IDM INSTITUTE FOR DISEASE MODELING

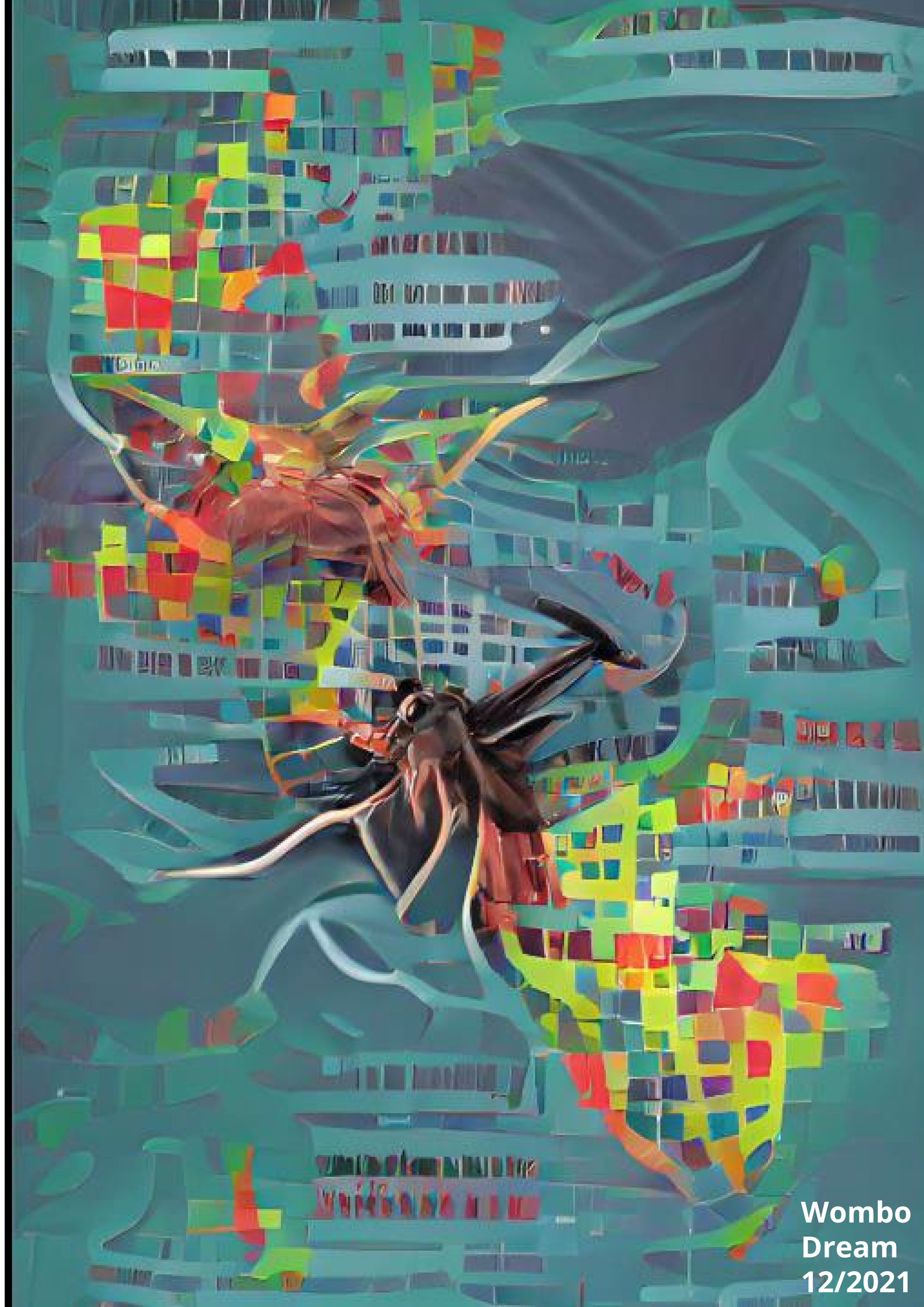
Annual Symposium
May 24th 2023
Seattle, WA, USA

Human movement and environmental barriers shape the emergence of dengue

Vinyas Harish, Felipe J. Colon-González, Filipe R.R. Moreira, Rory Gibb, Moritz U.G. Kraemer, Megan Davis, Robert C. Reiner Jr., David M. Pigott, T. Alex Perkins, Daniel J. Weiss, Isaac I. Bogoch, Gonzalo Vazquez-Prokopec, Pablo Manrique Saide, Gerson L. Barbosa, Ester C. Sabino, Kamran Khan, Nuno R. Faria, Simon I. Hay, Fabian Correa-Morales, Francisco Chiaravalloti-Neto, **Oliver J. Brady**

UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE



Wombo Dream 12/2021

Introduction

- Dengue is a emerging arboviral disease that has rapidly expanded over the last 40 years [1,2]
- Understanding dengue spread is vital for guiding public health interventions [3, 4]
- The re-expansion of virus circulation following cessation of *Aedes aegypti* eradication programs during the 1970s makes Latin America a unique natural experiment to study the drivers of dengue spread in the modern era [5, 6]
- However, regular, comprehensive dengue surveillance programs were rarely in place before 2000
    - Reports documenting early spread are opportunistic, incomplete, and potentially biased

∴ **Modelling is needed to fill these gaps & better understand spread at different time scales**

UNIVERSITY OF TORONTO

[1] Messina *et al.*, Nat Micro 2019
[2] Lindsay *et al.*, PTRSB 2021
[3] Imai *et al.*, PLoS NTD 2015
[4] Flasche *et al.*, PLoS Med 2016
[5] Messina *et al.*, Sci Data 2014
[6] Dick *et al.*, AJTMH 2012

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

Research questions

1. How can models help understand dengue spread in Latin America (Mexico & Brazil)?

2. How has dengue spread through these countries in the past and how will it spread in the future?

UNIVERSITY OF TORONTO

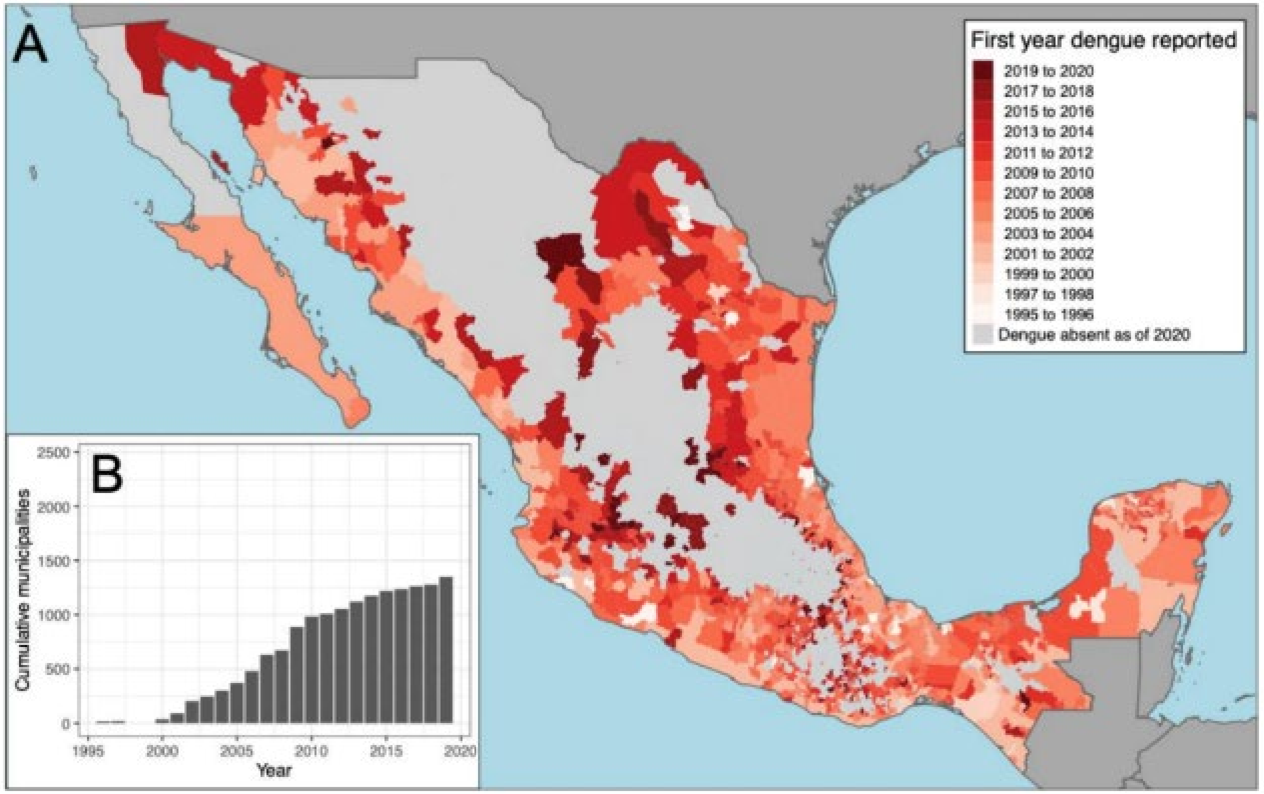
LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

Methods

Routine surveillance data - Mexico

- Dengue surveillance coordinated federally by the General Directorate for Epidemiologic Surveillance and operated through state epidemiology departments [7]

- Data captures the early spread process



[7] Hernandez-Avila *et al.*, PLoS One 2013

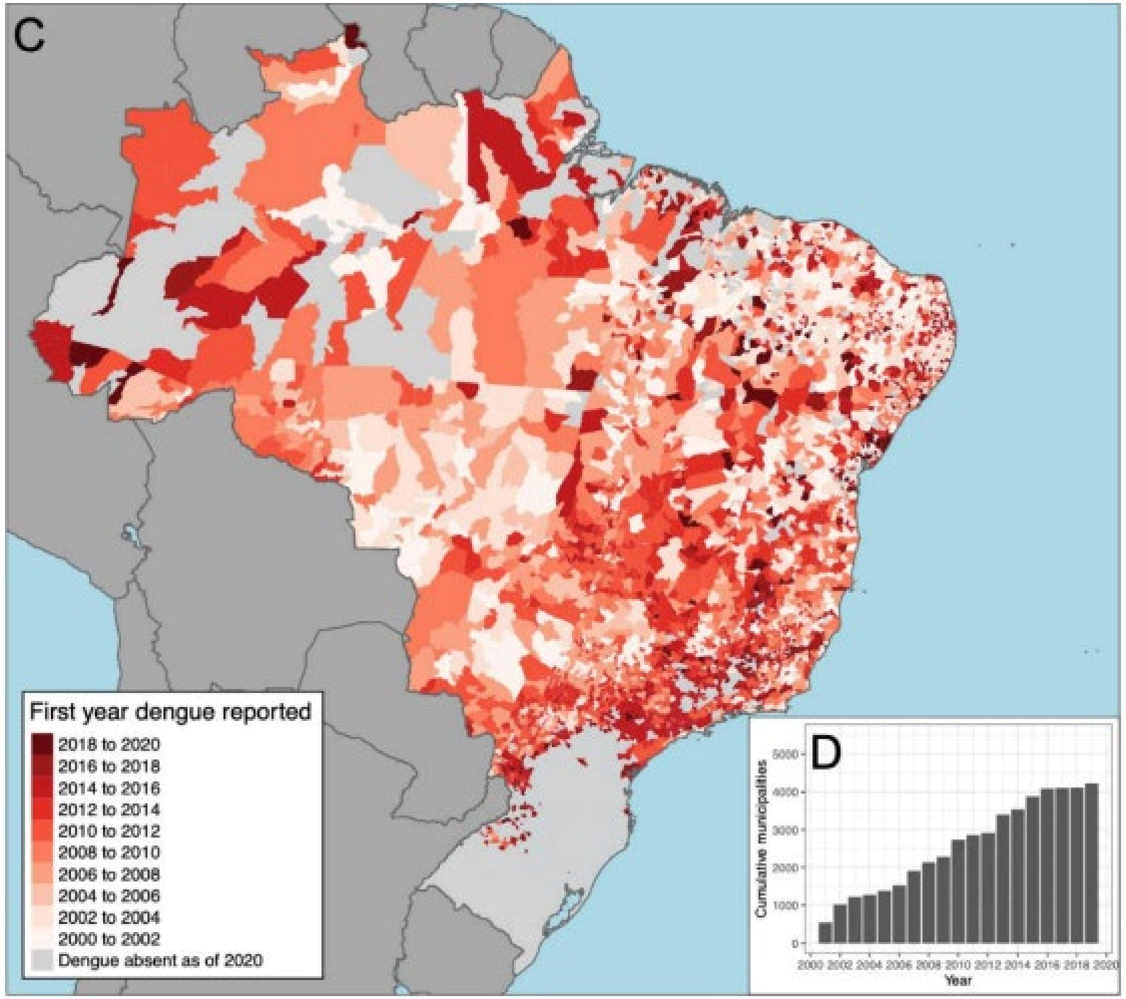
UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Routine surveillance data - Brazil

- Data on suspected and confirmed cases obtained from the Notifiable Diseases Information System (SINAN) from the Ministry of Health Information Department (DATASUS) [8]

- ~ 20% of municipalities invaded not captured in data → Use case for reconstruction



[8] Ministério da Saúde. DATASUS - Departamento de Informática do SUS. Informações de Saúde (TABNET). DENGUE - NOTIFICAÇÕES REGISTRADAS NO SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO - BRASIL (2001-2006; 2007-2013; 2014-2019)

UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Feature selection

Satellite remote sensing environmental:

- ~~Latitude~~ & longitude
- ~~*Aedes* suitability~~
- Daytime land surface temperature (mean / std)
- Nighttime land surface temperature (mean / std)
- Enhanced vegetation index (mean / std)
- Tasselled cap wetness ~~(mean /~~ std)
- Tasselled cap brightness (mean / std)
- Landcover

Aggregated to municipality level w Worldpop 2015 UN estimates

Centered & scaled

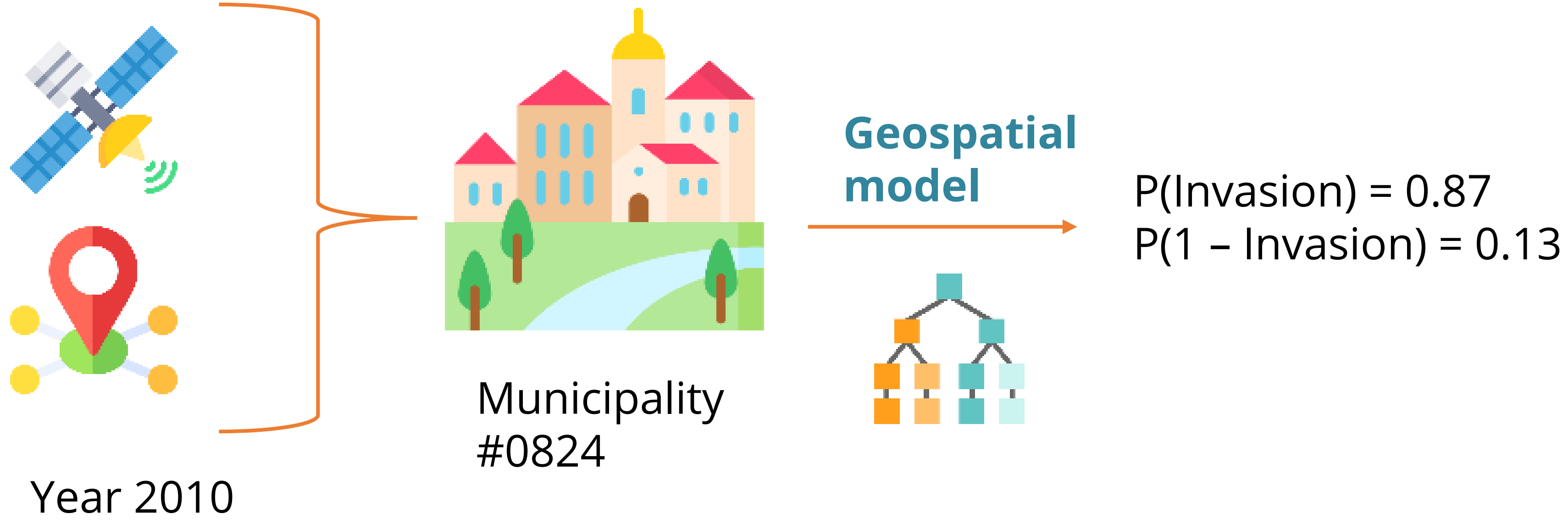
Mobility between invaded and uninvaded areas:

- ~~Great circle distance~~
- Adjacency
- ~~Gravity (log-transformed)~~
- Radiation (log-transformed)
- Friction surface (log-transformed)
- Flight data (log-transformed)
- Between-state migration (log-transformed)

Calculated using centroid of most populated 1 km x 1 km pixel from Worldpop 2015 UN estimates

Icons from Flaticon

Modelling approach – Stage 1 (Risk Prediction)



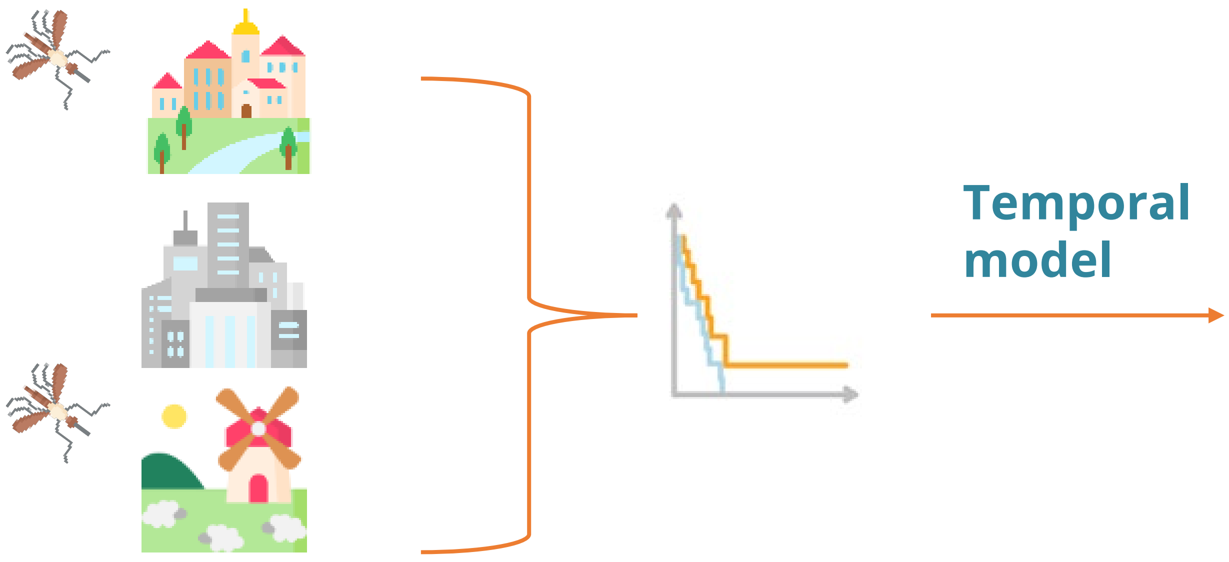
How to determine the number of municipalities to invade each year?

Icons from Flaticon

UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

Modelling approach – Stage 2 (Thresholding)



...

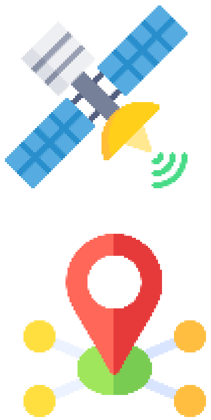
Year 2010

N municipalities invaded each year

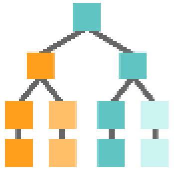
- **Temporal model** developed using survival analysis approach
  - Allowed us to convey uncertainty using different percentiles of predictions

- **Core assumption:** municipalities were "invaded" once they passed a set threshold number of annual cases, once the model predicts a municipality to be invaded (class 1) — it cannot be un-invaded (class 0)

Overall modelling approach



Year 2010



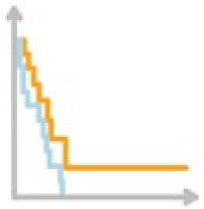
Geospatial model

P(Invasion) = 0.87

P(Invasion) = 0.23

P(Invasion) = 0.48

...



Temporal model

N municipalities invaded in 2010

Top *N* uninvaded municipalities by P(Invasion) =

Invaded in 2010

Icons from Flaticon

UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

How do we evaluate our approach at different timescales?

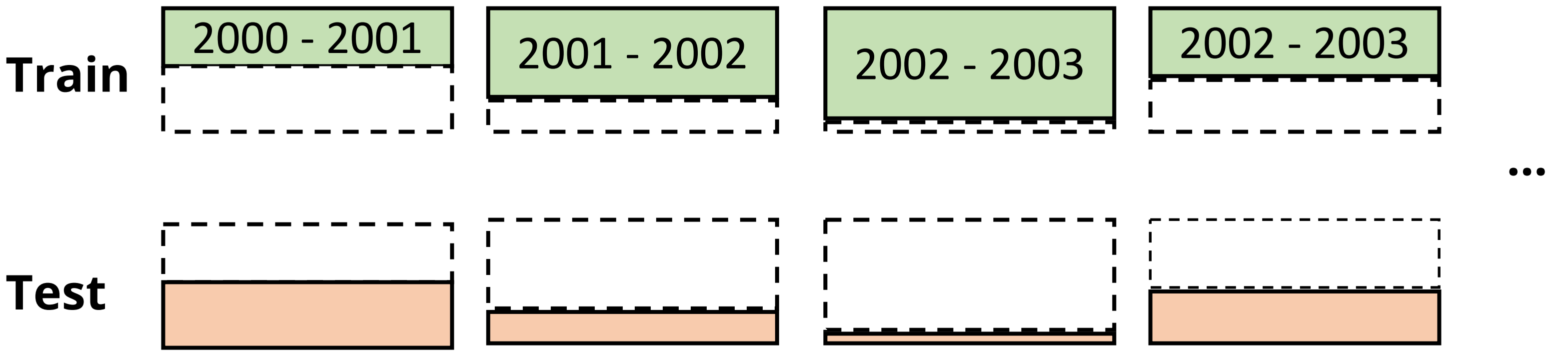
- Traditional cross-validation techniques split data into different folds of training and test sets at random
  - Test set AUC inflates predictive performance due to spatiotemporal autocorrelation [9]
  - Bespoke cross-validation techniques are preferred for many ecological/disease mapping applications [10]
- Gilbert *et al.* (2014) used a spatial cross-validation technique to evaluate the predictive performance of a model for H7N9 infection across Asian poultry markets [11]
- Here, we develop an approach for conducting **time-series cross-validation (TSCV)** at short and medium and long term time-scales

[9] Bahn & McGill, Oikos 2012
[10] Roberts et al. Ecography 2016
[11] Gilbert *et al.*, Nature Comms 2014

Time series cross-validation

Data from routine surveillance

Naïve cross-validation



UNIVERSITY OF TORONTO

Short-term time series cross-validation (S-TSCV)

Train Round 1 2000 | 2000 -2001

Test Round 1 2001 -2002

Incorrect model predictions are NOT carried forward

Train Round 2 2000 | 2000 -2001 | 2001 | 2001 -2002

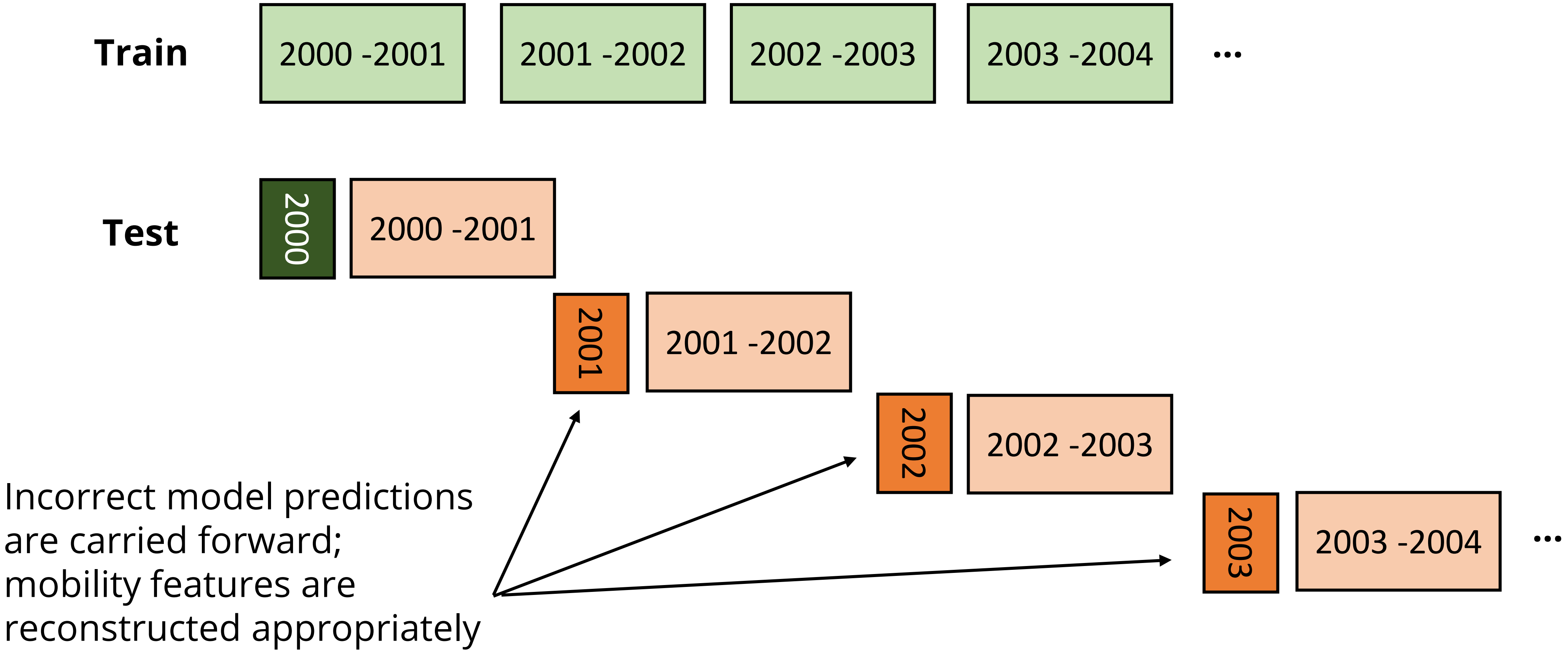
Test Round 2 2002 -2003

...

UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

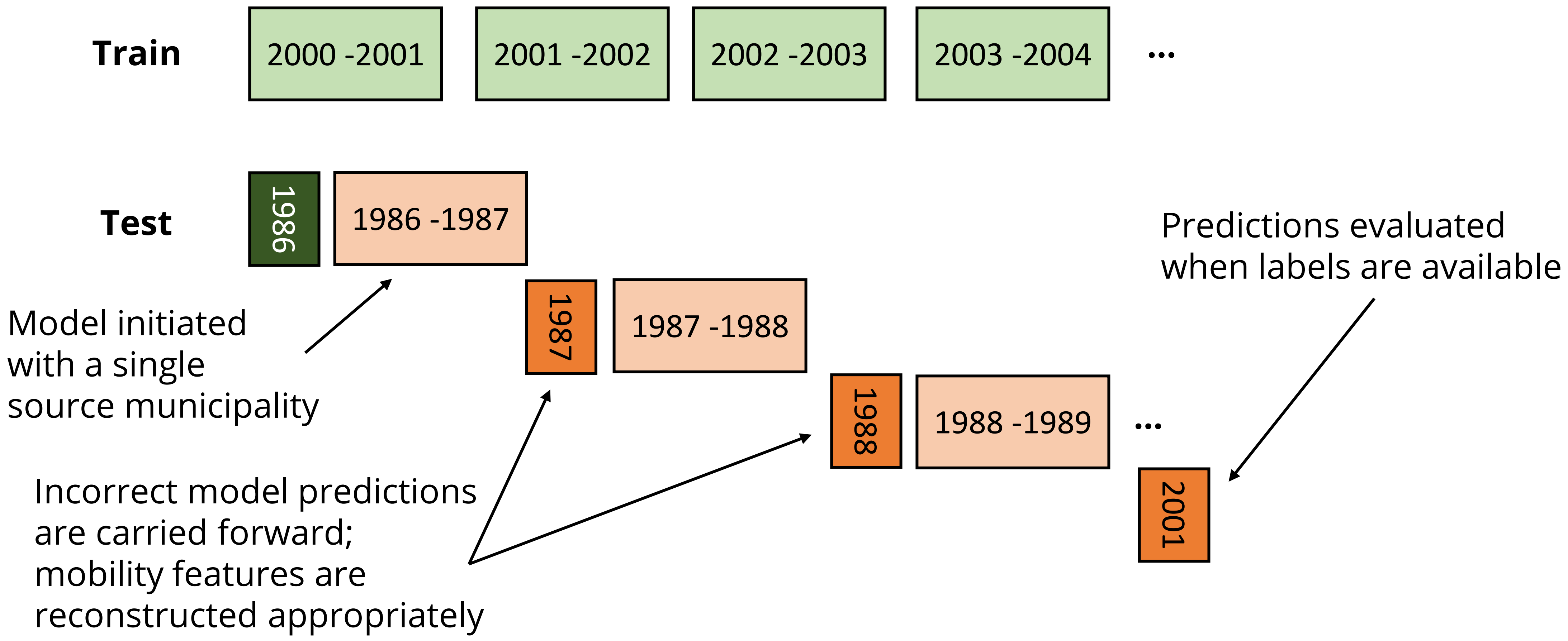
Medium-term time series cross-validation (M-TSCV)



UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

Long-term time series cross-validation (L-TSCV)



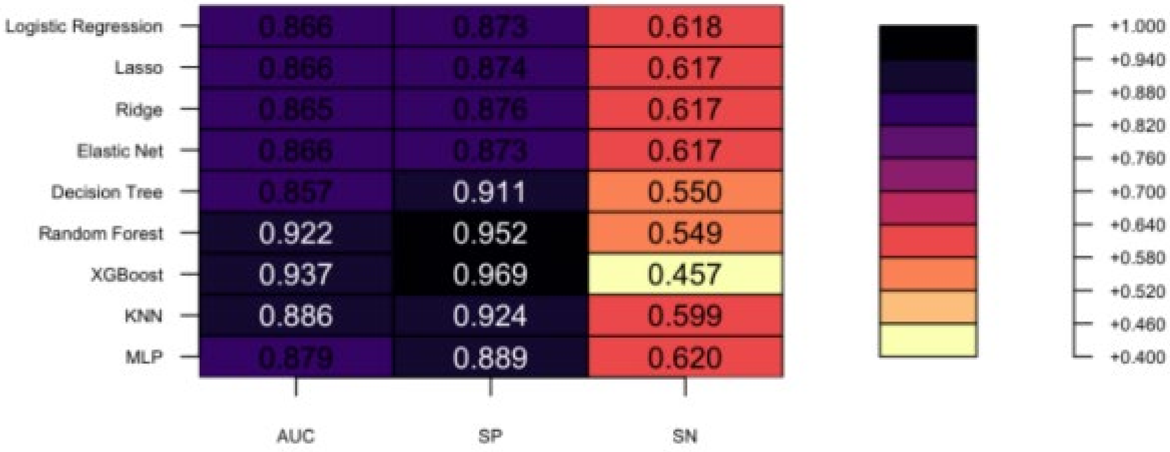
Results

UNIVERSITY OF TORONTO

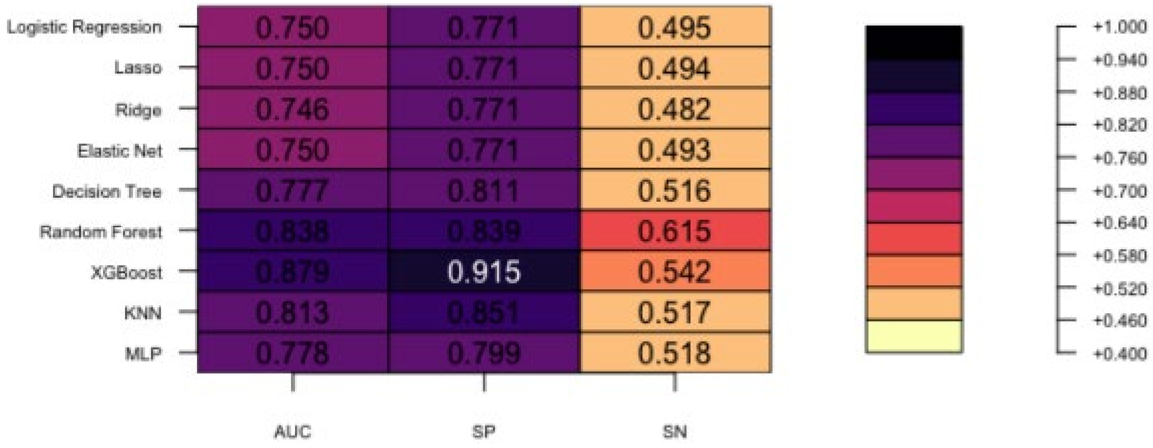
LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

Gradient boosted trees reign supreme (again)

Mexico



Brazil



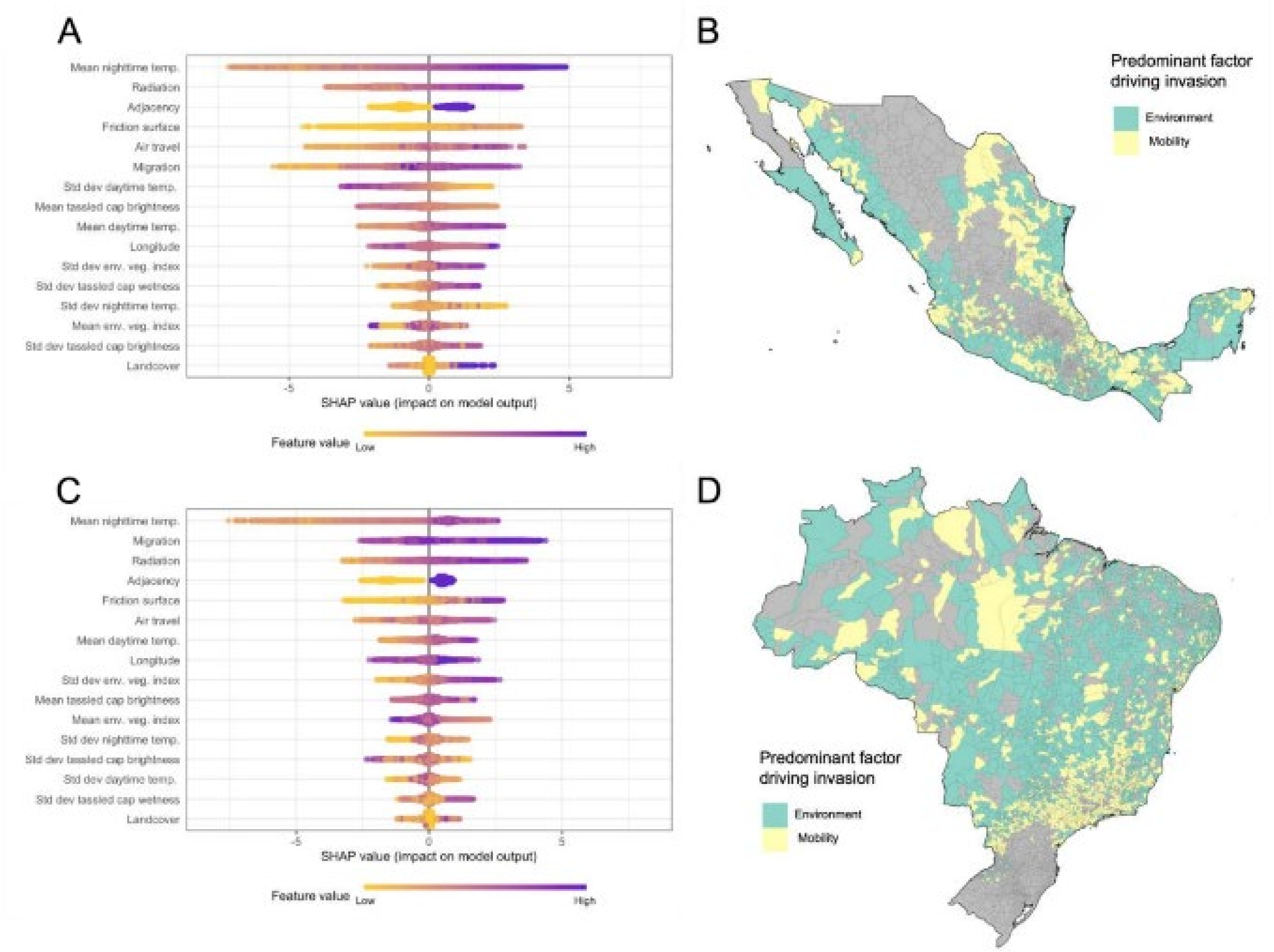
- Gradient boosted decision trees have been shown to capture complex functions without overfitting across numerous disease mapping problems (dengue, avian influenza, zoonotic disease emergence, etc.) [11-14]

[11] Gilbert *et al.* Nature Comms 2014
[12] Elith *et al.* Journal of Animal Ecology 2008
[13] Bhatt *et al.* Nature 2013
[14] Allen *et al.* Nature Comms 2017

UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

Feature importance reaffirms existing knowledge



- Shapley values computed on dataset of newly invaded municipalities [15]

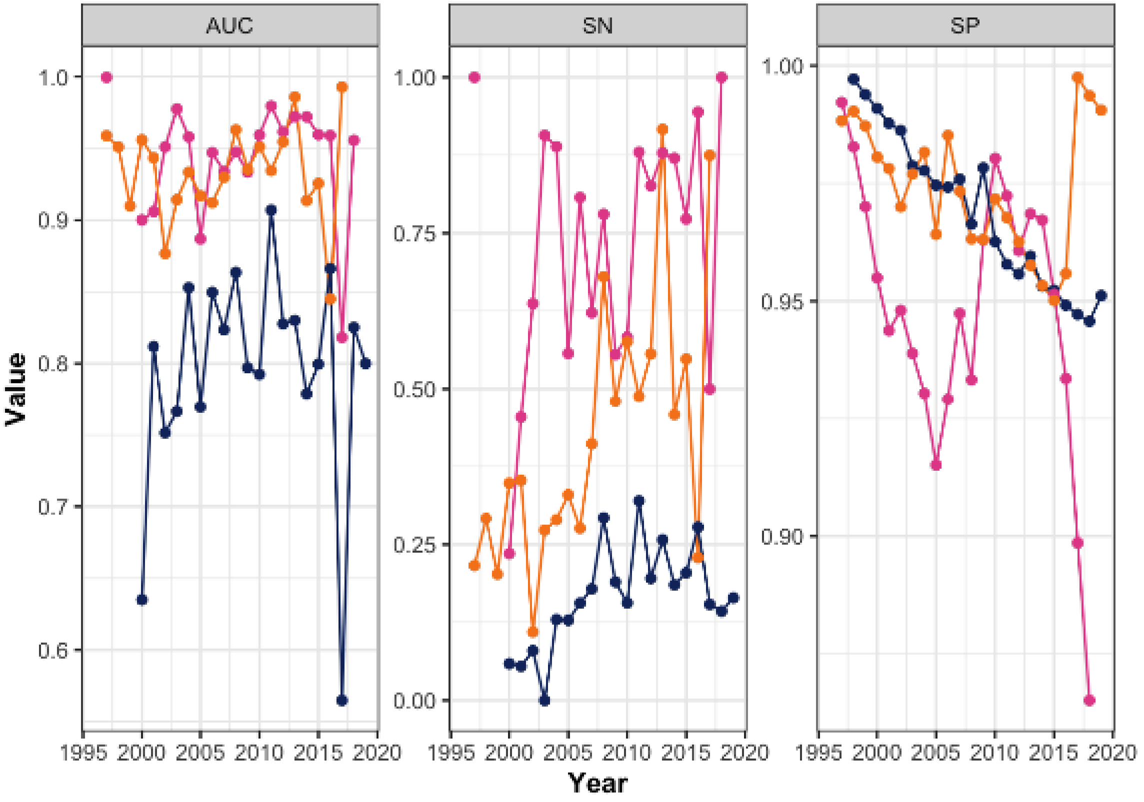
[15] Lundberg *et al.*, Nature Machine Intelligence 2020

UNIVERSITY OF TORONTO

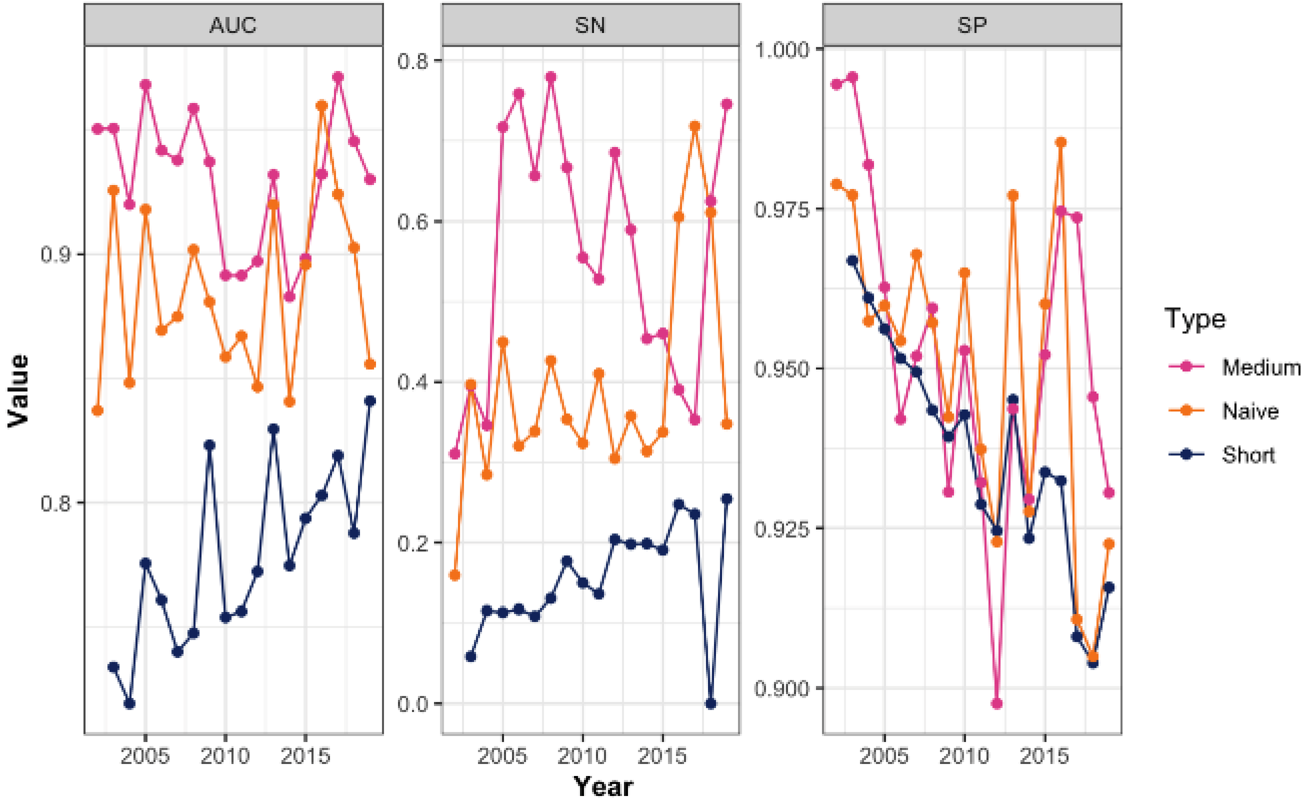
LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

Model stability varies across timescales

Mexico



Brazil

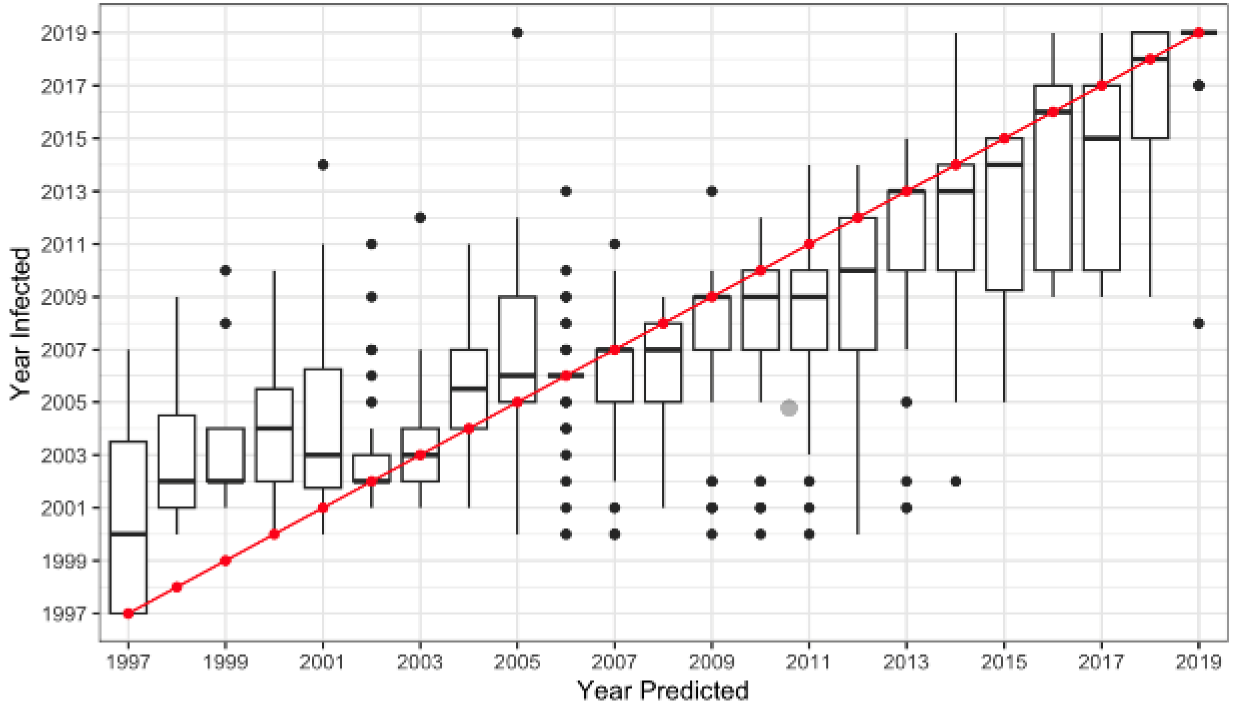


UNIVERSITY OF TORONTO

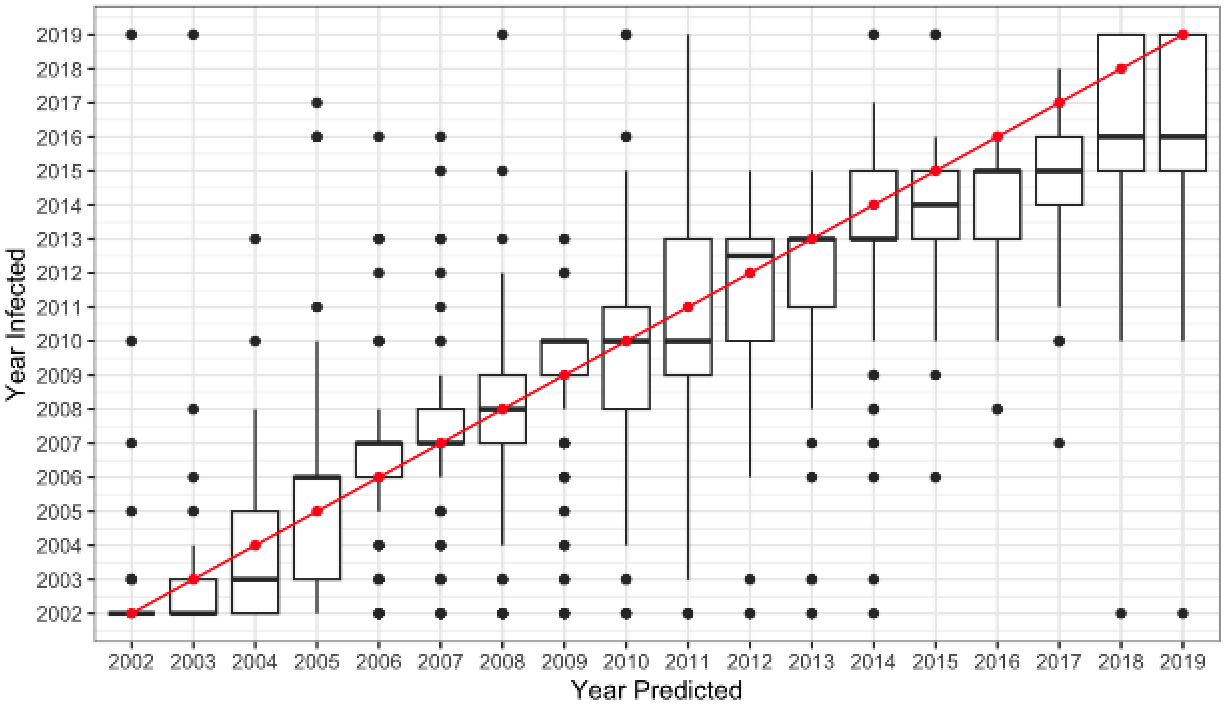
LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Medium-term predictions are well-calibrated

Mexico



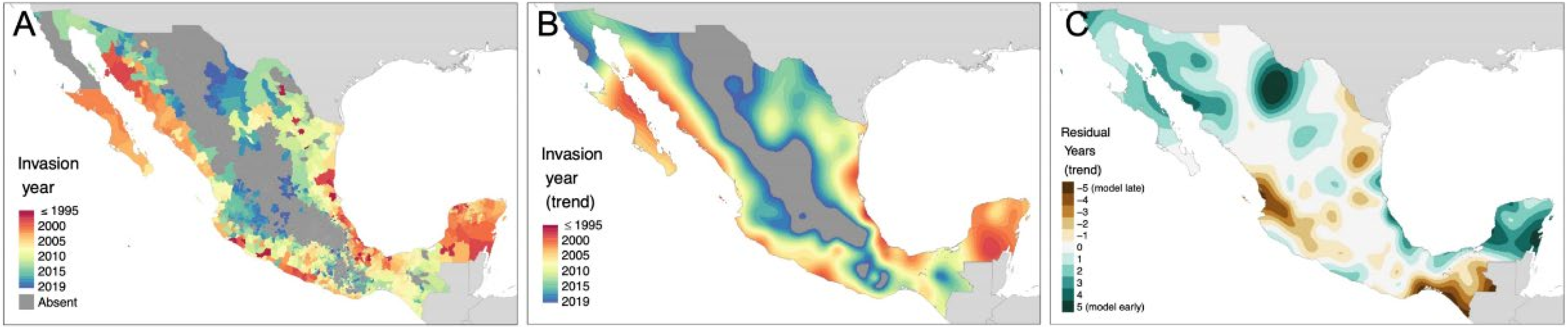
Brazil



UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

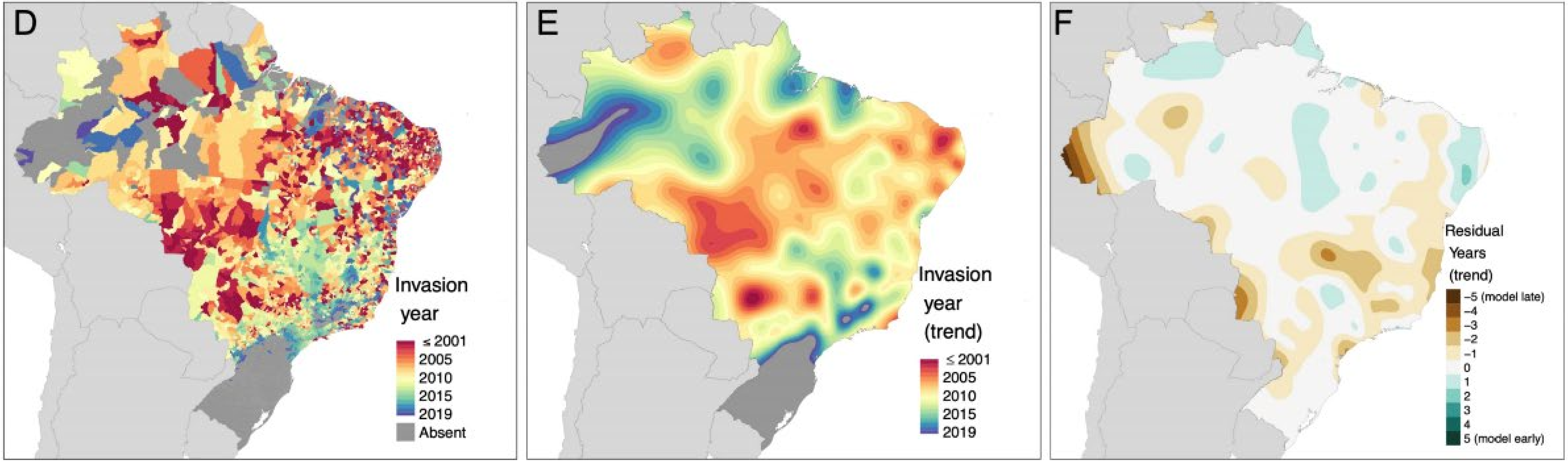
Contemporary predictions for Mexico



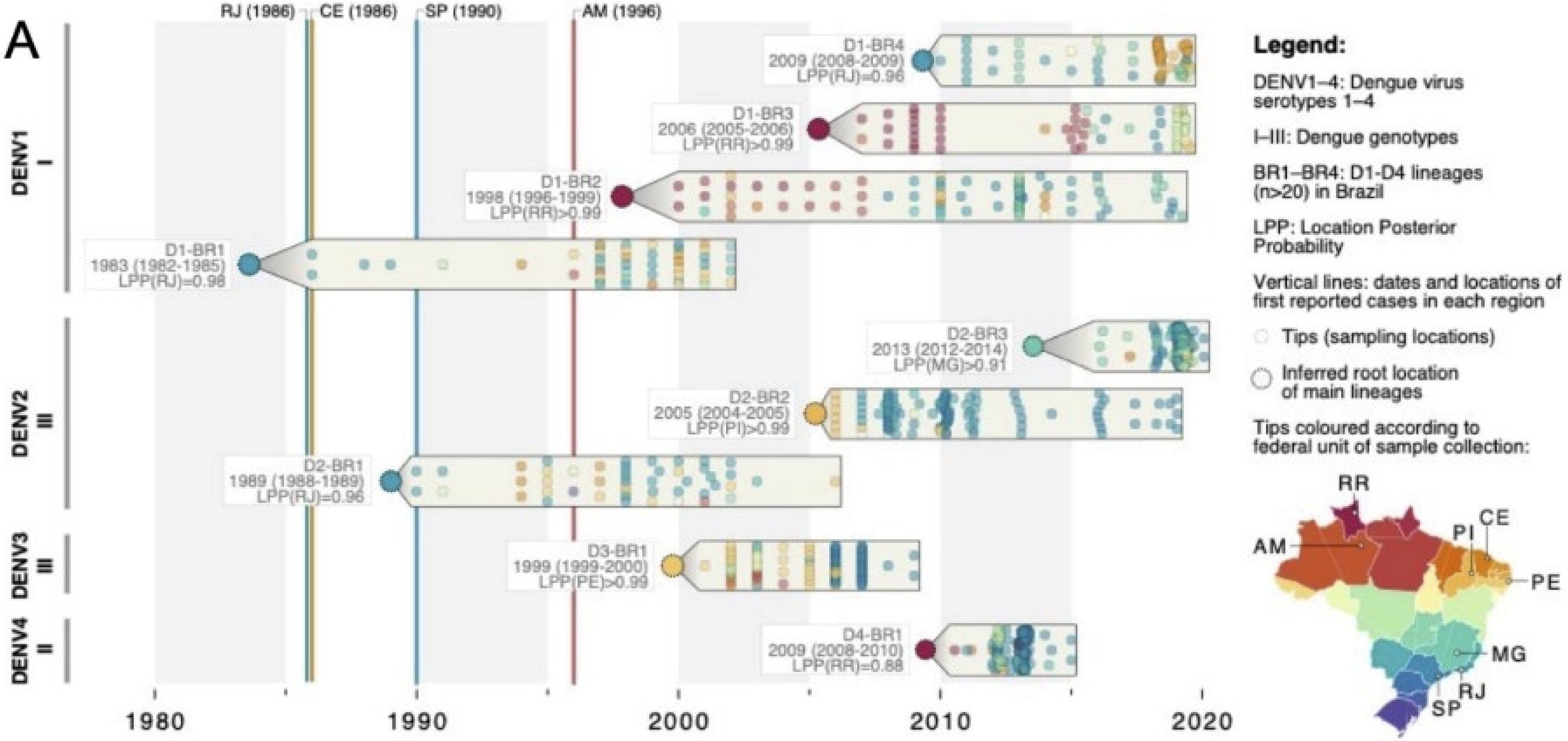
UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

Contemporary predictions for Brazil



Historical reconstruction unveils plausible spread pattern

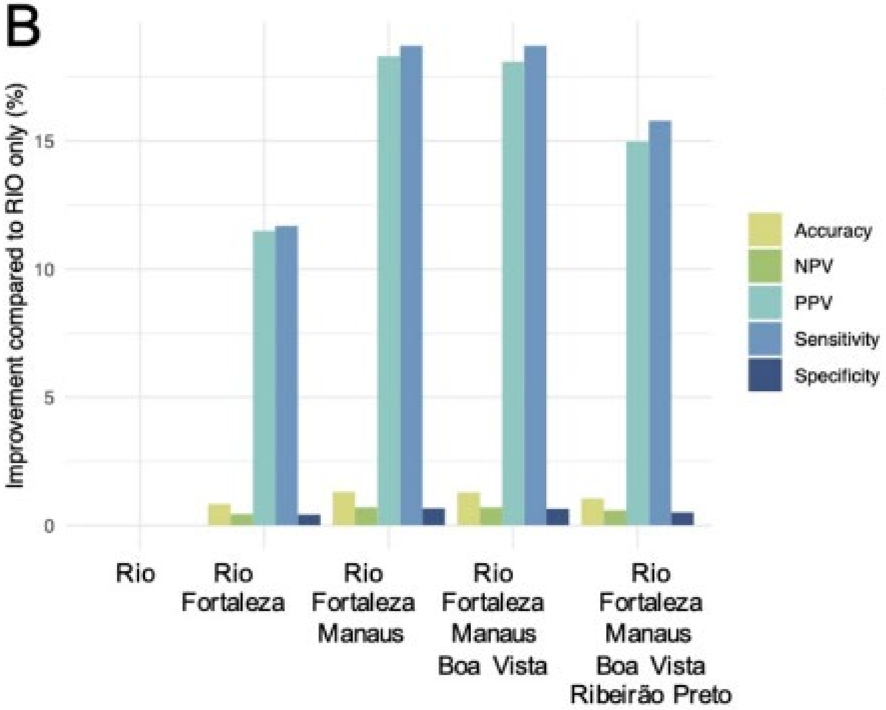


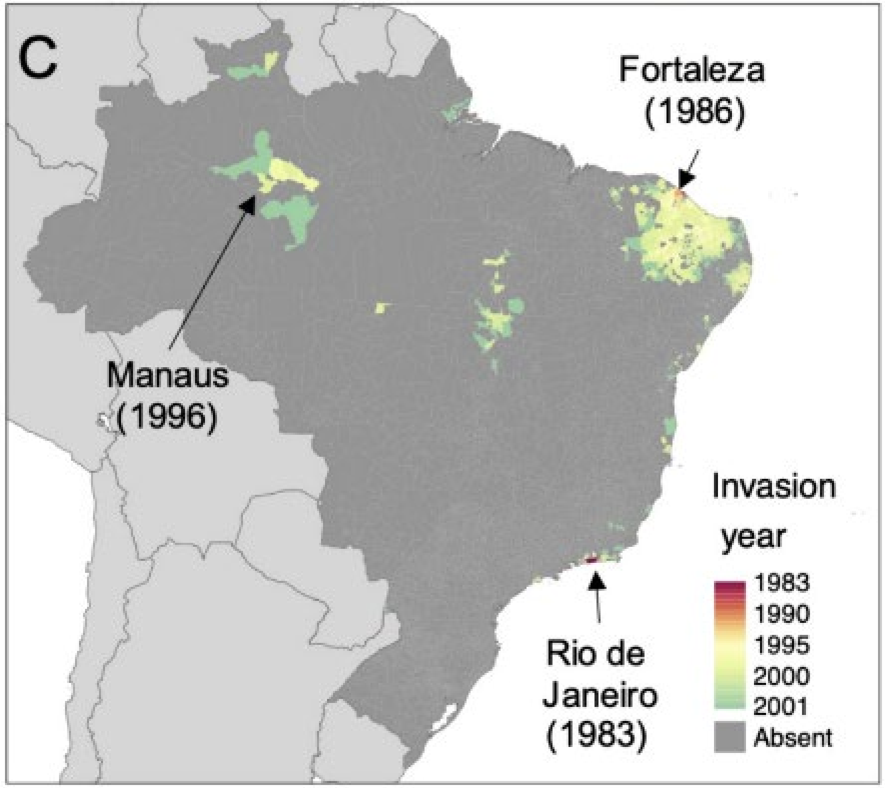
- Analysis led by: Filipe R.R. Moreira & Nuno Faria

UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

Historical reconstruction unveils plausible spread pattern



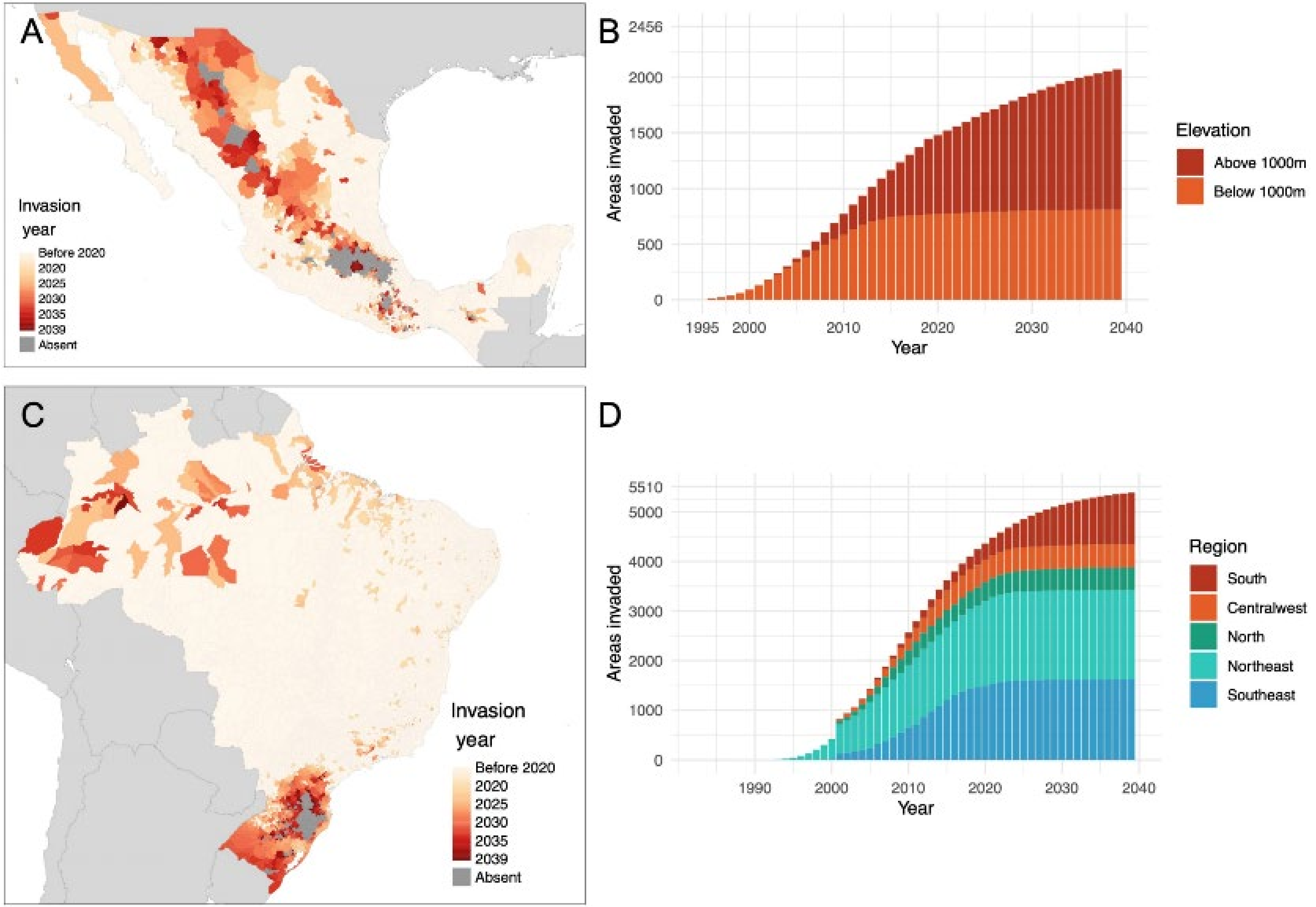


- Highest fidelity reconstruction obtained with 3 sources shown in C:
  - **ACC** → 0.873; **SN** → 0.369; **SP** → 0.926; **PPV** → 0.359; **NPV** → 0.930

UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

Future projections



Putting it all together

UNIVERSITY OF TORONTO

Wrap Up

27

UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

Key learnings and limitations

- Using well-understood climatic and mobility features, a two-step modelling approach combining GBDT and survival model for thresholding can accurately model dengue spread process on medium-term time horizon

- High-resolution maps of arrival times generated by models can fill historical surveillance gaps and offer context useful for interpreting a wide range of epidemiological data

- **Limitations:**
  - Assumed the continuous presence of dengue in a municipality once it arrives – but in reality transient invasion does occur
  - Assumed consistent surveillance capacity, but likely improving over time

Future directions

- Explore model transportability to other settings within and beyond Latin America
- Predict serotype-specific arrival times
- Incorporate features that capture international spread (e.g., international air travel)
- Experiment with Bayesian machine learning approaches that better capture uncertainty (e.g. Gaussian process boosting) [16]

[16] Sigrist, Arxiv 2020

Acknowledgements

VH funding:

CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada

Ontario

VECTOR INSTITUTE

OB funding:

wellcome

UKRI Medical Research Council

Centre for Arbovirus cadde Discovery, Diagnostics, Genomics & Epidemiology

KT funding:

BILL & MELINDA GATES foundation

Contact:

Vinyas – v.harish@mail.utoronto.ca
Oli - Oliver.Brady@lshtm.ac.uk

UNIVERSITY OF TORONTO

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE