Spatio-temporal modeling of arboviruses

MSc in Statistics

Project Fair

Prof. Nuno R. Faria
School of Public Health
Imperial College London

Viral genomic epidemiology
Epidemiology and viral epidemics
Outbreak preparedness

Dr. André V. Ribeiro Amaral
Department of Mathematics
Imperial College London

Spatio-temporal statistics
Bayesian statistics
Computational methods for inference

Introduction

Four potential project ideas:

Project 1. Linking dengue transmission and social vulnerability

Project 2. Modelling the dynamic behavior and interaction of dengue virus lineages

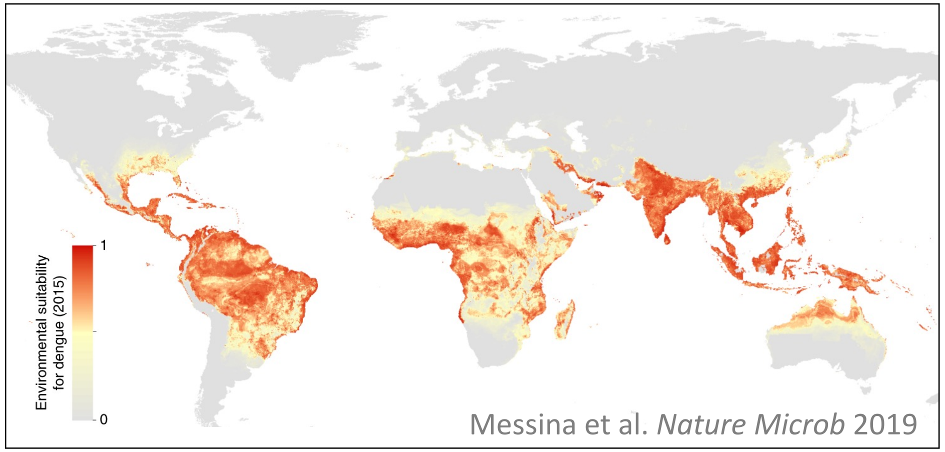
Project 3. Seasonal dengue forecasting

Project 4. Modeling yellow-fever incidence based on different types of transmission

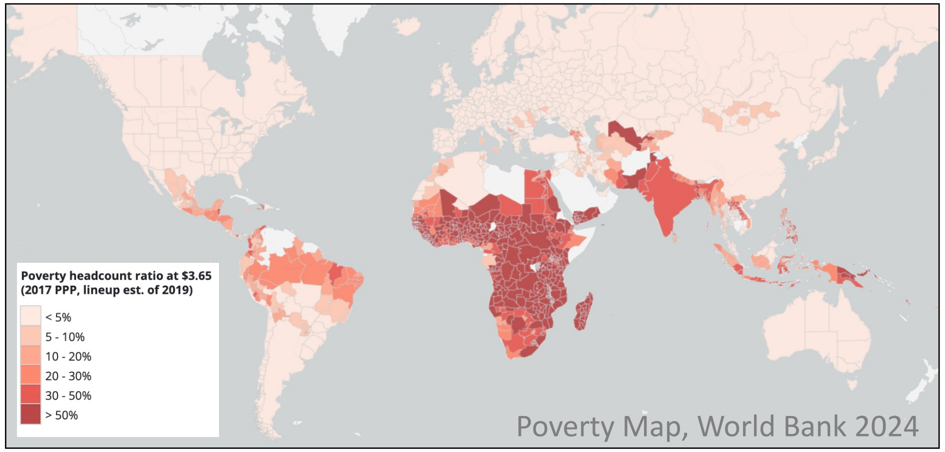
Linking dengue transmission and social vulnerability

Overarching Aim: to investigate the correlations between socioeconomic factors and dengue incidence.

Climate factors, such as temperature and precipitation, contribute to the proliferation of *Aedes aegypti*. However, social aspects, including urbanization and poverty levels, also play significant roles both in dengue transmission and in dengue surveillance.



Messina et al. *Nature Microb* 2019

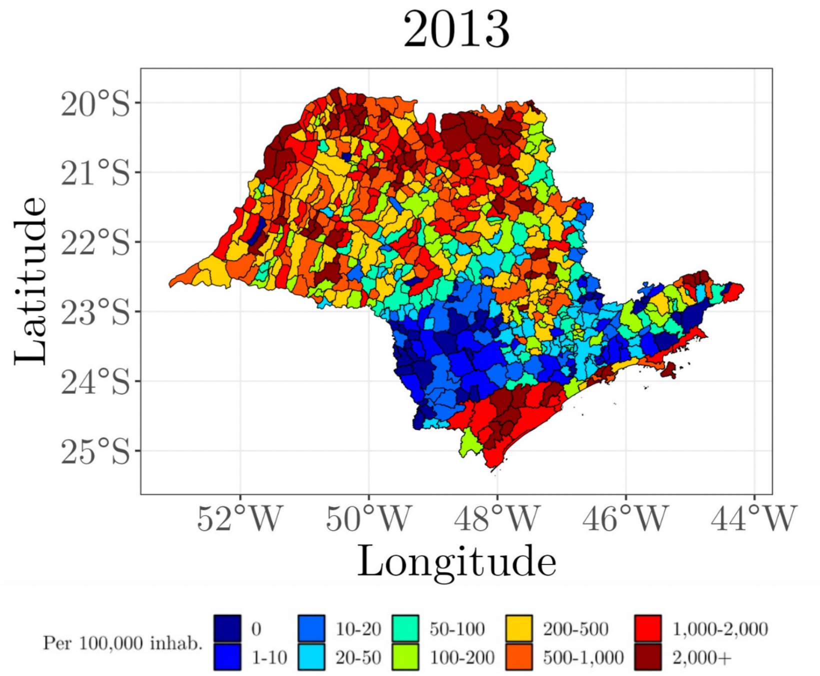


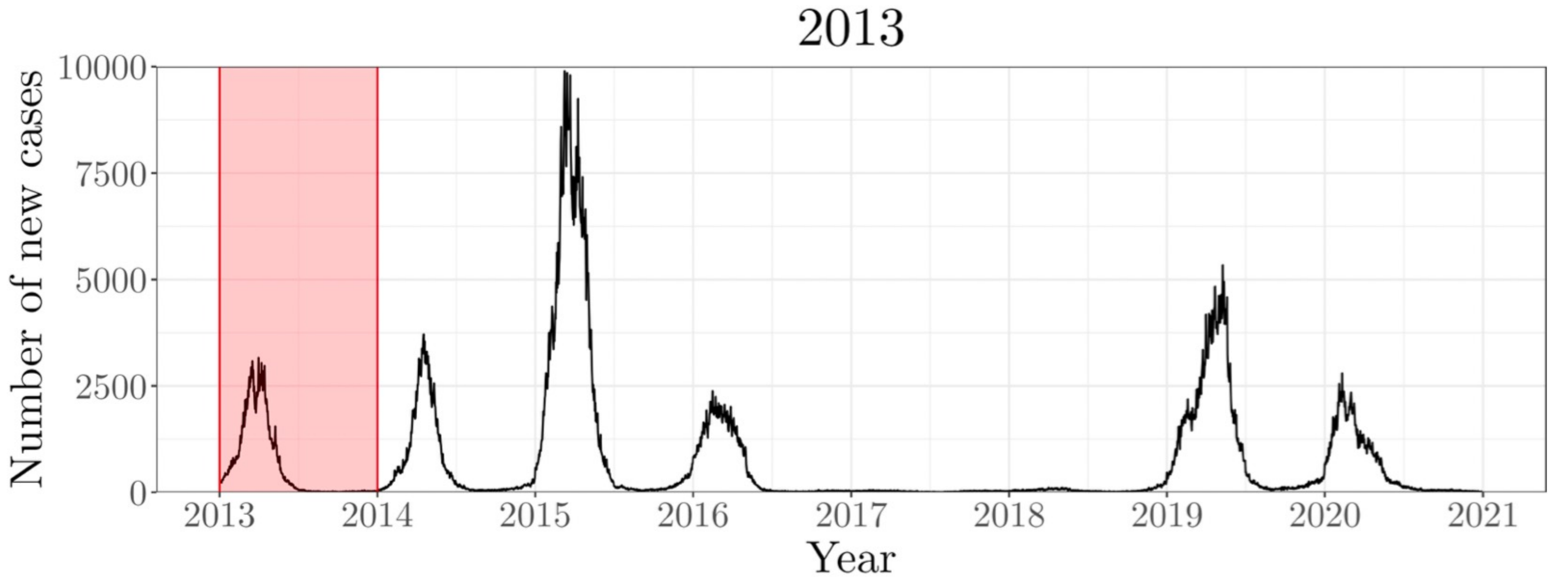
Poverty Map, World Bank 2024

Dengue incidence in São Paulo, Brazil

Data: 10-year municipality-level resolution dengue incidence data from São Paulo State, publicly available from SINAN, Brazil Ministry of Health. Municipality-level resolution poverty indicators available through collaboration with UK-Brazil CADDE Project.

Note: It is possible to evaluate link between dengue & socioeconomic vulnerability also at global scale.





CADDE Project 2024

Modelling the dynamic behavior and interaction of dengue virus lineages

Overarching Aim: to investigate the dynamic behavior and interaction of dengue virus lineages in hyperendemic settings in Brazil (and/or e.g. Thailand).

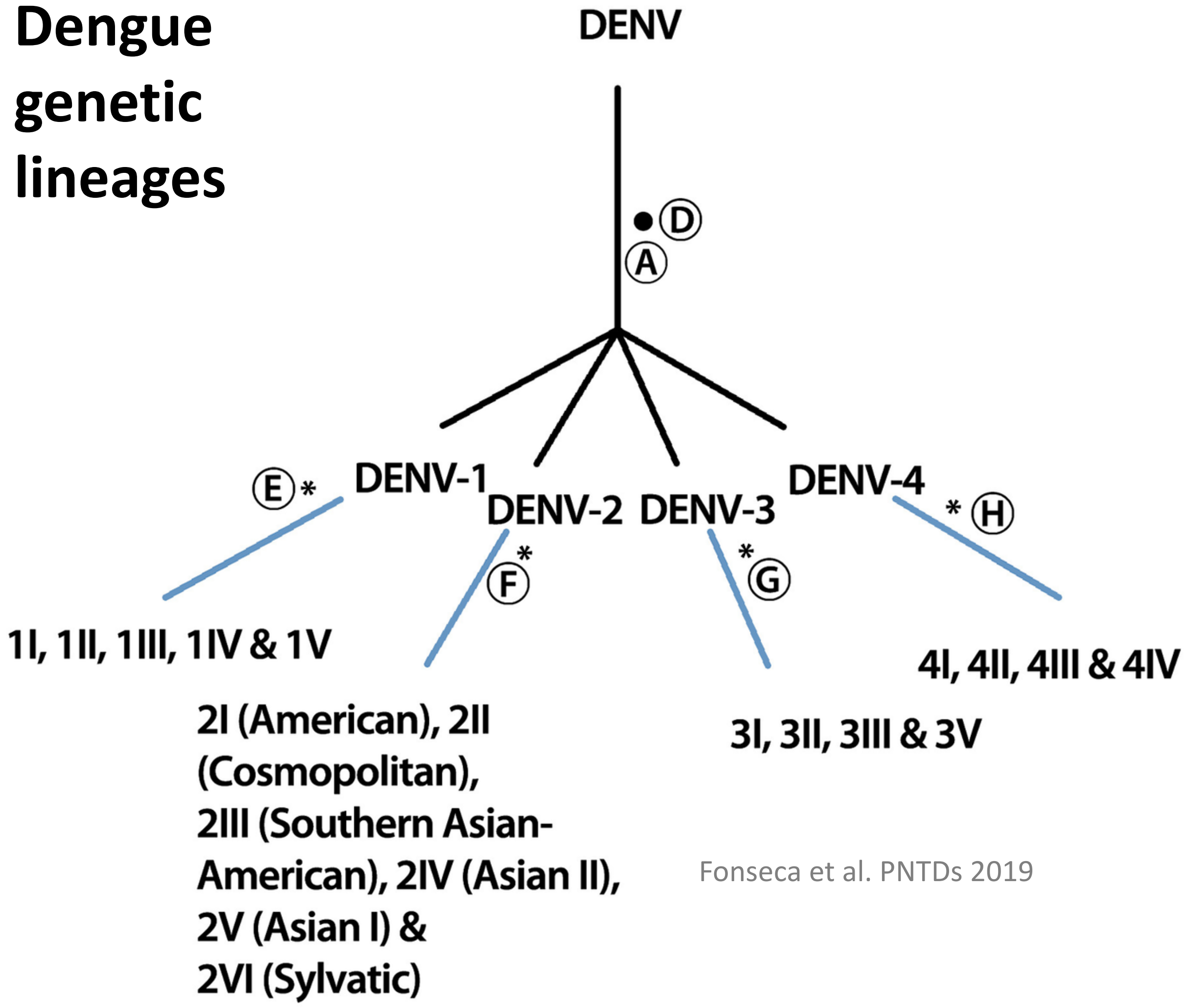
There are four dengue virus serotypes (DENV1-4), and >20 genotypes. Where detected, these lineages seem to go extinct and are replaced cyclically, with invasion of a new lineage typically associated with larger outbreaks.

While reinfection with the same serotype is rare in the short to mid-term, an individual who has previously experienced infection with a specific serotype does not develop immunity to others. These serotypes can circulate simultaneously (hyperendemicity).

The poor understanding of the dynamic behaviour and interaction between dengue serotypes makes it challenging to anticipate the occurrence of large outbreaks.

Dengue genetic and epidemiological data

Dengue genetic lineages



Fonseca et al. PNTDs 2019

NCBI Virus Database

Nucleotide (45,733) | Protein (44,115) | RefSeq Genome (4) | Select Columns

Accession	Organism Name	Submitters	Organization	Release Date
NC_002640 RefSeq	dengue virus type 4	Durbin,A.P., et al.	National Center for Biotec...	2001-01-03
NC_001477 RefSeq	dengue virus type 1	Puri,B., et al.	National Center for Biotec...	1997-02-28
NC_001474 RefSeq	dengue virus type 2	Kinney,R.M., et al.	National Center for Biotec...	1993-08-02

https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/

Open dengue database

OPEN DENGUE

A global database of public dengue case data

https://opendengue.org/project.html

Seasonal dengue forecasting

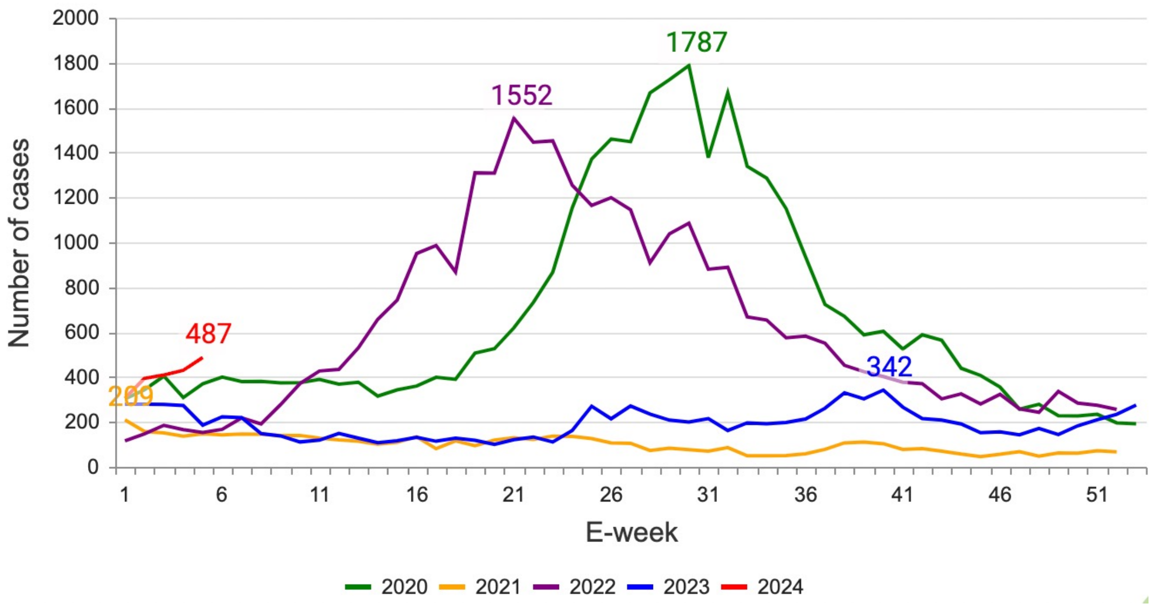
Overarching Aim: To develop an accurate forecasting model for dengue early warning system in Singapore and/or Brazil, using a wide range of weather factors as input variables.

Traditional machine learning models: Poisson regression models, hierarchical Bayesian models, autoregressive integrated moving average (ARIMA), least absolute shrinkage and selection operator (LASSO) regression, generalized additive models (GAMs).

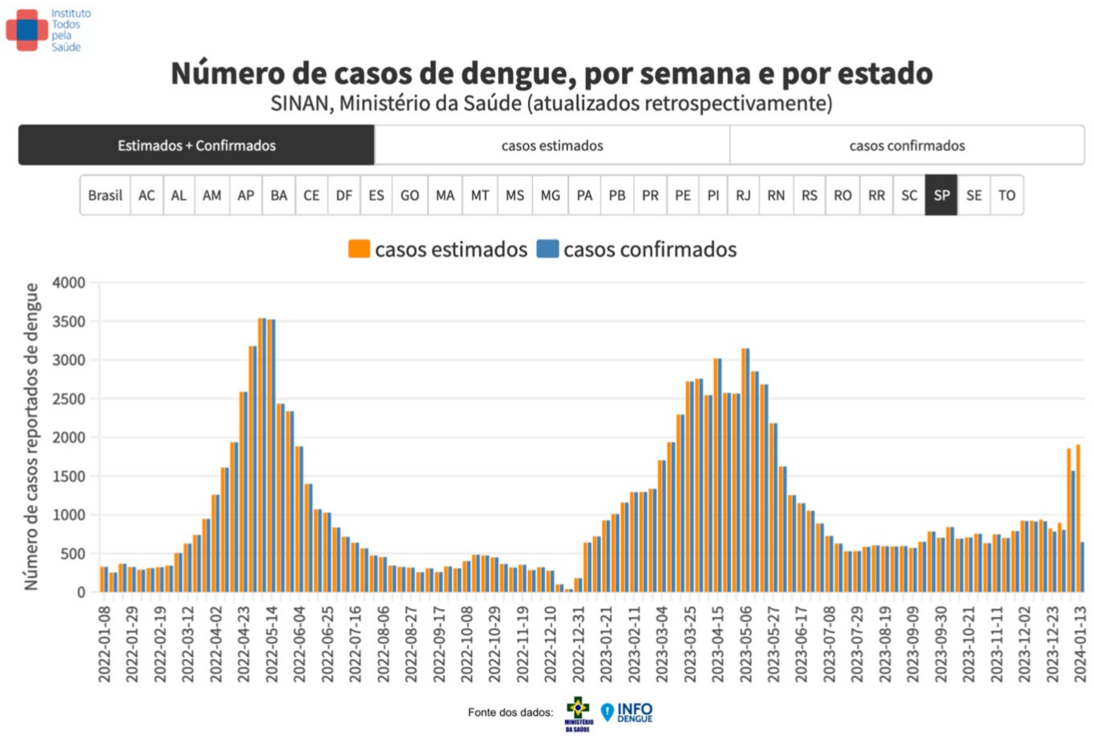
Deep learning models: Neural Networks (Artificial and Back-Propagation) and Long Short-Term Memory (LSTM) Models => more robust to missing data.

Seasonal dengue in Singapore and/or Brazil

Data: 5-year dengue cases per epidemiological week in Singapore and/or 10-year dengue cases per day São Paulo state, Brazil.



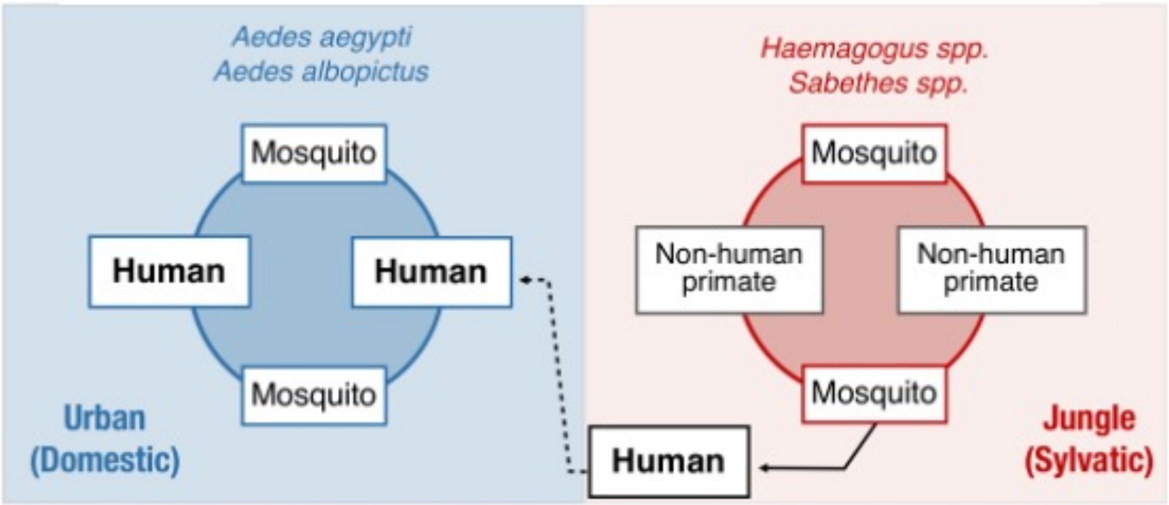
https://www.nea.gov.sg/dengue-zika/dengue/dengue-cases



https://www.Itps.com.br

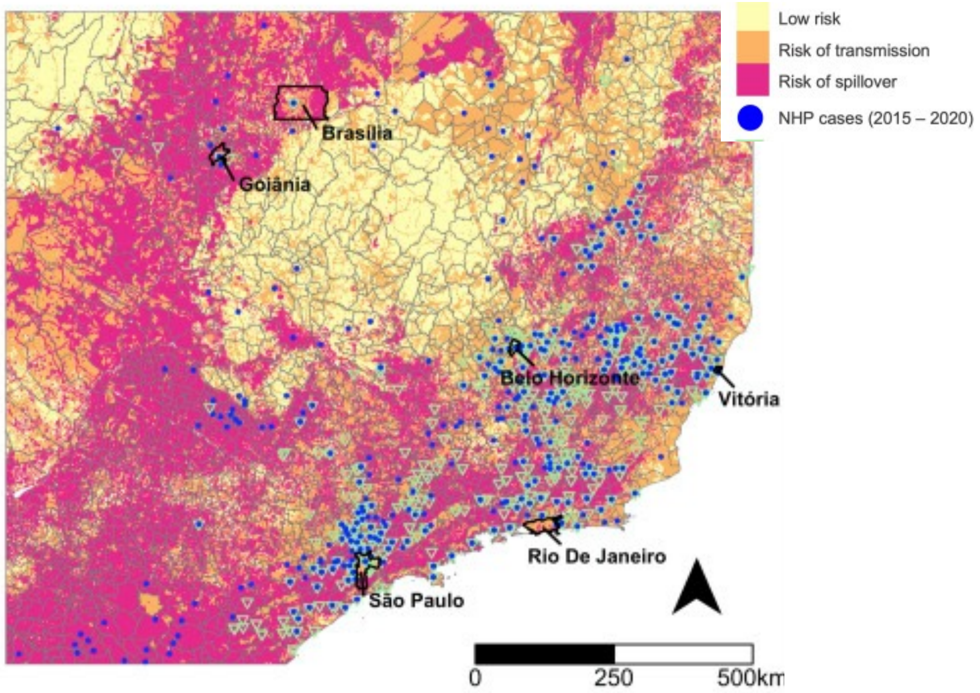
Modelling yellow fever transmission dynamics

Overarching Aim: to investigate the dynamic behavior of *Haemagogus leucocelanus,* the mosquito vector for sylvatic yellow fever in Brazil.



Faria et al. *Science* 2018

In South America, yellow-fever (YF) transmission occurs mainly via the non-urban cycle, in which non-human primates (NHP) are infected by mosquito vectors.

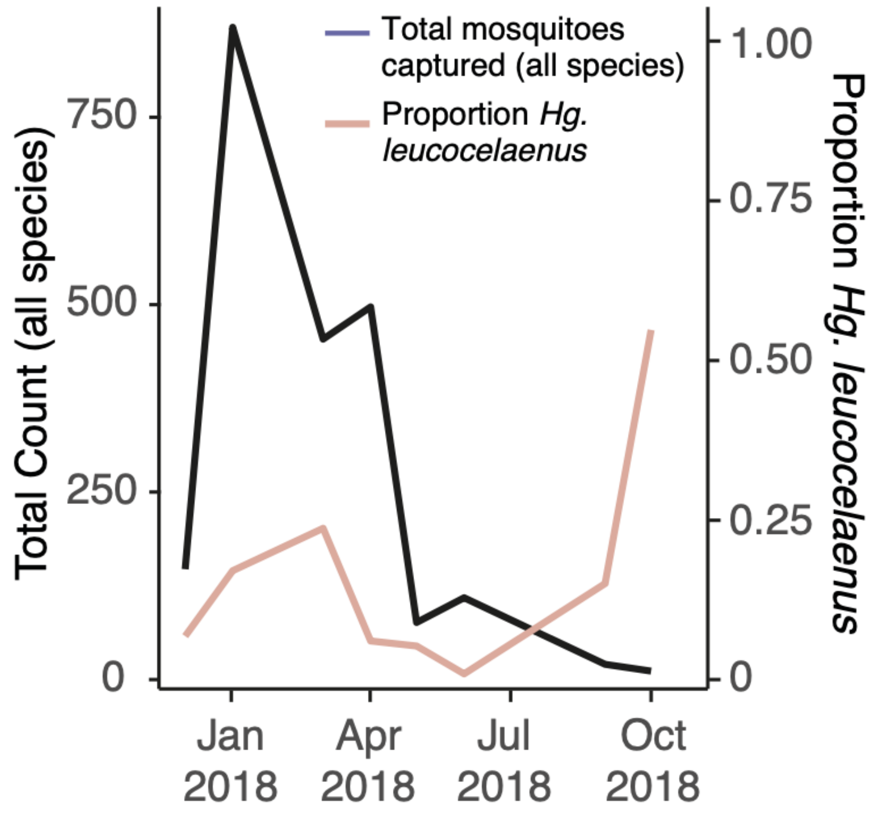


Li et al. *PNTDs* 2022

Modelling yellow fever transmission dynamics

Data. Environmental covariates (temperature, humidity, rain), and mosquito occurrence data (*Haemagogus leucocelanus,* the main vector for YF in southeast Brazil) captured by local public health teams in São Paulo State, Brazil.

Note: non-human primate cases, human cases, and vaccination data are also available for more detailed analysis.



de Deus *in prep* 2024

A final note on Statistics

$y = (y_1, \cdots, y_n)$	Observed data
$y \sim p(y \mid g^{-1}(\eta), \theta)$	Observational model
$\eta = X\beta + f_{\mathrm{GP}}$	Linear predictor
$f_{\mathrm{GP}} \sim \mathrm{GP}(\mu, \Sigma(\theta))$	Random effects (GPs)

This modeling scheme can be used for, e.g., “point patterns,” “areal data,” or “geostatistical data.”

Our goal is to estimate $g(\mathbb{E}(y \mid f_{\mathrm{GP}})) = X\beta + f_{\mathrm{GP}}$; however, it scales with $\mathcal{O}(n^3)$.

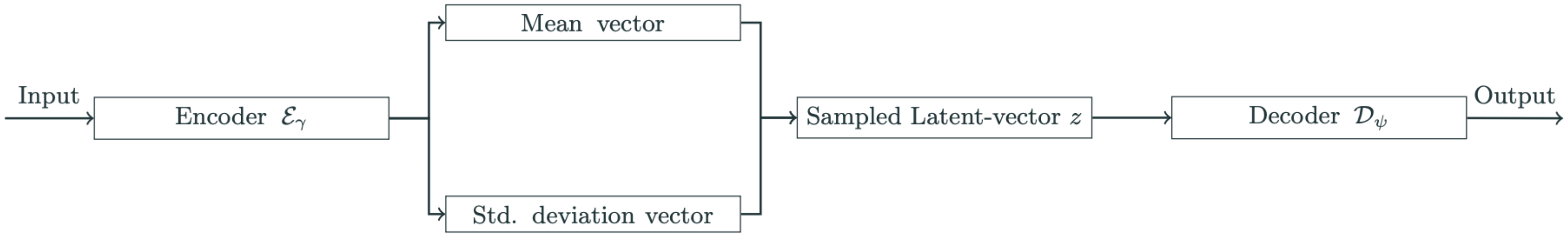
A final note on Statistics

$$
\begin{aligned}
y &= (y_1, \cdots, y_n) && \text{Observed data} \\
y &\sim p(y|g^{-1}(\eta), \theta) && \text{Observational model} \\
\eta &= X\beta + f_{\text{GP}} && \text{Linear predictor} \\
f_{\text{GP}} &\sim \text{GP}(\mu, \Sigma(\theta)) && \text{Random effects (GPs)}
\end{aligned}
$$

This modeling scheme can be used for, e.g., “point patterns,” “areal data,” or “geostatistical data.”

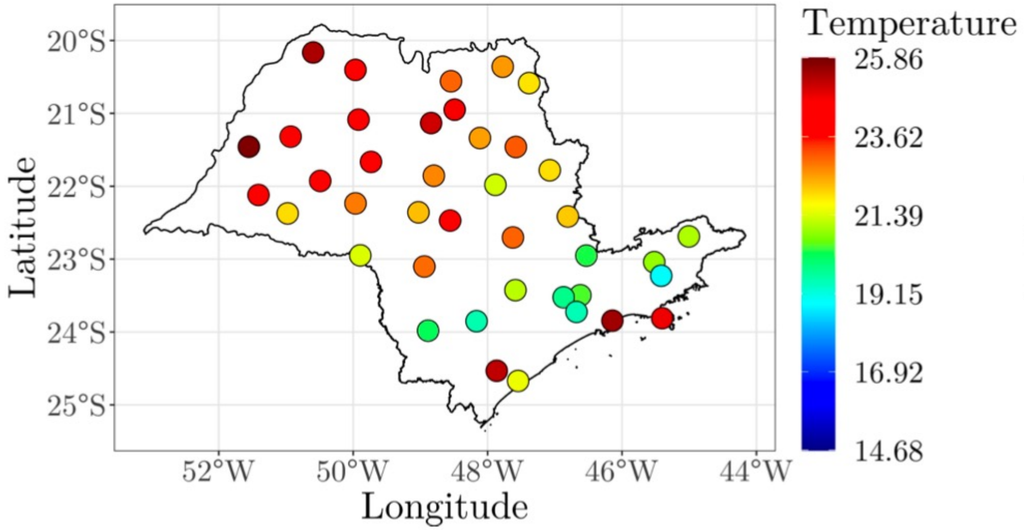
Our goal is to estimate $g(\mathbb{E}(y|f_{\text{GP}})) = X\beta + f_{\text{GP}}$; however, it scales with $\mathcal{O}(n^3)$.

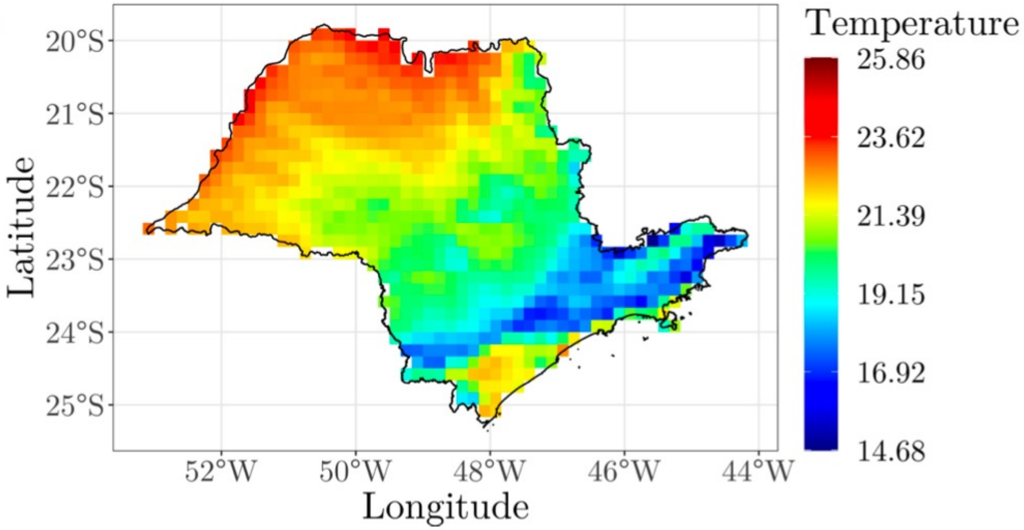
To overcome this issue, we substitute f_{GP} by f_{VAE}.



A final note on Statistics

For many of these problems, we may want to integrate various data sources and at different resolutions. Within the previously introduced framework,





$$y(x_i)|f,\mu,\sigma^2 \sim \mathcal{N}\left(\mu(x_i)+f(x_i),\sigma^2\right) \quad \text{and} \quad y(B_j)|f,\mu,\sigma^2 \sim \mathcal{N}\left(\frac{1}{|B_j|}\int_{B_j}[\mu(x)+f(x)]dx,\sigma^2\right)$$