRAN.R-project.org/package=WaveletComp.

[41] Trevor Hastie and Robert Tibshirani. *Generalized additive models*. Chapman & Hall/CRC, Boca Raton, Fla, 1999. ISBN 978-0-412-34390-2.

[42] Simon N. Wood. The R Journal: Mgcv: GAMs and generalized ridge regression for R. *R News*, 1 (2):20–25, 2001. ISSN 1609-3631.

[43] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, December 1974. ISSN 0018-9286. http://dx.doi.org/10.1109/TAC.1974.1100705. URL http://ieeexplore.ieee.org/document/1100705/.

[44] Gideon Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), March 1978. ISSN 0090-5364. http://dx.doi.org/10.1214/aos/1176344136. URL https://projecteuclid.org/journals/annals-of-statistics/volume-6/issue-2/Estimating-the-Dimension-of-a-Model/10.1214/aos/1176344136.full.

[45] Peter Craven and Grace Wahba. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31(4):377–403, December 1978. ISSN 0029-599X, 0945-3245. http://dx.doi.org/10.1007/BF01404567. URL http://link.springer.com/10.1007/BF01404567.

[46] Julian Besag, Jeremy York, and Annie Mollie. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20, March 1991. ISSN 0020-3157, 1572-9052. http://dx.doi.org/10.1007/BF00116466. URL http://link.springer.com/10.1007/BF00116466.

[47] Daniel Simpson, Håvard Rue, Andrea Riebler, Thiago G. Martins, and Sigrunn H. Sørbye. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, 32(1), February 2017. ISSN 0883-4237. http://dx.doi.org/10.1214/16-STS576. URL https://projecteuclid.org/journals/statistical-science/volume-32/issue-1/Penalising-Model-Component-Complexity--A-Principled-Practical-Approach-to/10.1214/16-STS576.full.

[48] A. Gasparrini, B. Armstrong, and M. G. Kenward. Distributed lag non-linear models. *Statistics in Medicine*, 29(21):2224–2234, September 2010. ISSN 02776715. http://dx.doi.org/10.1002/sim.3940. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.3940.

[49] Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(2):319–392, April 2009. ISSN 1369-7412, 1467-9868. http://dx.doi.org/10.1111/j.1467-9868.2008.00700.x. URL https://academic.oup.com/jrsssb/article/71/2/319/7092907.

[50] Estee Y. Cramer, Evan L. Ray, Velma K. Lopez, Johannes Bracher, Andrea Brennen, Alvaro J. Castro Rivadeneira, Aaron Gerding, Tilmann Gneiting, Katie H. House, Yuxin Huang, Dasuni Jayawardena, Abdul H. Kanji, Ayush Khandelwal, Khoa Le, Anja Mühlemann, Jarad Niemi, Apurv Shah, Ariane Stark, Yijin Wang, Nutcha Wattanachit, Martha W. Zorn, Youyang Gu, Sansiddh Jain, Nayana Bannur, Ayush Deva, Mihir Kulkarni, Srujana Merugu, Alpan Raval, Siddhant Shingi, Avtansh Tiwari, Jerome White, Neil F. Abernethy, Spencer Woody, Maytal Dahan, Spencer Fox, Kelly Gaither, Michael Lachmann, Lauren Ancel Meyers, James G. Scott, Mauricio Tec, Ajitesh Srivastava, Glover E. George, Jeffrey C. Cegan, Ian D. Dettwiller, William P. England, Matthew W. Farthing, Robert H. Hunter, Brandon Lafferty, Igor Linkov, Michael L. Mayo, Matthew D. Parno, Michael A. Rowland, Benjamin D. Trump, Yanli Zhang-James, Samuel Chen, Stephen V. Faraone, Jonathan Hess, Christopher P. Morley, Asif Salekin, Dongliang Wang, Sabrina M. Corsetti, Thomas M. Baer, Marisa C. Eisenberg, Karl Falb, Yitao Huang, Emily T. Martin, Ella McCauley, Robert L. Myers, Tom Schwarz, Daniel Sheldon, Graham Casey Gibson, Rose Yu, Liyao Gao, Yian Ma, Dongxia Wu, Xifeng Yan, Xiaoyong Jin, Yu-Xiang Wang, YangQuan Chen, Lihong Guo, Yanting Zhao, Quanquan Gu, Jinghui Chen, Lingxiao Wang, Pan Xu, Weitong Zhang, Difan Zou, Hannah Biegel, Joceline Lega, Steve McConnell, V. P. Nagraj, Stephanie L. Guertin, Christopher Hulme-Lowe, Stephen D. Turner, Yunfeng Shi, Xuegang Ban, Robert Walraven, Qi-Jun Hong, Stanley Kong, Axel van de Walle, James A. Turtle, Michal Ben-Nun, Steven Riley, Pete Riley, Ugur Koyluoglu, David DesRoches, Pedro Forli, Bruce Hamory, Christina Kyriakides, Helen Leis, John Milliken, Michael Moloney, James Morgan, Ninad Nirgudkar, Gokce Ozcan, Noah Piwonka,

Matt Ravi, Chris Schrader, Elizabeth Shakhnovich, Daniel Siegel, Ryan Spatz, Chris Stiefeling, Barrie Wilkinson, Alexander Wong, Sean Cavany, Guido España, Sean Moore, Rachel Oidtman, Alex Perkins, David Kraus, Andrea Kraus, Zhifeng Gao, Jiang Bian, Wei Cao, Juan Lavista Ferres, Chaozhuo Li, Tie-Yan Liu, Xing Xie, Shun Zhang, Shun Zheng, Alessandro Vespignani, Matteo Chinazzi, Jessica T. Davis, Kunpeng Mu, Ana Pastore Y Piontti, Xinyue Xiong, Andrew Zheng, Jackie Baek, Vivek Farias, Andreea Georgescu, Retsef Levi, Deeksha Sinha, Joshua Wilde, Georgia Perakis, Mohammed Amine Bennouna, David Nze-Ndong, Divya Singhvi, Ioannis Spantidakis, Leann Thayaparan, Asterios Tsiourvas, Arnab Sarker, Ali Jadbabaie, Devavrat Shah, Nicolas Della Penna, Leo A. Celi, Saketh Sundar, Russ Wolfinger, Dave Osthus, Lauren Castro, Geoffrey Fairchild, Isaac Michaud, Dean Karlen, Matt Kinsey, Luke C. Mullany, Kaitlin Rainwater-Lovett, Lauren Shin, Katharine Tallaksen, Shelby Wilson, Elizabeth C. Lee, Juan Dent, Kyra H. Grantz, Alison L. Hill, Joshua Kaminsky, Kathryn Kaminsky, Lindsay T. Keegan, Stephen A. Lauer, Joseph C. Lemaitre, Justin Lessler, Hannah R. Meredith, Javier Perez-Saez, Sam Shah, Claire P. Smith, Shaun A. Truelove, Josh Wills, Maximilian Marshall, Lauren Gardner, Kristen Nixon, John C. Burant, Lily Wang, Lei Gao, Zhiling Gu, Myungjin Kim, Xinyi Li, Guannan Wang, Yueying Wang, Shan Yu, Robert C. Reiner, Ryan Barber, Emmanuela Gakidou, Simon I. Hay, Steve Lim, Chris Murray, David Pigott, Heidi L. Gurung, Prasith Baccam, Steven A. Stage, Bradley T. Suchoski, B. Aditya Prakash, Bijaya Adhikari, Jiaming Cui, Alexander Rodríguez, Anika Tabassum, Jiajia Xie, Pinar Keskinocak, John Asplund, Arden Baxter, Buse Eylul Oruc, Nicoleta Serban, Sercan O. Arik, Mike Dusenberry, Arkady Epshteyn, Elli Kanal, Long T. Le, Chun-Liang Li, Tomas Pfister, Dario Sava, Rajarishi Sinha, Thomas Tsai, Nate Yoder, Jinsung Yoon, Leyou Zhang, Sam Abbott, Nikos I. Bosse, Sebastian Funk, Joel Hellewell, Sophie R. Meakin, Katharine Sherratt, Mingyuan Zhou, Rahi Kalantari, Teresa K. Yamana, Sen Pei, Jeffrey Shaman, Michael L. Li, Dimitris Bertsimas, Omar Skali Lami, Saksham Soni, Hamza Tazi Bouardi, Turgay Ayer, Madeline Adee, Jagpreet Chhatwal, Ozden O. Dalgic, Mary A. Ladd, Benjamin P. Linas, Peter Mueller, Jade Xiao, Yuanjia Wang, Qinxia Wang, Shanghong Xie, Donglin Zeng, Alden Green, Jacob Bien, Logan Brooks, Addison J. Hu, Maria Jahja, Daniel McDonald, Balasubramanian Narasimhan, Collin Politsch, Samyak Rajanala, Aaron Rumack, Noah Simon, Ryan J. Tibshirani, Rob Tibshirani, Valerie Ventura, Larry Wasserman, Eamon B. O'Dea, John M. Drake, Robert Pagano, Quoc T. Tran, Lam Si Tung Ho, Huong Huynh, Jo W. Walker, Rachel B. Slayton, Michael A. Johansson, Matthew Biggerstaff, and Nicholas G. Reich. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 119 (15):e2113561119, April 2022. ISSN 1091-6490. http://dx.doi.org/10.1073/pnas.2113561119.

[51] Felipe J. Colón-González, Leonardo Soares Bastos, Barbara Hofmann, Alison Hopkin, Quillon Harpham, Tom Crocker, Rosanna Amato, Iacopo Ferrario, Francesca Moschini, Samuel James, Sajni Malde, Eleanor Ainscoe, Vu Sinh Nam, Dang Quang Tan, Nguyen Duc Khoa, Mark Harrison, Gina Tsarouchi, Darren Lumbroso, Oliver J. Brady, and Rachel Lowe. Probabilistic seasonal dengue forecasting in Vietnam: A modelling study using superensembles. *PLOS Medicine*, 18 (3):e1003542, March 2021. ISSN 1549-1676. http://dx.doi.org/10.1371/journal.pmed.1003542. URL https://dx.plos.org/10.1371/journal.pmed.1003542.

[52] Michael A. Johansson, Karyn M. Apfeldorf, Scott Dobson, Jason Devita, Anna L. Buczak, Benjamin Baugher, Linda J. Moniz, Thomas Bagley, Steven M. Babin, Erhan Guven, Teresa K. Yamana, Jeffrey Shaman, Terry Moschou, Nick Lothian, Aaron Lane, Grant Osborne, Gao Jiang, Logan C. Brooks, David C. Farrow, Sangwon Hyun, Ryan J. Tibshirani, Roni Rosenfeld, Justin Lessler, Nicholas G. Reich, Derek A. T. Cummings, Stephen A. Lauer, Sean M. Moore, Hannah E. Clapham, Rachel Lowe, Trevor C. Bailey, Markel García-Díez, Marilia Sá Carvalho, Xavier Rodó, Tridip Sardar, Richard Paul, Evan L. Ray, Krzysztof Sakrejda, Alexandria C. Brown, Xi Meng, Osonde Osoba, Raffaele Vardavas, David Manheim, Melinda Moore, Dhananjai M. Rao, Travis C. Porco, Sarah Ackley, Fengchen Liu, Lee Worden, Matteo Convertino, Yang Liu, Abraham Reddy, Eloy Ortiz, Jorge Rivero, Humberto Brito, Alicia Juarrero, Leah R. Johnson, Robert B. Gramacy, Jeremy M. Cohen, Erin A. Mordecai, Courtney C. Murdock, Jason R. Rohr, Sadie J. Ryan,

Anna M. Stewart-Ibarra, Daniel P. Weikel, Antarpreet Jutla, Rakibul Khan, Marissa Poultney, Rita R. Colwell, Brenda Rivera-García, Christopher M. Barker, Jesse E. Bell, Matthew Biggerstaff, David Swerdlow, Luis Mier-y Teran-Romero, Brett M. Forshey, Juli Trtanj, Jason Asher, Matt Clay, Harold S. Margolis, Andrew M. Hebbeler, Dylan George, and Jean-Paul Chretien. An open challenge to advance probabilistic forecasting for dengue epidemics. *Proceedings of the National Academy of Sciences*, 116(48):24268–24274, November 2019. ISSN 0027-8424, 1091-6490. http://dx.doi.org/10.1073/pnas.1909865116. URL https://pnas.org/doi/full/10.1073/pnas.1909865116.

[53] Roberto Buizza. Introduction to the special issue on "25 years of ensemble forecasting". *Quarterly Journal of the Royal Meteorological Society*, 145(S1):1–11, September 2019. ISSN 0035-9009, 1477-870X. http://dx.doi.org/10.1002/qj.3370. URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3370.

[54] Johannes Bracher, Evan L. Ray, Tilmann Gneiting, and Nicholas G. Reich. Evaluating epidemic forecasts in an interval format. *PLOS Computational Biology*, 17(2):e1008618, February 2021. ISSN 1553-7358. http://dx.doi.org/10.1371/journal.pcbi.1008618. URL https://dx.plos.org/10.1371/journal.pcbi.1008618.

[55] Evan L. Ray, Logan C. Brooks, Jacob Bien, Matthew Biggerstaff, Nikos I. Bosse, Johannes Bracher, Estee Y. Cramer, Sebastian Funk, Aaron Gerding, Michael A. Johansson, Aaron Rumack, Yijin Wang, Martha Zorn, Ryan J. Tibshirani, and Nicholas G. Reich. Comparing trained and untrained probabilistic ensemble forecasts of COVID-19 cases and deaths in the United States. *International Journal of Forecasting*, 39(3):1366–1383, July 2023. ISSN 01692070. http://dx.doi.org/10.1016/j.ijforecast.2022.06.005. URL https://linkinghub.elsevier.com/retrieve/pii/S0169207022000966.

[56] Tilmann Gneiting and Matthias Katzfuss. Probabilistic Forecasting. *Annual Review of Statistics and Its Application*, 1(1):125–151, January 2014. ISSN 2326-8298, 2326-831X. http://dx.doi.org/10.1146/annurev-statistics-062713-085831. URL https://www.annualreviews.org/doi/10.1146/annurev-statistics-062713-085831.

[57] Evan L. Ray and Nicholas G. Reich. Prediction of infectious disease epidemics via weighted density ensembles. *PLOS Computational Biology*, 14(2):e1005910, February 2018. ISSN 1553-7358. http://dx.doi.org/10.1371/journal.pcbi.1005910. URL https://dx.plos.org/10.1371/journal.pcbi.1005910.

[58] Spencer Wadsworth, Jarad Niemi, and Nick Reich. Mixture distributions for probabilistic forecasts of disease outbreaks. 2023. http://dx.doi.org/10.48550/ARXIV.2310.11939. URL https://arxiv.org/abs/2310.11939. Publisher: arXiv Version Number: 1.

[59] Madhurima Panja, Tanujit Chakraborty, Sk Shahid Nadim, Indrajit Ghosh, Uttam Kumar, and Nan Liu. An ensemble neural network approach to forecast Dengue outbreak based on climatic condition. *Chaos, Solitons & Fractals*, 167:113124, February 2023. ISSN 09600779. http://dx.doi.org/10.1016/j.chaos.2023.113124. URL https://linkinghub.elsevier.com/retrieve/pii/S0960077923000255.

[60] Soudeep Deb and Sougata Deb. An Ensemble Method for Early Prediction of Dengue Outbreak. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(1):84–101, January 2022. ISSN 0964-1998, 1467-985X. http://dx.doi.org/10.1111/rssa.12714. URL https://academic.oup.com/jrsssa/article/185/1/84/7068440.

[61] Alessandro Sebastianelli, Dario Spiller, Raquel Carmo, James Wheeler, Artur Nowakowski, Ludmilla Viana Jacobson, Dohyung Kim, Hanoch Barlevi, Zoraya El Raiss Cordero, Felipe J Colón-González, Rachel Lowe, Silvia Liberata Ullo, and Rochelle Schneider. A reproducible ensemble machine learning approach to forecast dengue outbreaks. *Scientific Reports*, 14(1):3807, February 2024. ISSN 2045-2322. http://dx.doi.org/10.1038/s41598-024-52796-9. URL https://www.nature.com/articles/s41598-024-52796-9.

[62] Ryan Tibshirani and Logan Brooks. *quantgen: Tools for generalized quantile modeling*. 2022.

[63] Tilmann Gneiting and Adrian E Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477):359–378, March 2007. ISSN 0162-1459, 1537-274X. http://dx.doi.org/10.1198/016214506000001437. URL http://www.tandfonline.com/doi/abs/10.1198/016214506000001437.

[64] Nikos I. Bosse, Sam Abbott, Anne Cori, Edwin Van Leeuwen, Johannes Bracher, and Sebastian Funk. Scoring epidemiological forecasts on transformed scales. *PLOS Computational Biology*, 19 (8):e1011393, August 2023. ISSN 1553-7358. http://dx.doi.org/10.1371/journal.pcbi.1011393. URL https://dx.plos.org/10.1371/journal.pcbi.1011393.

[65] Julien Herzen, Francesco LÃ¤ssig, Samuele Giuliano Piazzetta, Thomas Neuer, LÃ©o Tafti, Guillaume Raille, Tomas Van Pottelbergh, Marek Pasieka, Andrzej Skrodzki, Nicolas Huguenin, Maxime Dumonal, Jan KoÅ›cisz, Dennis Bader, FrÃ©dÃ©rick Gusset, Mounir Benheddi, Camila Williamson, Michal Kosinski, Matej Petrik, and GaÃ«l Grosch. Darts: User-Friendly Modern Machine Learning for Time Series. *Journal of Machine Learning Research*, 23(124):1–6, 2022. URL http://jmlr.org/papers/v23/21-1177.html.

[66] Cristian Challú Kin G. Olivares Federico Garza, Max Mergenthaler Canseco. StatsForecast: Lightning fast forecasting with statistical and econometric models, 2022. URL https://github.com/Nixtla/statsforecast. Published: PyCon Salt Lake City, Utah, US 2022.

[67] Firuz Kamalov, Khairan Rajab, Aswani Kumar Cherukuri, Ashraf Elnagar, and Murodbek Safaraliev. Deep learning for Covid-19 forecasting: State-of-the-art review. *Neurocomputing*, 511: 142–154, October 2022. ISSN 09252312. http://dx.doi.org/10.1016/j.neucom.2022.09.005. URL https://linkinghub.elsevier.com/retrieve/pii/S0925231222010918.

[68] Azul Garza, Cristian Challu, and Max Mergenthaler-Canseco. TimeGPT-1, May 2024. URL http://arxiv.org/abs/2310.03589. arXiv:2310.03589 [cs].

[69] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need, August 2023. URL http://arxiv.org/abs/1706.03762. arXiv:1706.03762 [cs].

[70] Anastasios N. Angelopoulos, Emmanuel J. Candes, and Ryan J. Tibshirani. Conformal PID Control for Time Series Prediction, July 2023. URL http://arxiv.org/abs/2307.16895. arXiv:2307.16895 [cs, eess, stat].

[71] Simon Pollett, Michael A. Johansson, Nicholas G. Reich, David Brett-Major, Sara Y. Del Valle, Srinivasan Venkatramanan, Rachel Lowe, Travis Porco, Irina Maljkovic Berry, Alina Deshpande, Moritz U. G. Kraemer, David L. Blazes, Wirichada Pan-ngum, Alessandro Vespigiani, Suzanne E. Mate, Sheetal P. Silal, Sasikiran Kandula, Rachel Sippy, Talia M. Quandelacy, Jeffrey J. Morgan, Jacob Ball, Lindsay C. Morton, Benjamin M. Althouse, Julie Pavlin, Wilbert Van Panhuis, Steven Riley, Matthew Biggerstaff, Cecile Viboud, Oliver Brady, and Caitlin Rivers. Recommended reporting items for epidemic forecasting and prediction research: The EPIFORGE 2020 guidelines. *PLOS Medicine*, 18(10):e1003793, October 2021. ISSN 1549-1676. http://dx.doi.org/10.1371/journal.pmed.1003793. URL https://dx.plos.org/10.1371/journal.pmed.1003793.

[72] Gene H. Golub, Michael Heath, and Grace Wahba. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, 21(2):215–223, May 1979. ISSN 0040-1706, 1537-2723. http://dx.doi.org/10.1080/00401706.1979.10489751. URL http://www.tandfonline.com/doi/abs/10.1080/00401706.1979.10489751.

[73] Clifford M. Hurvich and Chih-Ling Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989. ISSN 0006-3444, 1464-3510. http://dx.doi.org/10.1093/biomet/76.2.297. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/76.2.297.

[74] C. MacLachlan, A. Arribas, K. A. Peterson, A. Maidens, D. Fereday, A. A. Scaife, M. Gordon, M. Vellinga, A. Williams, R. E. Comer, J. Camp, P. Xavier, and G. Madec. Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, 141(689):1072–1084, April 2015. ISSN 0035-9009, 1477-870X. http://dx.doi.org/10.1002/qj.2396. URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.2396.

[75] Alexandre S. Gagnon, Andrew B. G. Bush, and Karen E. Smoyer-Tomic. Dengue epidemics and the El Niño Southern Oscillation. *Climate Research*, 19(1):35–43, 2001. ISSN 0936577X, 16161572. URL http://www.jstor.org/stable/24866766. Publisher: Inter-Research Science Center.

[76] Chuanxi Li, Zhao Liu, Wen Li, Yuxi Lin, Liangyu Hou, Shuyue Niu, Yue Xing, Jianbin Huang, Yidan Chen, Shangchen Zhang, Xuejie Gao, Ying Xu, Can Wang, Qi Zhao, Qiyong Liu, Wei Ma, Wenjia Cai, Peng Gong, and Yong Luo. Projecting future risk of dengue related to hydrometeorological conditions in mainland China under climate change scenarios: a modelling study. *The Lancet Planetary Health*, 7(5):e397–e406, May 2023. ISSN 25425196. http://dx.doi.org/10.1016/S2542-5196(23)00051-7. URL https://linkinghub.elsevier.com/retrieve/pii/S2542519623000517.

[77] L. Brown, J. Medlock, and V. Murray. Impact of drought on vector-borne diseases–how does one manage the risk? *Public Health*, 128(1):29–37, January 2014. ISSN 1476-5616. http://dx.doi.org/10.1016/j.puhe.2013.09.006.

[78] Brendan J. Trewin, Hazel R. Parry, Daniel E. Pagendam, Gregor J. Devine, Myron P. Zalucki, Jonathan M. Darbro, Cassie C. Jansen, and Nancy A. Schellhorn. Simulating an invasion: unsealed water storage (rainwater tanks) and urban block design facilitate the spread of the dengue fever mosquito, Aedes aegypti, in Brisbane, Australia. *Biological Invasions*, 23(12):3891–3906, December 2021. ISSN 1387-3547, 1573-1464. http://dx.doi.org/10.1007/s10530-021-02619-z. URL https://link.springer.com/10.1007/s10530-021-02619-z.

[79] M. Trpis. Dry season survival of Aedes aegypti eggs in various breeding sites in the Dar es Salaam area, Tanzania. *Bulletin of the World Health Organization*, 47(3):433–437, 1972. ISSN 0042-9686.

[80] Osama M. E. Seidahmed and Elfatih A. B. Eltahir. A Sequence of Flushing and Drying of Breeding Habitats of Aedes aegypti (L.) Prior to the Low Dengue Season in Singapore. *PLOS Neglected Tropical Diseases*, 10(7):e0004842, July 2016. ISSN 1935-2735. http://dx.doi.org/10.1371/journal.pntd.0004842. URL https://dx.plos.org/10.1371/journal.pntd.0004842.

[81] Yawen Wang, Yuchen Wei, Kehang Li, Xiaoting Jiang, Conglu Li, Qianying Yue, Benny Chung-ying Zee, and Ka Chun Chong. Impact of extreme weather on dengue fever infection in four Asian countries: A modelling analysis. *Environment International*, 169:107518, November 2022. ISSN 01604120. http://dx.doi.org/10.1016/j.envint.2022.107518. URL https://linkinghub.elsevier.com/retrieve/pii/S0160412022004457.

[82] World Health Organization and UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases. *Global vector control response 2017-2030*. World Health Organization, Geneva, 2017. ISBN 978-92-4-151297-8. URL https://iris.who.int/handle/10665/259205.

[83] Nicolas Banholzer, Thomas Mellan, H. Juliette T. Unwin, Stefan Feuerriegel, Swapnil Mishra, and Samir Bhatt. A comparison of short-term probabilistic forecasts for the incidence of COVID-19 using mechanistic and statistical time series models, May 2023. URL http://arxiv.org/abs/2305.00933. arXiv:2305.00933 [cs, q-bio, stat].

[84] Taha Falatouri, Farzaneh Darbanian, Patrick Brandtner, and Chibuzor Udokwu. Predictive Analytics for Demand Forecasting – A Comparison of SARIMA and LSTM in Retail SCM. *Procedia Computer Science*, 200:993–1003, 2022. ISSN 18770509. http://dx.doi.org/10.1016/j.procs.2022.01.298. URL https://linkinghub.elsevier.com/retrieve/pii/S1877050922003076.

[85] Lorenzo Menculini, Andrea Marini, Massimiliano Proietti, Alberto Garinei, Alessio Bozza, Cecilia Moretti, and Marcello Marconi. Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices. *Forecasting*, 3(3):644–662, September 2021. ISSN 2571-9394. http://dx.doi.org/10.3390/forecast3030040. URL https://www.mdpi.com/2571-9394/3/3/40.

[86] Tim K. Tsang, Qiurui Du, Benjamin J. Cowling, and Cécile Viboud. An adaptive weight ensemble approach to forecast influenza activity in an irregular seasonality context. *Nature Communications*, 15(1):8625, October 2024. ISSN 2041-1723. http://dx.doi.org/10.1038/s41467-024-52504-1. URL https://www.nature.com/articles/s41467-024-52504-1.

[87] J. Welles Wilder. *New concepts in technical trading systems*. Trend Research, Greensboro, N.C, 1978. ISBN 978-0-89459-027-6.

[88] C. J. Clopper and E. S. Pearson. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*, 26(4):404–413, 1934. ISSN 0006-3444, 1464-3510. http://dx.doi.org/10.1093/biomet/26.4.404. URL https://academic.oup.com/biomet/article-lookup/doi/

10.1093/biomet/26.4.404.