GeoHealth

RESEARCH ARTICLE
10.1029/2023GH000870

Key Points:

- Meteorological and ecohydrological factors can accurately forecast Yellow Fever (YF) occurrence and incidence in Brazil weeks in the future
- Different predictors contribute to forecasted risk of YF occurrence and to forecasted incidence if cases are seen
- Temperature, precipitation, and previous YF burden were the most influential predictors in the forecasting models

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. L. Servadio,
serva024@umn.edu

Citation:

Servadio, J. L., Convertino, M., Fiecas, M., & Muñoz-Zanzi, C. (2023). Weekly forecasting of Yellow Fever occurrence and incidence via eco-meteorological dynamics. *GeoHealth*, *7*, e2023GH000870. https://doi.org/10.1029/2023GH000870

Received 27 MAY 2023
Accepted 11 OCT 2023

Author Contributions:

Conceptualization: Joseph L. Servadio, Claudia Muñoz-Zanzi
Data curation: Joseph L. Servadio, Matteo Convertino, Claudia Muñoz-Zanzi
Formal analysis: Joseph L. Servadio
Funding acquisition: Joseph L. Servadio, Claudia Muñoz-Zanzi
Methodology: Joseph L. Servadio, Matteo Convertino, Mark Fiecas, Claudia Muñoz-Zanzi
Supervision: Matteo Convertino, Claudia Muñoz-Zanzi

Weekly Forecasting of Yellow Fever Occurrence and Incidence via Eco-Meteorological Dynamics

Joseph L. Servadio[1,2], Matteo Convertino[3], Mark Fiecas[4], and Claudia Muñoz-Zanzi[2]

[1]Department of Biology, Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA, USA, [2]Division of Environmental Health Sciences, School of Public Health, University of Minnesota, Minneapolis, MN, USA, [3]fuTuRE EcoSystems Lab, Tsinghua SIGS, Tsinghua University, Shenzhen, China, [4]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA

Abstract Yellow Fever (YF), a mosquito-borne disease, requires ongoing surveillance and prevention due to its persistence and ability to cause major epidemics, including one that began in Brazil in 2016. Forecasting based on factors influencing YF risk can improve efficiency in prevention. This study aimed to produce weekly forecasts of YF occurrence and incidence in Brazil using weekly meteorological and ecohydrological conditions. Occurrence was forecast as the probability of observing any cases, and incidence was forecast to represent morbidity if YF occurs. We fit gamma hurdle models, selecting predictors from several meteorological and ecohydrological factors, based on forecast accuracy defined by receiver operator characteristic curves and mean absolute error. We fit separate models for data before and after the start of the 2016 outbreak, forecasting occurrence and incidence for all municipalities of Brazil weekly. Different predictor sets were found to produce most accurate forecasts in each time period, and forecast accuracy was high for both time periods. Temperature, precipitation, and previous YF burden were most influential predictors among models. Minimum, maximum, mean, and range of weekly temperature, precipitation, and humidity contributed to forecasts, with optimal lag times of 2, 6, and 7 weeks depending on time period. Results from this study show the use of environmental predictors in providing regular forecasts of YF burden and producing nationwide forecasts. Weekly forecasts, which can be produced using the forecast model developed in this study, are beneficial for informing immediate preparedness measures.

Plain Language Summary Yellow Fever (YF) is a persistent mosquito-borne disease affecting much of the tropics with zoonotic reservoirs. Its persistence, emphasized by a major epidemic affecting Brazil in 2016 and 2017, highlights the importance of predictions of when and where future disease burden will be seen. This study aimed to create a statistical model that can provide regular, weekly forecasts of YF burden in Brazil. Because YF, like many mosquito-borne diseases, is known to relate to environmental conditions, various static ecohydrological and dynamic meteorological conditions such as temperature, humidity, elevation, and vegetation were included as potential model predictors. The model was developed to forecast both a probability for whether YF cases will occur as well as an estimated number of cases if the disease does occur. The resulting forecast model can produce regular weekly forecasts, with the most accurate forecasts using conditions 2, 6, or 7 weeks prior across models. Use of the forecasting model developed in this study can inform locations within Brazil that are at greatest risk of high YF burden each week, allowing preventative measures to be targeted where and when they are most needed.

1. Introduction

Yellow Fever (YF) is a mosquito-borne viral disease endemic to Latin America and sub-Saharan Africa. It is spread by multiple genera of mosquito, including urban-dwelling *Aedes* mosquitoes as well as forest-dwelling *Haemagogus* and *Sabethes* mosquitoes in South America, and infects both humans and non-human primates (Litvoc et al., 2018; World Health Organization, 2018). Symptomatic cases of YF typically present flu-like symptoms such as fever, nausea, and body aches, which typically persist for less than a week (Litvoc et al., 2018). A subset of these cases then become severe, with symptoms including jaundice and hemorrhaging, and approximately 40% of severe cases are fatal due to organ damage (Johansson et al., 2014; Monath & Vasconcelos, 2015; Servadio et al., 2021; World Health Organization, 2018). Though there are no treatments for YF (Collins & Barrett, 2017), human cases can be prevented through a safe and effective vaccine (Gardner & Ryman, 2010).

Validation: Joseph L. Servadio, Matteo Convertino, Mark Fiecas, Claudia Muñoz-Zanzi
Visualization: Joseph L. Servadio, Matteo Convertino, Mark Fiecas, Claudia Muñoz-Zanzi
Writing – original draft: Joseph L. Servadio
Writing – review & editing: Joseph L. Servadio, Matteo Convertino, Mark Fiecas, Claudia Muñoz-Zanzi

Because YF cannot be eradicated, preventative measures such as vaccination campaigns, mosquito control, or public health message are needed to reduce morbidity and mortality. In recent decades, tens of thousands of YF cases have been reported globally, particularly in the countries that are endemic for YF, including 13 in North and South America (World Health Organization, 2021). Due to common underreporting of cases (Monath & Vasconcelos, 2015), actual burden of YF is likely even higher than reported.

In order to prioritize immediate responses to YF risk, forecasts for YF risk are needed a short time into the future, such as several weeks. Among arthropod-borne viruses, it is common for studies to project burden more than 10 years into the future (World Health Organization, 2013), primarily with a focus on the effects of global climate change (Semenza & Suk, 2018), rather than forecasts weeks to months into the future. Forecasts weeks to months into the future are typically more common for human-to-human communicable diseases (Chae et al., 2018; Lutz et al., 2019), particularly during an epidemic or pandemic (Roosa et al., 2020; Shanafelt et al., 2018), though there exist some examples of studies aiming to forecast more immediate risk of mosquito-borne diseases (Chen, Chu, Chen, & Cook, 2018; Chen, Ong, et al., 2018). Providing disease forecasts weeks in advance instead of projections years in advance motivates immediate decisions for prevention of new cases by targeting locations in greatest need (Chen, Ong, et al., 2018; Fischer et al., 2016), which can be particularly beneficial in resource constrained settings. Simply tracking human cases is inadequate for future preparedness, as delays in reporting of cases that have occurred and have been detected lag current new infections.

One group of mechanisms affecting risk of YF and other mosquito-borne diseases includes environmental patterns (Chen, Chu, et al., 2018). Multiple environmental factors, including meteorological conditions such as temperature and precipitation as well as ecohydrological conditions such as ecological biomes and elevation, have been shown to relate to risk of different mosquito-borne diseases (Ruiz et al., 2006; Servadio et al., 2018; Tesla et al., 2018), including YF (Hamlet et al., 2018; Hamrick et al., 2017; Servadio et al., 2022). Environmental factors can affect risk of YF as they relate to mosquito populations through life cycles (Bayoh & Lindsay, 2003) and pathogen incubation (Gage et al., 2008), nonhuman primates through habitat suitability (Aristizabal et al., 2018), and humans through behavioral adjustments (Aspvik et al., 2018) and mobility patterns. Some of these factors, such as elevation and ecological biomes, are not expected to fluctuate over short periods of time such as days or weeks, while others, such as temperature and precipitation, do experience such fluctuations. These are likely to have differing effects on the urban-dwelling *Aedes* mosquitoes as compared to the sylvatic *Haemagogus* and *Sabethes* mosquitoes (Servadio et al., 2022). These factors are of particular interest to informing YF preparedness compared to other mosquito-borne diseases due to zoonotic spillover from non-human primates through mosquito vectors.

The need for preparation is highlighted by recent outbreaks of YF, which have occurred in nations across South America (Barros & Boecken, 1996; de Goes et al., 1976) and Africa (Addy et al., 1986; "Outbreak news. Yellow fever, Uganda," 2011). A large outbreak was observed in Brazil beginning in December 2016 (Goldani, 2017; Moreira-Soto et al., 2018), following the Zika outbreak that previously impacted Eastern Brazil (Lowe et al., 2018a). In December 2016 and the 6 months that followed, Brazil observed more YF cases than had been observed over the previous decade. Outbreak cases were largely seen in southeastern Brazil, including Rio de Janeiro and São Paulo (Possas et al., 2018), which is in contrast to previous years where cases were primarily seen in Western Brazil, near the Amazon rainforest (Hamrick et al., 2017). These southeastern locations that experienced the greatest burden were not previously recommended locations for YF vaccination prior to travel (Centers for Disease Control and Prevention, 2018). Following this outbreak, YF burden currently persists in much of the Americas, motivating continued surveillance and prevention.

This study aimed to evaluate the ability of meteorological and ecohydrological conditions to produce weekly forecasts of YF occurrence and incidence, using municipalities of Brazil as a case study. Statistical models were developed aiming to maximize the accuracy of forecasts using various meteorological and ecohydrological conditions as predictors. The resulting models in this study produced municipality-level forecasts each week for the probability of YF occurrence and, if it occurs, the estimated incidence. Because the forecasts are expected to show different probabilities of disease occurrence and different estimated incidences across municipalities, those with highest probability of occurrence or estimated burden can be prioritized for preventative measures.

2. Methods and Materials

2.1. Data Sources

2.1.1. Yellow Fever Incidence

This study focuses on forecasting in Brazil, the largest nation in South America by both land area and population. It is subdivided into 26 states plus a federal district, containing 5,570 municipalities (Instituto Brasileiro de Geografia Estastistica, 2018). Laboratory-confirmed YF case data between January 2000 and March 2018 were provided by the Pan American Health Organization and the Brazilian Ministry of Health, with all cases having an associated municipality and date of report. While a small number of cases were identified as tourists, information was unavailable regarding whether cases were associated with domestic or international travel and therefore imported. We aggregated case data by week and conducted all analyses by municipality-week. In total, 952 weeks of data were available, with week one corresponding to 3–9 January 2000 and week 952 corresponding to 26 March to 1 April 2018.

Annual population data per municipality are publicly available through the Brazilian Institute for Geography and Statistics (Instituto Brasileiro de Geografia Estastistica, 2018) for each year between 2000 and 2018, with the exceptions of 2007 and 2010, for which we singly imputed using the arithmetic mean of the two adjacent years' populations. Using YF case data and annual population values, we calculated the incidence per 100,000 population for each municipality each week. Previous weeks' occurrence and incidence were also considered as potential model predictors to account for potential temporal autocorrelation in YF occurrence and incidence. The terms "occurrence" and "incidence" in this study have distinct definitions: occurrence is defined as the binary outcome of observing any YF cases, and incidence is defined as YF cases per 100,000 population.

A safe and highly effective vaccine is available for YF that confers lifelong immunity (Gotuzzo et al., 2013). Vaccine coverage data were available from Shearer et al. (2018b), representing proportions of populations vaccinated by municipality for 2000, 2010, and 2016. For other years, we assigned the vaccine coverage of the nearest of the 3 available years. We considered vaccination as a potential predictor in producing forecasts.

2.1.2. Ecohydrological and Meteorological Predictors

We selected environmental determinants as potential predictors based on established relationships with YF dynamics, whether through direct studies of YF and similar vector-borne diseases, connections to mosquito population dynamics, or connections to disease incubation (Table 1). Dynamic meteorological predictors considered for modeling include precipitation (Childs et al., 2019; de Almeida et al., 2019; Kaul et al., 2018; Zhao et al., 2018), temperature (Childs et al., 2019; Gage et al., 2008), and humidity (Dickens et al., 2018; Opayele et al., 2017). Static ecohydrological predictors include water drainage (Hayes et al., 1976), elevation, vegetation, and ecological biome (Hamrick et al., 2017).

Temperature, precipitation, and humidity data are publicly available from the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) from the United States National Aeronautics and Space Administration (NASA) (Gelaro et al., 2017; Global Modeling and Assimilation Office (GMAO), 2015). This data set contains hourly temperature in Kelvin (degrees Celsius + 273.15), hourly precipitation in millimeters per hour, and hourly humidity in kilograms of water per kilogram of air worldwide in 0.5° by 0.625° grids (Gelaro et al., 2017; Global Modeling and Assimilation Office (GMAO), 2015). We aggregated the data from each grid into daily averages for temperature and humidity and daily cumulative totals for precipitation, and then the grids were matched to Brazilian municipalities through a spatial join in ArcGIS Pro version 2.2.0 (Environmental Systems Research Institute (ESRI), 2018) using shapefiles for municipalities available from the Database of Global Administrative Areas version 3.6 (Database of Global Administrative Areas, 2018), providing daily averages (temperature, humidity) and totals (precipitation) for each municipality between 1 January 2000 and 31 March 2018.

Previous studies have shown that meteorological conditions can affect YF incidence in several ways. Various studies have found associations between maxima (de Almeida et al., 2019; Servadio et al., 2018), minima (Laneri et al., 2019; Ogashawara et al., 2019), means (de Almeida et al., 2019), and ranges (Paaijmans et al., 2009) of meteorological conditions and mosquito-borne diseases. We considered all four in this study by using the minimum, maximum, range between the two, and mean of the daily averages and totals described previously for each week. We used the same form (minimum, maximum, mean, range) within a model, favoring parsimony and

GeoHealth

10.1029/2023GH000870

Table 1
Summary of Predictors Considered in Model Fitting for Yellow Fever Forecasting

Variable	Source	Temporal resolution	Variations considered, if applicable
Previous Yellow Fever occurrence	PAHO/Brazilian Ministry of Health	Weekly	Lagged between 2 and 8 weeks
Previous Yellow Fever incidence	PAHO/Brazilian Ministry of Health	Weekly	Lagged between 2 and 8 weeks
Vaccine coverage	Shearer et al. (2018b)	Annually, for 2000, 2010, and 2016 only	
Temperature	MERRA-2, NASA	Weekly	Minimum, maximum, mean, range of daily averages; Lagged between 2 and 8 weeks; Linear and quadratic
Precipitation	MERRA-2, NASA	Weekly	Minimum, maximum, mean, range of daily averages; Lagged between 2 and 8 weeks; Linear and quadratic
Humidity	MERRA-2, NASA	Weekly	Minimum, maximum, mean, range of daily averages; Lagged between 2 and 8 weeks; Linear and quadratic
Drainage density		One-time	
Elevation	ASTER, NASA	One-time	
Vegetation	MODIS, "MODISTools" R package	Monthly	
Terrestrial biomes	World Wildlife Foundation	One-time	
Month		Month of the calendar year	

interpretability despite potential constraints on forecasting ability. Additionally, meteorological conditions have been found to show nonlinear associations with mosquito-borne disease. This can occur with temperature, with mosquito life cycles and virus incubation inhibited by cold or hot temperatures (Ciota & Keyel, 2019; Marinho et al., 2016), as well as with precipitation, where rainfall can create mosquito breeding grounds or wash away larval habitats (de Thoisy et al., 2020; Lowe et al., 2018b). Therefore, we considered both linear and quadratic terms for temperature, precipitation, and humidity to account for potential nonlinear effects.

Standing water creates opportunities for mosquito breeding (Centers for Disease Control and Prevention, n.d.), and river drainage can create opportunities for standing water to exist, therefore impacting risk of mosquito-borne diseases (Hayes et al., 1976). Drainage density was modeled for Brazil using elevation and hydrological characteristics of the area as of 2017. Digital elevation model data from NASA Radar Shuttle Topographic Mission (Farr et al., 2007) were used to extract both drainage basins and water streams using routine hydrogeomorphological extraction (Tarboton et al., 1991). Drainage density is then defined as the ratio of the length of rivers and streams to total land area (Montgomery & Dietrich, 1989; Tarboton et al., 1992) associated with water basins with a minimum area of 1,000 km^2. The water basins were matched to municipalities in ArcGIS Pro in order to compute a spatially-weighted average drainage density for each municipality.

Previous research has shown that lower elevations can relate to increased risk of mosquito-borne diseases (Hamrick et al., 2017). Locations with high elevation have long periods of lower temperatures, making them less suitable for mosquito habitats (Watts et al., 2017). Elevation data are publicly available through the Advanced Spaceborne Thermal Emission and Reflection Radiometer Digital Elevation model, from NASA and Japan's Ministry of Economy, Trade, and Industry (NASA & Ministry of Economics, Trade, and Industry, 2019). These data are represented as the elevation above sea level in meters in 30-m by 30-m grids. These grids were matched to municipalities to represent each municipality's average elevation.

Vegetation serves as a proxy for urbanization and development as well as habitat opportunities for tree dwelling mosquitoes (Hamlet et al., 2021). Vegetation data are available from NASA Moderate Resolution Imaging Spectroradiometer (MODIS) in the form of a normalized difference vegetation index (NDVI). These are accessible from the R package "MODISTools" in the form of NDVI values per municipality for each month (Tuck et al., 2014).

Terrestrial biomes, which have been previously related to YF risk (Hamrick et al., 2017), are publicly available from the World Wildlife Foundation (Olson et al., 2001), which we matched to municipalities using a spatial join to represent the most prevalent biome within each municipality. Because the majority of YF cases were seen in only two biomes observed in Brazil (Tropical and subtropical moist broadleaf forests; and Tropical and subtropical grasslands, savannas, and shrublands), the remaining four biomes were collapsed into a single category.

2.1.3. Temporal Lags

We considered temporal lags for the three meteorological variables as well as the two variables for previous weeks' YF incidence and occurrence within each municipality to allow time between production of forecasts and implementation of potential prevention strategies as well as to consider delayed effects of meteorological changes. As expected for vector-borne diseases, the effects of meteorological events on mosquito populations, and therefore disease incidence, are assumed not to occur instantaneously (Choi et al., 2016; Kakarla et al., 2019). Temporal lags are defined as the number of weeks between the week associated with the observed data and the week being targeted for forecasting. For example, a lag period of l weeks means that, in order to produce a forecast for week t, data from week $t - l$ were used as predictors. We considered lag periods between 2 and 8 weeks, using the same lag period for the three weather variables and the two variables for previous YF for model parsimony and interpretation. The choice of 2–8 weeks considered the time needed to respond to produced forecasts as well as forecasting within a window of time where responses are practical.

2.2. Model Fitting

In model fitting, we aimed to develop models that produce most accurate forecasts of weekly YF occurrence and incidence, using the previously described ecohydrological factors for each municipality-week as potential predictors (Table 1). We also considered month of forecast, collapsing June through November due to low occurrence. Including month as a potential predictor considers the possibility of including seasonal trends as well as

annual patterns that are not related to ecometeorological factors, which is a common consideration when creating prediction models for environmentally sensitive infectious diseases (Imai et al., 2015; Lowe et al., 2013). These can include school terms or vacations, tourism seasons, or human mobility for holidays.

Due to the outbreak that began in December 2016, we split the data into two time periods and analyzed them separately: January 2000 to November 2016 (referred to herein as the endemic time period) and December 2016 to March 2018 (referred to herein as the epidemic time period). Splitting the full time series at the start of the epidemic allowed us to compare whether disease dynamics outside of and within an outbreak are explained by similar factors. Such differences have been noted in previous studies (Servadio et al., 2022). Additionally, because the majority (84%) of cases (and 77% of weeks with cases) were observed after December 2016, a single model for the entire time period carries the risk of being driven by fitting these later months well, while underrepresenting the earlier years. There also exist differences in locations where YF was seen; during the endemic period, cases were primarily observed in northwestern Brazil, and during the epidemic period, cases were primarily observed in southeastern Brazil.

For each time period, we split the data into training and testing data sets. During the endemic time period, the training data set spanned January 2000 to December 2013, and the testing data set spanned January 2014 to November 2016. The epidemic time period only spans 16 months, between December 2016 and March 2018, creating challenges in assigning a cutoff to separate a training and testing period that allows the testing period to cover more than one season. To preserve the ratio between testing and training times across the two time periods, we randomly selected 3 months for testing data: February 2017, November 2017, and March 2018.

Between Januay 2000 and March 2018, 466 municipalities reported at least one YF case. Of these, 462 are present among the 5,570 municipalities present in the Database of Global Administrative Areas and therefore able to be matched to environmental data; only these municipalities were included in model fitting. This was done for both the endemic and epidemic time periods in order to increase confidence that a lack of reported cases represented a true lack of cases rather than epidemiologic silence, in which cases may exist but were not detected (Doyle et al., 2002; Servadio et al., 2022). We then applied results of the fitted models to all ecometeorological data to produce forecasts nationwide.

The model form was a Gamma hurdle model (Anderson, 2014; Zuur & Ieno, 2016), which consists of two distinct steps. The first step, the binomial step, fits all the incidence data as a binary outcome for occurrence in a logistic regression model with a logit link. The second step, the Gamma step, uses only the data where incidence is greater than zero in a Gamma regression model with a natural log link to predict incidence values. This model takes the form

$$\begin{aligned} \mathrm{YF}_{m,t} | \mathrm{YF}_{m,t} > 0 &\sim \mathrm{Gamma}(a, b) \\ \mathrm{logit}\left(E\left[P(\mathrm{YF}_{m,t} > 0)\right]\right) &= \alpha + \sum_i \beta_i X_{i,m,t} \\ \ln\left(E\left[\mathrm{YF}_{m,t} | \mathrm{YF}_{m,t} > 0\right]\right) &= \gamma + \sum_j \delta_j X_{j,m*,t*} \end{aligned} \quad (1)$$

where $\mathrm{YF}_{m,t}$ represents incidence of Yellow Fever per 100,000 population in municipality m at time t, α and γ represent the intercepts of the binomial and Gamma steps, respectively, β and δ represent the sets of coefficients corresponding to predictors denoted X, and $m*$ and $t*$ represent the subset of municipality and time combinations where $\mathrm{YF}_{m,t} > 0$. The interpretation of the output from Gamma model is predicted incidence under the condition that incidence is nonzero. Model fits from a Gamma hurdle model consist of two sets of predictors: one set of predictors from the binomial step, describing predictors' associations with observing nonzero incidence, and another set of predictors from the Gamma step, describing predictors' associations with observing higher YF incidence when incidence is nonzero. Similarly, the resulting forecasts from this model consist of two distinct values: (a) a probability of observing any YF occurrence and (b) an estimated incidence that will be seen if incidence is nonzero.

We chose to use a hurdle model a priori due to the high percentage of municipality-weeks observing no YF cases, and we selected a Gamma model for the incidence step prior to model fitting because nonzero incidence values resemble a Gamma distribution by being continuous, strictly positive, and exhibiting a heavy right tail.

The Gamma distribution, similarly defined to be positive and continuous, has been used previously to describe disease incidence (Servadio et al., 2020) and growth (Champredon & Dushoff, 2015; Chowell et al., 2016). Gamma hurdle models are not commonly used for disease incidence, but have been used in ecology applications (Saunders et al., 2019). The hurdle model in general has been used for infectious disease research previously for relating meteorological conditions to disease risk (Harris et al., 2019), and the motivations for using a hurdle model are conceptually similar to those for using a zero-inflated Poisson model (Imai et al., 2014; Wang et al., 2014).

We underwent model fitting and selection of the candidate predictors separately for the binomial and Gamma steps of the model, which allows the two steps of the same model to have distinct sets of predictors. This provides flexibility in predictor choice, ultimately allowing best forecasts. The form of the meteorological conditions (minimum, maximum, mean, range) and lag periods were also allowed to differ between the two steps. We considered all combinations of inclusion/exclusion of the previously described candidate predictors, excluding combinations that include quadratic terms without appropriate linear terms. We considered a total of 5,182 combinations of parameters for each combination of meteorological variable form (minimum, maximum, mean, range) and lag period (2–8 weeks).

2.3. Model Evaluation

To determine the best-fitting model for forecasting, we evaluated forecasts from the model produced from the training data within the testing time period. We evaluated forecasts of the binomial step through Receiver Operator Characteristic curves and maximizing the associated area under the curve (AUC), where we compared forecasted probabilities of occurrence in the testing period to the observed occurrences. The AUC, constrained between zero and one, represents how reliably the binomial step can predict whether any YF cases will be seen.

We evaluated the Gamma step by minimizing the mean absolute error (MAE), comparing predicted nonzero forecasts in the testing period to the observed nonzero incidence values. The MAE is defined as

$$\text{MAE} = \frac{1}{\sum y_{m,w} > 0} \sum |\hat{y}_{m,w,>0} - y_{m,w,>0}| \tag{2}$$

where $y_{m,w,>0}$ represents the observed incidence values in municipality m on week w among only those that greater than zero and $\hat{y}_{m,w,>0}$ represent the corresponding fitted values for municipality m on week w from the model. The MAE represents the average difference, in cases per 100,000 population, between observed and predicted incidence when the incidence is greater than zero. It is important to note that MAE only applies when any YF cases are seen because the Gamma step is conditioned on seeing incidence greater than zero. As a result, predicted incidence can only be compared to observed incidence when observed incidence is greater than zero. Previous works forecasting infectious diseases and examining environmental patterns relating to mosquito-borne diseases have used both AUC (Hamlet et al., 2018; Kaul et al., 2018; Liu et al., 2014) and MAE (Benedum et al., 2020; Mohamed et al., 2022; Patil & Pandya, 2021) as model diagnostics. All analyses were run using R version 4.0.3 (R Core Team, 2019), with code and data available at (Servadio, 2023).

2.4. Producing Forecasts

After selecting best-fitting models for both time periods, we applied the fitted regression equations to the ecohydrological and meteorological data described previously to produce forecasts. For demonstrative purposes, we produced forecasts between February 2000 and April 2018, based on available data. Within forecasts, every municipality-week is assigned two forecasted values: a probability of any YF occurrence, and an estimated incidence if YF does occur. Because the two values are produced independently, forecasted incidence is not influenced by forecasted probability of occurrence, even if this probability is low. Mapping both values allows locations to be identified that are most likely to experience any YF burden and/or experience high burden.

3. Results

A total of 2,079 confirmed YF cases were seen between January 2000 and March 2018. The time series of weekly YF cases throughout the entire study period shows the severity of the 2016–2017 outbreak, having notably higher

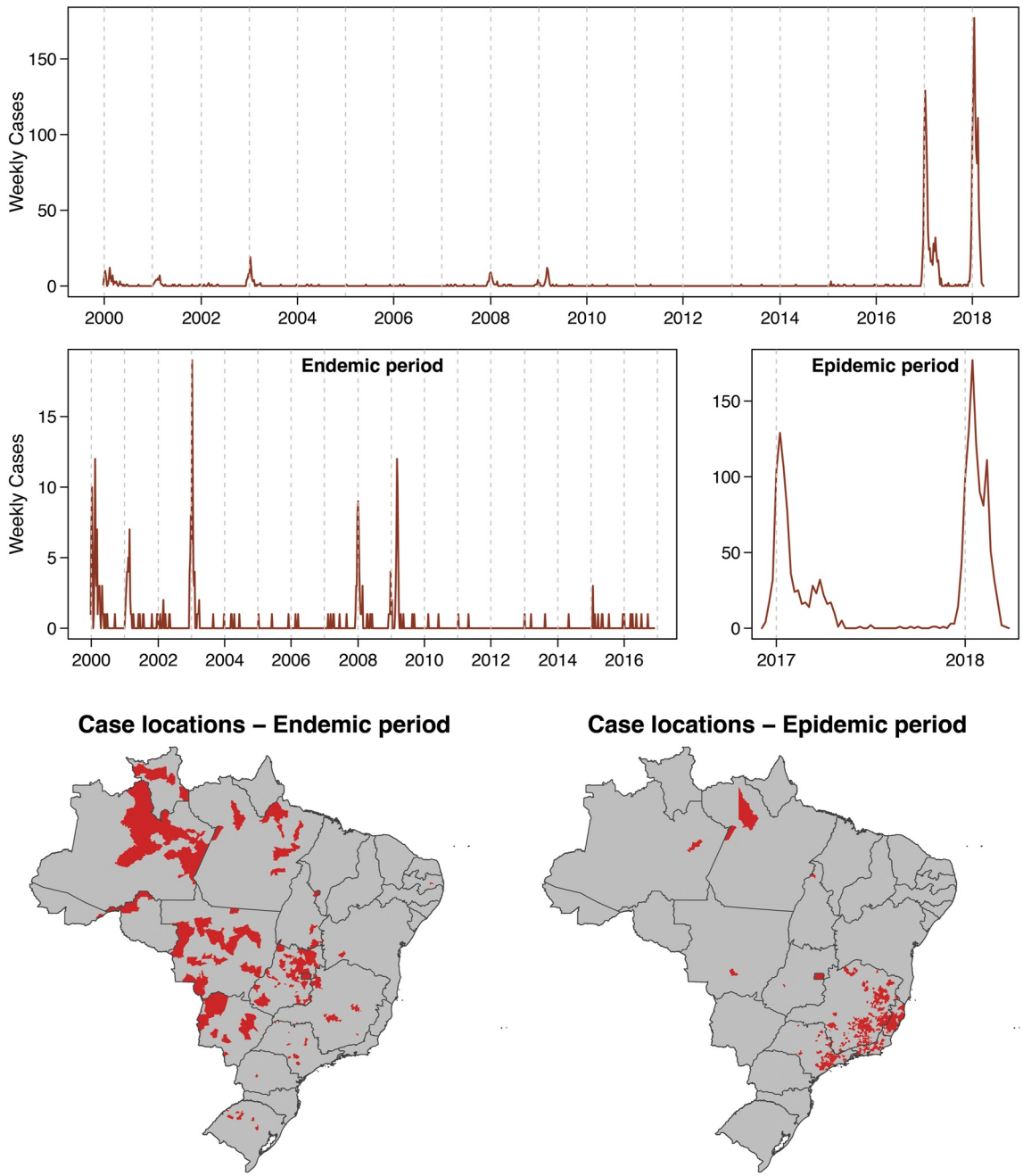

Figure 1. Weekly Yellow Fever case counts across Brazil during the entire time period of observation, January 2000 to March 2018 (top panel); the endemic period, January 2000 to November 2016 (middle left panel); and the epidemic period, December 2016 to March 2018 (middle right panel). Vertical lines indicate the first week of each calendar year. Municipalities seeing cases during the endemic period (lower left) and epidemic period (lower right) are shown. Map source: base map is publicly available from the Database of Global Administrative Areas, license is available at gadm.org/license.html.

case counts compared to previous years (Figure 1). Seasonal patterns of YF are also seen throughout all years of observation, with cases predominantly occurring during the early months of each year. Tables S1 and S2 in Supporting Information S1 show descriptive statistics of the YF incidence data and the predictors considered in the models for both the endemic and epidemic time periods.

Table 2
Estimated Model Parameters for the Binomial and Gamma Steps of the Hurdle Model for the Endemic Time Period, January 2000 to November 2016

Binomial step			Gamma step		
Parameter	Estimate	95% confidence interval	Parameter	Estimate	95% confidence interval
Intercept	−35.618	[−45.946, −26.290]	Intercept	−41.095	[−64.372, −17.818]
Min Temp	2.377	[1.542, 3.212]	Max Temp	3.45	[1.608, 5.292]
Min Temp^2	−0.049	[−0.067, −0.031]	Max Temp^2	−0.064	[−0.099, −0.029]
Previous Incidence	0.059	[0.037, 0.081]	Max Precipitation	20.491	[−1.328, 42.31]
			Max Humidity	−134.259	[−245.865, −22.653]
Lag period	6-week		Month—February	0.101	[−0.438, 0.640]
Model fit	AUC = 0.851		Month—March	0.219	[−0.253, 0.691]
			Month—April	0.027	[−0.665, 0.719]
			Month—May	0.284	[−0.565, 1.133]
			Month—June–November	0.417	[−0.242, 1.076]
			Month—December	−0.203	[−0.903, 0.497]
			Previous cases	−0.027	[−1.374, 1.32]
			Elevation	0.002	[0.002, 0.002]
			Biome—Grasslands	−0.315	[−0.683, 0.053]
			Biome—Other	−1.072	[−1.94, −0.204]
			Proportion Vaccinated	−0.201	[−0.303, −0.099]
			Lag period	7-week	
			Model fit	MAE = 7.009	

Note. Lag period and model fit diagnostics are shown in the bottom rows.

3.1. Model Results

3.1.1. Endemic Time Period

In the best fitting hurdle model for the endemic time period, the binomial step included as predictors minimum temperature as a linear and quadratic term as well as previous incidence. A 6-week lag period was used. The AUC value was 0.851. The model shows a quadratic relationship between minimum temperature 6 weeks prior and risk of YF occurrence, with probability of occurrence maximized at a minimum temperature near 24.5°C. Increased incidence 6 weeks prior was associated with higher probability of observing YF occurrence (Table 2, Figure S1 in Supporting Information S1).

The Gamma step for the endemic time period's model included as predictors maximum temperature as a linear and quadratic term, maximum precipitation, maximum humidity, month, previous occurrence of YF, elevation, biome, and proportion of the population vaccinated. A 7-week lag period was used. The MAE value for this model was 7.009. The quadratic relationship between temperature and incidence shows highest forecasted incidence when maximum weekly temperature is near 27°C. Higher maximum precipitation, lower maximum humidity, the months of June through November, observing no previous cases 7 weeks prior, higher elevation, belonging to broadleaf forest biome, and lower population vaccination were all associated with higher forecasted incidence of YF if any YF is to be seen (Table 2, Figure S2 in Supporting Information S1).

3.1.2. Epidemic Time Period

In the best fitting hurdle model for the epidemic time period, the binomial step included as predictors mean precipitation, mean humidity, month, previous YF occurrence, drainage density, and population vaccinated. A lag period of 2 weeks was used, and the AUC value was 0.902. Higher mean precipitation, lower mean humidity, occurrence of YF 2 weeks prior, lower drainage density, and lower population vaccination were associated with higher forecasted probability of YF occurrence. Weeks in January had the highest forecasted probability compared to all other months, with weeks in December through April having higher forecasted probabilities compared to weeks between May and November (Table 3, Figure S3 in Supporting Information S1).

GeoHealth

10.1029/2023GH000870

Table 3
Estimated Model Parameters for the Binomial and Gamma Steps of the Hurdle Model for the Epidemic Time Period, December 2016 to March 2018

Binomial step			Gamma step		
Parameter	Estimate	95% confidence interval	Parameter	Estimate	95% confidence interval
Intercept	0.628	[−0.311, 1.567]	Intercept	−0.759	[−1.596, 0.078]
Mean precipitation	183.448	[98.915, 267.981]	Range of Temperature	−0.117	[−0.213, −0.021]
Mean Hum	−86.41	[−139.244, −33.576]	Range of Precipitation	−19.317	[−69.865, 31.231]
Month-February	−0.481	[−0.697, −0.265]	Range of Precipitation^2	970.615	[−1,111.599, 3,052.829]
Month-March	−1.321	[−1.603, −1.039]	Range of Humidity	148.803	[−255.974, 553.58]
Month-April	−1.778	[−2.15, −1.406]	Range of Humidity^2	−43,993.645	[−122,758.964, 34,771.674]
Month-May	−3.599	[−4.432, −2.766]	Previous Incidence	0.023	[0.015, 0.031]
Month-June–November	−5.246	[−6.11, −4.382]	Drainage density	15.063	[9.23, 20.896]
Month-December	−2.049	[−2.388, −1.71]	NDVI	0.314	[0.238, 0.39]
Previous occurrence	2.624	[2.42, 2.828]			
Drainage density	−1.592	[−6.741, 3.557]	Lag period	2-week	
Proportion vaccinated	−0.282	[−0.329, −0.235]	Model fit	MAE = 4.860	
Lag period	2-week				
Model fit	AUC = 0.902				

Note. Lag period and model fit diagnostics are shown in the bottom rows.

The Gamma step for the epidemic period included as predictors range of temperature, range of precipitation as a linear and quadratic term, range of humidity as a linear and quadratic term, previous incidence, drainage density, and vegetation. A 2-week lag period was used, and the MAE value was 4.860. Parabolic relationships were seen between both precipitation and humidity and forecasted incidence, with forecasted incidence minimized when the range of precipitation was near 0.01 mm and maximized when the range of humidity was near 0.002 kg of water per kg of air. Lower range of temperature, higher YF incidence 2 weeks prior, higher drainage density, and higher vegetation were associated with a higher forecasted incidence of YF if YF occurs. Weeks located between December and May were associated with higher incidence compared to weeks found between June and November (Table 3, Figure S4 in Supporting Information S1).

The parameter estimates in Tables 2 and 3 and individual effects shown in Figures S1–S4 in Supporting Information S1 are interpreted to show, based on the particular model, the individual effect of each predictor. The parameters, when combined, produce forecasts of future YF burden as described previously. Because the aim of model selection was accurate forecasts rather than identifying unconfounded effects of individual predictors, interpreting individual parameters is subject to Table 2 fallacy (Westreich & Greenland, 2013).

We compared the selected models, with the best fits, to a collection of the 50 best-fitting models for each time period as a robustness check. Tables S3 and S4 in Supporting Information S1 show the frequency at which different predictors were retained within the best-fitting models for the endemic and epidemic periods respectively. In the endemic time period, the most commonly included predictors for the binomial step were mean temperature as a linear and quadratic term (26 times) and mean precipitation as a linear and quadratic term (26 times), with the most common lag period being 5 weeks (24 times); the most commonly included predictors for the Gamma step were month, biome, elevation (50 times each) and vaccine coverage (40 times), with the most common lag period being 7 weeks (39 times). In the epidemic time period, the most commonly included predictors for the binomial step were month, previous occurrence, and vaccine coverage (50 times each); the most commonly included predictors for the Gamma step were range of temperature (50 times), range of precipitation as a linear and quadratic term (49 and 42 times, respectively) previous incidence (43 times), and vegetation (50 times). All 50 best-fitting models for both steps included a lag period of 2 weeks (Tables S3 and S4 in Supporting Information S1). The directions of the parameter estimates were also largely consistent with those seen in Tables 2 and 3. Among the 50 worst-fitting models for both time periods, the only predictor from the selected models in Tables 2 and 3 that was commonly included was vaccine coverage, which appeared in the 50 worst-fitting models for the Gamma step of the endemic period (Table S5 in Supporting Information S1).

To show the utility of the selected predictors in providing accurate forecasts, the selected models were compared to both the null and saturated models, containing no predictors and maximum number of predictors respectively. In the endemic period, the null model produces an AUC of 0.5 by definition and an MAE of 7.283, and the saturated model produces an AUC of 0.618 and an MAE of 7.073. In the epidemic period, the null model produces an AUC of 0.5 and an MAE of 5.057, and the saturated model produces an AUC of 0.710 and an MAE of 5.166. Among the 50 best-fitting models in the endemic period, AUC values ranged between 0.835 and 0.851. And MAE values ranged between 7.009 and 7.053. Among the best-fitting models in the epidemic period, AUC values ranged between 0.897 and 0.902 and MAE values ranged between 4.860 and 4.879. All best-fitting models included at most 12 of the 14 possible predictors. Most of the binomial steps from the set of best fitting models consisted of 5–7 predictors (Table S3 in Supporting Information S1), and most of the Gamma steps consisted of 8–11 predictors out of a maximum of 14 (Table S4 in Supporting Information S1).

3.2. Forecast Examples

We produced example forecasts by applying the models described in Tables 2 and 3 to the observed data throughout Brazil. The forecasts show two numeric values for each municipality in each week: a forecasted probability of observing any YF cases and a forecasted number of cases per 100,000 population if any YF is seen. Because the binomial and Gamma steps of the models produce these forecasts separately, incidence forecasts are produced even when the forecasted probability of YF occurrence is low. The maps shown in Figure 2 are examples of maps that can be produced weekly in order to show locations that are at the greatest risk of observing YF cases or observing higher incidence compared to others in Brazil. These examples include a randomly selected week from the testing data that was used for model fitting as well as a week that occurs after the testing data to show a true forecast to a week that did not have data involved in the model fitting process. These weeks are the weeks

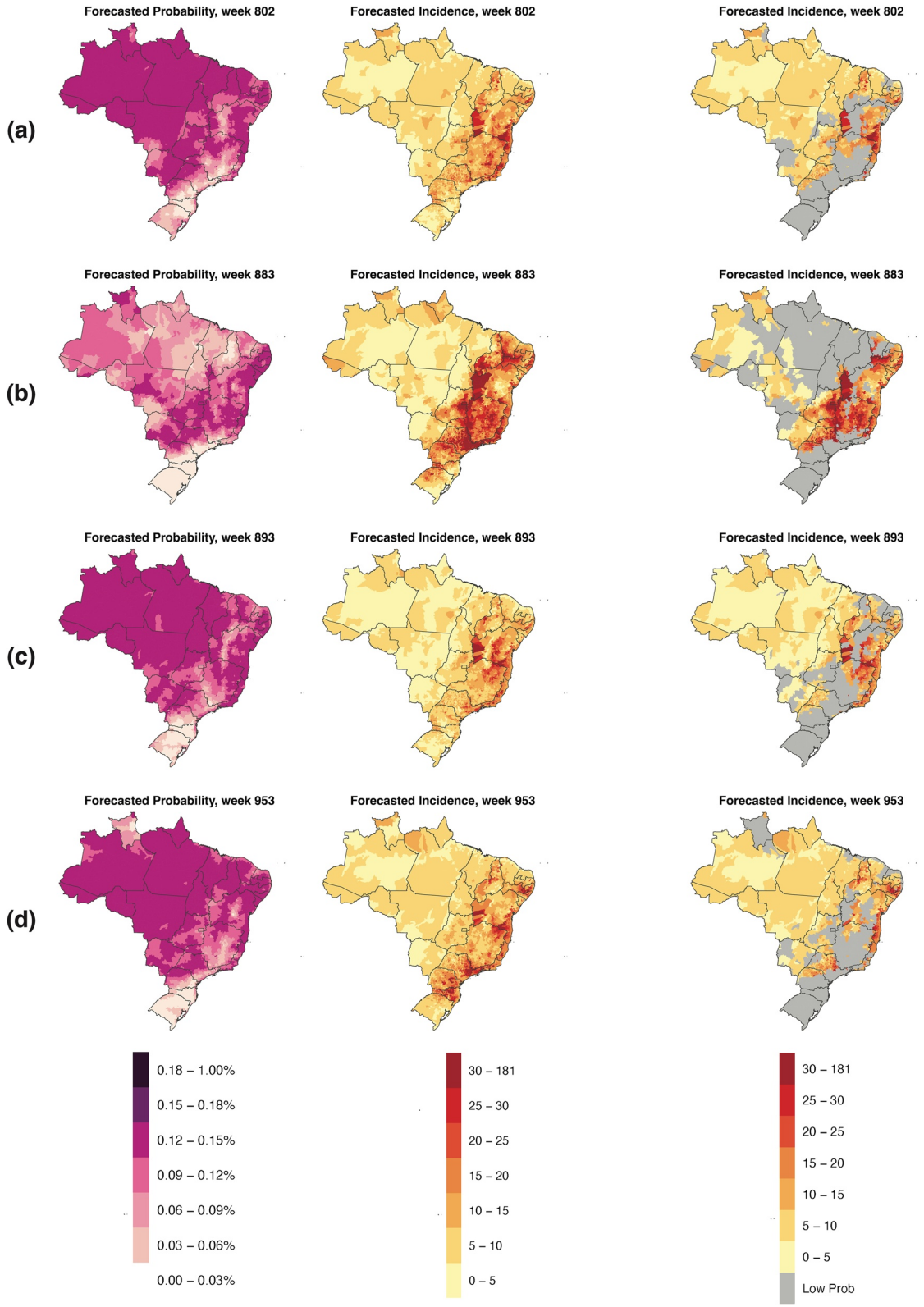
(a)
Forecasted Probability, week 802
Forecasted Incidence, week 802
Forecasted Incidence, week 802
(b)
Forecasted Probability, week 883
Forecasted Incidence, week 883
Forecasted Incidence, week 883
(c)
Forecasted Probability, week 893
Forecasted Incidence, week 893
Forecasted Incidence, week 893
(d)
Forecasted Probability, week 953
Forecasted Incidence, week 953
Forecasted Incidence, week 953
0.18 – 1.00%
0.15 – 0.18%
0.12 – 0.15%
0.09 – 0.12%
0.06 – 0.09%
0.03 – 0.06%
0.00 – 0.03%
30 – 181
25 – 30
20 – 25
15 – 20
10 – 15
5 – 10
0 – 5
30 – 181
25 – 30
20 – 25
15 – 20
10 – 15
5 – 10
0 – 5
Low Prob

Figure 2.

of 11–17 May 2015 (week 802, during the testing data for the endemic model); 5–11 December 2016 (week 883, following the testing data for the endemic model); 13–19 February 2017 (week 893, during the testing data for the epidemic model); and 2–8 April 2018 (week 953, following the testing data for the epidemic model).

The example forecasts show different spatial patterns between forecasted probability of occurrence and forecasted incidence due to the fact that different sets of parameters contributed to each forecast. Spatial trends in forecasted probability of occurrence and forecasted incidence can be seen in relation to the model predictors in Supporting Information S1 (Figures S5–S8 in Supporting Information S1). For example, Figure 2a shows higher probability of YF occurrence throughout the North and much of the Northwest and Center-West regions, with the exception of parts of Bahia, Piaui, Mato Grosso, and Mato Grosso do Sul states. These areas of high probability also show minimum temperatures near 24.5°C, the value indicating highest probability (Figures S1 and S5 in Supporting Information S1). Though most of the South and Southeast regions (Figure 2a) have lower probabilities of observing YF, much of the Southeast region, along with parts of Bahia state, contained municipalities with the highest forecasted incidence if YF occurs. Many of these areas have higher elevation or low population vaccination, both of which were associated with higher incidence (Figure S5 in Supporting Information S1). The trends seen in Figure 2 and Figures S5–S8 in Supporting Information S1, using the selected forecast models, are consistent with trends seen among the 50 best-fitting models (Figure S9 in Supporting Information S1).

To show both the trajectories of forecasts over time as well as the precision of forecasts, we selected 10 municipalities at random that represent 10 distinct states and displayed their forecasted probabilities of YF occurrence and forecasted incidence, along with 95% prediction intervals, over the entire testing time period of the endemic period and the entirety of the epidemic period (Figure 3). The time series plots show that seasonality can be seen in the forecasts, particularly in the forecasted incidence during the endemic period and in forecasted probability of occurrence during the epidemic period. Prediction intervals were generally narrow, through higher forecasted incidence values particularly showed less precision (Figure 3). Across the 50 best-fitting models, the longitudinal trajectories varied across models, with groups of models producing very similar forecasts (Figure S10 in Supporting Information S1). These differences among the 50 best-fitting models, however, were small. Larger differences in forecasts were seen comparing the selected best-fitting models to the set of worst-fitting models (Figure S11 in Supporting Information S1).

To assess accuracy of forecasts, we compared whether municipalities with high forecasted probability of occurrence did observe YF cases as well as whether forecasted incidence was more commonly an overestimate or underestimate. In the endemic time period, among the 258 municipality-weeks that observed any YF occurrence, 76 had forecasted probabilities above the 75th percentile. In the epidemic time period, among the 866 municipality-weeks that observed any YF occurrence, 139 had forecasted probabilities above the 75th percentile. Based on the fact that false negatives occurred more frequently than true positives if the 75th percentile of forecasts is used as a threshold above which YF cases are expected, the binomial steps of each model appear to be more likely to underestimate YF risk than overestimate. Among forecasted incidence values in the endemic time period, 159 municipality-weeks (62%) overestimated incidence, and 99 (38%) underestimated. In the epidemic time period, 646 municipality-weeks (75%) overestimated incidence, and 220 (25%) underestimated. Overestimations were more common compared to underestimations, so the Gamma steps of each model appear more likely to overestimate YF incidence rather than underestimate.

3.3. Global Sensitivity and Uncertainty Analyses

Full results of sensitivity analyses are reported in Supporting Information S1. A sensitivity analysis was conducted to examine the length of the training data sets used for model fitting. Keeping the end of the training data set for the endemic time period constant, the first week of the training data was moved forward to determine if a shorter training period would impact model performance, evaluated by the model fit criteria (AUC and MAE).

Figure 2. Examples of Yellow Fever (YF) forecast maps produced for (a) 11–17 May 2015 (week 802, predicted using data from 30 March to 5 April 2015 for the binomial step and 23–29 March 2015 for the Gamma step); (b) 5–11 December 2016 (week 883, predicted using data from 24 to 30 October 2016 for the binomial step and 17–23 October 2016 for the Gamma step); (c) 13–19 February 2017 (week 893, predicted using data from 30 January to 5 February 2017 for the binomial and Gamma steps); and (d) 2–8 April 2018 (week 953, predicted using data from 19 to 25 March 2018 for the binomial and Gamma steps). Forecasts show probability of observing any YF cases (left), estimated incidence (middle), and estimated incidence among municipalities where probability of occurrence was above the median value (right). Forecasts were generated using the models described in Tables 2 and 3. Map source: base map is publicly available from the Database of Global Administrative Areas, license is available at gadm.org/license.html.

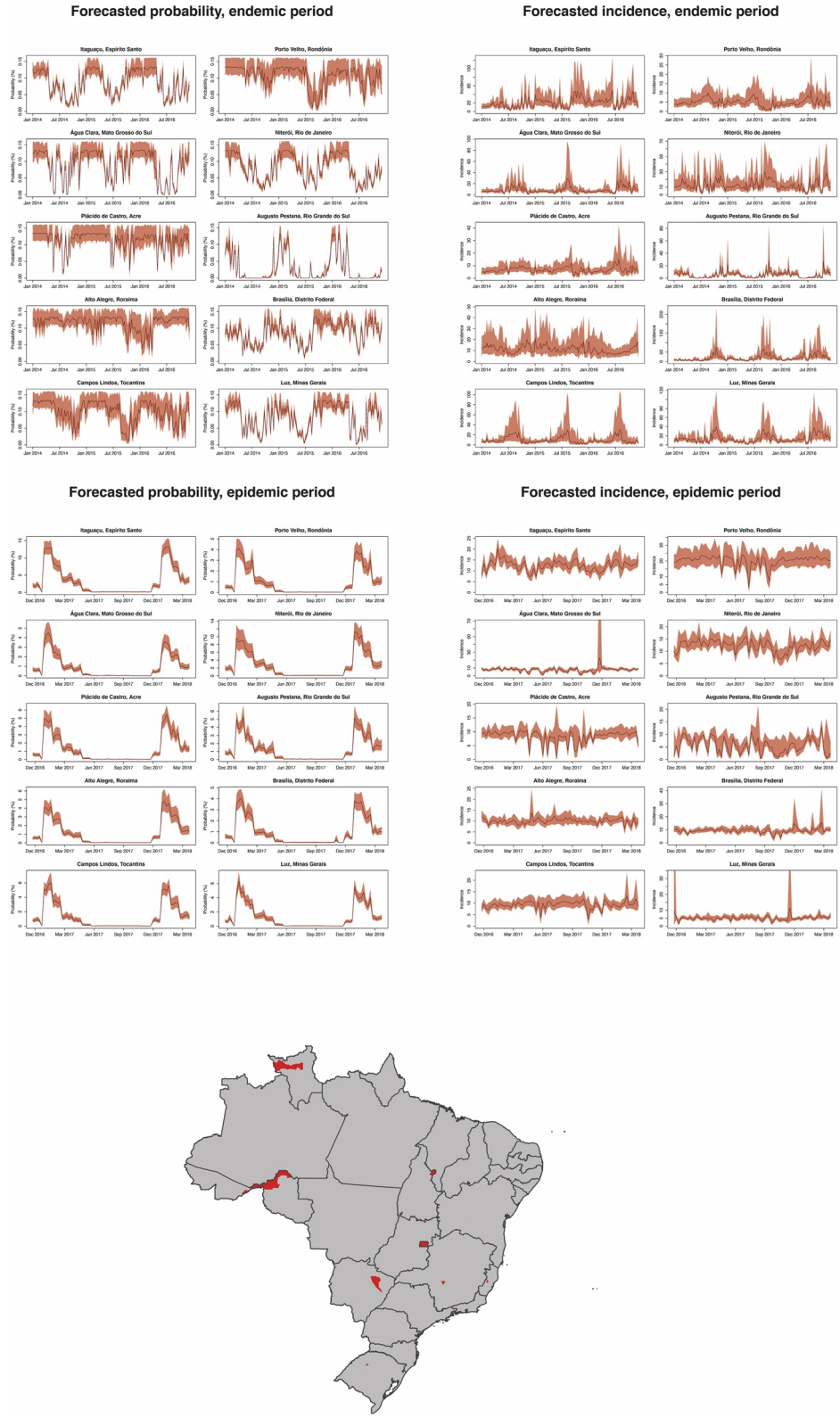

Figure 3. Time series of forecasted probabilities of Yellow Fever occurrence (left column) and incidence (right column) for selected municipalities during the entire testing period for the endemic time period (top row) and the entire epidemic time period (bottom row). 95% prediction intervals are included to show precision of forecasts. The map shows the locations of the selected municipalities, highlighted in red. Map source: base map is publicly available from the Database of Global Administrative Areas, license is available at gadm.org/license.html.

A decrease in AUC was observed in the binomial step for the endemic time period when the training data begin after January 2001. Small changes were seen in MAE of the Gamma step for the endemic time period, with a small increase in MAE when the training data began after July 2000. Keeping the start of the training period at January 2000 and shortening the end of the training period showed that AUC steadily increased as the training period became longer, and MAE stayed consistent when the training period spanned until February 2009 or farther (Figure S12 in Supporting Information S1). Earlier weeks may bear high influence because of higher case counts in the earlier weeks of observation (Figure 1).

Because the training data for the epidemic time period consists of nonconsecutive months, 100 randomly drawn combinations of three nonconsecutive months were assessed to see if results are sensitive to which months were drawn. Changes in AUC and MAE across drawn training/testing periods showed high sensitivity to the choice of training and testing data (Table S6 in Supporting Information S1). Repeating these sensitivity analyses with the sets of best-fitting models showed similarities in trends for training period duration in the endemic period (Figure S12 in Supporting Information S1) and that the combination of months selected for testing data can lead to notable differences in model performance across the various models, though no particular months stand out as best or worst for inclusion in the testing data (Figure S13 in Supporting Information S1).

Forecasting abilities of the models were also assessed across different months to determine if forecasts during certain times of year are most or least accurate. This was done by evaluating the AUC and MAE using only specific months in the testing data, omitting months where no YF cases were seen. The endemic model showed marginally higher accuracy of probability forecasts between July and December and substantially lower accuracy of incidence forecasts in October, December and January. The epidemic model showed highest forecasting accuracy for probability of YF occurrence in March and highest forecasting accuracy for incidence in November (Figure S14 in Supporting Information S1).

Within both the endemic and epidemic models, the binomial and Gamma steps of each were found to produce most accurate forecasts using different lag periods. To consider stakeholder preferences in aligning the lag periods of the two models or using specific lag periods, model performance was compared using lag periods between 2 and 8 weeks. For both time periods, the AUC from the binomial step can show differences across lag periods; the differences in MAE from the Gamma steps are less pronounced (Table S7 in Supporting Information S1).

A global sensitivity analysis was conducted to determine which, if any, variables are most influential in each model step. We calculated first and second-order Sobol indices (Sobol, 1993; Wu et al., 2013) for each model step to measure relative influence. Sobol indices use variance decomposition in order to identify influential predictors or groups of predictors. First-order indices represent individual parameters' influence, and second-order indices combine first-order indices with each parameters' influence in conjunction with each of the other predictors. In the endemic period, temperature was the most influential predictor in both model steps. In the epidemic period, previous YF burden was commonly most influential along with precipitation and drainage density (Table S8 in Supporting Information S1, Figure 4).

4. Discussion

This study aimed to develop forecasting models to predict weekly YF occurrence and incidence in Brazil using meteorological and ecohydrological predictors, particularly targeting short-term forecasts to inform immediate preventative measures. Dynamic variables characterizing weather, particularly temperature, were shown to be influential predictors of both occurrence and incidence of YF among Brazilian municipalities; precipitation, water drainage, and previous YF occurrence were also influential predictors.

Different sets of predictors provided the most accurate forecasts in the endemic (January 2000 to November 2016) and epidemic (December 2016 to March 2018) time periods, which represent time before and after the start of the 2016–2017 outbreak. Little has been done to compare forecasts of the same disease in the same location both before and during an epidemic, but changes in data quality and disease spread during an outbreak are likely to influence model fits. The models demonstrated high predictive accuracy as defined through AUC and MAE values with the testing data, suggesting that environmental factors can be reliably used for forecasting YF burden in Brazil. Seasonality is a known characteristic of YF incidence, with cases primarily occurring between January and April (Cavalcante & Tauil, 2016). The strong predictive ability of these environmental factors for future YF burden may in part result from a shared seasonality. It is clear, however, that the models are

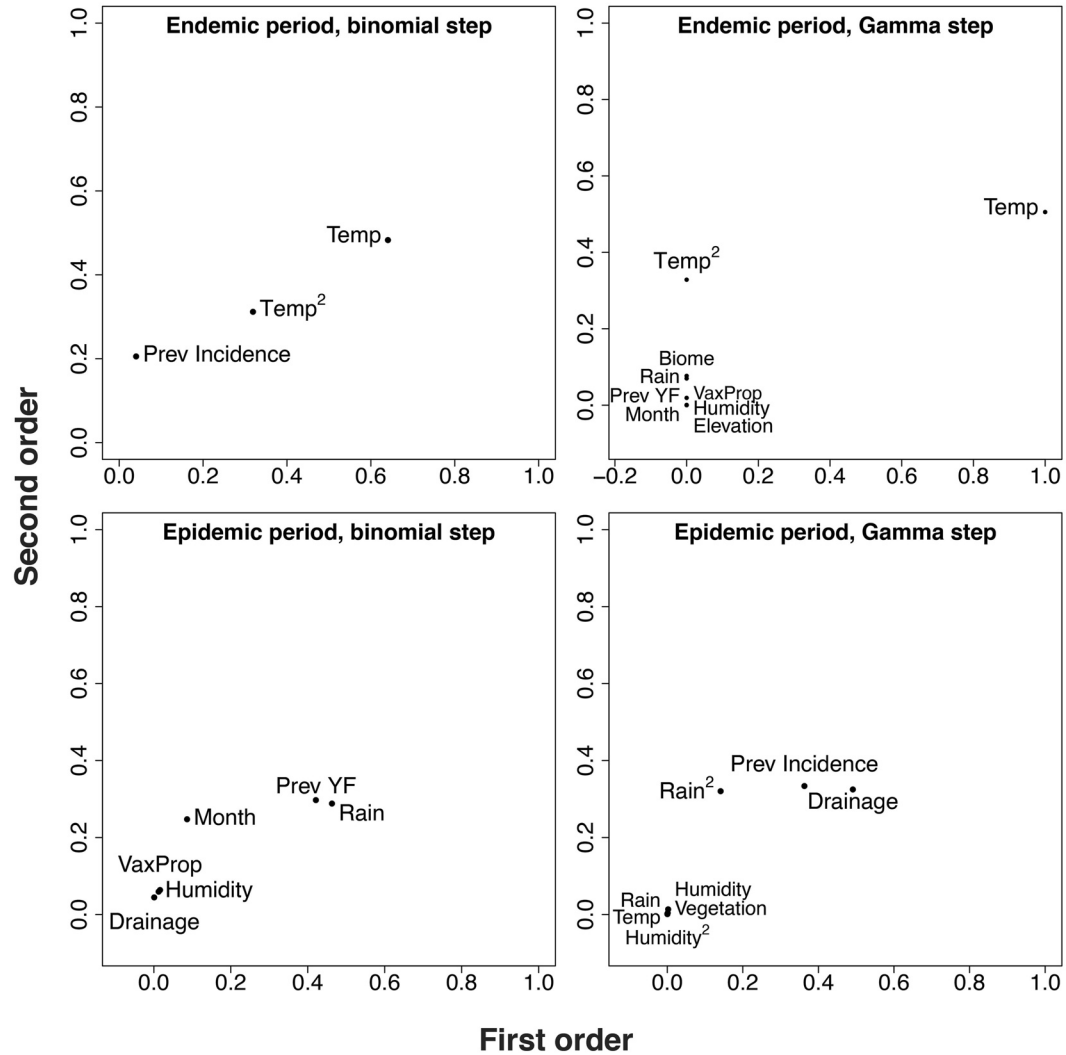

Figure 4. First and second order Sobol Indices for Yellow Fever forecast models. Higher values of Sobol indices indicate greater relative influence for a predictor over model predictions: first and second order quantify the independent and interactive contribution of predictors to model prediction and variability, respectively.

not only detecting seasonality, as noted by the forecasts shown in Figure 3. These forecasts across 10 example municipalities do not only show identical seasonal patterns. Further, the existence of relationships between the predictors and YF burden beyond simple seasonality are evident by the inclusion of predictors beyond month of year and lagged case/incidence. Despite YF occurring more frequently between January and April in Brazil and mosquitoes spreading YF laying eggs more in certain months (Alencar et al., 2013), the endemic period does not consist solely of highly pronounced seasons (Figure 1). A purely seasonal model would likely produce a worse fit compared to the models presented in Tables 2 and 3.

Differences seen in the optimal parameter sets and lag periods across the endemic and epidemic time periods suggest that some environmental factors may contribute differently to YF risk during an outbreak compared to outside of an outbreak. This may also relate to spatial differences in YF burden during the two time periods; cases during the endemic time period were primarily seen in the North and Central-West regions, while cases during the epidemic time period were primarily seen in the Southeast region (Figure 1). In addition to spatial differences, differences in transmission types exist between the two time periods. Cases in the endemic period were primarily from sylvatic transmission, with *Haemagogus* and *Sabethes* mosquitoes as primary vectors. During the epidemic, urban transmission also potentially occurred, where *Aedes* mosquitoes were significant vectors (Cunha et al., 2020). One previous study split the epidemic into two periods and found different contributions of environmental and anthropogenic factors during different times of the same epidemic (de Thoisy et al., 2020).

Differences in lag times are also noteworthy, with 6–7 weeks lags providing the best forecasts in the endemic period, and 2-week lags providing the best forecasts in the epidemic period. Benefits of shorter lag periods when predicting an ongoing epidemic have been noted previously (Convertino et al., 2021). In this study, the shorter lag period during the epidemic may result from higher reliance of previous cases. In the endemic period, cases resulted from sylvatic transmission, and predictors relating to mosquito life cycles may have been most important for forecast accuracy. Life cycles of the mosquito genera that transmit YF typically span up to several weeks (Dégallier et al., 1998; United States Environmental Protection Agency, 2023), making longer time lags more relevant to the impact of ecometeorological factors on YF risk through the mechanism of mosquito life cycles. Different lag periods may also result from mechanistic lags varying among predictors. While we chose to keep consistent lags to consider practical implementation, other studies have found varying time lags between disease risk and environmental conditions, spanning weeks to months (Davis et al., 2018; Horta et al., 2014; Lowe et al., 2018b; Schuster et al., 2011). In the epidemic period, outbreak cases likely included urban transmission, where recent human cases are important. In the context of model fitting, it may follow that a 1-week lag period would have a better model fit, but we did not include this in analyses because it is impossible to obtain case data and weather data from week $t-1$ and produce a forecast for week t until week t has already started. YF case data available for week $t-1$ during week t are further likely subject to incompleteness (McGough et al., 2020). Methods have been proposed to account for this (Beesley et al., 2022), but this is outside the scope of this study.

Differences in predictor sets between the two steps (binomial and Gamma) within the same hurdle model suggests that there may be different contributing factors to seeing any YF cases and to seeing a greater incidence when observing cases. The same candidate predictors were considered for both steps, and the predictor sets that produced the most accurate forecasts are shown in Tables 2 and 3. This was also seen with the temporal lag periods; three different lag periods were observed among the four model components (Tables 2 and 3). Previous research has noted the complex dynamics between environmental factors and vector-borne disease risk (Pascual & Bouma, 2009; Tabachnick, 2010). Results from this study support this notion that the relationships between environmental characteristics and disease risk are not straightforward and may change in different circumstances, which has been observed in Zika research (Harris et al., 2019). Some predictors largely relate to each other, though little multicollinearity was seen among continuous predictors included in the final models presented in Tables 2 and 3. The highest correlations were seen between mean humidity and mean precipitation ($\rho = 0.578$) as well as between range of temperature and range of humidity ($\rho = 0.572$) (Table S9 in Supporting Information S1).

Based on the Sobol indices for model parameters, temperature was the most influential predictor for the model for the endemic period, and rainfall, drainage, and previous YF burden were shown to be most influential predictors in the model for the epidemic period (Figure 4). Connections between these variables have been made previously to YF (de Almeida et al., 2019; De Paiva et al., 2019), other mosquito-borne diseases (Akter et al., 2020), and mosquito vectors in general (Dickens et al., 2018). Both linear and quadratic terms were included for at least one of temperature, precipitation, and humidity in three of the four component models, supporting hypotheses of nonlinear trends connecting disease risk and environmental factors (Githeko et al., 2000; Harris et al., 2019). The existence of both time varying meteorological predictors and static ecohydrological predictors as influential predictors in the models highlights the importance of both factors as likely contributors to YF burden.

Across the models and analyses presented in this study, we identified a collection of factors that successfully predict future YF burden. Most of these are consistent with previous findings regarding their link to risk of YF or other mosquito-borne viruses. Temperature in particular has a well-established link (Alencar et al., 2021; Possas et al., 2018), with previous studies also finding nonlinear relationships with disease burden (Ciota & Keyel, 2019; de Thoisy et al., 2020) and examining extreme temperatures (Marinho et al., 2016). Similarly, precipitation has been previously linked to disease burden, both linearly (Possas et al., 2018; Silva-Inacio et al., 2020) and nonlinearly (de Thoisy et al., 2020) as we found. This has also been observed for humidity (Alencar et al., 2021; de Almeida et al., 2019). Our study considered weekly minimum, maximum, mean, and range, in constrast to other studies selecting one. Other studies have, similar to ours, noted relationships between YF burden and vegetation (de Almeida et al., 2019; Hamlet et al., 2018) as well as landscape types (Andreo et al., 2021; Wilk-da-Silva et al., 2020), which are similar to the biomes we used. Elevation has been previously inversely linked to suitability of mosquito habitats (Kumm et al., 1946). Drainage density with river networks is not commonly used in disease prediction, but issues of standing water in urban environments has been examined previously (Prestes-Carneiro et al., 2023; Souza et al., 2017).

Robustness analyses comparing the selected models shown in Tables 2 and 3 to their corresponding 50 best-performing models show that the variables included in these models are substantial and robust in their contributions to YF forecasting. The frequent inclusion of the predictors from Tables 2 and 3 among these 50 models as well as agreement in their direction of association (Tables S3 and S4 in Supporting Information S1), along with similarities in their forecast outputs (Figures S9 and S10 in Supporting Information S1) show their overall importance, while also showing that other predictors can be included or excluded to produce similar quality forecasts.

While the contributions of these predictors are well confirmed by these robustness analyses, the individual parameter estimates should be interpreted with caution because they may represent associations that are not directly causal or mechanistic in nature, which is a consequence of the study design. For example, the Gamma step of the endemic model predicts higher incidence if YF is not seen 7 weeks prior (Table 2). Rather than interpreting this finding to mean that absence of YF cases is harmful, not seeing YF cases 7 weeks in the past may be indicative of a higher pool of susceptible individuals or past conditions where YF transmission was not sustained, but can intensify affects from other factors. An example of the latter case is a drought removing mosquito breeding grounds followed by rainfall having increased opportunities to create puddles in previously dried up lands. Precipitation and drought are capable of both creating and reducing mosquito breeding grounds (Caldwell et al., 2021; Lowe et al., 2018b). It is also likely that some models contain predictors that contribute redundant prediction information. We omitted interaction terms to avoid unnecessary complexity in these models, but this allows part of the contribution of one predictor in forecasting to be represented by another predictor, even if inclusion of both is beneficial for forecast accuracy.

From the models generated in this study, regular weekly forecasts can be produced that can identify specific municipalities that would be most likely to benefit from immediate actions to prevent YF infections. By updating the meteorological and environmental data used in model fitting, forecasts can be continually produced nationwide. Extrapolating the forecasts beyond the 462 municipalities where YF was seen between January 2000 and March 2018 allows all municipalities to be included in produced forecasts and therefore be targeted for preventions.

The model fit criteria, AUC for the binomial step and MAE for the Gamma step, showed that the models fit in this study showed high predictive accuracy. AUC values of 0.851 and 0.902 in the endemic and epidemic time periods respectively indicate that the binomial steps of both models can reliably predict whether any YF cases will be seen. The MAE values of 7.009 and 4.860 indicate the average difference, in cases per 100,000 people, between observed and predicted incidence values when incidence is greater than zero. Comparisons of the AUC and MAE between the selected models and null models show a notably larger increase in AUC between the selected model and null model compared to the reduction in MAE between the selected model and null model. However, it is challenging to definitively compare AUC and MAE because they represent fits of different data types. A low sample size of municipality-weeks with positive incidence contributes to the challenge of forecasting incidence; doing so when nonzero incidence is more frequent may prove more successful.

The model forms used in other works predicting YF risk bear similarities and differences to the Gamma hurdle model used in this study. Many other studies use binary regression in order to compare occurrence and absence of YF (Hamlet et al., 2018; Hamrick et al., 2017; Kaul et al., 2018; Kraemer et al., 2017), while similar methods such as regression trees (Shearer et al., 2018a) are used in other works. Others, rather than using statistical models, use mathematical or mechanistic models (Bonin et al., 2018; Childs et al., 2019; Glover & White, 2020), including in conjunction with a statistical model (Jean et al., 2020). While many of these methods produce either predictions of YF occurrence or predicted disease burden, the hurdle model developed in this study provides both pieces of information, allowing more informed decisions to be made.

The objective of the modeling in this study was to produce accurate weekly forecasts for YF in order to predict locations where it is likely to be seen along with an estimate of incidence if seen. Similarly motivated work has been seen with dengue, where precipitation and temperature were used to distinguish between epidemic and non-epidemic years in seven cities in Brazil using a machine learning approach (Stolerman et al., 2019). Other work sought one-time associations, where weather patterns were used along with human, primate, and mosquito population data to estimate spillover of YF from nonhuman primates to humans in Brazil using a mechanistic and statistical model together (Childs et al., 2019).

4.1. Implications for Preparedness

These analyses were developed to consider use for preparedness planning, whether by government or research-based groups. This is primarily done through the produced forecasts that can be generated, similar to those in Figure 2. Knowledge of locations that are more likely to see any YF cases or greater incidence of YF can be useful for conserving resources dedicated to public health messaging, preparing medical facilities to expect cases, or other preventative measures. Given the likely changes in YF dynamics in Brazil after March 2018 as a result of the outbreak eventually concluding, implementing a forecast model such as this would require an update. Repeating the analyses with more recent data would allow the methods and findings of this study to most effectively inform policy and preparedness efforts.

The example forecasts shown in Figure 2 show some complexities of providing both forecasted probability of occurrence and forecasted incidence. The forecasts for the week of 11–17 May 2015 (Figure 2a) show that municipalities with the lowest probability of occurrence, but highest estimated incidence, are found throughout the Southeast and South regions. These municipalities saw conditions where observing YF cases is unlikely compared to other locations, but observing high incidence is likely if cases are seen. These competing statements produce communication challenges; policy recommendations will rely on tradeoffs between prioritizing probability of occurrence or incidence. Combining these pieces of information into one map, as shown in Figure 2 (right column), may prove useful in allowing municipalities with high estimated incidence but low probability of observing nonzero incidence, to be shown as low-risk. The probability threshold, however, is an arbitrary decision that should be made in conjunction with stakeholders.

Presenting regular forecasts to stakeholders for preparedness is an inherently challenging activity, as current forecast communication are typically individualized, and there exist few real-world examples (George et al., 2019; Tushar & Reich, 2017). Implementing the forecasting models developed in this study and then presenting regular, interpretable forecasts requires computational resources and communication strategies tailored to stakeholders. This is discussed in more detail in Supporting Information S1.

4.2. Limitations and Future Directions

While this study aimed to focus on environmental predictors and made use of publicly available data, other factors not accounted for in these analyses could yield more accurate predictions and explain more of the dynamics of YF spread (Harris et al., 2019), though potentially with a loss of usability. Other works have used human mobility (Kraemer et al., 2017), mosquito movements, and nonhuman primate movement (Childs et al., 2019; Kaul et al., 2018). However, these data are difficult to acquire at a high quality, particularly across a large nation such as Brazil and over several years of observation. Within the environmental predictors we used, other methods of applying them exist beyond minimum, maximum, mean, and range of daily values each week. Numerous possibilities exist, such as apply diurnal temperature ranges, cumulative time within particular temperature ranges, or numbers of days with precipitation. Even without these data, the models presented in this study were shown to produce accurate forecasts as measured by AUC and MAE.

In aiming to determine the utility of primarily environmental predictors and focusing on YF determinants that are not potential targets of intervention, certain determinants of YF risk related to human activities are not accounted for in the models. These can include socioeconomic status (Jean et al., 2020), and specific human mobility patterns (Kraemer et al., 2017). However, while these factors can be highly influential in YF risk prediction, the results of this study, focusing solely on the environmental predictors, showed that accurate forecasts can still be achieved, particularly when forecasting YF occurrence, as evidenced by high AUC values.

Using the models generated in this study today, where one set of parameter estimates applies to the entire nation and the entire time period of observation, relies on the assumption that all municipalities of Brazil are currently capable of seeing YF cases. There may exist characteristics of municipalities outside the 462 used in model fitting that lead to not being susceptible (Servadio et al., 2022). We also assumed that the ability to detect and report YF cases is uniform across all municipalities and that YF can be observed and reported, even if it had not been reported between 2000 and 2018. The forecasts produced in Figure 2 among municipalities outside the 462 used in model fitting assume that YF was not observed in those locations. It is likely that these abilities vary across locations as well as over time. The overall sparsity of YF data, appearing in less than a tenth of municipalities, creates difficulty in assessing spatial variation in forecast ability. Further, different areas of Brazil see

different YF vectors, with urban areas having primarily *Aedes* mosquitoes and western parts of the country seeing more *Haemagogus* and *Sabethes* mosquitoes. However, the municipalities used in model fitting represent several geographic areas of Brazil, suggesting they may be adequately representative of the nation from an environmental perspective. Additionally, use of the model today would likely require an extension to repeat the methods using recent data following the end of the YF outbreak. Doing so would allow comparisons of YF dynamics before, during, and after a large outbreak.

5. Conclusions

Weekly occurrence and incidence of YF can be forecast accurately using meteorological and ecohydrological predictors as well as previous cases. Specifically, forecasts based on temperature and incorporating previous YF occurrence were the most accurate both before and after the start of a major epidemic. The results from these models are useable for informing measures to prevent YF cases by providing weekly forecasts for the probability of YF occurrence and estimated YF incidence. Basing forecasts on environmental predictors, as done in this study, allows seasonality of YF to be considered and identifies environmental determinants that may be associated to ecological controls. The methods used in this study are translatable to similar vector-borne diseases and demonstrate the use of forecasting models for informing public health science and practice.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Sample data and relevant code for this study can be found at github.com/jlservadio/YF_Forecast and Zenodo (Servadio, 2023).

Publicly available data used in this study include population data by municipality (Instituto Brasileiro de Geografia Estastistica, 2018), files "Estimativas_2000" through "Estimativas_2018"; temperature, precipitation, and humidity data (Gelaro et al., 2017; Global Modeling and Assimilation Office (GMAO), 2015); shape and locations of municipalities (Database of Global Administrative Areas, 2018); elevation (NASA & Ministry of Economics, Trade, and Industry, 2019); vegetation (Tuck et al., 2014); and biomes (Olson et al., 2001). Previously published data used in this study include vaccine coverage (Shearer et al., 2018b).

Acknowledgments

The authors thank the Pan American Health Organization and Brazilian Ministry of Health for providing the Yellow Fever data used in this study. The authors also thank Fernando Nardi and Antonio Annis at the University for Foreigners of Perugia, Italy, for providing the drainage density data. The authors acknowledge the use of the resources from the Minnesota Supercomputing Institute and the Pennsylvania State University Institute for Computational and Data Sciences for data management and model fitting. This study was funded by a letter of agreement from the Pan American Health Organization and the Doctoral Dissertation Fellowship from the University of Minnesota Graduate School. M. C. acknowledges support of the Shenzhen Pengcheng Talent Award.

References

Addy, P. A., Minami, K., & Agadzi, V. K. (1986). Recent yellow fever epidemics in Ghana (1969-1983). *East African Medical Journal*, *63*(6), 422–434.

Akter, R., Hu, W., Gatton, M., Bambrick, H., Naish, S., & Tong, S. (2020). Different responses of dengue to weather variability across climate zones in Queensland, Australia. *Environmental Research*, *184*, 109222. https://doi.org/10.1016/j.envres.2020.109222

Alencar, J., de Mello, C. F., Leite, P. J., Bastos, A. Q., Silva, S. O. F., Serdeiro, M., et al. (2021). Oviposition activity of *Haemagogus leucocelaenus* (Diptera: Culicidae) during the rainy and dry seasons, in areas with yellow fever virus circulation in the Atlantic Forest, Rio de Janeiro, Brazil. *PLoS One*, *16*(12), e0261283. https://doi.org/10.1371/JOURNAL.PONE.0261283

Alencar, J., Morone, F., De Mello, C. F., Dégallier, N., Lucio, P. S., Da Serra-Freire, N. M., & Guimarães, A. É. (2013). Flight height preference for oviposition of mosquito (Diptera: Culicidae) vectors of sylvatic yellow fever virus near the hydroelectric reservoir of Simplício, Minas Gerais, Brazil. *Journal of Medical Entomology*, *50*(4), 791–795. https://doi.org/10.1603/ME12120

Anderson, S. C. (2014). Gamma hurdle models - Fitting and interpreting Gamma hurdle models. Retrieved from https://seananderson.ca/2014/05/18/gamma-hurdle/

Andreo, V., Porcasi, X., Guzman, C., Lopez, L., & Scavuzzo, C. M. (2021). Spatial distribution of *Aedes aegypti* oviposition temporal patterns and their relationship with environment and dengue incidence. *Insects*, *12*(10), 919. https://doi.org/10.3390/INSECTS12100919

Aristizabal, J. F., Lévêque, L., Chapman, C. A., & Serio-Silva, J. C. (2018). Impacts of temperature on behaviour of the Mexican endangered black howler monkey *Alouatta pigra* Lawrence, 1933 (Primates: Atelidae) in a fragmented landscape. *Acta Zoologica Bulgarica*, *70*(3), 377–382.

Aspvik, N. P., Viken, H., Ingebrigtsen, J. E., Zisko, N., Mehus, I., Wisløff, U., & Stensvold, D. (2018). Do weather changes influence physical activity level among older adults? – The generation 100 study. *PLoS One*, *13*(7), e0199463. https://doi.org/10.1371/journal.pone.0199463

Barros, M. L., & Boecken, G. (1996). Jungle yellow fever in the central Amazon. *Lancet (London, England)*, *348*(9032), 969–970. https://doi.org/10.1016/s0140-6736(05)65392-5

Bayoh, M. N., & Lindsay, S. W. (2003). Effect of temperature on the development of the aquatic stages of *Anopheles gambiae* sensu stricto (Diptera: Culicidae). *Bulletin of Entomological Research*, *93*(5), 375–381. https://doi.org/10.1079/ber2003259

Beesley, L. J., Osthus, D., & Del Valle, S. Y. (2022). Addressing delayed case reporting in infectious disease forecast modeling. *PLoS Computational Biology*, *18*(6), e1010115. https://doi.org/10.1371/JOURNAL.PCBI.1010115

Benedum, C. M., Shea, K. M., Jenkins, H. E., Kim, L. Y., & Markuzon, N. (2020). Weekly dengue forecasts in Iquitos, Peru; San Juan, Puerto Rico; and Singapore. *PLoS Neglected Tropical Diseases*, *14*(10), 1–26. https://doi.org/10.1371/JOURNAL.PNTD.0008710

Bonin, C. R. B., Fernandes, G. C., dos Santos, R. W., & Lobosco, M. (2018). A qualitatively validated mathematical-computational model of the immune response to the yellow fever vaccine. *BMC Immunology*, *19*(1), 15. https://doi.org/10.1186/s12865-018-0252-1

Caldwell, J. M., LaBeaud, A. D., Lambin, E. F., Stewart-Ibarra, A. M., Ndenga, B. A., Mutuku, F. M., et al. (2021). Climate predicts geographic and temporal variation in mosquito-borne disease dynamics on two continents. *Nature Communications*, *12*(1), 1233. https://doi.org/10.1038/S41467-021-21496-7

Cavalcante, K. R. L. J., & Tauil, P. L. (2016). Epidemiological characteristics of yellow fever in Brazil, 2000-2012. *Epidemiologia e Servicos de Saude: Revista do Sistema Unico de Saude do Brasil*, *25*(1), 11–20. https://doi.org/10.5123/S1679-49742016000100002

Centers for Disease Control and Prevention. (n.d.). Mosquito life cycle.

Centers for Disease Control and Prevention. (2018). Areas with risk of Yellow Fever virus transmission in South America. Retrieved from https://www.cdc.gov/yellowfever/maps/south_america.html

Chae, S., Kwon, S., & Lee, D. (2018). Predicting infectious disease using deep learning and big data. *International Journal of Environmental Research and Public Health*, *15*(8), 1596. https://doi.org/10.3390/ijerph15081596

Champredon, D., & Dushoff, J. (2015). Intrinsic and realized generation intervals in infectious-disease transmission. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1821), 20152026. https://doi.org/10.1098/rspb.2015.2026

Chen, Y., Chu, C. W., Chen, M. I. C., & Cook, A. R. (2018). The utility of LASSO-based models for real time forecasts of endemic infectious diseases: A cross country comparison. *Journal of Biomedical Informatics*, *81*, 16–30. https://doi.org/10.1016/j.jbi.2018.02.014

Chen, Y., Ong, J. H. Y., Rajarethinam, J., Yap, G., Ng, L. C., & Cook, A. R. (2018). Neighbourhood level real-time forecasting of dengue cases in tropical urban Singapore. *BMC Medicine*, *16*(1), 129. https://doi.org/10.1186/s12916-018-1108-5

Childs, M. L., Nova, N., Colvin, J., & Mordecai, E. A. (2019). Mosquito and primate ecology predict human risk of yellow fever virus spillover in Brazil. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1782), 20180335. https://doi.org/10.1098/rstb.2018.0335

Choi, Y., Tang, C. S., McIver, L., Hashizume, M., Chan, V., Abeyasinghe, R. R., et al. (2016). Effects of weather factors on dengue fever incidence and implications for interventions in Cambodia. *BMC Public Health*, *16*(1), 241. https://doi.org/10.1186/s12889-016-2923-2

Chowell, G., Viboud, C., Simonsen, L., & Moghadas, S. M. (2016). Characterizing the reproduction number of epidemics with early subexponential growth dynamics. *Journal of the Royal Society Interface*, *13*(123), 20160659. https://doi.org/10.1098/rsif.2016.0659

Ciota, A. T., & Keyel, A. C. (2019). The role of temperature in transmission of zoonotic arboviruses. *Viruses*, *11*(11), 1013. https://doi.org/10.3390/V11111013

Collins, N. D., & Barrett, A. D. T. (2017). Live attenuated Yellow Fever 17D vaccine: A legacy vaccine still controlling outbreaks in modern day. *Current Infectious Disease Reports*, *19*(3), 14. https://doi.org/10.1007/s11908-017-0566-9

Convertino, M., Reddy, A., Liu, Y., & Munoz-Zanzi, C. (2021). Eco-epidemiological scaling of Leptospirosis: Vulnerability mapping and early warning forecasts. *Science of the Total Environment*, *799*, 149102. https://doi.org/10.1016/J.SCITOTENV.2021.149102

Cunha, M. S., Tubaki, R. M., de Menezes, R. M. T., Pereira, M., Caleiro, G. S., Coelho, E., et al. (2020). Possible non-sylvatic transmission of yellow fever between non-human primates in São Paulo city, Brazil, 2017-2018. *Scientific Reports*, *10*(1), 15751. https://doi.org/10.1038/s41598-020-72794-x

Database of Global Administrative Areas. (2018). GADM data. version 3.6. [Dataset]. Retrieved from https://gadm.org/download_country_v3.html

Davis, J. K., Vincent, G. P., Hildreth, M. B., Kightlinger, L., Carlson, C., & Wimberly, M. C. (2018). Improving the prediction of arbovirus outbreaks: A comparison of climate-driven models for West Nile virus in an endemic region of the United States. *Acta Tropica*, *185*, 242–250. https://doi.org/10.1016/J.ACTATROPICA.2018.04.028

de Almeida, M. A. B., dos Santos, E., Cardoso, J. D. C., da Silva, L. G., Rabelo, R. M., & Bicca-Marques, J. C. (2019). Predicting Yellow Fever through species distribution modeling of virus, vector, and monkeys. *EcoHealth*, *16*(1), 95–108. https://doi.org/10.1007/s10393-018-1388-4

Dégallier, N., Sa Filho, G. C., Monteiro, H. A. O., Castro, F. C., Da Silva, O. V., Brandão, R. C. F., et al. (1998). Release-recapture experiments with canopy mosquitoes in the genera Haemagogus and Sabethes (Diptera: Culicidae) in Brazilian Amazonia. *Journal of Medical Entomology*, *35*(6), 931–936. https://doi.org/10.1093/JMEDENT/35.6.931

de Goes, P., Guimaraes, J. C., Machado, R. D., Lobo, G. G., Andrade, C. M., Bastos, R. A., & de Paola, D. (1976). On an outbreak of sylvan yellow fever verified in the State of Goias (Brazil) in the period from 1972--1973. *Anais de Microbiologia*, *22*, 9–34.

De Paiva, C. A., Oliveira, A. P. D. S., Muniz, S. S., Calijuri, M. L., Dos Santos, V. J., & Alves, S. D. C. (2019). Determination of the spatial susceptibility to yellow fever using a multicriteria analysis. *Memorias do Instituto Oswaldo Cruz*, *114*(3), e180509. https://doi.org/10.1590/0074-02760180509

de Thoisy, B., Silva, N. I. O., Sacchetto, L., de Souza Trindade, G., & Drumond, B. P. (2020). Spatial epidemiology of yellow fever: Identification of determinants of the 2016-2018 epidemics and at-risk areas in Brazil. *PLoS Neglected Tropical Diseases*, *14*(10), 1–22. https://doi.org/10.1371/JOURNAL.PNTD.0008691

Dickens, B. L., Sun, H., Jit, M., Cook, A. R., & Carrasco, L. R. (2018). Determining environmental and anthropogenic factors which explain the global distribution of *Aedes aegypti* and *Ae. Albopictus*. *BMJ Global Health*, *3*(4), e000801. https://doi.org/10.1136/bmjgh-2018-000801

Doyle, T. J., Glynn, M. K., & Groseclose, S. L. (2002). Completeness of notifiable infectious disease reporting in the United States: An analytical literature review. *American Journal of Epidemiology*, *155*(9), 866–874. https://doi.org/10.1093/aje/155.9.866

Environmental Systems Research Institute (ESRI). (2018). ArcGIS Pro.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., et al. (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, *45*(2), 2004. https://doi.org/10.1029/2005RG000183

Fischer, L. S., Santibanez, S., Hatchett, R. J., Jernigan, D. B., Meyers, L. A., Thorpe, P. G., & Meltzer, M. I. (2016). CDC Grand Rounds: Modeling and public health decision-making. *MMWR. Morbidity and Mortality Weekly Report*, *65*(48), 1374–1377. https://doi.org/10.15585/mmwr.mm6548a4

Gage, K. L., Burkot, T. R., Eisen, R. J., & Hayes, E. B. (2008). Climate and vectorborne diseases. *American Journal of Preventive Medicine*, *35*(5), 436–450. https://doi.org/10.1016/j.amepre.2008.08.030

Gardner, C. L., & Ryman, K. D. (2010). Yellow fever: A reemerging threat. *Clinics in Laboratory Medicine. NIH Public Access*, *30*(1), 237–260. https://doi.org/10.1016/j.cll.2010.01.001

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

George, D. B., Taylor, W., Shaman, J., Rivers, C., Paul, B., O'Toole, T., et al. (2019). Technology to advance infectious disease forecasting for outbreak management. *Nature Communications*, *10*(1), 1–4. https://doi.org/10.1038/s41467-019-11901-7

Githeko, A. K., Lindsay, S. W., Confalonieri, U. E., & Patz, J. A. (2000). Climate change and vector-borne diseases: A regional analysis. *Bulletin of the World Health Organization*, *78*(9), 1136–1147. https://doi.org/10.1590/S0042-96862000000900009

Global Modeling and Assimilation Office (GMAO). (2015). MERRA-2 tavg1_2d_flx_Nx: 2d,1-hourly, time-averaged, single-level, assimilation, surface flux diagnostics V5.12.4 [Dataset]. Goddard Earth Sciences Data and Information Services Center (GES DISC). https://doi.org/10.5067/7MCPBJ41Y0K6

Glover, A., & White, A. (2020). A vector-host model to assess the impact of superinfection exclusion on vaccination strategies using dengue and yellow fever as case studies. *Journal of Theoretical Biology*, *484*, 110014. https://doi.org/10.1016/j.jtbi.2019.110014

Goldani, L. Z. (2017). Yellow fever outbreak in Brazil, 2017. *Brazilian Journal of Infectious Diseases*, *21*(2), 123–124. https://doi.org/10.1016/J.BJID.2017.02.004

Gotuzzo, E., Yactayo, S., & Córdova, E. (2013). Efficacy and duration of immunity after yellow fever vaccination: Systematic review on the need for a booster every 10 years. *The American Journal of Tropical Medicine and Hygiene*, *89*(3), 434–444. https://doi.org/10.4269/AJTMH.13-0264

Hamlet, A., Gaythorpe, K. A. M., Garske, T., & Ferguson, N. M. (2021). Seasonal and inter-annual drivers of yellow fever transmission in South America. *PLoS Neglected Tropical Diseases*, *15*(1), 1–18. https://doi.org/10.1371/JOURNAL.PNTD.0008974

Hamlet, A., Jean, K. K., Perea, W., Yactayo, S., Biey, J., Van Kerkhove, M., et al. (2018). The seasonal influence of climate and environment on yellow fever transmission across Africa. *PLoS Neglected Tropical Diseases*, *12*(3), e0006284. https://doi.org/10.1371/journal.pntd.0006284

Hamrick, P. N., Aldighieri, S., Machado, G., Leonel, D. G., Vilca, L. M., Uriona, S., & Schneider, M. C. (2017). Geographic patterns and environmental factors associated with human yellow fever presence in the Americas. *PLoS Neglected Tropical Diseases*, *11*(9), e0005897. https://doi.org/10.1371/journal.pntd.0005897

Harris, M., Caldwell, J. M., & Mordecai, E. A. (2019). Climate drives spatial variation in Zika epidemics in Latin America. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1909), 20191578. https://doi.org/10.1098/rspb.2019.1578

Hayes, R. O., Francy, D. B., Lazuick, J. S., Smith, G. C., & Jones, R. H. (1976). Arbovirus surveillance in six states during 1972. *The American Journal of Tropical Medicine and Hygiene*, *25*(3), 463–476. https://doi.org/10.4269/ajtmh.1976.25.463

Horta, M. A., Bruniera, R., Ker, F., Catita, C., & Ferreira, A. P. (2014). Temporal relationship between environmental factors and the occurrence of dengue fever. *International Journal of Environmental Health Research*, *24*(5), 471–481. https://doi.org/10.1080/09603123.2013.865713

Imai, C., Armstrong, B., Chalabi, Z., Mangtani, P., & Hashizume, M. (2015). Time series regression model for infectious disease and weather. *Environmental Research*, *142*, 319–327. https://doi.org/10.1016/j.envres.2015.06.040

Imai, C., Brooks, W. A., Chung, Y., Goswami, D., Anjali, B. A., Dewan, A., et al. (2014). Tropical influenza and weather variability among children in an urban low-income population in Bangladesh. *Global Health Action*, *7*(1), 24413. https://doi.org/10.3402/gha.v7.24413

Instituto Brasileiro de Geografia Estastistica. (2018). Population estimates [Dataset]. Retrieved from https://www.ibge.gov.br/estatisticas-novo-portal/sociais/populacao/9103-estimativas-de-populacao.html?edicao=17283&t=downloads

Jean, K., Hamlet, A., Benzler, J., Cibrelus, L., Gaythorpe, K. A. M., Sall, A., et al. (2020). Eliminating yellow fever epidemics in Africa: Vaccine demand forecast and impact modelling. *PLoS Neglected Tropical Diseases*, *14*(5), e0008304. https://doi.org/10.1371/journal.pntd.0008304

Johansson, M. A., Vasconcelos, P. F. C., & Staples, J. E. (2014). The whole iceberg: Estimating the incidence of yellow fever virus infection from the number of severe cases. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *108*(8), 482–487. https://doi.org/10.1093/trstmh/tru092

Kakarla, S. G., Caminade, C., Mutheneni, S. R., Morse, A. P., Upadhyayula, S. M., Kadiri, M. R., & Kumaraswamy, S. (2019). Lag effect of climatic variables on dengue burden in India. *Epidemiology and Infection*, *147*, e170. https://doi.org/10.1017/S0950268819000608

Kaul, R. B., Evans, M. V., Murdock, C. C., & Drake, J. M. (2018). Spatio-temporal spillover risk of yellow fever in Brazil. *Parasites & Vectors*, *11*(1), 488. https://doi.org/10.1186/s13071-018-3063-6

Kraemer, M. U. G., Faria, N. R., Reiner, R. C., Golding, N., Nikolay, B., Stasse, S., et al. (2017). Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: A modelling study. *The Lancet Infectious Diseases*, *17*(3), 330–338. https://doi.org/10.1016/S1473-3099(16)30513-8

Kumm, H. W., Osorno-Mesa, E., & Boshell-Manrique, J. (1946). Periodicity in the annual incidence of reported cases of yellow fever during the past fifty years. *The American Journal of Tropical Medicine and Hygiene*, *43*(1), 13–28. https://doi.org/10.1093/oxfordjournals.aje.a119048

Laneri, K., Cabella, B., Prado, P. I., Coutinho, R. M., & Kraenkel, R. A. (2019). Climate drivers of malaria at its southern fringe in the Americas. *PLoS One*, *14*(7), e0219249. https://doi.org/10.1371/journal.pone.0219249

Litvoc, M. N., Novaes, C. T. G. F., & Lopes, M. I. B. F. (2018). Yellow fever. *Revista da Associacao Medica Brasileira*, *64*(2), 106–113. https://doi.org/10.1590/1806-9282.64.02.106

Liu, H. N., Gao, L. D., Chowell, G., Hu, S. X., Lin, X. L., Li, X. J., et al. (2014). Time-specific ecologic niche models forecast the risk of hemorrhagic fever with renal syndrome in Dongting Lake District, China, 2005–2010. *PLoS One*, *9*(9), e106839. https://doi.org/10.1371/JOURNAL.PONE.0106839

Lowe, R., Bailey, T. C., Stephenson, D. B., Jupp, T. E., Graham, R. J., Barcellos, C., & Carvalho, M. S. (2013). The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Statistics in Medicine*, *32*(5), 864–883. https://doi.org/10.1002/SIM.5549

Lowe, R., Barcellos, C., Brasil, P., Cruz, O. G., Honório, N. A., Kuper, H., & Carvalho, M. S. (2018a). The Zika virus epidemic in Brazil: From discovery to future implications. *International Journal of Environmental Research and Public Health*, *15*(1), 96. https://doi.org/10.3390/ijerph15010096

Lowe, R., Gasparrini, A., Van Meerbeeck, C. J., Lippi, C. A., Mahon, R., Trotman, A. R., et al. (2018b). Nonlinear and delayed impacts of climate on dengue risk in Barbados: A modelling study. *PLoS Medicine*, *15*(7), e1002613. https://doi.org/10.1371/JOURNAL.PMED.1002613

Lutz, C. S., Huynh, M. P., Schroeder, M., Anyatonwu, S., Dahlgren, F. S., Danyluk, G., et al. (2019). Applying infectious disease forecasting to public health: A path forward using influenza forecasting examples. *BMC Public Health*, *19*(1), 1659. https://doi.org/10.1186/s12889-019-7966-8

Marinho, R. A., Beserra, E. B., Bezerra-Gusmão, M. A., Porto, V. D. S., Olinda, R. A., & dos Santos, C. A. C. (2016). Effects of temperature on the life cycle, expansion, and dispersion of *Aedes aegypti* (Diptera: Culicidae) in three cities in Paraiba, Brazil. *Journal of Vector Ecology: Journal of the Society for Vector Ecology*, *41*(1), 1–10. https://doi.org/10.1111/JVEC.12187

McGough, S. F., Johansson, M. A., Lipsitch, M., & Menzies, N. A. (2020). Nowcasting by Bayesian Smoothing: A flexible, generalizable model for real-time epidemic tracking. *PLoS Computational Biology*, *16*(4), e1007735. https://doi.org/10.1371/JOURNAL.PCBI.1007735

Mohamed, J., Mohamed, A. I., & Daud, E. I. (2022). Evaluation of prediction models for the malaria incidence in Marodijeh Region, Somaliland. *Journal of Parasitic Diseases: Official Organ of the Indian Society for Parasitology*, *46*(2), 395–408. https://doi.org/10.1007/S12639-021-01458-Y

Monath, T. P., & Vasconcelos, P. F. C. C. (2015). Yellow fever. *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology*, *64*, 160–173. https://doi.org/10.1016/j.jcv.2014.08.030

Montgomery, D. R., & Dietrich, W. E. (1989). Source areas, drainage density, and channel initiation. *Water Resources Research*, *25*(8), 1907–1918. https://doi.org/10.1029/WR025I008P01907

Moreira-Soto, A., Torres, M. C., Lima de Mendonça, M. C., Mares-Guia, M. A., dos Santos Rodrigues, C. D., Fabri, A. A., et al. (2018). Evidence for multiple sylvatic transmission cycles during the 2016-2017 yellow fever virus outbreak, Brazil. *Clinical Microbiology and Infection*, *24*(9), 1019.e1–1019.e4. https://doi.org/10.1016/j.cmi.2018.01.026

NASA, & Ministry of Economics, Trade, and Industry. (2019). ASTER Global Digital Elevation Model Version 3 [Dataset]. https://doi.org/10.5067/ASTER/ASTGTM.003

Ogashawara, I., Li, L., & Moreno-Madriñán, M. J. (2019). Spatial-temporal assessment of environmental factors related to dengue outbreaks in São Paulo, Brazil. *GeoHealth*, *3*(8), 202–217. https://doi.org/10.1029/2019GH000186

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., et al. (2001). Terrestrial ecoregions of the world: A new map of life on earth. *BioScience*, *51*(11), 933–938. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Opayele, A. V., Adekunle, A. J., Ibrahim, K. T., & Olaleye, D. O. (2017). Influence of meteorological variables on diversity and abundance of mosquito vectors in two livestock farms in Ibadan, Nigeria: Public health implications. *Journal of Mosquito Research*, *7*(9), 70. https://doi.org/10.5376/jmr.2017.07.0009

2011). Outbreak news. Yellow fever, Uganda. *Releve Epidemiologique Hebdomadaire*, *86*(5), 37–38.

Paaijmans, K. P., Read, A. F., & Thomas, M. B. (2009). Understanding the link between malaria risk and climate. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(33), 13844–13849. https://doi.org/10.1073/pnas.0903423106

Pascual, M., & Bouma, M. J. (2009). Do rising temperatures matter? *Ecology*, *90*(4), 906–912. https://doi.org/10.1890/08-0730.1

Patil, S., & Pandya, S. (2021). Forecasting dengue hotspots associated with variation in meteorological parameters using regression and time series models. *Frontiers in Public Health*, *9*, 798034. https://doi.org/10.3389/FPUBH.2021.798034

Possas, C., Lourenço-de-Oliveira, R., Tauil, P. L., de Paula Pinheiro, F., Pissinatti, A., da Cunha, R. V., et al. (2018). Yellow fever outbreak in Brazil: The puzzle of rapid viral spread and challenges for immunisation. *Memorias do Instituto Oswaldo Cruz. NLM (Medline)*, *113*(10), e180278. https://doi.org/10.1590/0074-02760180278

Prestes-Carneiro, L. E., Barbosa Souza, A., Belussi, G. L., Grande, G. H. D., Bertacco, E. A. M., Vieira, A. G., & Flores, E. F. (2023). Dengue outbreaks in a city with recent transmission in São Paulo state, Brazil. *International Journal of Environmental Health Research*, 1–14. https://doi.org/10.1080/09603123.2023.2199972

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J. M., et al. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling*, *5*, 256–263. https://doi.org/10.1016/j.idm.2020.02.002

Ruiz, D., Poveda, G., Vélez, I. D., Quiñones, M. L., Rúa, G. L., Velásquez, L. E., & Zuluaga, J. S. (2006). Modelling entomological-climatic interactions of Plasmodium falciparum malaria transmission in two Colombian endemic-regions: Contributions to a National Malaria Early Warning System. *Malaria Journal*, *5*(1), 66. https://doi.org/10.1186/1475-2875-5-66

Saunders, S. P., Ries, L., Neupane, N., Isabel Ramírez, M., García-Serrano, E., Rendón-Salinas, E., & Zipkin, E. F. (2019). Multiscale seasonal factors drive the size of winter monarch colonies. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(17), 8609–8614. https://doi.org/10.1073/pnas.1805114116

Schuster, G., Ebert, E. E., Stevenson, M. A., Corner, R. J., & Johansen, C. A. (2011). Application of satellite precipitation data to analyse and model arbovirus activity in the tropics. *International Journal of Health Geographics*, *10*(1), 8. https://doi.org/10.1186/1476-072X-10-8

Semenza, J. C., & Suk, J. E. (2018). Vector-borne diseases and climate change: A European perspective. *FEMS Microbiology Letters*, *365*(2), fnx244. https://doi.org/10.1093/FEMSLE/FNX244

Servadio, J. (2023). jlservadio/YF_Forecast (version V1) [Software]. Zenodo. https://doi.org/10.5281/zenodo.8302174

Servadio, J. L., Machado, G., Alvarez, J., de Ferreira Lima Júnior, F. E., Vieira Alves, R., Convertino, M., & Convertino, M. (2020). Information differences across spatial resolutions and scales for disease surveillance and analysis: The case of Visceral Leishmaniasis in Brazil. *PLoS One*, *15*(7), e0235920. https://doi.org/10.1371/journal.pone.0235920

Servadio, J. L., Muñoz-Zanzi, C., & Convertino, M. (2021). Estimating case fatality risk of severe Yellow Fever cases: Systematic literature review and meta-analysis. *BMC Infectious Diseases*, *21*(1), 1–12. https://doi.org/10.1186/S12879-021-06535-4

Servadio, J. L., Muñoz-Zanzi, C., & Convertino, M. (2022). Environmental determinants predicting population vulnerability to high yellow fever incidence. *Royal Society Open Science*, *9*(3), 220086. https://doi.org/10.1098/RSOS.220086

Servadio, J. L., Rosenthal, S. R., Carlson, L., & Bauer, C. (2018). Climate patterns and mosquito-borne disease outbreaks in South and Southeast Asia. *Journal of Infection and Public Health*, *11*(4), 566–571. https://doi.org/10.1016/j.jiph.2017.12.006

Shanafelt, D. W., Jones, G., Lima, M., Perrings, C., & Chowell, G. (2018). Forecasting the 2001 foot-and-mouth disease epidemic in the UK. *EcoHealth*, *15*(2), 338–347. https://doi.org/10.1007/s10393-017-1293-2

Shearer, F. M., Longbottom, J., Browne, A. J., Pigott, D. M., Brady, O. J., Kraemer, M. U. G. G., et al. (2018a). Existing and potential infection risk zones of yellow fever worldwide: A modelling analysis. *Lancet Global Health*, *6*(3), e270–e278. https://doi.org/10.1016/S2214-109X(18)30024-X

Shearer, F. M., Moyes, C. L., Pigott, D. M., Brady, O. J., Marinho, F., Deshpande, A., et al. (2018b). Data from: Global yellow fever vaccination coverage from 1970 to 2016: An adjusted retrospective analysis [Dataset]. Dryad. https://doi.org/10.5061/dryad.5ps8j

Silva-Inacio, C. L., de Paiva, A. A. P., de Araújo, J. M. G., & Ximenes, M. D. F. F. D. M. (2020). Ecological relationships of *Haemagogus spegazzinii* (Diptera: Culicidae) in a semiarid area of Brazil. *Revista da Sociedade Brasileira de Medicina Tropical*, *53*, 1–8. https://doi.org/10.1590/0037-8682-0502-2020

Sobol, I. (1993). Sensitivity analysis for non-linear mathematical model. *Mathematical Modelling and Computational Experiments*, *1*, 407–414.

Souza, R. L., Mugabe, V. A., Paploski, I. A. D., Rodrigues, M. S., Moreira, P. S. D. S., Nascimento, L. C. J., et al. (2017). Effect of an intervention in storm drains to prevent *Aedes aegypti* reproduction in Salvador, Brazil. *Parasites & Vectors*, *10*(1), 328. https://doi.org/10.1186/s13071-017-2266-6

Stolerman, L. M., Maia, P. D., Nathan Kutz, J., & Kutz, J. N. (2019). Forecasting dengue fever in Brazil: An assessment of climate conditions. *PLoS One*, *14*(8), e0220106. https://doi.org/10.1371/journal.pone.0220106

Tabachnick, W. J. (2010). Challenges in predicting climate and environmental effects on vector-borne disease episystems in a changing world. *Journal of Experimental Biology*, *213*(6), 946–954. https://doi.org/10.1242/jeb.037564

Tarboton, D. G., Bras, R. L., & Rodriguez-Iturbe, I. (1991). On the extraction of channel networks from digital elevation data. *Hydrological Processes*, *5*(1), 81–100. https://doi.org/10.1002/HYP.3360050107

Tarboton, D. G., Bras, R. L., & Rodriguez-Iturbe, I. (1992). A physical basis for drainage density. *Geomorphology*, *5*(1–2), 59–76. https://doi.org/10.1016/0169-555X(92)90058-V

Tesla, B., Demakovsky, L. R., Mordecai, E. A., Ryan, S. J., Bonds, M. H., Ngonghala, C. N., et al. (2018). Temperature drives Zika virus transmission: Evidence from empirical and mathematical models. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1884), 20180795. https://doi.org/10.1098/rspb.2018.0795

Tuck, S. L., Phillips, H. R. P., Hintzen, R. E., Scharlemann, J. P. W., Purvis, A., & Hudson, L. N. (2014). MODISTools – Downloading and processing MODIS remotely sensed data in R. *Ecology and Evolution*, *4*(24), 4658–4668. https://doi.org/10.1002/ECE3.1273

Tushar, A., & Reich, N. G. (2017). flusight: Interactive visualizations for infectious disease forecasts. *Journal of Open Source Software*, *2*(13), 231. https://doi.org/10.21105/JOSS.00231

United States Environmental Protection Agency. (2023). Mosquito life cycle. Retrieved from https://www.epa.gov/mosquitocontrol/mosquito-life-cycle

Wang, C., Jiang, B., Fan, J., Wang, F., & Liu, Q. (2014). A study of the dengue epidemic and meteorological factors in Guangzhou, China, by using a zero-inflated Poisson regression model. *Asia-Pacific Journal of Public Health*, *26*(1), 48–57. https://doi.org/10.1177/1010539513490195

Watts, A. G., Miniota, J., Joseph, H. A., Brady, O. J., Kraemer, M. U. G., Grills, A. W., et al. (2017). Elevation as a proxy for mosquito-borne Zika virus transmission in the Americas. *PLoS One*, *12*(5), e0178211. https://doi.org/10.1371/JOURNAL.PONE.0178211

Westreich, D., & Greenland, S. (2013). The table 2 fallacy: Presenting and interpreting confounder and modifier coefficients. *American Journal of Epidemiology*, *177*(4), 292–298. https://doi.org/10.1093/AJE/KWS412

Wilk-da-Silva, R., Mucci, L. F., Ceretti-Junior, W., de Castro Duarte, A. M. R., Marrelli, M. T., & Medeiros-Sousa, A. R. (2020). Influence of landscape composition and configuration on the richness and abundance of potential sylvatic yellow fever vectors in a remnant of Atlantic Forest in the city of São Paulo, Brazil. *Acta Tropica*, *204*, 105385. https://doi.org/10.1016/j.actatropica.2020.105385

World Health Organization. (2013). *Sustaining the drive to overcome the global impact of neglected tropical diseases*. WHO. World Health Organization.

World Health Organization. (2018). Yellow fever. Retrieved from http://www.who.int/en/news-room/fact-sheets/detail/yellow-fever

World Health Organization. (2021). Countries with risk of yellow fever transmission and countries requiring yellow fever vaccination. (May 2021). Retrieved from https://www.who.int/publications/m/item/countries-with-risk-of-yellow-fever-transmission-and-countries-requiring-yellow-fever-vaccination-(may-2021)

Wu, J., Dhingra, R., Gambhir, M., & Remais, J. V. (2013). Sensitivity analysis of infectious disease models: Methods, advances and their application. *Journal of the Royal Society Interface*, *10*(86), 20121018. https://doi.org/10.1098/rsif.2012.1018

Zhao, S., Stone, L., Gao, D., & He, D. (2018). Modelling the large-scale yellow fever outbreak in Luanda, Angola, and the impact of vaccination. *PLoS Neglected Tropical Diseases*, *12*(1), e0006158. https://doi.org/10.1371/journal.pntd.0006158

Zuur, A. F., & Ieno, E. N. (2016). *Beginner's guide to zero-inflated models with R*. Highland Statistics Ltd.