The time- and space-varying roles of human mobility in shaping urban dengue epidemics

Cathal Mills*[1, 2], Geziel dos Santos de Sousa[3, 4], Antonio Silva Lima Neto[5, 7], Vasco Furtado[6, 7], Sen Pei[8], Moritz U.G. Kraemer[2, 9], 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]Mestrado Profissional em Gestão em Saúde, Universidade Estadual do Ceará, Fortaleza, Brazil
[4]Programa de Pós-Graduação em Desenvolvimento e Meio Ambiente, Universidade Federal do Ceará, Fortaleza, Brazil
[5]Secretaria Executiva de Vigilância em Saúde, Secretaria da Saúde do Ceará, Fortaleza, Brazil
[6]Empresa de Tecnologia da Informação do Ceará, Governo do Estado do Ceará, Fortaleza, Brazil
[7]Laboratório de Ciência de Dados e Inteligência Artificial, Universidade de Fortaleza, Fortaleza, Brazil
[8]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, United States of America
[9]Department of Biology, University of Oxford, Oxford, United Kingdom

Abstract

Dengue outbreaks continue to cause large morbidity and mortality globally. Previous work has shown that human movement alongside weather, climatic, ecological, and socioeconomic factors influence outbreak size, persistence, and geographical spread. However, the relative importance of human movement is unclear for the establishment, rapid expansion, and persistence of dengue in urban settings. Relatedly, the extent to which neighbourhoods differentially influence outbreaks over time is unclear. To address these gaps, we developed a multi-model framework that integrates Bayesian hierarchical modelling and data-driven, deep-learning-based approaches to parameter inference and out-of-sample probabilistic predictions of dengue in Fortaleza, Brazil. We apply our framework to dengue outbreaks at the neighbourhood level, using epidemiological surveillance data alongside public transportation data. We use information criteria, scoring rules, and explainability metrics to measure the models' performance and ultimately explain how human mobility shapes disease dynamics. We find that human mobility is a consistent driver of dengue outbreaks over the five-year period studied. We also find that human movement is a more important driver of transmission dynamics between neighbourhoods than transmission within them, both within and across dengue seasons, and that spatially, the relative importance of different neighbourhoods to transmission elsewhere is relatively constant over time. Our framework highlights the time- and space-varying roles of human mobility and is applicable to outbreaks of other infectious diseases and to other questions of relative epidemic drivers.

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

1 Introduction

Dengue is a vector-borne disease caused by the dengue virus (DENV) that has four distinct serotypes and is spread by *Aedes* mosquitoes, transmitted between humans primarily by *Ae. aegypti* and to a lesser extent, *Ae. albopictus*. Dengue is the most rapidly expanding mosquito-borne disease, approximately half of the world's population is currently at risk of infection [1,2], and the incidence and distribution are expanding [3]. Faced with the growing public health challenges associated with severe dengue, there has been an increasing effort to improve our understanding of the biological, human, and environmental influences on dengue epidemics around the world. Modelling is one way to provide such understanding, and a wide range of approaches have been used, including statistical, mechanistic (or mathematical), machine learning (ML), and deep learning frameworks. Given the climatic sensitivities across both the pathogen (DENV) and mosquito vector [4,5,6], a common research objective has been to quantify how transmission

dynamics have been shaped by recent climatic conditions and their interactions with human (e.g. urbanisation, human mobility, and socioeconomic factors[7,8,9]) and epidemiological influences (e.g. serotype replacement, serotype competition, and population immunity[10,11]).

In recent years, human movement has been identified as an important driver in dengue epidemic dynamics. In Pakistan, a mechanistic modelling study used mobile phone data and climate data to reveal that the inclusion of human mobility improved predictions of the spread and timing of dengue epidemics[12]. Mechanistic approaches have similarly been used in southern China and Singapore to model dengue spread within and across cities[13,14,15,16]. Statistical frameworks, such as Bayesian hierarchical models, have generally incorporated the effects of human mobility using either human movement matrices, distance-based approximations, or neighbourhood-based approximations[8,17]. In the city of Fortaleza in Brazil, both mobility-informed artificial neural networks and mechanistic models were employed at a fine spatial resolution (neighbourhood level). The inclusion of human mobility (bus transportation) data again yielded improvements in prediction accuracy[18]. A separate predictive modelling study[19] used data from the COVID-19 pandemic to identify human mobility as a crucial factor in shaping dengue epidemic dynamics in Fortaleza. At a longer-term horizon, results from Harish et al.[20], based on a hierarchical model of geographical expansion, indicated important synergistic roles for short- and long-distance human movement, alongside suitable environmental conditions, for the geographical expansion of dengue in Brazil and Mexico.

However, while the increasing availability of mobility data (in different guises) has enabled such modelling studies approaches (which extend beyond mobility approximations[21]), we still have a poor understanding of how the relative importance of human movement in dengue outbreaks varies, within and across seasons and across space. Such time-varying effects may be important, as genomic studies indicate that once dengue is established, reimportations remain important[22]. We consider here when, where, and how human movement shapes urban dengue outbreaks at the neighbourhood level in a city (Fortaleza, Brazil) where dengue is endemic for over 30 years. Research questions we ask are: In an urban context, how does the role of human mobility change through time? How does human mobility shape the epidemic dynamics once the outbreak is under way? How does the relative importance of human mobility change across years? How do neighbourhoods differentially impact the city-wide dynamics of dengue over time? We investigate these questions using mobility-informed statistical and deep learning models, alongside statistically principled multi-model development and new evaluation procedures. Our methodology for mobility-informed modelling allows us to infer the high-resolution time- and space-varying relative importance of human mobility for driving urban dengue outbreak dynamics, and quantify the impacts of human mobility on source-sink dynamics of neighbourhoods. Looking ahead, our framework can provide the insights necessary for public health authorities to design more targeted and optimally timed public health interventions, both for dengue and other infectious diseases.

2 Materials and methods

2.1 Study area and data

Fortaleza is a coastal city in north-east Brazil with a population of approximately 2.4 million people[23]. It is the fifth-largest city in Brazil and consists of 119 neighbourhoods/bairros (Figure 1A).

Dengue is endemic in Fortaleza. The first recorded dengue outbreak occurred in 1986 and was caused by the introduction of DENV-1. DENV-2 was first introduced into Fortaleza in 1992 , with the first major epidemic occurring in 1994 when approximately 660,000 dengue cases were reported in Fortaleza (approximately 84% of all cases in Brazil) driven in part by the suspension of mosquito control activities in late 1993 due to a cholera outbreak[24,25]. By 2012, all four serotypes of DENV had been detected in the city. Female *Ae. aegypti* are the primary vector of DENV transmission across Brazil. Suitable conditions, for the development of their larval habitat and DENV transmission, are provided by the highly urbanised Fortaleza neighbourhoods and the equatorial savannah climate (with distinct tropical rainy and dry seasons)[7,15,26].

To measure human mobility, we used previously published data for the public bus transportation system in Fortaleza; the most widely used mode of public transport with an estimated 700,00 people using the system every day[27,28,29]. The mobility data were available on a weekly basis for 2015. Using the coordinates of the origin and destination location, we obtained a directed measure of the number of flows ($f_w^{(i,j)}$) from neighbourhood i to neighbourhood j for week w. In the absence of further available data, we assumed the weekly mobility matrix (i.e the directed adjacency matrix) from 2015 applied to corresponding weeks in all other years (2007–2014). The mobility data were described extensively in previous work[18,27,28,29].

We sourced data on newly reported dengue cases from[18], which used reported cases from the Sistema de

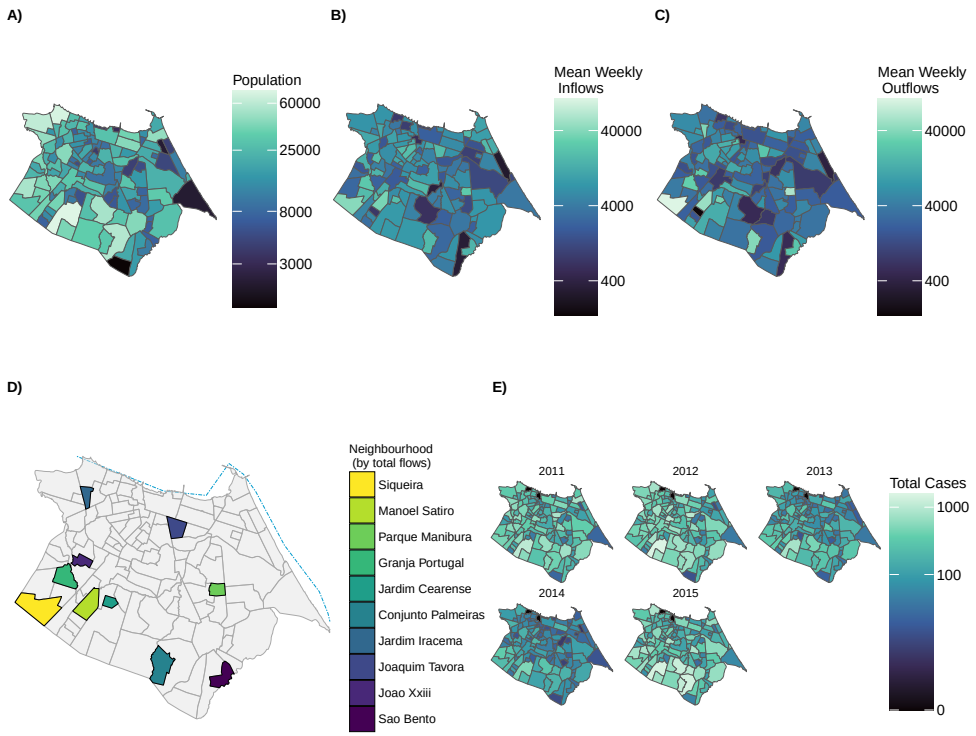

Figure 1: **Demography, human movement, and dengue cases in Fortaleza, Brazil:** Fortaleza is a coastal city in north-east Brazil, with a population of approximately 2.4 million people [23]. Within Fortaleza, there are 119 neighbourhoods (known as bairros), and all our data and analyses are at the neighbourhood level. **A)** The population (visualised on a logarithmic scale) of each of the 119 neighbourhoods, **B)** and **C)**: The mean weekly total number (visualised on a logarithmic scale) of inflows and outflows respectively for each of the 119 neighbourhoods. **D)** The neighbourhoods with the top 10 total number of inflows and outflows (visualised on a logarithmic scale) across all 52 weeks. Neighbourhoods are ordered in descending order from highest to lowest flows, and the blue line visualises the coastline of Fortaleza. **E)** The cumulative reported dengue cases (visualised on a logarithmic scale) in each neighbourhood per year of the testing period (2011–2015). The time series of cases within each year is visualised Figure 3A–B, while a summary of cases per week are visualised in Figure SI 3.

Monitoramento Diário de Agravos—SIMDA (Daily Disease Occurrence Monitoring System). These cases arise from diagnoses in hospitals or primary health care units. Note that dengue cases in Brazil can be confirmed solely on the basis of clinical and epidemiological criteria, without laboratory tests. The data were reported on a weekly basis (using diagnosis dates) for each of the 119 neighbourhoods according to the cases' home addresses.

For meteorological variables, we calculated daily mean temperature (at a 2 metre level) and daily total precipitation for Fortaleza using data from the ERA5-Land reanalysis dataset [30]. Due to data resolution, the same climatic conditions were used for each neighbourhood, and are generally characterised by year-round stable, suitable temperatures for dengue transmission and distinct rainy and dry seasons (Figure SI 1).

2.2 Model development and fitting periods

To guard against optimistic bias and overfitting, we used data from 2007 to 2010 inclusive as a model development period. During this period, we selected combinations of covariates, developed model structures, and made other model choices based on information criteria (see below).

We then focused on data from 2011 to 2015 inclusive to test model performance and draw insights. Rolling windows allowed us to estimate spatiotemporal variations in mobility effects without explicitly modelling the time variation for each observation, helping to guard against non-identifiability issues which could be caused by time-varying effects for each time point and neighbourhood.

After exploratory analyses (2007 to 2010), we chose a 17-week (approximately four-month) rolling window to allow sufficient learning of parametric relationships in our Bayesian hierarchical model, without losing excessive temporal resolution. The model learned about mobility influences within each rolling window, a procedure which also prevents

averaging over potential data distribution shifts (or equivalently, changing mobility-incidence-rate relationships). For our deep learning models, we fitted these models using all information up to a given time t, due to the need for larger volumes of data to estimate covariate effects (e.g., mobility influences). Unless otherwise stated, all evaluation occurs in the rolling windows.

2.3 Bayesian hierarchical models

We developed Bayesian hierarchical approaches to modelling y_{it}, the reported neighbourhood-level (i) weekly (t) case counts of dengue. We used a Negative Binomial (NegBin) likelihood to allow for potential overdispersion in the case data.

We incorporated the graph structures (from the weekly mobility data) using a lagged mobility-weighted sum of all other neighbourhoods' past dengue incidence rates (DIRs) per 100,000, which provides an estimated force of dengue importations. This approach was a computationally cheap (albeit limited, see Discussion) way to incorporate spatiotemporal structure that was informed by human mobility data. We used mobility flows and (reported) DIR from the previous week as inflow of infected individuals is less likely to result in new reported human cases within the reporting week (given the intermediate role of the mosquito vector, viral replication rates within the host and vector, and possible reporting delays).

We developed and fitted a range of candidate models (Supplementary Material A.2) in R version 4.4.0 using the cmdstanr interface to Stan, a platform for Bayesian inference[31,32,33]. Using the R package loo, we estimated the out-of-sample generalisability of each model using ELPD-LOO; leave-one-out (LOO) estimates of the expected log pointwise predictive density (ELPD)[34]. Based on results for the model development period, we selected the following candidate model which assumes a fixed mobility parameter (across space) and a space-varying autoregressive parameter:

$$y_{it} \sim NegBin(\mu_{it}, \phi), \tag{1}$$

$$\log(\mu_{it}) = \log(P_{it}) + \log(\eta_{it}), \tag{2}$$

$$\log(\eta_{it}) = \alpha + \beta_{\text{Mobility}} \cdot \left(\sum_{j \in \mathcal{N}(i)} \frac{\mathrm{f}_{t-1}^{(j,i)}}{\sum_{h \in \mathcal{N}(i)} \mathrm{f}_{t-1}^{(h,i)}} \cdot \log(\eta_{j,t-1}) \right) + \beta_{\text{AR[i]}} \cdot \log(\eta_{i,t-1}) + \beta_{\text{Year}[t]} + \beta_{\text{Nbr}[i]}, \tag{3}$$

$$\beta_{\text{Mobility}} \sim N(0,1), \tag{4}$$

$$\beta_{\text{AR}} \sim N(0,1), \tag{5}$$

$$\alpha \sim N(0,1), \tag{6}$$

$$\beta_{\text{Year}} \sim N(0,1), \tag{7}$$

$$\beta_{\text{Nbr}} \sim N(0,1), \tag{8}$$

$$\phi^{-1} \sim \text{Half-Normal}(0,1), \tag{9}$$

where μ_{it} denotes the mean parameter of the Negative Binomial distribution and ϕ is the reciprocal overdispersion parameter. We used a population offset $\log(P_{it})$ such that η_{it} represented the DIR per 100,000 population. We then modelled the (logarithm of) DIR with i) an intercept α, ii) an estimated force of dengue importations – a mobility effect controlled by fixed parameter β_{Mobility} and the lagged (logarithm of) DIR of all other neighbourhoods, $\mathcal{N}(i)$, weighted by the proportion of lagged inflows from such other neighbourhoods $\mathrm{f}_{t-1}^{(j,i)}$, iii) a potentially non-stationary and neighbourhood-specific autoregressive process of order one, controlled by $\beta_{\text{AR[i]}}$ (which allows for within-neighbourhood transmission caused by factors beyond mobility), iv) a year-to-year random effect (i.e. β_{Year} is a varying intercept for each year), and v) a neighbourhood-specific random effect (i.e. β_{Nbr} is a varying intercept for each neighbourhood). Modelling the effects on a DIR scale allowed us to disentangle each of these effects from population artefacts (e.g. large versus small neighbourhoods). Our goal was to understand the relative importance of human-mobility-driven transmission versus local transmission and the mobility-driven roles of different neighbourhoods in epidemic spread. We also explored explicitly accounting for climatic drivers (weekly total precipitation and average temperature), but the lack of neighbourhood-specific data meant that, causally, climatic covariates could not provide information for climate-driven local transmission dynamics (see Discussion).

A fixed parameter (within each window) for mobility does not assume that the effects of mobility (on each neighbourhood's DIR) are homogeneous across space and time, as i) both the human movement and human case data vary across space and time, and ii) the model can be fitted in rolling windows (which allows spatiotemporal variations in the autoregressive term and the levels at which inflows of infected individuals generate new infections).

2.3.1 Mobility effects across space and over time

Using this model (eqs (1)–(9)), we inferred parameter-based estimates of i) how human mobility varies in terms of its relative importance within and across neighbourhoods over time, and ii) how human mobility impacts the time-varying, relative contributions (i.e. source-sink dynamics) that neighbourhoods make to the epidemics of other neighbourhoods.

We computed the following ratio, $r_i(t)$, for each neighbourhood i at time t:

$$r_i(t) := \frac{\beta_{Mobility} \cdot \sum_{j \in \mathcal{N}(i)} \left(w_t^{(j,i)} \cdot \log(\eta_{j,t-1}) \right)}{\beta_{\mathrm{AR}[i]} \cdot \log(\eta_{i,t-1})}, \tag{10}$$

where $w_t^{(j,i)} = \frac{\mathrm{f}_{t-1}^{(j,i)}}{\sum_{h \in \mathcal{N}(i)} \mathrm{f}_{t-1}^{(h,i)}}$ is the weight assigned to neighbourhood i for the estimated inflows of infected travellers from neighbourhood j.

Eq. (10) measures the estimated time-varying relative importance of human mobility; contributions (to neighbourhood i's epidemic) from all other neighbourhoods, relative to its own autoregressive epidemic process, with values of $r_i(t)$ greater (less) than 1 indicating a greater (lower) relative importance of mobility-driven case importations. Using DIRs enabled comparison of the relative importance of human mobility across neighbourhoods of potentially different sizes. We compared how $r_i(t)$ differed across space (i) and over time (t), which helped us investigate the question of whether mobility's relative importance in driving dengue seasons is temporally homogeneous across space – i.e. does the importance follow similar patterns within and across seasons in different neighbourhoods?

We also estimated source-sink dynamics by deriving formulae for $\widetilde{C_j}$. This variable was standardised per window (across neighbourhoods) and reflects the relative contributions of a given neighbourhood j to the dengue epidemic dynamics in all other neighbourhoods. Values of $\widetilde{C_j}$ greater (less) than 0 indicate that a given neighbourhood has contributed more (less) to transmission elsewhere, relative to the mean contribution of neighbourhoods.

$$\begin{aligned}
C_j(t) &:= \sum_{i \in \mathcal{N}(j)} \beta_{Mobility} \cdot w_t^{(j,i)} \cdot \log(\eta_{j,t-1}), \\
\overline{C_j} &:= \frac{1}{T} \sum_{k=1}^{T} C_j(t_k), \quad \overline{C} := \frac{1}{U} \sum_{j=1}^{U} \overline{C_j}, \quad \sigma(\overline{C}) := \sqrt{\frac{1}{U} \sum_{j=1}^{U} (\overline{C_j} - \overline{C})^2}, \\
\widetilde{C_j} &:= \frac{\overline{C_j} - \overline{C}}{\sigma(\overline{C})},
\end{aligned} \tag{11}$$

where T is the length of the rolling window (i.e. the 17-week period to which the model was fitted), U is the number (119) of neighbourhoods, $|\mathcal{N}(j)| = U - 1$ is the number of neighbours in $\mathcal{N}(j)$, and $w_t^{(j,i)}$ is the weight assigned for the proportion of outflows that leave neighbourhood j to visit neighbourhood i. Then, by tracking $\widetilde{C_j}$ in each rolling window, we measured how the relative contribution of a single neighbourhood j (to all other neighbourhoods) varies over time. We standardised per window to quantify *relative* contributions (source/sink) of neighbourhoods during different stages of dengue seasons (i.e. accounting for different epidemic magnitudes in different windows).

We assessed the importance and roles of human mobility via these derived estimates (Eqs (10) – (11)), alongside information criteria (ELPD LOO) and out-of-sample predictive performance for one-step-ahead forecasts using the Weighted Interval Score (WIS)[35] (a proper scoring rule which balances sharpness and calibration, see below).

2.4 Data-driven, deep learning models

We also developed data-driven, deep learning models (denoted as *DL models* in visualisations). The objective was to draw more robust conclusions by leaning on different approaches to modelling the effects of human mobility.

We implemented two foundational time series models (with and without human mobility effects), both based on TimeGPT; a transformer-based, foundational time series model that was pre-trained on more than 100 billion data points from a range of domains[36,37]. In both models, we modelled the logarithm of incident cases y_{it} in neighbourhood i at time t as our response for two reasons; i) we constrained our model's predictions to be positive, and ii) we measured one-step-ahead probabilistic forecasts using a recommended logarithmic scale[38]. In only the mobility-informed deep learning model, we used a covariate of a mobility-weighted sum of other neighbourhoods' past cases $m_{i,t}$ (on a logarithmic scale), which entered via

$$m_{i,t} = \sum_{j \in \mathcal{N}(i)} \frac{\mathrm{f}_{t-1}^{(j,i)}}{\sum_{h \in \mathcal{N}(i)} \mathrm{f}_{t-1}^{(h,i)}} \cdot \log(y_{j,t-1}), \tag{12}$$

where $\mathcal{N}(i)$ and $\mathrm{f}_{t-1}^{(j,i)}$ are defined as before. In both models (with and without the inclusion of human mobility effects), we further included the neighbourhood as a factor variable (i.e. a spatial effect) and the neighbourhood's most recently observed cases $\log(y_{i,t-1})$ (on a logarithmic scale, i.e. similar to an autoregressive covariate). Unless otherwise stated, we use the mobility-informed model when discussing the deep learning model's results.

These foundational time series models facilitate probabilistic predictions (via quantile regression with conformal inference) in the form of predictive quantiles of the forecast/predictive distribution. Therefore, to ensure comparability when assessing models' probabilistic predictions, we used the WIS (via the scoringutils R package [39]), which is an approximation to the Continuous Ranked Probability Score (CRPS) for quantile-based forecasts, which we represented using 23 predictive quantiles, as in previous forecasting work [40,41].

A key challenge of any deep learning model is explainability [42]. To understand why our deep learning models made its *individual predictions*, we used SHAP (SHapley Additive exPlanations) values. These improve upon feature importance metrics which can only measure overall importance to the model and cannot measure the directional influences of each feature. In contrast, SHAP values, based on a concept from game theory, measure the contribution of each feature/covariate to the model's individual predictions [43].

We denote $\mathrm{SHAP}_i^{(AR)}(t)$ and $\mathrm{SHAP}_i^{(Mob)}(t)$ as the SHAP values (i.e. estimated contributions) of the autoregressive influence, $\log(y_{i,t-1})$ and mobility-weighted inflows, m_{it} to the model's forecast (of the logarithm of cases) for neighbourhood i and time t. We can then derive an estimate for the relative importance of human mobility that is analogous to $r_i(t)$ (Eq (10)) from the Bayesian hierarchical model:

$$rSHAP_i(t) := \frac{\mathrm{SHAP}_i^{(Mob)}(t) - \log(P_{it})}{\mathrm{SHAP}_i^{(AR)}(t) - \log(P_{it})}, \tag{13}$$

where P_{it} is defined as before as the population offset (population divided by 100,000) needed to convert to (the logarithm of) DIR scale.

This formulation, eq. (13), again, can measure the *estimated* relative importance of human mobility in individual neighbourhoods with potentially different population sizes. Of course, we caveat our approach with the causal limitation that SHAP values measure feature contributions to the models' individual forecast. There is no explainability metric that can guarantee a correct causal result. Although our multi-model, triangulation-like approach [44] may mitigate this causal limitation, recent research has demonstrated causal misalignments when using SHAP values for deep learning models in epidemiological settings (even when the models are well-calibrated and highly predictive) [45]. In our setting, a SHAP value does not measure the contribution to observed case data nor to the true latent infection process. Furthermore, unlike our Bayesian hierarchical models which are generative and more mechanistic, the SHAP and rSHAP values of our deep learning models are not constrained by even a semi-mechanistic structure. For example, the SHAP value for the mobility-weighted influence $\mathrm{SHAP}_i^{(Mob)}(t)$ may be positive when $m_{it} = 0$ and could even be negative (when indicating that the base forecast should be decreased when m_{it} takes a particular, e.g. low, value). It is also possible that another unmodelled factor beyond mobility drives the difference between the SHAP values.

2.5 Evidence synthesis

Using both modelling approaches (Bayesian and deep learning), we generated one-step-ahead probabilistic forecasts of dengue cases (on a logarithmic scale, as recommended for epidemic forecasts [38]), and measured out-of-sample predictive performance of the mobility-informed models versus i) their analogous models with the mobility effect omitted and ii) a naive, baseline forecasting model [41]. This one-step-ahead probability distribution is equivalent to the predictive likelihood. Improvements over reference models, as measured by relative WIS, rWIS (WIS for given model divided by WIS for an analogous model without mobility or baseline model), for different rolling time windows and neighbourhoods, allowed us to quantify the extent to which mobility-driven importations were important for generating the observed incidence data. For the Bayesian hierarchical model, we also measured improvements (upon the analogous model without mobility) using LOO estimates of the ELPD, thus measuring how much mobility effects improve (estimated) out-of-sample generalisability.

We interpret $r_i(t)$ and $rSHAP_i(t)$ from our Bayesian hierarchical model and deep learning model as proxies for the relative importance of human mobility over time in driving epidemics. We compared values for $r_i(t)$ and $rSHAP_i(t)$ using Pearson's [46] and Spearman's correlation [47], categorisation into binary and four-level categories, and t-tests with p-values (Bonferroni-adjusted [48] when comparing multiple time series across neighbourhoods). Drawing comparison between these quantities ensured that we did not over-inflate the insights from a single modelling approach about the relative importance of human movement.

To understand the likely mechanisms behind values for $r_i(t)$ and $rSHAP_i(t)$, we visualised relationships with factors such as epidemic stage (measured by the cumulative fraction of annual cases observed by a given week), year, and neighbourhood. We further developed simple generalised additive models (GAMs) for $r_i(t)$ and $rSHAP_i(t)$, consisting of a Gaussian likelihood with covariates of neighbourhood-specific smooth functions (thin-plate splines) for epidemic stage, annual temporal and neighbourhood-specific spatial random effects (to account for e.g. varying reporting rates), and the most recent observed cases (on a logarithmic) scale, $y_{i,t-1}$. We compared alternative GAM models using the minimised generalised cross-validation score (GCV)[49].

Due to its generative structure, the Bayesian hierarchical model permits estimates of how an individual neighbourhood j (via journeys of its infected residents/commuters) impact the epidemics in other neighbourhoods, as measured by $\widetilde{C_j}$. We quantified possible causes behind the values of $\widetilde{C_j}$ in different rolling windows using a similar procedure; visualisations and local contextualisation complemented by follow-up GAMs chosen based on their GCV. The candidate GAMs here are, by construction of $\widetilde{C_j}$, conceptually simpler as $\widetilde{C_j}$ is a standardised per-rolling-window quantity, and so temporal effects (for e.g. annual reporting rates) are likely less relevant.

After model fitting, using Pearson's correlation coefficients[46], we also performed a simple correlation analysis of the full DIR time series (2007–2015) for the 119 neighbourhoods to understand the spatial synchrony across dengue seasons. Finally, to better understand the structure of the mobility network, for each week, we used i) modularity from the Louvain algorithm[50] (a community detection algorithm that works by hierarchical clustering), ii) modularity from the Infomap algorithm[51] (a community detection algorithm for directed, weighted graphs), and iii) Shannon's entropy[52] of the mobility flow matrix. Higher values of modularity and lower values of entropy indicate strong clustering (i.e. low mixing) and high concentration of flows in particular locations.

3 Results

The 119 neighbourhoods of Fortaleza have different population sizes (ranging from 1,342 to 76,044, Figure 1A) and therefore exhibit different human movement patterns (as approximated by the bus transportation data, Figure 1B–D) and consequently, may have different daytime (or ambient) population sizes due to inflows or outflows caused by commuter journeys in the city. In general, pairwise total flows between neighbourhoods were similar throughout the year (Figure SI 2), and there was a strong relationship (Pearson's correlation coefficient $r = 0.84$, p-value < 0.0001) between a neighbourhoods' total inflows and total outflows.

Across the studied period (2007–2015), the most severe annual dengue seasons were observed in 2008, 2011, 2012, and 2015 (Figures 1E and SI 3B), with peaks occurring almost across all neighbourhoods at similar times (Figures SI 3B, 3A). We visualised (Figure SI 3A) and quantified this high spatial similarity of time series, reflective of epidemic synchrony, using pairwise Pearson's correlation coefficients[46] between individual neighbourhoods (Median $r = 0.74$, Q1 = 0.66, Q3 = 0.81).

On their own, these exploratory results suggest, but do not confirm, roles for human mobility in driving dengue epidemics in Fortaleza. It is plausible that the highly synchronous dengue seasons across pairs of neighbourhoods (Figure SI 3A) are caused by single initial seeding events (as infectious hosts move from one neighbourhood to another) and subsequent similar epidemic trajectories.

3.1 Mobility consistently drives urban dengue outbreaks

Across our statistical and deep learning models, the inclusion of the effects of an estimated force of dengue importations (via weighted estimates of inflows, see Materials and methods) almost consistently improved model performance, as measured by probabilistic forecast performance metrics (rWIS, Figure 2A-C) and information criteria (ELPD LOO, Figure 2D). We observed these trends when we compared mobility-informed models to our baseline model which propagates the most recent observed cases as the next (median) forecast, and to analogous models which omitted the effects of the estimated force of dengue importations (Figures 2D and 3G). These trends were generally consistent i) across the 119 neighbourhoods, including in boundary neighbourhoods along the south and west of the city where our restriction to city-level mobility data could induce boundary effects (Figure 2B) and ii) over time, including during both peaks and troughs of dengue cases (Figure 2A and C). The only locations where we observed any predictive improvement from using the baseline model versus the mobility-informed Bayesian model were in Jacarecanga and Cristo Redentor; coastal/beach neighbourhoods which reported zero total dengue cases (Figure 1E). We found largely similar outperformance for the mobility-informed deep learning model (in 116 of 119 locations, Figure 2B). As a further model validation, the more generative (or semi-mechanistic) mobility-informed Bayesian hierarchical model consistently outperformed (within and across all dengue seasons) the corresponding data-driven

deep learning model in terms of probabilistic predictive performance (WIS), suggesting greater out-of-sample model generalisability (Figure 1A and C).

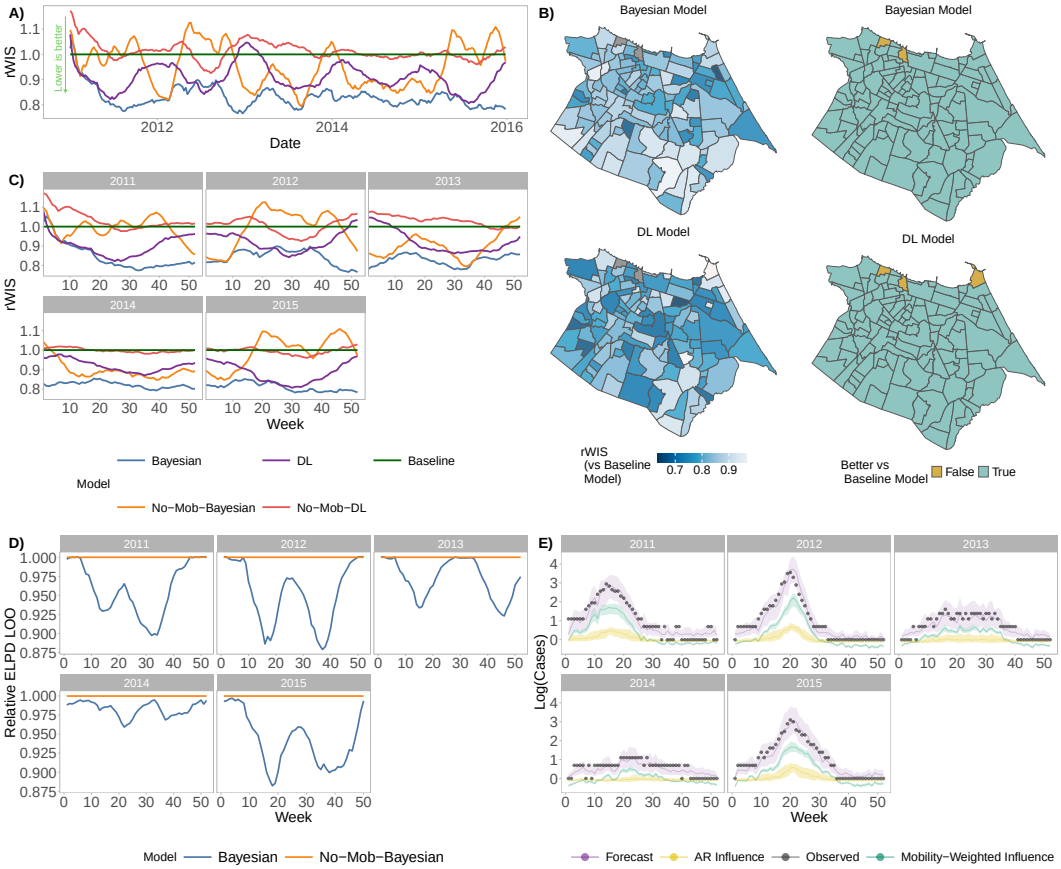

Figure 2: **Capturing human mobility effects improves model performance:** In approximate four-month trailing windows (17 weeks) of the testing period (2011-2015 inclusive), we visualise model performance for each of our four models across Bayesian hierarchical and deep learning (DL) approaches and compare performance to an epidemiologically naive baseline model (see Materials and methods). Any date shown represents the maximum date within the trailing window. **A)** Relative weighted interval score (rWIS) over time for each of the four models, where rWIS was calculated (relative to a baseline model, see Section 2.5) out of sample across all neighbourhoods. Lower values of the rWIS indicate superior predictive performance for out-of-sample case data (on a logarithmic scale). **B)** rWIS values for each of the 119 neighbourhoods across all trailing windows and a binary classification of whether the Bayesian/DL model outperformed the baseline (i.e. rWIS < 1). The grey shaded neighbourhoods on the north coastline are Jacarecanga and Cristo Redentor; neighbourhoods which reported zero total dengue cases across all seasons. **C)** Analogous to A), rWIS over time but using separate facets to capture model performance within each dengue season. **D)** Information criteria results for mobility-informed Bayesian hierarchical model. Higher ELPD LOO values indicate superior model generalisability. However, for visualisation of relative model performance over time (irrespective of the magnitude of observed data), we derived relative ELPD LOO values by dividing the ELPD LOO values for the Bayesian model by those for the analogous model without mobility (No-Mob-Bayesian). This was possible, as all ELPD LOO values were negative (across all windows and both models). Therefore lower values of this relative ELPD LOO indicate superior estimated model generalisability. **E)** Interquartile ranges of observed and forecasted cases (on a logarithmic scale) alongside the SHAP contributions from the neighbourhood-level autoregressive (AR) process and the mobility-weighted influence of the neighbours. Higher values of the SHAP values indicate greater contributions to the forecasted value.

3.2 Relative importance of human mobility over time to individual neighbourhoods

How the relative importance of human mobility changes throughout dengue outbreaks in urban settings remains poorly understood. We hinted above that human movement could play its most important role at the early stages of a dengue season (when seeding events can occur as a result of between-neighbourhood movement). Thereafter, human mobility could play less important and potentially negligible roles in sustaining transmission, once dengue is circulating widely in neighbourhoods of the city.

Using our Bayesian model, eqs (1)–(9), we assessed relative importance of human mobility for each neighbourhood

via our estimates of $r_i(t)$ – the ratio of (mobility-informed) case importations versus the neighbourhood's own autoregressive generation of new cases from past cases. Temporally, we estimated different trends in the relative importance of human mobility in different dengue seasons across the individual neighbourhoods (Figure 3A-C, Figure SI 5). While the rWIS values indicated spatially consistent (Figure 3G) and temporally consistent (Figure 2A,C, and D) outperformance of the mobility-informed model relative to the analogous model without human mobility, the estimates of $r_i(t)$ were spatially inconsistent within the annual dengue seasons (Figure SI 5C and Figure SI 6). As $r_i(t)$ varied substantially across space and time, we used both a binary classification (Figure 3C) and a four-level classification (Figure 3E) of $r_i(t)$ to further investigate the reasons/mechanisms behind variations in $r_i(t)$. Across all observations (all neighbourhoods and time points), the times of estimated greater relative importance of human mobility ($r_i(t)$) most frequently coincided with times near the midpoint (Figure 3D) or epidemic peak (Figure SI 5, SI 6) of an annual dengue season. Despite these general observations, there is no distinct neighbourhood-level evidence of such systematic pattern.

To test the robustness of model-based conclusions, we assessed the relationship between the Bayesian and deep learning models' findings. First, we observed the same spatial and temporal outperformance of the predictions for the mobility-informed deep learning models, relative to analogous models without the mobility effect (as per rWIS, see Figure 2A-C and Figure 3G). We also compared the Bayesian-model-derived estimates $r_i(t)$ with those from the SHAP values for the mobility-weighted influence, $\text{SHAP}_i^{(Mob)}(t)$, on the deep learning model's predictions. The SHAP values indicate i) greater relative importance of the mobility-weighted sum of neighbours' cases, compared to the neighbourhood's own autoregressive process of case generation, at all times where cases are greater than zero, and ii) also that the mobility-weighted sum scales more drastically with the annual epidemics than the autoregressive process (Figure 2E). This suggests a potential greater relative importance of human mobility (versus the autoregressive process) for the persistence of dengue in these urban environments.

Of course, $r_i(t)$ and the SHAP mobility influence estimates are not directly comparable, as the SHAP value scales with the magnitude of the data (e.g. Figure 2E), unlike our Bayesian hierarchical model (which models DIRs using population offsets). Therefore, we used the relative SHAP quantity, $r\text{SHAP}_i(t)$, computed as the ratio between the mobility-weighted influence (to DIR) and autoregressive influence SHAP values (see Materials and methods). We found that there was no one-to-one, linear relationship between $r_i(t)$ and $r\text{SHAP}_i(t)$ (Figure SI 5), perhaps due to the lack of a mechanistic or generative structure in our deep learning model which could constrain the range of plausible values for $r\text{SHAP}_i(t)$. However, times with greater relative importance of human mobility from the Bayesian model were consistently associated with higher SHAP values for the mobility effect (Figures 3H, SI 5, SI 6), particularly in years with larger dengue outbreaks. Using t-tests, we formally tested this relationship, between $r_i(t)$ and $r\text{SHAP}_i(t)$ across all observations and for each of the 119 neighbourhoods individually (see Figure 3F for Bonferroni-adjusted[48] p-values). For both models, we estimated similar trends for the timings of maximal importance of human mobility (middle stages of season, Figures 3E-F, SI 5, and SI 6).

The maximal relative importance of mobility was most often observed at the midpoint of the annual dengue season (Figure 3E-F), and the relative importance values generally tracked the seasonal epidemic curve (Figure SI 6), yet this estimated pattern varied across seasons (Figure SI 7) and across neighbourhoods (Figures SI 8–SI 9). Alongside the consistent high values of $r_i(t)$ (above 1), these results collectively suggest a potential longer-term trend of human-mobility-driven spread (in urban environments) being consistently more important than internal case generation across all stages of the dengue season, and likely most important in early-to-mid stages of the dengue season, both in more and less severe dengue seasons. Mechanistically, this implies that human-mobility-driven case importations/amplifications may account for sustained circulation of (reported) dengue cases, particularly at seasonal peaks (Figure SI 6).

To further investigate the plausibility of these findings, we explored the human movement patterns using various mixing metrics, such as modularity (from the Louvain and Infomap algorithms) and Shannon's entropy for the mobility matrix (Figure SI 10). We estimated consistently low modularity (and the Infomap algorithm did not detect any communities) which indicated a well-mixed mobility network in Fortaleza, at least at a neighbourhood level of analysis. Similarly, the entropy was high across all weeks (Median = 7.96, minimum = 7.87, and maximum = 8.06), relative to the maximum achievable of $\log(119^2) = 9.56$. This high entropy indicates more mixing and more evenly spread flows throughout Fortaleza. Collectively, the results indicate that high mixing may have generated the consistently high importance of human movement and the widespread epidemics.

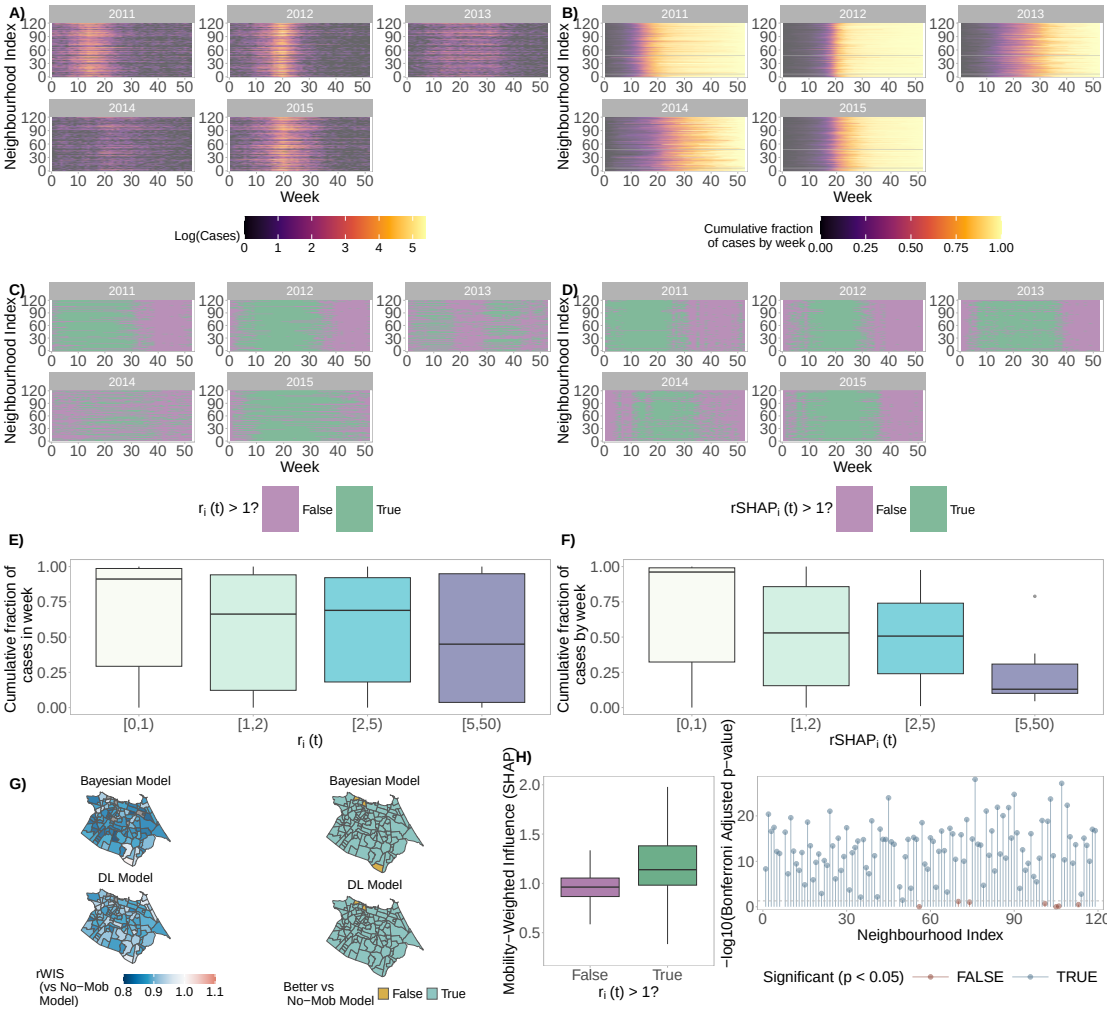

Figure 3: **The time-varying relative importance of human mobility:** Across 2011–2015 and the 119 neighbourhoods, we visualise the estimated relative importance of human-movement-based importations of dengue cases, versus the neighbourhoods' own internal generation of new dengue cases. **A)** and **B)** visualise the observed cases (on a logarithmic scale) and cumulative fraction of annual cases observed for individual weeks across 2011-2015. Neighbourhoods are ordered in ascending order by their total annual number of flows (inflows and outflows). **C)** and **D** show the estimated relative importance from our Bayesian hierarchical and deep learning models respectively, where values of $r_i(t)$ and $r\text{SHAP}_i(t)$ greater than one indicate greater relative importance of the case importations (see Materials and methods). Figure SI 5C-D is the analogous continuous version of these estimates of $r_i(t)$ and $r\text{SHAP}_i(t)$. **E)** and **F)** visualise the relationship between dengue season stage (measured by the cumulative fraction of cases) versus the values of $r_i(t)$ and $r\text{SHAP}_i(t)$. **E)** Across the 119 neighbourhoods, the relative predictive performance of the Bayesian and deep learning model is visualised, using the relative weighted interval score (rWIS) compared against the corresponding models without a mobility effect. **G)** and **H)** capture the relationship between the model-based relative importance estimates $r\text{SHAP}_i(t)$ and a binary version of $r_i(t)$ (= 1 if > 1) using a boxplot and visualisation of the Bonferroni-adjusted p-values[48] from a one-sided t-test of whether $r\text{SHAP}_i(t)$ were significantly greater when $r_i(t) > 1$. The box and whiskers of each boxplot cover the interquartile range.

3.3 Space-varying human-mobility-driven influences of neighbourhoods

Different neighbourhoods in a city can play different roles in daily life (e.g. serving as commuter hubs, retail epicentres, or residential areas), and therefore play different roles in seeding and/or sustaining the dengue epidemics of other neighbourhoods. We used our Bayesian hierarchical model, eqs (1)–(9), in rolling windows alongside the derived quantity $\widetilde{C_j}$, eq. (11), to investigate which neighbourhoods were key contributors (i.e. sources) to the dengue outbreaks in *other* neighbourhoods, and to determine whether these neighbourhoods played consistent source roles within each season and across all seasons. As $\widetilde{C_j}$ was standardised per window (see Materials and methods), we could compare the neighbourhoods' (relative) contributions at different stages of dengue seasons. This comparison of neighbourhoods' contributions over time differs from above, where we were concerned with the estimated time-varying relative importance of human mobility to each neighbourhood's dengue season, as here we are focused on source-sink dynamics. $\widetilde{C_j}$ does not represent a neighbourhood's overall importance for the public health burden in dengue, but rather the human-mobility-driven relative influence of neighbourhood j on the epidemic dynamics in all other neighbourhoods.

We found a strong, spatially consistent pattern of high-mobility neighbourhoods consistently being the greatest contributors, in rolling windows, to the epidemics of other neighbourhoods. This pattern was replicated across different stages of individual dengue seasons (Figure 4A), and across all dengue seasons (Figure 4B–C). There was generally little variability in the standardised contributions of individual neighbourhoods across the 258 rolling windows (Figures 4C and SI 11). Indeed, a GAM for median $\widetilde{C_j}$ values with a very simple Gaussian linear model structure, neighbourhood-specific random effect, and intercept (i.e. no temporal effects) explained 84% of the variation in $\widetilde{C_j}$ (R^2).

Spatially, while intuitive, it is not guaranteed that neighbourhoods with high mobility flows are consistently the greatest contributors. The contributions of generally high-mobility neighbourhoods could be influenced by i) weekly variability in pairwise mobility patterns and/or ii) the stage of the epidemic (as a high-mobility neighbourhood could, in some windows, have relatively few infectious hosts, relative to other neighbourhoods). In our case, both factors, i) and ii), are apparently not applicable as we found temporally homogeneous mobility patterns (Figure SI 2) and high epidemic synchrony in the city (Figures SI 3 and SI 4). These findings are limited by our dependence on bus transportation data only (see Discussion). For example, the neighbourhood of Siqueira had the greatest inflows (Figure 1B), outflows (Figure 1C), and estimated contributions to other neighbourhoods (Figure 4A–C), and is home to a bus terminal, which could inflate this neighbourhood's estimated influence on other neighbourhoods.

The only exception to the high relative influence of high-mobility neighbourhoods was during the "peak" of the low-incidence 2013 season, when their estimated relative contribution to other neighbourhoods was lower than lower-mobility neighbourhoods. This phenomenon may be partially explained by the highest cases during this period not being in the highest-mobility neighbourhoods (Figure SI 12) and the lower synchrony of the outbreaks in the low-incidence season of 2013 (Figure SI 4, Median Pairwise Pearson's Correlation Coefficient [46] $r = 0.36$, Q1 = 0.22, Q3 = 0.49), yet the high-mobility neighbourhoods did not have the lowest cases during this wave (Figure SI 12) and this spatial mobility influence trend was not replicated in the next lower-incidence season of 2014. This phenomenon could, of course, be a modelling limitation as the ELPD LOO indicates poorer generalisability, compared to a model without mobility, during this wave (Figure 2D)

4 Discussion

Human movement is widely understood as an important driver of many respiratory, vector-borne, waterborne, sexually transmitted and other infectious diseases [53,54,55], shaping the propagation of infectious diseases across space and time. However, for dengue epidemics in large cities, it was less clear how i) the relative importance of human mobility and ii) the human-mobility-driven source-sink dynamics of different neighbourhoods, vary within and across individual dengue seasons. Here, we provided multi-model methods to investigate these questions, illustrating our application in high spatial resolution for urban dengue outbreaks in Fortaleza, Brazil.

Understanding *when* human mobility is most important in an outbreak has practical consequences for public health policy. This can inform the optimal timing for public health interventions (e.g. vector control strategies) for endemic, epidemic, and pandemic pathogens. Understanding *where* human movement is seeding, expanding and/or sustaining outbreaks of other geographies can enable more targeted and cost-effective interventions. For example, in our study, while we estimated that human mobility was consistently important across neighbourhoods throughout a dengue season, we identified a high relative importance i) of human mobility at the middle stages and peaks of dengue seasons and ii) of particular neighbourhoods to geographic spread in each dengue season. Targeting vector control

interventions to reduce human-vector contact within such times and neighbourhoods could have an outsized, positive impact on bringing the epidemic under control.

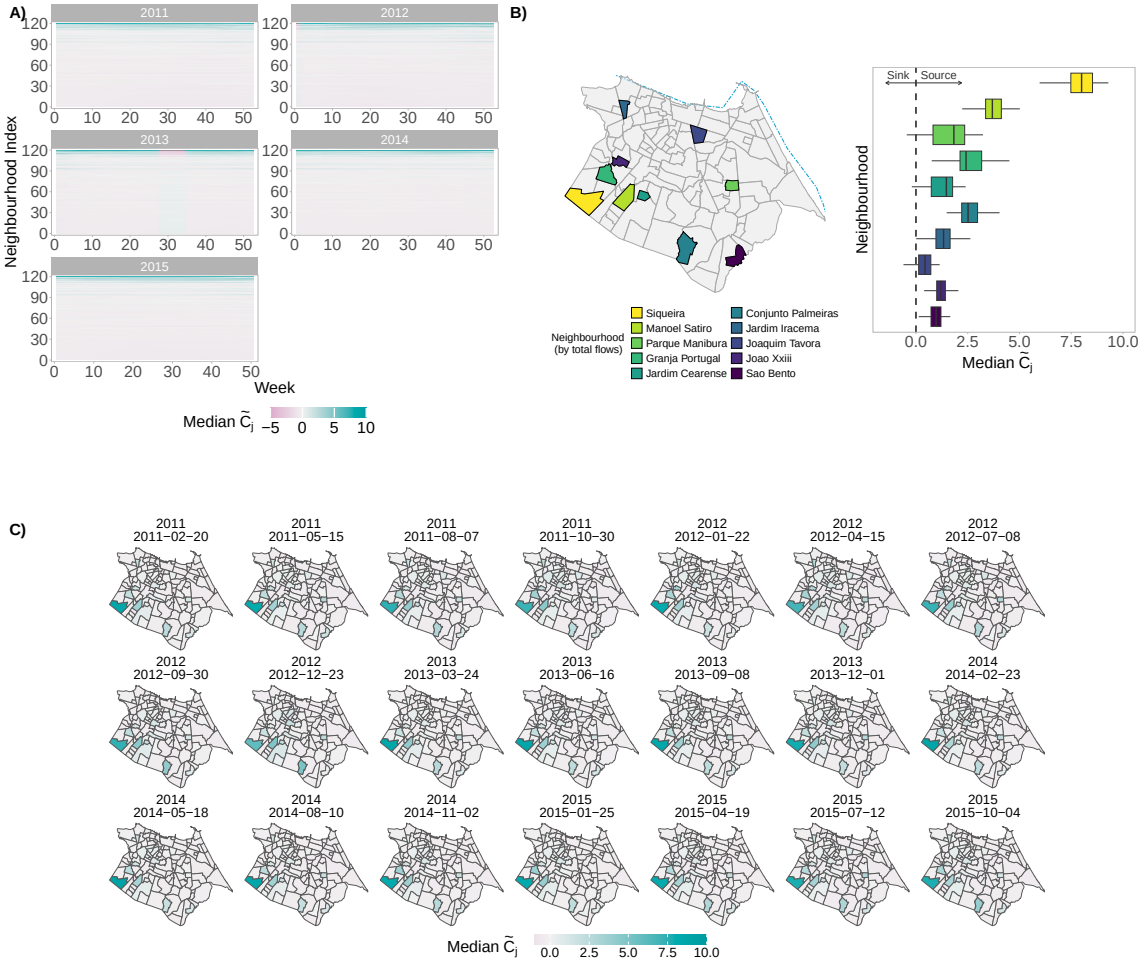

Figure 4: **Space-varying source-sink dynamics within and across dengue seasons:** Estimated relative (standardised) contributions of individual neighbourhoods to the epidemics of all other neighbourhoods, where estimates are based on the fits of our Bayesian hierarchical model, eqs (1)–(9), and the derived quantity $\widetilde{C_j}$, eq. (11). The model was fitted in rolling windows of length 17 weeks to allow for sufficient data for model training, while allowing for potentially time-varying effects. **A)** and **C)** are the median $\widetilde{C_j}$ for different neighbourhoods in different rolling windows across 2011–2015, where the week/date visualised represents the maximum date of the rolling window. Neighbourhoods in A) are ordered in ascending order from 1 to 119 by their total annual number of flows (inflows and outflows). Note that in **C)**, we visualised results for only every 12th rolling window (i.e. every 12 weeks) for legibility, and this results in different limits to the legend colours (compared to A). **B)** is a boxplot (right) for the neighbourhoods with the top ten total annual number of flows (left), where for each neighbourhood, all rolling window median $\widetilde{C_j}$ values are summarised. The box and whiskers cover the interquartile range.

Beyond public health interventions, such insights could also be used to inform design of human movement data collection strategies. Routine mobility data collection and sharing, particularly in low- and middle-income countries (LMICs), is challenging[21] (due to economic, political, and logistical barriers) and often infeasible, therefore suggesting that more cost-effective and targeted data collection is needed. Although our focus here has not been on forecasting (see[18,40,56,57,58]) nor invasion dynamics (see[20,59,60]), we have demonstrated consistently strong (out-of-sample) predictive ability of our models via the enhancement of spatiotemporal predictions of dengue incidence by using human mobility data. Therefore, incorporating the spatial structure induced by human mobility in invasion and forecasting models is likely important for more accurate forecasts that can be used for short-term outbreak response and longer-term seasonal planning.

Our results complement previous work using similar and other data streams to investigate other questions about the influences of human mobility on shaping the spread of infectious diseases. In addition to public transportation data, researchers have used human movement data from air and water travel[61], population censuses[62], and mobile phones[63,64] to investigate questions of disease spread on subnational, national, and global scales. Future work may

look at understanding how best to integrate these heterogeneous mobility data sources with our methods and analyses (potentially extending research questions to roles of long-distance human movements), as well as the net information gains (including corresponding costs and potential data biases) from such additional data collection. These net information gains will depend on the setting-specific research questions (e.g. early warning detection, nowcasting, or forecasting) and the time and region, yet quantifying the information gains is crucial for encouraging and guiding data collection. This could be a stepping stone towards addressing our aforementioned barriers caused by the lack of subnational human mobility data, particularly in resource-limited LMICs[21,65]. Understanding data information gains is also important as different model-based proxies[21] and mobility data streams[65] can induce biases in the modelling estimates that are ultimately used to inform public health authorities, although there are recent advances in the accuracy of mobility data proxies (e.g. gravity-like models[66]) and integrating heterogeneous mobility data streams[67].

The consistently high relative importance of human mobility (and peaks coinciding with seasonal epidemic peaks) and of specific source neighbourhoods makes intuitive sense within the context of a large, highly urbanised, well-mixed, and commuter-centric city such as Fortaleza. Mechanistically, once the seasonal urban dengue outbreak becomes established, the movement of humans and thus, provision of infectious hosts, apparently allows for the epidemic surges (i.e. expansion) and sustained spread across neighbourhoods. In previous work, arbovirus risk (for dengue, Zika, and chikungunya) in Fortaleza (albeit focused on 2015 onwards) had estimated spatial clusters of cases[68], including for clusters of neighbourhoods over longer distances (up to 6.8 km). The authors hypothesised that human mobility likely allowed infected individuals to regularly import viruses to distant areas in Fortaleza. Similarly, infestation surveys, which estimate *Ae. aegypti* density, frequently indicate a high level of vector adaptation across Fortaleza[24]. Interpreted alongside a consistent role for human movement, this suggests that, even in less suitable climatic conditions (e.g. dry season), the mosquito vector could be present near homogeneously across space and time. This example underlines the importance of understanding the history of dengue transmission when applying our framework and interpreting results in other settings. In recently affected cities (e.g. Buenos Aires, Argentina), the likely fluctuations in mosquito vector density (as the effects of climate change materialise[60,69]) could modulate mobility effects and yield very different influences of human movement on urban dengue epidemic dynamics.

Meanwhile, for COVID-19 Aguilar et al.[70] compared hierarchical cities with sprawled/well-mixed cities, and found that epidemic spread was less rapid in sprawled cities. Interpreting our results, albeit for a different infectious disease, alongside these findings, we suspect that the high mixing across the city (Figure SI 10) may have supported the relatively long dengue seasons frequently observed in Fortaleza (Figure SI 3B). This may explain why we estimated a high relative importance of mobility in multiple individual dengue seasons. Further work could investigate whether these results generalise in other environments – different mobility patterns, different spatial scales of cities, and different socio-environmental settings.

Our findings align with research that documented consistently superior overall predictive performance of mobility-informed models in Fortaleza[18,19], other cities (e.g. 12 cities in Guangzhou province, China[13]), and at national levels (e.g. in Pakistan[12] and Vietnam[8]). Our research also builds on such research by disentangling the relative importance and influences of human movement for different stages of the epidemic and for different neighbourhoods. For future studies, our unitless measures of the time-varying relative importance of human mobility ($r_i(t)$, eq. (10), and $r\text{SHAP}_i(t)$, eq. (13)) and of different neighbourhoods ($\widetilde{C_j}$, eq. (11)) could be useful. These could be paired with new formulae for other unitless measures of epidemic properties such as time-varying reproduction numbers $R(t)$ (e.g. for dengue[71,72]) to further quantify how the relative importance of human mobility and the influence of different neighbourhoods evolve with varying epidemic strength (as measured by $R(t)$, the number of new infections generated by a typical individual, see[73]).

The models that we develop are new for modelling the time- and space-varying roles and effects of human movement on dengue epidemics, yet the models remain conceptually simple. The key advance is the ability to derive mathematical quantities (together with robust uncertainties) for the importance and influence of human mobility from a generative, Bayesian model, and to blend the results for these quantities with those using our derived formulae for game-theoretic quantities (i.e. SHAP and rSHAP values). This multi-model procedure for extracting spatiotemporal insights (by testing theories and synthesising evidence across disciplines) enhances robustness of causal interpretations[44], and our conceptual advances are generalisable to other settings (e.g. other infectious diseases) and other research questions of relative importance (e.g. climatic versus socioenvironmental influences).

Our analysis has several data limitations and there are many future avenues of research. First, we used weekly bus transportation data from 2015 to represent movements for all years. Although there is potential for representation bias from any form of mobility data[65] and we advocate for greater availability of more regular data, previous analyses have shown that these data are representative for a wide range of explanatory and predictive research questions (e.g. crime rates, dengue epidemics and police allocation)[18,27,28,29]. Similarly, there were no substantial time variations in flows

throughout the weeks of 2015 (Figure SI 2) and our mobility-informed models consistently enhanced performance across information and predictive performance criteria, hinting at the generalisability of these data (beyond 2015) and at the potential value of intermittent collection and sharing of such data. However, the aggregated bus data only capture shorter-distance movements (as opposed to importations from inward travel to the city) are not linked to demographic or social identifiers, nor specific disease status, and future work could explore the additional value or differences in results from integrating other individual-level human movement indicators (e.g. mobile phone data) and long-distance movement indicators (e.g. flight data). Such data sources may also have greater variability throughout the dengue season and provide valuable additional information for understanding the dynamics of invasion and geographical spread. Our models did not explicitly account for within-neighbourhood mobility, as bus transportation data were more suitable (within-neighbourhood bus trips were rare) for our studied research question – comparing imported cases with those generated within neighbourhoods, which means that within-neighbourhood cases can therefore account for human, biological, and environmental effects. Future work, likely relying on shorter-range movement data (compared to bus transportation data), could explore the relative importance of within-neighbourhood mobility over space and time. This would be instructive for better understanding of local community transmission in large informal settlements, which could be supported by the movement of infectious individuals on foot or motorcycles, coupled with the spread of mosquito breeding sites.

Further data limitations include our usage of observed dengue case data which are an imperfect measure of true dengue infections. The case data may be sensitive to reporting delays and/or underreporting in time or space (e.g. due to health centre availability or reporting capacity/propensity), which we attempted to proxy by using neighbourhood-specific random effects. These issues may also be exacerbated by the high spatial resolution of our study, yet this is balanced against the benefits of i) avoiding potential spatial aggregation artefacts/effects and ii) providing high-resolution insights for targeted policy questions.

Meanwhile, we did not have access to serotype or seroprevalence data. This may limit the robustness of our illustrative application as all four DENV serotypes have circulated in Fortaleza and 2011 and 2012 outbreaks were caused by different serotypes (the reintroduction and introduction of DENV-1 and DENV-4 respectively), which was the first time that a major Brazilian city experienced a serotype switch in two consecutive annual dengue epidemics. In 2015, DENV-1 was dominant once again. Overall, our results are likely not confounded by more recent introductions of Zika, and chikungunya viruses (which would induce misclassifications, yet these were almost simultaneously first reported in 2015)[68]. We also used only the one-week-lagged cases and did not model the mosquito population directly, and therefore we only approximated the true causal dynamical processes (e.g. biological mechanisms) of DENV transmission. Further causal limitations include the limitations of explainability metrics (such as SHAP values) for deep learning models when disentangling causal effects[42,45] (see Materials and methods). In future work, contingent on the availability of new data streams (seroprevalence, genomic, higher spatial resolution climate, or mosquito catch data), researchers could consider alternative networked mechanistic (or biomathematical) and/or deep learning models. These models could enhance the causal realism with explicit time- and space-varying properties (e.g. generation time and serial interval distributions[71,72]). Similarly, models could incorporate additional human influences (e.g. population immunity[10], access to clean drinking water[7,74] and circulating serotypes[8,11]), climatic drivers (e.g. interactions with precipitation, drought, humidity, or temperature[7,75]) and/or mosquito factors (e.g. mosquito density[6,76]).

In conclusion, our theoretical advances (evidence synthesis across multiple models) and findings (influences of human movement) can be a foundation for i) future studies that improve our understanding of the complex processes of dengue epidemics in different settings and ii) future public health policy that targets limited resources for dengue epidemic control.

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] World Health Organization. Disease Outbreak News; Dengue – Global situation. Technical report, May 2024. URL https://www.who.int/emergencies/disease-outbreak-news/item/2024-DON518.

[3] Sadie J. Ryan, Colin J. Carlson, Erin A. Mordecai, and Leah R. Johnson. Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. *PLoS Neglected Tropical Diseases*, 13(3):e0007213, March 2019. ISSN 1935-2727. doi: 10.1371/journal.pntd.0007213. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6438455/.

[4] Miranda Chan and Michael A. Johansson. The incubation periods of dengue viruses. *PLOS ONE*, 7(11):e50972, November 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0050972. URL https://dx.plos.org/10.1371/journal.pone.0050972.

[5] Erin A. Mordecai, Krijn P. Paaijmans, Leah R. Johnson, Christian Balzer, Tal Ben-Horin, Emily De Moor, Amy McNally, Samraat Pawar, Sadie J. Ryan, Thomas C. Smith, and Kevin D. Lafferty. Optimal temperature for malaria transmission is dramatically lower than previously predicted. *Ecology Letters*, 16(1):22–30, January 2013. ISSN 1461-023X, 1461-0248. doi: 10.1111/ele.12015. URL https://onlinelibrary.wiley.com/doi/10.1111/ele.12015.

[6] John H. Huber, Marissa L. Childs, Jamie M. Caldwell, and Erin A. Mordecai. Seasonal temperature variation influences climate suitability for dengue, chikungunya, and Zika transmission. *PLOS Neglected Tropical Diseases*, 12(5):e0006451, May 2018. ISSN 1935-2735. doi: 10.1371/journal.pntd.0006451. URL https://dx.plos.org/10.1371/journal.pntd.0006451.

[7] 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. doi: 10.1016/S2542-5196(20)30292-8. URL https://linkinghub.elsevier.com/retrieve/pii/S2542519620302928.

[8] 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. doi: 10.1038/s41467-023-43954-0. URL https://www.nature.com/articles/s41467-023-43954-0.

[9] M. U. G. Kraemer, T. A. Perkins, D. a. T. Cummings, R. Zakar, S. I. Hay, D. L. Smith, and R. C. Reiner. Big city, small world: density, contact rates, and transmission of dengue across Pakistan. *Journal of The Royal Society Interface*, 12(111):20150468, October 2015. doi: 10.1098/rsif.2015.0468. URL https://royalsocietypublishing.org/doi/full/10.1098/rsif.2015.0468. Publisher: Royal Society.

[10] 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.

[11] Emilie Finch, Adam Kucharski, Shuzhen Sim, Lee-Ching Ng, and Rachel Lowe. Climate variation and serotype competition drive dengue outbreak dynamics in Singapore, September 2024. URL http://medrxiv.org/lookup/doi/10.1101/2024.09.17.24313793.

[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. doi: 10.1073/pnas.1504964112. URL https://pnas.org/doi/full/10.1073/pnas.1504964112.

[13] Qinghui Zeng, Xiaolin Yu, Haobo Ni, Lina Xiao, Ting Xu, Haisheng Wu, Yuliang Chen, Hui Deng, Yingtao Zhang, Sen Pei, Jianpeng Xiao, and Pi Guo. Dengue transmission dynamics prediction by combining metapopulation networks and Kalman filter algorithm. *PLOS Neglected Tropical Diseases*, 17(6):e0011418, June 2023. ISSN 1935-2735. doi: 10.1371/journal.pntd.0011418. URL https://dx.plos.org/10.1371/journal.pntd.0011418.

[14] Guanghu Zhu, Tao Liu, Jianpeng Xiao, Bing Zhang, Tie Song, Yonghui Zhang, Lifeng Lin, Zhiqiang Peng, Aiping Deng, Wenjun Ma, and Yuantao Hao. Effects of human mobility, temperature and mosquito control on the spatiotemporal transmission of dengue. *Science of The Total Environment*, 651:969–978, February 2019. ISSN

00489697. doi: 10.1016/j.scitotenv.2018.09.182. URL https://linkinghub.elsevier.com/retrieve/pii/S004896971833626X.

[15] Emanuele Massaro, Daniel Kondor, and Carlo Ratti. Assessing the interplay between human mobility and mosquito borne diseases in urban environments. *Scientific Reports*, 9(1):16911, November 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-53127-z. URL https://www.nature.com/articles/s41598-019-53127-z.

[16] Mathew V. Kiang, Mauricio Santillana, Jarvis T. Chen, Jukka-Pekka Onnela, Nancy Krieger, Kenth Engø-Monsen, Nattwut Ekapirat, Darin Areechokchai, Preecha Prempree, Richard J. Maude, and Caroline O. Buckee. Incorporating human mobility data improves forecasts of Dengue fever in Thailand. *Scientific Reports*, 11(1): 923, January 2021. ISSN 2045-2322. doi: 10.1038/s41598-020-79438-0. URL https://www.nature.com/articles/s41598-020-79438-0.

[17] Sophie A. Lee, Christopher I. Jarvis, W. John Edmunds, Theodoros Economou, and Rachel Lowe. Spatial connectivity in mosquito-borne disease models: a systematic review of methods and assumptions. *Journal of The Royal Society Interface*, 18(178):20210096, May 2021. doi: 10.1098/rsif.2021.0096. URL https://royalsocietypublishing.org/doi/full/10.1098/rsif.2021.0096. Publisher: Royal Society.

[18] Rafael Bomfim, Sen Pei, Jeffrey Shaman, Teresa Yamana, Hernán A. Makse, José S. Andrade, Antonio S. Lima Neto, and Vasco Furtado. Predicting dengue outbreaks at neighbourhood level using human mobility in urban areas. *Journal of The Royal Society Interface*, 17(171):20200691, October 2020. ISSN 1742-5689, 1742-5662. doi: 10.1098/rsif.2020.0691. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2020.0691.

[19] Jorge L. B. Araújo, Rafael Bomfim, Cesar I. N. Sampaio Filho, Luciano P. G. Cavalcanti, Antonio S. Lima Neto, José S. Andrade, and Vasco Furtado. The impact of COVID-19 mobility restrictions on dengue transmission in urban areas. *PLOS Neglected Tropical Diseases*, 18(11):e0012644, November 2024. ISSN 1935-2727. doi: 10.1371/journal.pntd.0012644. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11627415/.

[20] Vinyas Harish, Felipe J. Colón-González, Filipe R. R. Moreira, Rory Gibb, Moritz U. G. Kraemer, Megan Davis, Robert C. Reiner, David M. Pigott, T. Alex Perkins, Daniel J. Weiss, Isaac I. Bogoch, Gonzalo Vazquez-Prokopec, Pablo Manrique Saide, Gerson L. Barbosa, Ester C. Sabino, Kamran Khan, Nuno R. Faria, Simon I. Hay, Fabián Correa-Morales, Francisco Chiaravalloti-Neto, and Oliver J. Brady. Human movement and environmental barriers shape the emergence of dengue. *Nature Communications*, 15(1):4205, May 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-48465-0. URL https://www.nature.com/articles/s41467-024-48465-0. Publisher: Nature Publishing Group.

[21] Jack Wardle, Sangeeta Bhatia, Moritz U.G. Kraemer, Pierre Nouvellet, and Anne Cori. Gaps in mobility data and implications for modelling epidemic spread: A scoping review and simulation study. *Epidemics*, 42:100666, March 2023. ISSN 17554365. doi: 10.1016/j.epidem.2023.100666. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436523000026.

[22] Henrik Salje, Justin Lessler, Irina Maljkovic Berry, Melanie C. Melendrez, Timothy Endy, Siripen Kalayanarooj, Atchareeya A-Nuegoonpipat, Sumalee Chanama, Somchai Sangkijporn, Chonticha Klungthong, Butsaya Thaisomboonsuk, Ananda Nisalak, Robert V. Gibbons, Sopon Iamsirithaworn, Louis R. Macareo, In-Kyu Yoon, Areerat Sangarsang, Richard G. Jarman, and Derek A. T. Cummings. Dengue diversity across spatial and temporal scales: Local structure and the effect of host population size. *Science*, 355(6331): 1302–1306, March 2017. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaj9384. URL https://www.science.org/doi/10.1126/science.aaj9384.

[23] Brazilian Institute of Geography and Statistics. 2022 Brazilian Census, 2022. URL https://censo2022.ibge.gov.br/panorama/.

[24] Benjamin MacCormack-Gelles, Antonio S. Lima Neto, Geziel S. Sousa, Osmar J. Nascimento, Marcia M. T. Machado, Mary E. Wilson, and Marcia C. Castro. Epidemiological characteristics and determinants of dengue transmission during epidemic and non-epidemic years in Fortaleza, Brazil: 2011-2015. *PLOS Neglected Tropical Diseases*, 12(12):e0006990, December 2018. ISSN 1935-2735. doi: 10.1371/journal.pntd.0006990. URL https://dx.plos.org/10.1371/journal.pntd.0006990.

[25] Daniela Cristina Sensato Monteiro, Natália Vasconcelos De Souza, Jane Cavalcante Amaral, Kaynan Bezerra De Lima, Fernanda Montenegro Carvalho De Araújo, Izabel Letícia Cavalcante Ramalho, Victor Emanuel Pessoa Martins, Jeová Keny Baima Colares, Luciano Pamplona De Góes Cavalcanti, and Danielle Malta Lima. Dengue: 30 years of cases in an endemic area. *Clinics*, 74:e675, 2019. ISSN 18075932. doi: 10.6061/clinics/2019/e675. URL https://linkinghub.elsevier.com/retrieve/pii/S1807593222006342.

[26] Oliver J. Brady and Simon I. Hay. The Global Expansion of Dengue: How *Aedes aegypti* Mosquitoes Enabled the First Pandemic Arbovirus. *Annual Review of Entomology*, 65(1):191–208, January 2020. ISSN 0066-4170, 1545-4487. doi: 10.1146/annurev-ento-011019-024918. URL https://www.annualreviews.org/doi/10.1146/annurev-ento-011019-024918.

[27] Carlos Caminha, Vasco Furtado, Vladia Pinheiro, and Caio Silva. Micro-interventions in urban transportation from pattern discovery on the flow of passengers and on the bus network. In *2016 IEEE International Smart Cities Conference (ISC2)*, pages 1–6, Trento, Italy, September 2016. IEEE. ISBN 978-1-5090-1846-8. doi: 10.1109/ISC2.2016.7580776. URL http://ieeexplore.ieee.org/document/7580776/.

[28] Carlos Caminha and Vasco Furtado. Impact of human mobility on police allocation. In *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 125–127, Beijing, China, July 2017. IEEE. ISBN 978-1-5090-6727-5. doi: 10.1109/ISI.2017.8004886. URL http://ieeexplore.ieee.org/document/8004886/.

[29] Carlos Caminha, Vasco Furtado, Tarcisio H. C. Pequeno, Caio Ponte, Hygor P. M. Melo, Erneson A. Oliveira, and José S. Andrade. Human mobility in large cities as a proxy for crime. *PLOS ONE*, 12(2):e0171609, February 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0171609. URL https://dx.plos.org/10.1371/journal.pone.0171609.

[30] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia De Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, July 2020. ISSN 0035-9009, 1477-870X. doi: 10.1002/qj.3803. URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3803.

[31] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2025. URL https://www.R-project.org/.

[32] Stan. RStan: the R interface to Stan, 2023. URL https://mc-stan.org/.

[33] Jonah Gabry, Rok Češnovar, Andrew Johnson, and Steve Bronder. *cmdstanr: R Interface to 'CmdStan'*. 2024. URL https://mc-stan.org/cmdstanr/.

[34] Aki Vehtari, Jonah Gabry, Mans Magnusson, Yuling Yao, Paul-Christian Bürkner, Topi Paananen, and Andrew Gelman. loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models, 2024. URL https://mc-stan.org/loo/.

[35] 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. doi: 10.1371/journal.pcbi.1008618. URL https://dx.plos.org/10.1371/journal.pcbi.1008618.

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

[37] 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].

[38] 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. doi: 10.1371/journal.pcbi.1011393. URL https://dx.plos.org/10.1371/journal.pcbi.1011393.

[39] Nikos I. Bosse, Hugo Gruson, Anne Cori, Edwin van Leeuwen, Sebastian Funk, and Sam Abbott. Evaluating Forecasts with scoringutils in R, November 2024. URL http://arxiv.org/abs/2205.07090. arXiv:2205.07090 [stat].

[40] Cathal Mills, Moritz U.G. Kraemer, and Christl A. Donnelly. Interdisciplinary modelling and forecasting of dengue, October 2024. URL http://medrxiv.org/lookup/doi/10.1101/2024.10.18.24315690.

[41] 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. doi: 10.1073/pnas.2113561119.

[42] Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, and Francisco Herrera. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, 99:101805, November 2023. ISSN 15662535. doi: 10.1016/j.inffus.2023.101805. URL https://linkinghub.elsevier.com/retrieve/pii/S1566253523001148.

[43] Scott Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions, November 2017. URL http://arxiv.org/abs/1705.07874. arXiv:1705.07874 [cs].

[44] Gemma Hammerton and Marcus R. Munafò. Causal inference with observational data: the need for triangulation of evidence. *Psychological Medicine*, 51(4):563–578, March 2021. ISSN 0033-2917. doi: 10.1017/S0033291720005127. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8020490/.

[45] Helen Coupland, Neil Scheidwasser, Alexandros Katsiferis, Megan Davies, Seth Flaxman, Naja Hulvej Rod, Swapnil Mishra, Samir Bhatt, and H. Juliette T. Unwin. Exploring the Potential and Limitations of Deep Learning and Explainable AI for Longitudinal Life Course Analysis, March 2025. URL https://www.researchsquare.com/article/rs-5582136/v1. ISSN: 2693-5015.

[46] Karl Pearson. Note on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London Series I*, 58:240–242, January 1895. URL https://ui.adsabs.harvard.edu/abs/1895RSPS...58..240P. ADS Bibcode: 1895RSPS...58..240P.

[47] C. Spearman. The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, 15(1):72, January 1904. ISSN 00029556. doi: 10.2307/1412159. URL https://www.jstor.org/stable/1412159?origin=crossref.

[48] Olive Jean Dunn. Multiple Comparisons among Means. *Journal of the American Statistical Association*, 56(293):52–64, March 1961. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.1961.10482090. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1961.10482090.

[49] 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. doi: 10.1007/BF01404567. URL http://link.springer.com/10.1007/BF01404567.

[50] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, October 2008. ISSN 1742-5468. doi: 10.1088/1742-5468/2008/10/P10008. URL http://arxiv.org/abs/0803.0476. arXiv:0803.0476 [physics].

[51] M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, January 2008. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0706851105. URL http://arxiv.org/abs/0707.0609. arXiv:0707.0609 [physics].

[52] C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, July 1948. ISSN 00058580. doi: 10.1002/j.1538-7305.1948.tb01338.x. URL https://ieeexplore.ieee.org/document/6773024.

[53] M. Naser Lessani, Zhenlong Li, Fengrui Jing, Shan Qiao, Jiajia Zhang, Bankole Olatosi, and Xiaoming Li. Human mobility and the infectious disease transmission: a systematic review. *Geo-spatial Information Science*, 27(6):1824–1851, November 2024. ISSN 1009-5020, 1993-5153. doi: 10.1080/10095020.2023.2275619. URL https://www.tandfonline.com/doi/full/10.1080/10095020.2023.2275619.

[54] M. U. G. Kraemer, N. Golding, D. Bisanzio, S. Bhatt, D. M. Pigott, S. E. Ray, O. J. Brady, J. S. Brownstein, N. R. Faria, D. A. T. Cummings, O. G. Pybus, D. L. Smith, A. J. Tatem, S. I. Hay, and R. C. Reiner. Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. *Scientific Reports*, 9(1):5151, March 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-41192-3. URL https://www.nature.com/articles/s41598-019-41192-3.

[55] Suttikiat Changruenngam, Dominique J. Bicout, and Charin Modchang. How the individual human mobility spatio-temporally shapes the disease transmission dynamics. *Scientific Reports*, 10(1):11325, July 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-68230-9. URL https://www.nature.com/articles/s41598-020-68230-9.

[56] 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. doi: 10.1371/journal.pmed.1003542. URL https://dx.plos.org/10.1371/journal.pmed.1003542.

[57] 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. doi: 10.1038/s41598-024-52796-9. URL https://www.nature.com/articles/s41598-024-52796-9.

[58] Eduardo Correa Araujo, Luiz Max Carvalho, Fabiana Ganem, Luã Bida Vacaro, Leonardo S. Bastos, Laís Picinini Freitas, Marcio Bastos, Ramila Alencar, Lucas Bianchi, Raúl Capellán, Xiang Chen, Oswaldo Cruz, Americo Cunha, Haridas K. Das, Chloe Fletcher, Raquel Martins Lana, Rachel Lowe, Daniela Lührsen, Giovenale Moirano, Paula Moraga, Lucas M. Stolerman, Fernanda Valente, Cláudia Torres Codeço, and Flávio C. Coelho. Leveraging probabilistic forecasts for dengue preparedness and control: the 2024 Dengue Forecasting Sprint in Brazil, May 2025. URL http://medrxiv.org/lookup/doi/10.1101/2025.05.12.25327419.

[59] Thales Bermann, Tatiana Schaffer Gregianini, Fernanda Marques Godinho, Ludmila Fiorenzano Baethgen, Fernanda Letícia Martiny, Erica Bortoli Möllmann, Amanda Pellenz Ruivo, Franciellen Machado dos Santos, Caroline do Nascimento Ferreira, Gustavo Barbosa de Lima, Danielly Alves Mendes Barbosa, Filipe Zimmer Dezordi, Milena Bauermann, Taina Machado Selayaran, Elgion Lucio da Silva Loreto, Felipe Gomes Naveca, André Luiz Sá de Oliveira, Ana Beatriz Gorini da Veiga, Marcelo Henrique Santos Paiva, de Lima Tulio Campos, Gonzalo Bello, Matheus Filgueira Bezerra, Felipe Valter de Oliveira, Rafael Maciel-de Freitas, Tiago Gräf, Richard Steiner Salvato, and Gabriel Luz Wallau. Increasing dengue outbreaks in temperate Brazil is linked to Aedes aegypti invasion and infestation level driving widespread virus transmission, February 2025. URL https://www.medrxiv.org/content/10.1101/2025.02.24.25322773v1. Pages: 2025.02.24.25322773.

[60] Claudia T. Codeco, Sara S. Oliveira, Danielle A.C. Ferreira, Thais I.S. Riback, Leonardo S. Bastos, Raquel M. Lana, Iasmim F. Almeida, Vinicius B. Godinho, Oswaldo G. Cruz, and Flavio C. Coelho. Fast expansion of dengue in Brazil. *The Lancet Regional Health - Americas*, 12:100274, August 2022. ISSN 2667193X. doi: 10.1016/j.lana.2022.100274. URL https://linkinghub.elsevier.com/retrieve/pii/S2667193X22000916.

[61] Ilaria Dorigatti, Arran Hamlet, Ricardo Aguas, Lorenzo Cattarino, Anne Cori, Christl A Donnelly, Tini Garske, Natsuko Imai, and Neil M Ferguson. International risk of yellow fever spread from the ongoing outbreak in Brazil, December 2016 to May 2017. *Eurosurveillance*, 22(28):30572, July 2017. ISSN 1025-496X. doi: 10.2807/1560-7917.ES.2017.22.28.30572. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5545764/.

[62] A. J. Tatem. Mapping population and pathogen movements. *International Health*, 6(1):5–11, March 2014. ISSN 1876-3413, 1876-3405. doi: 10.1093/inthealth/ihu006. URL https://academic.oup.com/inthealth/article-lookup/doi/10.1093/inthealth/ihu006.

[63] Amy Wesolowski, Caroline O. Buckee, Kenth Engø-Monsen, and C. J. E. Metcalf. Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data. *Journal of Infectious Diseases*, 214(suppl 4):S414–S420, December 2016. ISSN 0022-1899, 1537-6613. doi: 10.1093/infdis/jiw273. URL https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/jiw273.

[64] Pierre Nouvellet, Sangeeta Bhatia, Anne Cori, Kylie E. C. Ainslie, Marc Baguelin, Samir Bhatt, Adhiratha Boonyasiri, Nicholas F. Brazeau, Lorenzo Cattarino, Laura V. Cooper, Helen Coupland, Zulma M. Cucunuba, Gina Cuomo-Dannenburg, Amy Dighe, Bimandra A. Djaafara, Ilaria Dorigatti, Oliver D. Eales, Sabine L. Van Elsland, Fabricia F. Nascimento, Richard G. FitzJohn, Katy A. M. Gaythorpe, Lily Geidelberg, William D. Green, Arran Hamlet, Katharina Hauck, Wes Hinsley, Natsuko Imai, Benjamin Jeffrey, Edward Knock, Daniel J. Laydon,

John A. Lees, Tara Mangal, Thomas A. Mellan, Gemma Nedjati-Gilani, Kris V. Parag, Margarita Pons-Salort, Manon Ragonnet-Cronin, Steven Riley, H. Juliette T. Unwin, Robert Verity, Michaela A. C. Vollmer, Erik Volz, Patrick G. T. Walker, Caroline E. Walters, Haowei Wang, Oliver J. Watson, Charles Whittaker, Lilith K. Whittles, Xiaoyue Xi, Neil M. Ferguson, and Christl A. Donnelly. Reduction in mobility and COVID-19 transmission. *Nature Communications*, 12(1):1090, February 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-21358-2. URL https://www.nature.com/articles/s41467-021-21358-2.

[65] Natalya Kostandova, Ronan Corgel, Shweta Bansal, Sophie Bérubé, Eimear Cleary, Chelsea Hansen, Matt D. T. Hitchings, Bernardo García-Carreras, Lauren Gardner, Moritz U. G. Kraemer, Shengjie Lai, Yao Li, Amanda C. Perofsky, Giulia Pullano, Jonathan M. Read, Gabriel Ribeiro Dos Santos, Henrik Salje, Saki Takahashi, Cécile Viboud, Jasmine Wang, Derek A. T. Cummings, and Amy Wesolowski. Improving mobility data for infectious disease research. *Nature Human Behaviour*, March 2025. ISSN 2397-3374. doi: 10.1038/s41562-025-02151-3. URL https://www.nature.com/articles/s41562-025-02151-3.

[66] Oriol Cabanas-Tirapu, Lluís Danús, Esteban Moro, Marta Sales-Pardo, and Roger Guimerà. Human mobility is well described by closed-form gravity-like models learned automatically from data. *Nature Communications*, 16 (1):1336, February 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-56495-5. URL https://www.nature.com/articles/s41467-025-56495-5.

[67] Natalya Kostandova, Christine Prosperi, Simon Mutembo, Chola Nakazwe, Harriet Namukoko, Bertha Nachinga, Shengjie Lai, Andrew J. Tatem, Qianwen Duan, Elliot N. Kabalo, Kabondo Makungo, Gershom Chongwe, Innocent Chilumba, Gloria Musukwa, Kalumbu H. Matakala, Mutinta Hamahuwa, Webster Mufwambi, Japhet Matoba, Irene Mutale, Kenny Situtu, Edgar Simulundu, Phillimon Ndubani, Alvira Z. Hasan, Shaun A. Truelove, Amy K. Winter, Andrea C. Carcelen, Bryan Lau, William J. Moss, and Amy Wesolowski. Comparing and integrating human mobility data sources for measles transmission modeling in Zambia. *PLOS Global Public Health*, 5(5):e0003906, May 2025. ISSN 2767-3375. doi: 10.1371/journal.pgph.0003906. URL https://journals.plos.org/globalpublichealth/article?id=10.1371/journal.pgph.0003906. Publisher: Public Library of Science.

[68] Lilit Kazazian, Antonio S. Lima Neto, Geziel S. Sousa, Osmar José Do Nascimento, and Marcia C. Castro. Spatiotemporal transmission dynamics of co-circulating dengue, Zika, and chikungunya viruses in Fortaleza, Brazil: 2011–2017. *PLOS Neglected Tropical Diseases*, 14(10):e0008760, October 2020. ISSN 1935-2735. doi: 10.1371/journal.pntd.0008760. URL https://dx.plos.org/10.1371/journal.pntd.0008760.

[69] Christovam Barcellos and Rachel Lowe. Expansion of the dengue transmission area in Brazil: the role of climate and cities. *Tropical medicine & international health: TM & IH*, 19(2):159–168, February 2014. ISSN 1365-3156. doi: 10.1111/tmi.12227.

[70] Javier Aguilar, Aleix Bassolas, Gourab Ghoshal, Surendra Hazarie, Alec Kirkley, Mattia Mazzoli, Sandro Meloni, Sayat Mimar, Vincenzo Nicosia, José J. Ramasco, and Adam Sadilek. Impact of urban structure on infectious disease spreading. *Scientific Reports*, 12(1):3816, March 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-06720-8.

[71] Claudia T. Codeço, Daniel A.M. Villela, and Flavio C. Coelho. Estimating the effective reproduction number of dengue considering temperature-dependent generation intervals. *Epidemics*, 25:101–111, December 2018. ISSN 17554365. doi: 10.1016/j.epidem.2018.05.011. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436517300907.

[72] Cathal Mills, Tarek Alrefae, William S. Hart, Moritz U. G. Kraemer, Kris V. Parag, Robin N. Thompson, Christl A. Donnelly, and Ben Lambert. Renewal equations for vector-borne diseases, September 2024. URL http://arxiv.org/abs/2409.18726. arXiv:2409.18726 [q-bio].

[73] Jonathan Dushoff and Sang Woo Park. Speed and strength of an epidemic intervention. *Proceedings of the Royal Society B: Biological Sciences*, 288(1947):rspb.2020.1556, 20201556, March 2021. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2020.1556. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2020.1556.

[74] 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.

[75] Cathal Mills and Christl A. Donnelly. Climate-based modelling and forecasting of dengue in three endemic departments of Peru. *PLOS Neglected Tropical Diseases*, 18(12):e0012596, December 2024. ISSN 1935-2735. doi: 10.1371/journal.pntd.0012596. URL https://dx.plos.org/10.1371/journal.pntd.0012596.

[76] Erin A. Mordecai, Jeremy M. Cohen, Michelle V. Evans, Prithvi Gudapati, Leah R. Johnson, Catherine A. Lippi, Kerri Miazgowicz, Courtney C. Murdock, Jason R. Rohr, Sadie J. Ryan, Van Savage, Marta S. Shocket, Anna Stewart Ibarra, Matthew B. Thomas, and Daniel P. Weikel. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. *PLOS Neglected Tropical Diseases*, 11(4):e0005568, April 2017. ISSN 1935-2735. doi: 10.1371/journal.pntd.0005568. URL https://dx.plos.org/10.1371/journal.pntd.0005568.

[77] Estee Y. Cramer, Yuxin Huang, Yijin Wang, Evan L. Ray, Matthew Cornell, Johannes Bracher, Andrea Brennen, Alvaro J. Castro Rivadeneira, Aaron Gerding, Katie House, Dasuni Jayawardena, Abdul Hannan Kanji, Ayush Khandelwal, Khoa Le, Vidhi Mody, Vrushti Mody, Jarad Niemi, Ariane Stark, Apurv Shah, Nutcha Wattanchit, Martha W. Zorn, Nicholas G. Reich, US COVID-19 Forecast Hub Consortium, Tilmann Gneiting, Anja Mühlemann, Youyang Gu, Yixian Chen, Krishna Chintanippu, Viresh Jivane, Ankita Khurana, Ajay Kumar, Anshul Lakhani, Prakhar Mehrotra, Sujitha Pasumarty, Monika Shrivastav, Jialu You, Nayana Bannur, Ayush Deva, Sansiddh Jain, Mihir Kulkarni, Srujana Merugu, Alpan Raval, Siddhant Shingi, Avtansh Tiwari, Jerome White, Aniruddha Adiga, Benjamin Hurt, Bryan Lewis, Madhav Marathe, Akhil Sai Peddireddy, Przemyslaw Porebski, Srinivasan Venkatramanan, Lijing Wang, Maytal Dahan, Spencer Fox, Kelly Gaither, Michael Lachmann, Lauren Ancel Meyers, James G. Scott, Mauricio Tec, Spencer Woody, Ajitesh Srivastava, Tianjian Xu, Jeffrey C. Cegan, Ian D. Dettwiller, William P. England, Matthew W. Farthing, Glover E. George, Robert H. Hunter, Brandon Lafferty, Igor Linkov, Michael L. Mayo, Matthew D. Parno, Michael A. Rowland, Benjamin D. Trump, Samuel Chen, Stephen V. Faraone, Jonathan Hess, Christopher P. Morley, Asif Salekin, Dongliang Wang, Yanli Zhang-James, Thomas M. Baer, Sabrina M. Corsetti, Marisa C. Eisenberg, Karl Falb, Yitao Huang, Emily T. Martin, Ella McCauley, Robert L. Myers, Tom Schwarz, Graham Casey Gibson, Daniel Sheldon, Liyao Gao, Yian Ma, Dongxia Wu, Rose Yu, Xiaoyong Jin, Yu-Xiang Wang, Xifeng Yan, YangQuan Chen, Lihong Guo, Yanting Zhao, Jinghui Chen, Quanquan Gu, Lingxiao Wang, Pan Xu, Weitong Zhang, Difan Zou, Ishanu Chattopadhyay, Yi Huang, Guoqing Lu, Ruth Pfeiffer, Timothy Sumner, Dongdong Wang, Liqiang Wang, Shunpu Zhang, Zihang Zou, Hannah Biegel, Joceline Lega, Fazle Hussain, Zeina Khan, Frank Van Bussel, Steve McConnell, Stephanie L Guertin, Christopher Hulme-Lowe, V. P. Nagraj, Stephen D. Turner, Benjamín Bejar, Christine Choirat, Antoine Flahault, Ekaterina Krymova, Gavin Lee, Elisa Manetti, Kristen Namigai, Guillaume Obozinski, Tao Sun, Dorina Thanou, Xuegang Ban, Yunfeng Shi, Robert Walraven, Qi-Jun Hong, Axel Van De Walle, Michal Ben-Nun, Steven Riley, Pete Riley, James Turtle, Duy Cao, Joseph Galasso, Jae H. Cho, Areum Jo, David DesRoches, Pedro Forli, Bruce Hamory, Ugur Koyluoglu, 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, Julie S. Ivy, Maria E. Mayorga, Jessica Mele, Erik T. Rosenstrom, Julie L. Swann, Andrea Kraus, David Kraus, Jiang Bian, Wei Cao, Zhifeng Gao, Juan Lavista Ferres, Chaozhuo Li, Tie-Yan Liu, Xing Xie, Shun Zhang, Shun Zheng, Matteo Chinazzi, Alessandro Vespignani, Xinyue Xiong, Jessica T. Davis, Kunpeng Mu, Ana Pastore Y Piontti, Jackie Baek, Vivek Farias, Andreea Georgescu, Retsef Levi, Deeksha Sinha, Joshua Wilde, Andrew Zheng, Omar Skali Lami, Amine Bennouna, David Nze Ndong, Georgia Perakis, Divya Singhvi, Ioannis Spantidakis, Leann Thayaparan, Asterios Tsiourvas, Shane Weisberg, Ali Jadbabaie, Arnab Sarker, Devavrat Shah, Leo A. Celi, Nicolas D. Penna, Saketh Sundar, Abraham Berlin, Parth D. Gandhi, Thomas McAndrew, Matthew Piriya, Ye Chen, William Hlavacek, Yen Ting Lin, Abhishek Mallela, Ely Miller, Jacob Neumann, Richard Posner, Russ Wolfinger, Lauren Castro, Geoffrey Fairchild, Isaac Michaud, Dave Osthus, Daniel Wolffram, Dean Karlen, Mark J. Panaggio, Matt Kinsey, Luke C. Mullany, Kaitlin Rainwater-Lovett, Lauren Shin, Katharine Tallaksen, Shelby Wilson, Michael Brenner, Marc Coram, Jessie K. Edwards, Keya Joshi, Ellen Klein, Juan Dent Hulse, Kyra H. Grantz, Alison L. Hill, Kathryn Kaminsky, Joshua Kaminsky, Lindsay T. Keegan, Stephen A. Lauer, Elizabeth C. Lee, Joseph C. Lemaitre, Justin Lessler, Hannah R. Meredith, Javier Perez-Saez, Sam Shah, Claire P. Smith, Shaun A. Truelove, Josh Wills, Lauren Gardner, Maximilian Marshall, Kristen Nixon, John C. Burant, Jozef Budzinski, Wen-Hao Chiang, George Mohler, Junyi Gao, Lucas Glass, Cheng Qian, Justin Romberg, Rakshith Sharma, Jeffrey Spaeder, Jimeng Sun, Cao Xiao, Lei Gao, Zhiling Gu, Myungjin Kim, Xinyi Li, Yueying Wang, Guannan Wang, Lily Wang, Shan Yu, Chaman Jain, Sangeeta Bhatia, Pierre Nouvellet, Ryan Barber, Emmanuela Gaikedu, Simon Hay, Steve Lim, Chris Murray, David Pigott, Robert C. Reiner, Prasith Baccam, Heidi L. Gurung, Steven A. Stage, Bradley T. Suchoski, Chung-Yan Fong, Dit-Yan Yeung, Bijaya Adhikari, Jiaming Cui, B. Aditya Prakash, Alexander Ro-

dríguez, Anika Tabassum, Jiajia Xie, John Asplund, Arden Baxter, Pinar Keskinocak, Buse Eylul Oruc, Nicoleta Serban, Sercan O. Arik, Mike Dusenberry, Arkady Epshteyn, Elli Kanal, Long T. Le, Chun-Liang Li, Tomas Pfister, Rajarishi Sinha, Thomas Tsai, Nate Yoder, Jinsung Yoon, Leyou Zhang, Daniel Wilson, Artur A. Belov, Carson C. Chow, Richard C. Gerkin, Osman N. Yogurtcu, Mark Ibrahim, Timothee Lacroix, Matthew Le, Jason Liao, Maximilian Nickel, Levent Sagun, Sam Abbott, Nikos I. Bosse, Sebastian Funk, Joel Hellewell, Sophie R. Meakin, Katharine Sherratt, Rahi Kalantari, Mingyuan Zhou, Morteza Karimzadeh, Benjamin Lucas, Thoai Ngo, Hamidreza Zoraghein, Behzad Vahedi, Zhongying Wang, Sen Pei, Jeffrey Shaman, Teresa K. Yamana, Dimitris Bertsimas, Michael L. Li, Saksham Soni, Hamza Tazi Bouardi, Madeline Adee, Turgay Ayer, Jagpreet Chhatwal, Ozden O. Dalgic, Mary A. Ladd, Benjamin P. Linas, Peter Mueller, Jade Xiao, Jurgen Bosch, Austin Wilson, Peter Zimmerman, Qinxia Wang, Yuanjia Wang, Shanghong Xie, Donglin Zeng, Jacob Bien, Logan Brooks, Alden Green, 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, John M. Drake, Eamon B. O'Dea, Yaser Abu-Mostafa, Rahil Bathwal, Nicholas A. Chang, Pavan Chitta, Anne Erickson, Sumit Goel, Jethin Gowda, Qixuan Jin, HyeongChan Jo, Juhyun Kim, Pranav Kulkarni, Samuel M. Lushtak, Ethan Mann, Max Popken, Connor Soohoo, Kushal Tirumala, Albert Tseng, Vignesh Varadarajan, Jagath Vytheeswaran, Christopher Wang, Akshay Yeluri, Dominic Yurk, Michael Zhang, Alexander Zlokapa, Robert Pagano, Chandini Jain, Vishal Tomar, Lam Ho, Huong Huynh, Quoc Tran, Velma K. Lopez, Jo W. Walker, Rachel B. Slayton, Michael A. Johansson, Matthew Biggerstaff, and Nicholas G. Reich. The United States COVID-19 Forecast Hub dataset. *Scientific Data*, 9(1):462, August 2022. ISSN 2052-4463. doi: 10.1038/s41597-022-01517-w. URL https://www.nature.com/articles/s41597-022-01517-w.