Implications of climatic extremes, socioeconomic inequalities, and human mobility for predicting dengue across spatial scales in Vietnam

Sophie Belman, Chloe Fletcher, Rachel Lowe

Global Health Resilience Group
Barcelona Supercomputing Center

Identifying compound drivers and improving forecasts

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Aim:

To evaluate and compare the implications of climatic extremes, socioeconomic inequalities and human mobility, and the interactions between them, on predicting dengue dynamics across spatial scales in Vietnam, focusing on the province and district level.

Objectives:

1. To identify and compare additive and **interacting climatic, socioeconomic and environmental drivers of dengue** outbreaks at the province and district level in Vietnam, and evaluate trade-offs in the **predictive performance across spatial scales**.
2. To compare different methods of **accounting for spatial structure, including human mobility, in dengue prediction models** at the district level in Vietnam and whether they are more important in endemic (South) or emerging (North) regions.
3. To identify an **optimal district-level model for predicting dengue outbreak risk** in Vietnam to inform existing early warning system initiatives.

E4WARNING

Climatic, socioeconomic and mobility drivers in Vietnam

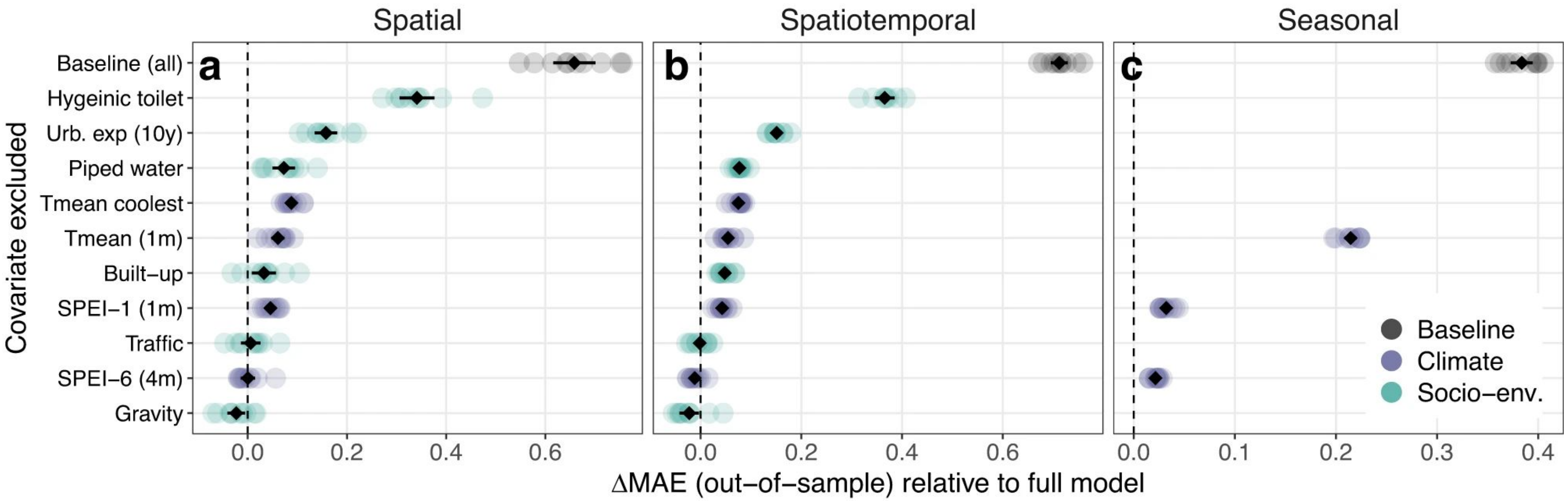


Figure: Influence on out-of-sample MAE when excluding each covariate from full model using three 5-fold cross validation hold-out designs (spatial, spatiotemporal, seasonal). Values above zero indicate an increase in prediction error relative to the full model when excluded (i.e. positive influence on predictive accuracy).

Source: Gibb et al. (2023) ***Nature Comms***

Climatic, socioeconomic and mobility drivers in Vietnam

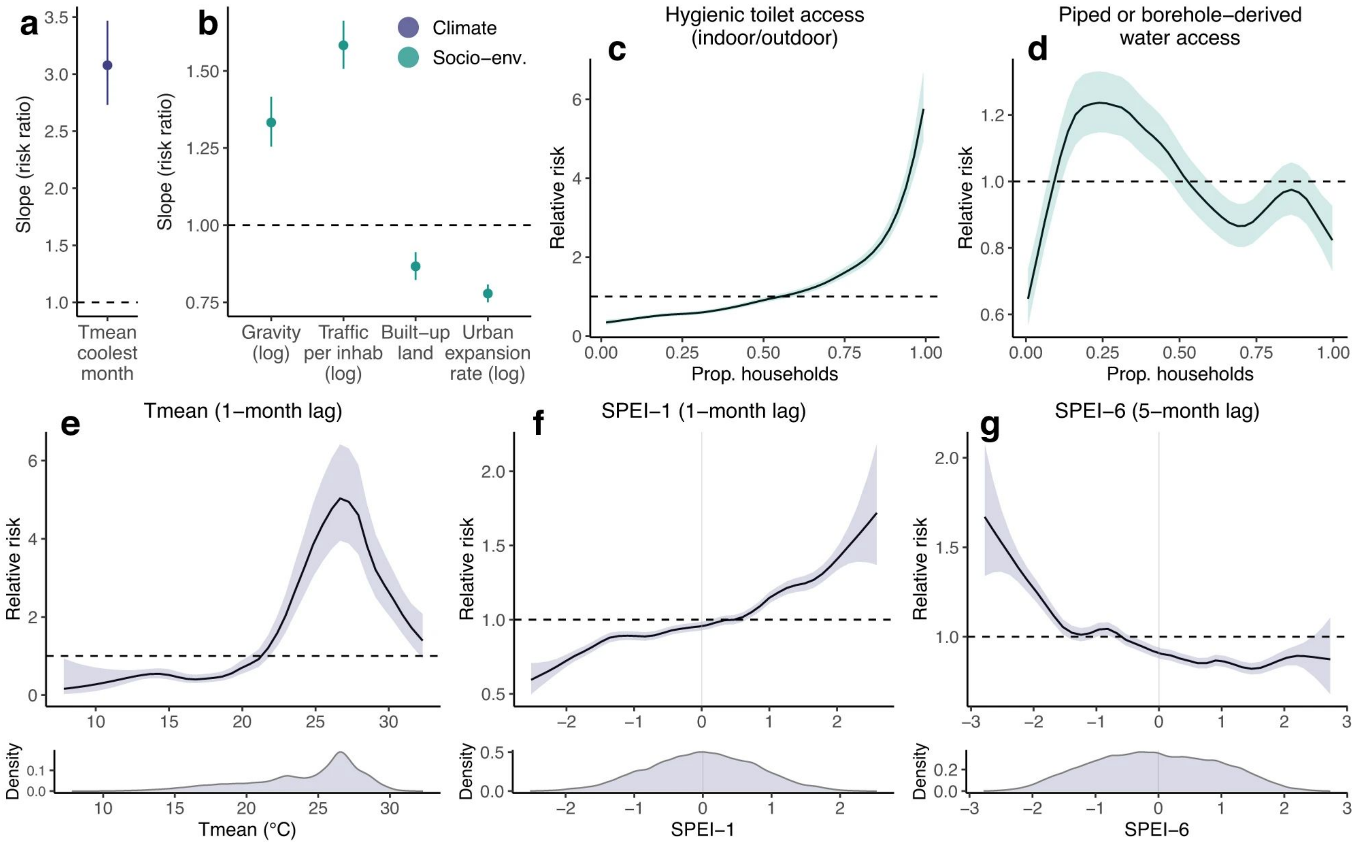


Figure: Linear and nonlinear fixed effects of climatic, socioeconomic, and mobility covariates on dengue incidence in Vietnam using a Bayesian hierarchical mixed-effects model, represented as risk ratios for scaled or log-transformed covariates. Results show the mean and 95% credible interval.

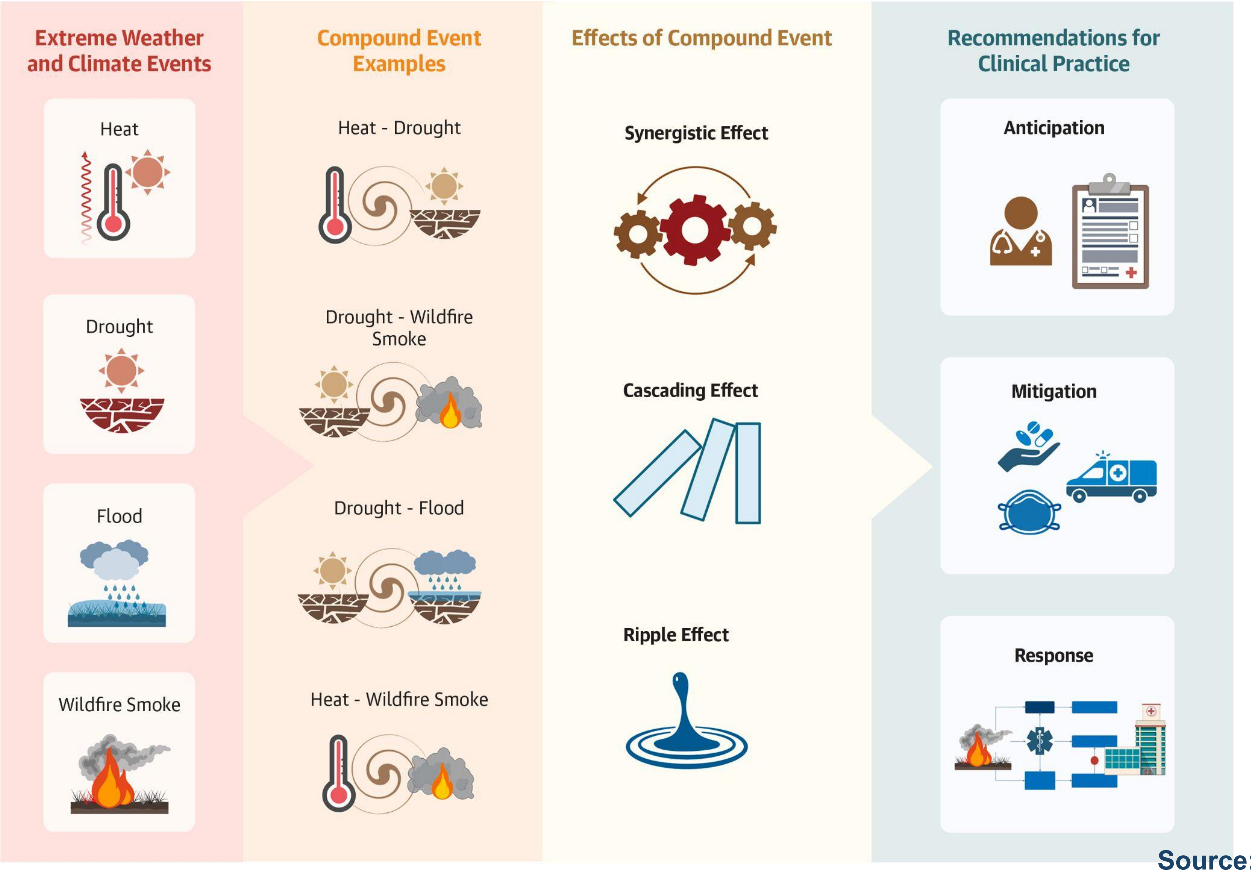
Source: Gibb et al. (2023) *Nature Comms*

Modelling the effects of compound climatic events on dengue risk

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

E4WARNING

Effects of compound climatic extremes on disease risk



Source: Chen et al. (2025) *JACC*

Compound climate extremes on dengue risk in Barbados

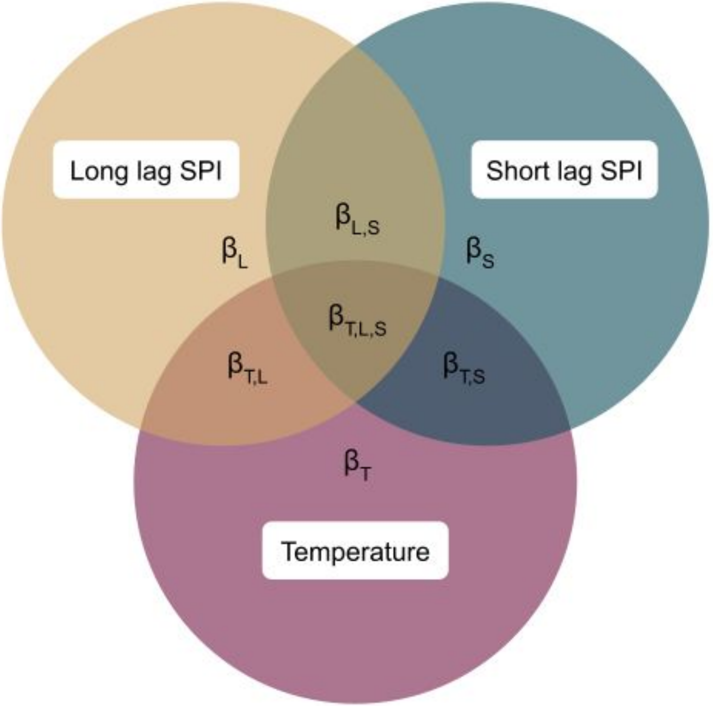


Figure: Visual representation of climate coefficients in the model where T = temperature, L = long-lag SPI, S = short-lag SPI. The equation includes individual effects (β_T, β_L and β_S), coupled interaction effects ($\beta_{T,L}$, $\beta_{T,S}$ and $\beta_{L,S}$) and three-way interaction effects ($\beta_{T,L,S}$).

Source: Fletcher et al. (2025) *Lancet Planet Health*

Compound climate extremes on dengue risk in Barbados

t	=	temporal index
T	=	temperature
L	=	long-lag SPI variable
S	=	short-lag SPI variable
a(t)	=	annual index
m(t)	=	month index

$$y_t \mid \mu_t, \kappa \sim \mathrm{NegBin}(\mu_t, \kappa)$$

$$\log(\mu_t) = \log(\rho_t) + \log(p_{a(t)})$$

$$\underbrace{\log(\rho_t)}_{\text{dengue incidence rate}} = \alpha + \underbrace{\beta_T X_T + \beta_L X_L + \beta_S X_S + \beta_{T,L} X_T X_L + \beta_{T,S} X_T X_S + \beta_{L,S} X_L X_S + \beta_{T,L,S} X_T X_L X_S}_{\text{climate covariates}} + \underbrace{\eta \log(\rho_{t-4} + 1)}_{\text{lagged dengue incidence rate}} + \underbrace{\delta_{m(t)} + \gamma_{a(t)}}_{\text{temporal random effects}}$$

Source: Fletcher et al. (2025) *Lancet Planet Health*

Compound climate extremes on dengue risk in Barbados

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

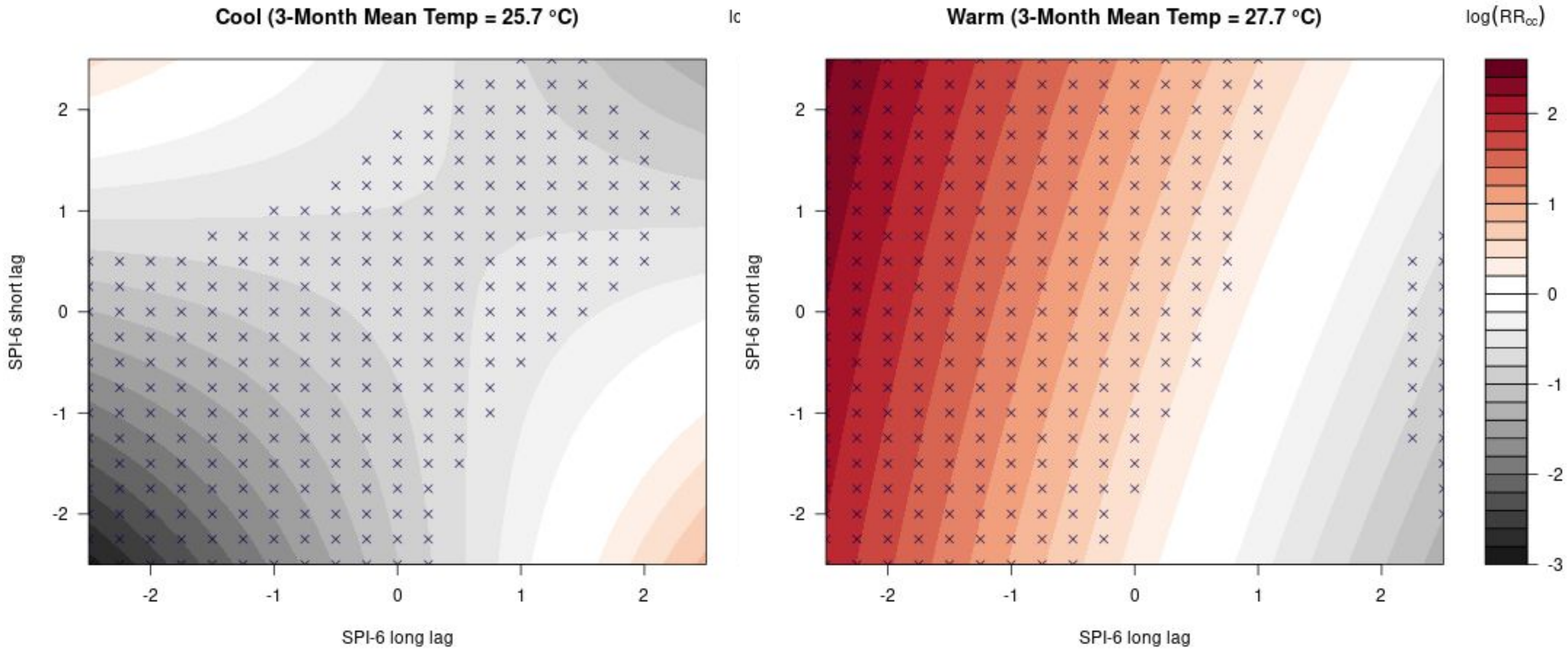
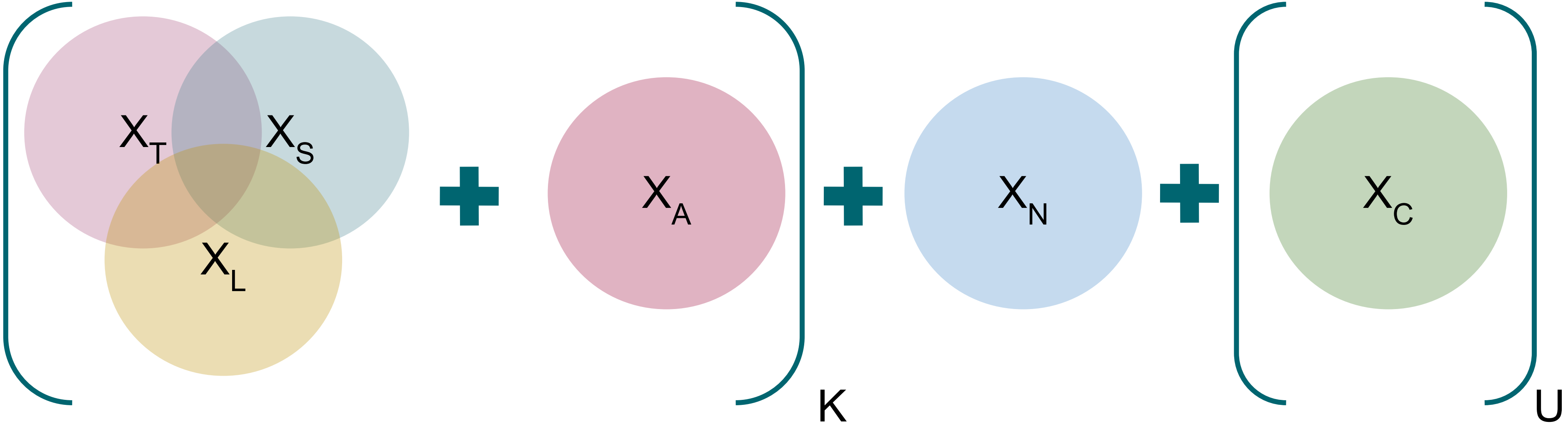


Figure: Mean compound contribution of climate interactions on dengue incidence (colour) where X indicates estimates with 95% credible intervals not including zero. We tested different long-lag and short-lag SPI-6 combinations under cool and warm conditions.

Source: Fletcher et al. (2025) *Lancet Planet Health*

Compound climate extremes on dengue risk in Brazil

Barcelona Supercomputing Center
Centro Nacional de Supercomputación



T =	6-month mean temp anomaly	(lagged 1 month)
L =	12-month SPEI	(lagged 3 months)
S =	3-month SPEI	(lagged 1 month)
A =	6-month mean abs temp	(lagged 1 month)
N =	3-month Niño 3.4 index (ONI)	(lagged 7 months)
C =	Pre-2018 binary index	-
K =	Köppen classification	-
U =	State	-

+ $(\text{Weekly RE})_U$ and Spatial RE

Source: Fletcher & Mila et al. (in prep)

Compound climate extremes on dengue risk in Brazil

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

t	=	temporal index
a(t)	=	annual index
w(t)	=	week index
s	=	spatial index
U(s)	=	state index
A	=	absolute temperature
T	=	temperature anomaly
L	=	long-lag drought index
S	=	short-lag drought index
N	=	niño index
C	=	post-2017 index (binary)

$$y_{s,t} \mid \mu_{s,t}, \kappa \sim \mathrm{NegBin}(\mu_{s,t}, \kappa)$$

$$\log(\mu_{s,t}) = \log(\rho_{s,t}) + \log(p_{s,a(t)})$$

population offset

$$\log(\rho_{s,t}) = \alpha + (\beta_T X_T + \beta_L X_L + \beta_S X_S + \beta_{T,L} X_T X_L + \beta_{T,S} X_T X_S + \beta_{L,S} X_L X_S + \beta_{T,L,S} X_T X_L X_S + \beta_A X_A)_K + \beta_N X_N + \beta_{C,U(s)} X_{C,U(s)} + \delta_{w(t),U(s)} + \gamma_{a(t)} + u + v$$

dengue incidence rate

climate and oceanic covariates

spatio-temporal random effects

Source: Fletcher & Mila et al. (in prep)

Compound climate extremes on dengue risk in Brazil

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

t	=	temporal index
a(t)	=	annual index
w(t)	=	week index
s	=	spatial index
U(s)	=	state index
A	=	absolute temperature
T	=	temperature anomaly
L	=	long-lag drought index
S	=	short-lag drought index
N	=	niño index
C	=	post-2017 index (binary)

$$y_{s,t} \mid \mu_{s,t}, \kappa \sim \mathrm{NegBin}(\mu_{s,t}, \kappa)$$

$$\log(\mu_{s,t}) = \log(\rho_{s,t}) + \log(p_{s,a(t)})$$

population offset

$$\log(\rho_{s,t}) = \alpha + (\beta_T X_T + \beta_L X_L + \beta_S X_S + \beta_{T,L} X_T X_L + \beta_{T,S} X_T X_S + \beta_{L,S} X_L X_S +$$

$$\beta_{T,L,S} X_T X_L X_S + \beta_A X_A)_K + \beta_N X_N + \beta_{C,U(s)} X_{C,U(s)} + \delta_{w(t),U(s)} \cancel{+ \gamma_{a(t)}} + u + v$$

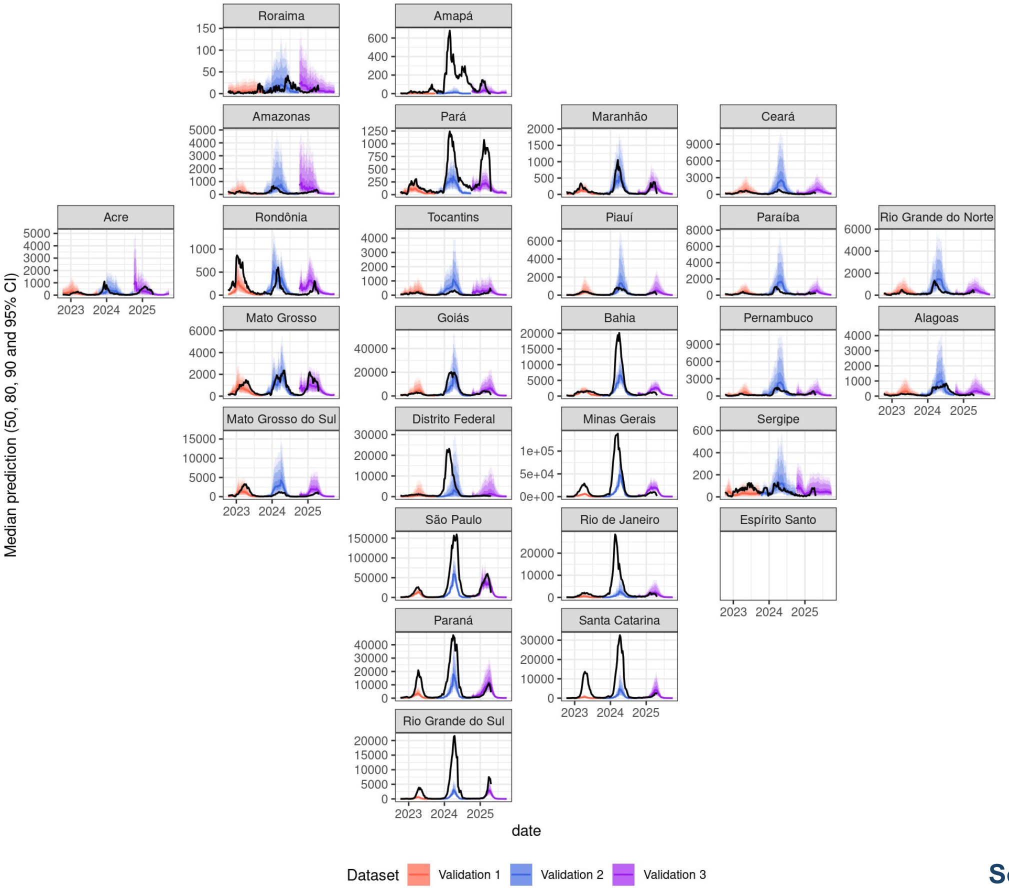
dengue incidence rate

climate and oceanic covariates

spatio-temporal random effects

Source: Fletcher & Mila et al. (in prep)

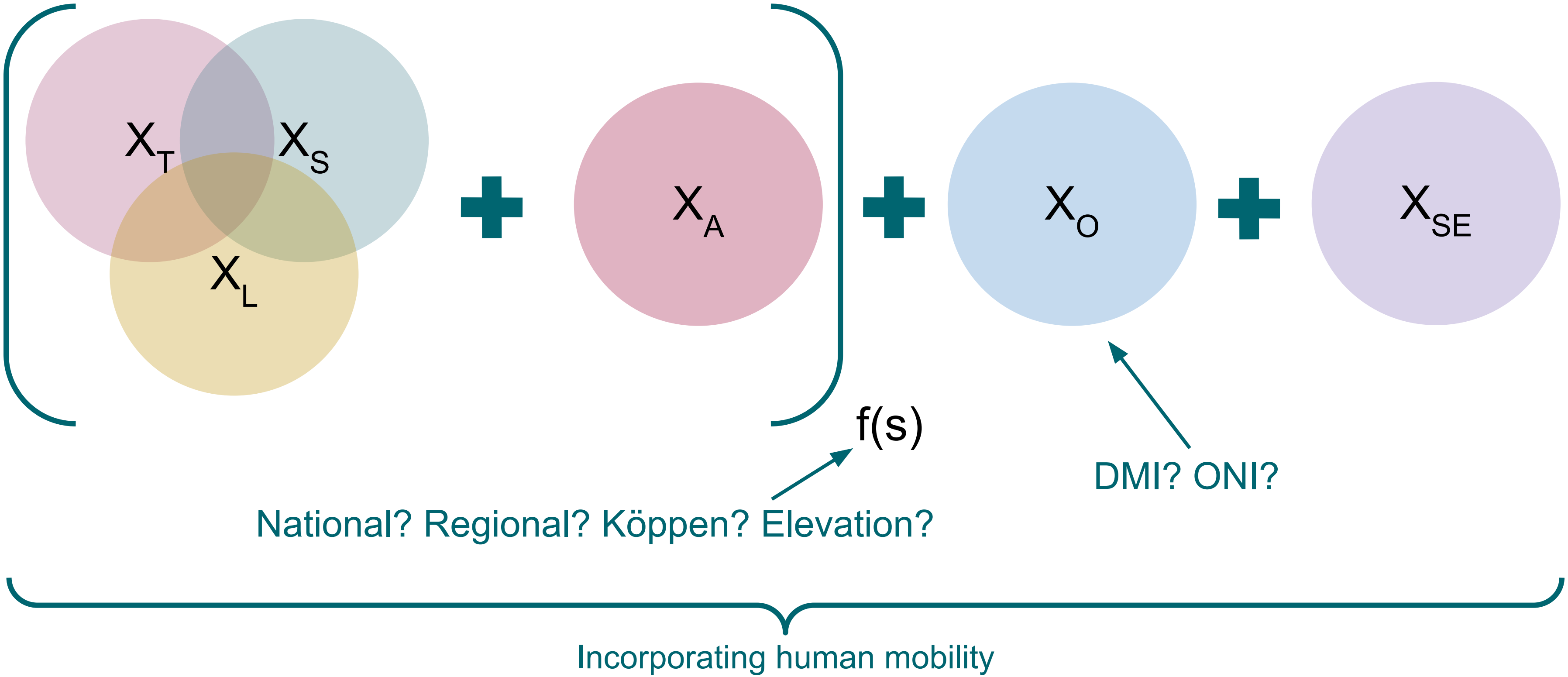
Forecasting weekly state-level dengue risk in Brazil



Source: Fletcher & Mila et al. (in prep)

Dengue modelling framework in Vietnam

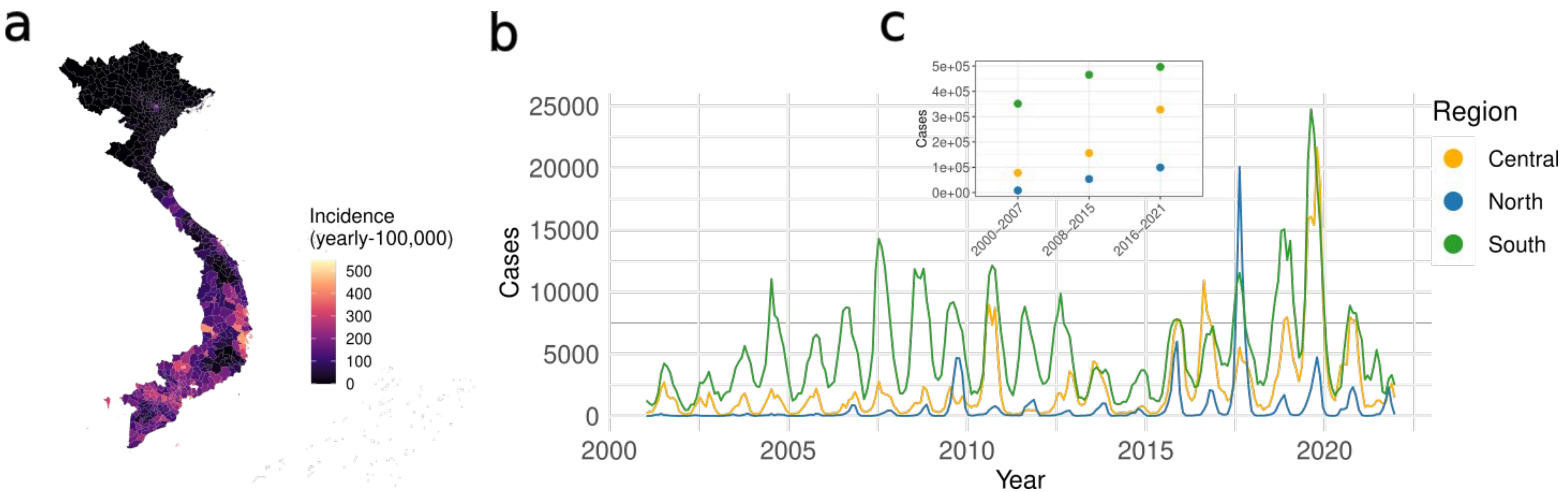
Barcelona Supercomputing Center
Centro Nacional de Supercomputación



Source: Belman & Fletcher et al. (in prep)

Dengue models in Vietnam in emerging versus endemic regions

Dengue virus cases in Vietnam across 23 years



Increasing in the Central and North/ more stable in the South

Random effects by region

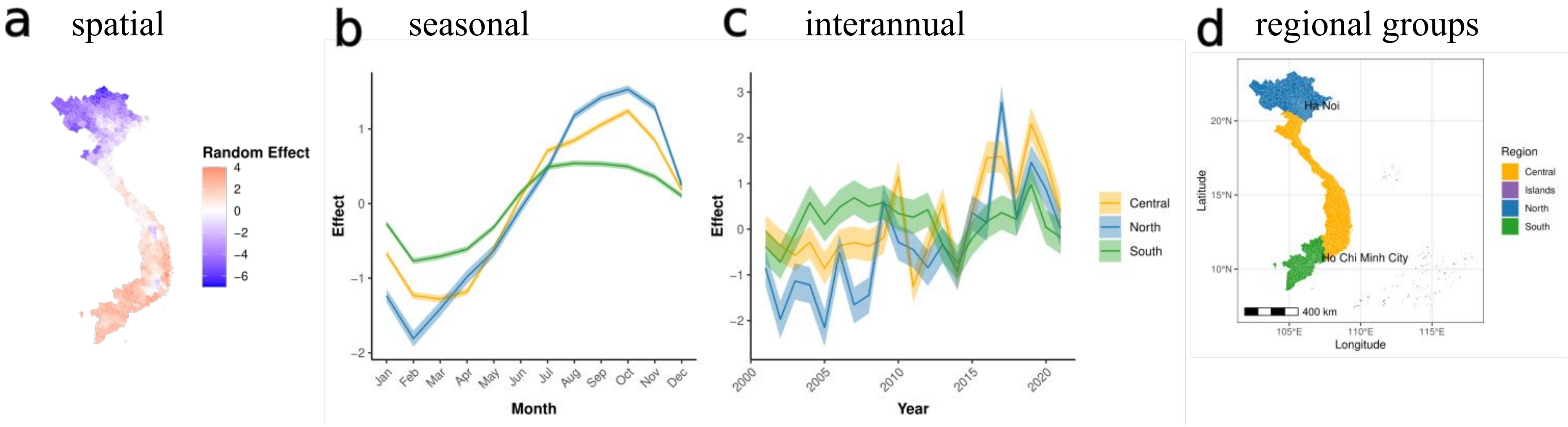


Figure 2. Random effects from model B3. B3 is the district (N=669) level monthly model of dengue cases in Vietnam from 2001-2021. *Random effects include: (a) spatial, (b) seasonal replicated by region, and (c) interannual replicated by region including Central (yellow), North (blue), and South (green) as regions. (d) indicates the regional stratifications we used including North (blue), Central (yellow), South (green), and islands (purple).*

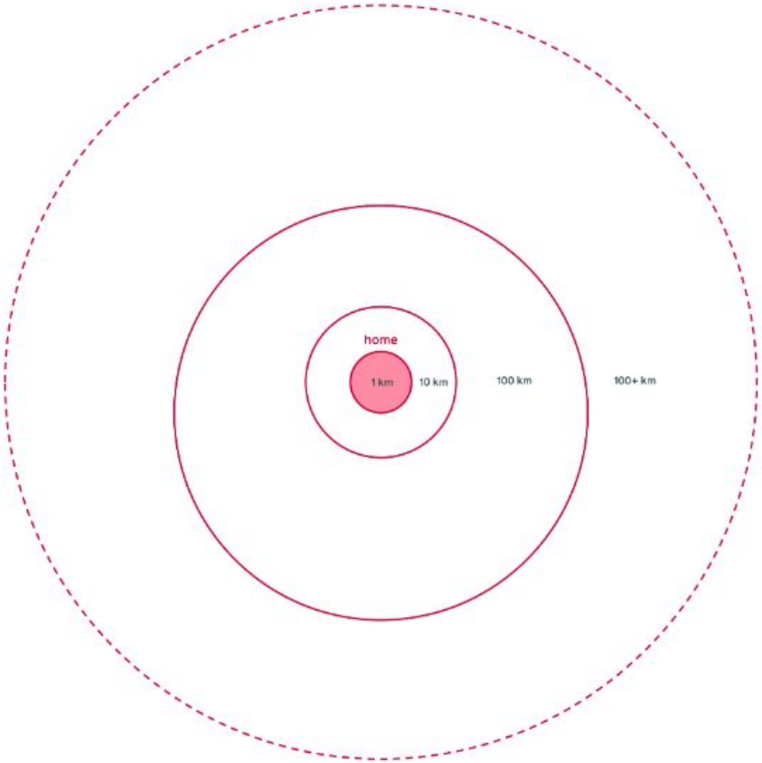
Disentangling the role of human mobility towards improving models in different regions

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

E4WARNING

Movement Distribution Maps

Barcelona Supercomputing Center
Centro Nacional de Supercomputación



Step 1: Assign people to geographic areas given nighttime location

Step 2: Calculate movement distribution
selecting a random location update for each person during the day and how far it is

Step 3: Assign Distance Categories
calculate the percentage of people in each of four distance categories
0km, 0-10km, 10-100km and 100km+.

Step 4: Add noise
Add noise to ensure anonymity

Meta
Data for Good

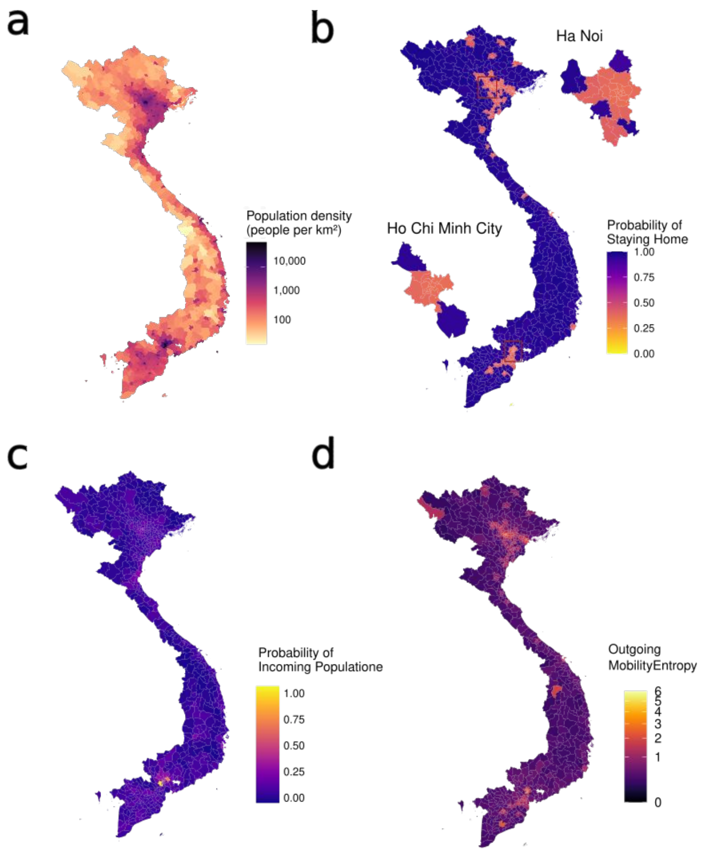
https://dataforgood.facebook.com/dfg/tools/movement-distribution-maps

Movement Distribution to Probability Matrix

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

1. **Recording Movement:** Movement distribution data takes a random location for each person/day and determine how far it is from the nighttime location.
2. **Distances:** Assign 4 distances (0km, 1-9km, 10-99km, 100+)
3. **Reported Fractions:** For each distance for each day the fraction (sum 1) of pings is reported.
4. **Reattribute to District Distances:** Merge with distance matrix to re-attribute fractions depending upon whether the distance is within the home district or a neighboring district. Assign fractions to each distance pair and divide by the number of districts at the distance to produce a probability for each row.

Adapt human mobility matrix into to covariates



(a) population density per 100,000 people per district in Vietnam (km^2)

(b) probability of staying within the home district whereby a high probability (>0.75) (purple) indicates not very much movement, and a low probability <0.25 (orange) indicates a low probability of staying in the home district. The insets demonstrate the high mobility between districts within Hà Nội and Hồ Chí Minh City

(c) probability of people coming into each district (normalized to the maximum number of incoming people)

(d) outgoing Shannon mobility entropy where a higher number indicates more evenness and number of locations visited and a lower number indicates fewer locations visited

Model formulations to understand spatial structure

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Is spatial autocorrelation a proxy for structured human mobility patterns?

$$y_{s,t} \mid \mu_{s,t} \sim \text{NegBin}(\mu_{s,t}, \kappa)$$

$$\log(\mu_{s,t}) = \log(p_{s,a(t)}) + \log(\rho_{s,t})$$

$\mu_{s,t}$: product of the population $p_{s,a(t)}$ (per 100,000 people) in a given spatial unit, s, (district or province) and year a(t), and the disease incidence $\rho_{s,t}$

α : intercept

$\delta_{m(t)}$: seasonal effect

$\gamma_{r,a(t)}$: regionally replicated interannual effect

$u_s + v_s$: spatial autocorrelation

<u>Including spatial autocorrelation</u>

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + u_s + v_s$$

<u>Including spatial covariates</u>

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + staying\_probability$$

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + incoming\_probability$$

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + Shannon\ mobility\ entropy$$

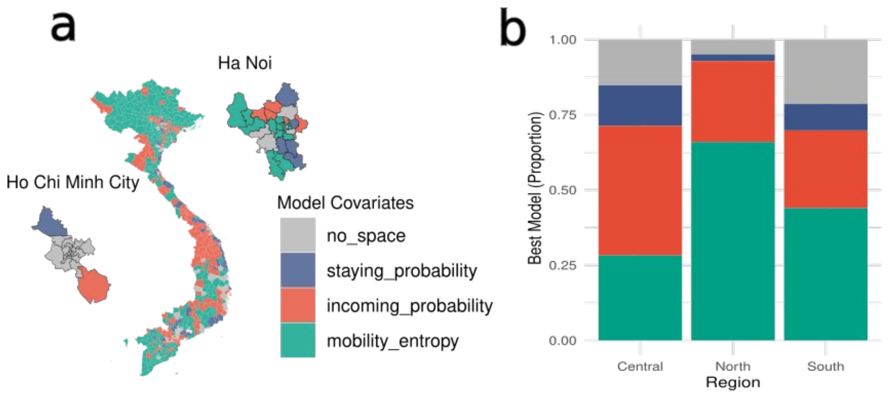
<u>Including spatial covariates + autocorrelation</u>

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + u_s + v_s + staying\_probability^*$$

**example, include all combinations of 3 covariates*

Mobility covariate models less important in Ho Chi Minh City (endemic region)

Ranked each model type with mean absolute error (by district)



<u>Including spatial covariates</u>

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + staying\_probability$$

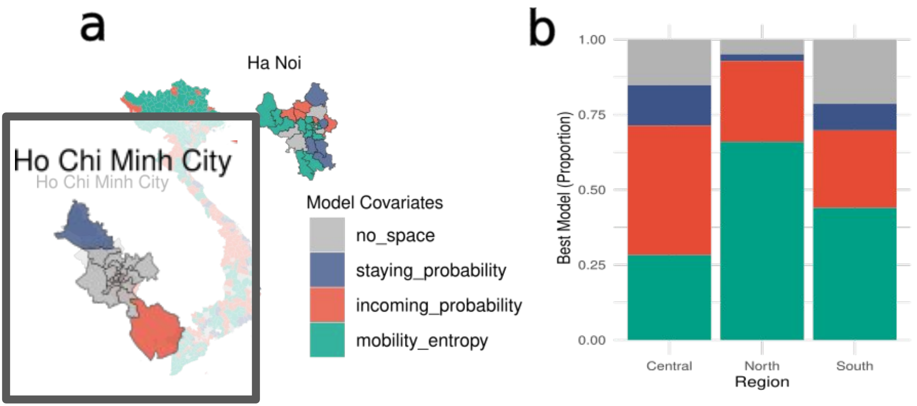
$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + incoming\_probability$$

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + Shannon\ mobility\ entropy$$

Mobility covariate models perform less well in Ho Chi Minh City (endemic region)

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Ranked each model type with mean absolute error (by district)



Including spatial covariates

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + staying\_probability$$

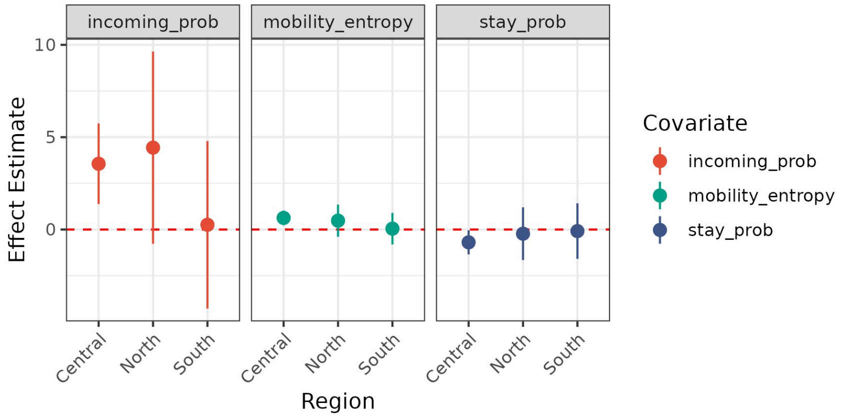
$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + incoming\_probability$$

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + Shannon\ mobility\ entropy$$

*more grey in the south indicates that models without spatial effect are the best performing [seasonal effects more important?]

Mobility covariates strongest effect in Central region

Ranked each model type with mean absolute error (by district)



<u>Including spatial covariates</u>

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + staying\_probability$$

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + incoming\_probability$$

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + Shannon\ mobility\ entropy$$

North and South both crossing 0 whereas mobility has a stronger effect on Central

Replace adjacency matrix with other formulations

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Spatial connectivity types

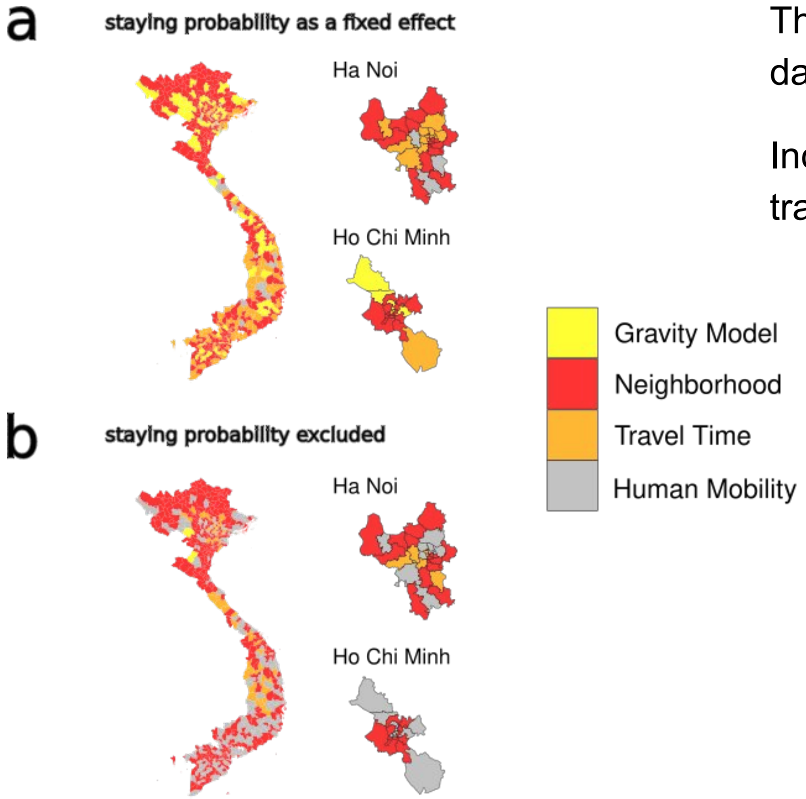
- Human mobility
- Gravity
- Travel Time
- Neighborhood (adjacency matrix)

<u>Including spatial covariates + autocorrelation</u>

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + \boldsymbol{u_s} + \boldsymbol{v_s} + staying\_probability*$$

**example, include all combinations of 3 covariates*

Ranked each model type by district



The best model is probably some combination of these datasets:

Including staying probability covariate improves ranking of travel time and gravity models (a)

<u>Including spatial covariates + autocorrelation</u>

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + u_s + v_s + staying\_probability^*$$

**example, include all combinations of 3 covariates*

Next Steps

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

1. Develop individual district and province level forecasting models using long-short-lag interaction framework including stratification by region, koppen classification, or elevation.

2. Explore incorporation of additional drivers including oceanic indices, human mobility, and socioeconomic variables.

3. Perform rolling origin cross-validation on best models with and without human mobility, by (i) national level (ii) regional stratifications, across spatial resolutions.

E4WARNING

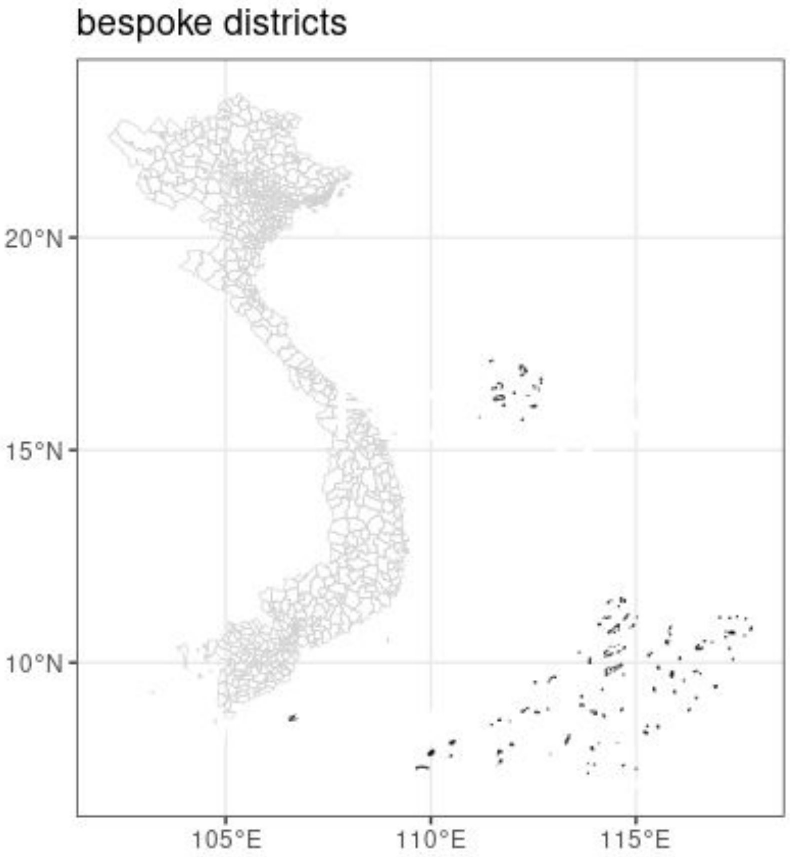
Disputed Islands

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

1. Paracel Islands (70355): occupied by China since 1974 despite being claimed by Vietnam

2. Spratly Islands (70698): ongoing dispute in South China Sea claimed by China, Taiwan, Vietnam, The Philippines, Malaysia, and Brunei

3. Con Son islands in Con Dao District in Ba Ria - Vung Tau Province (70711): ex-french colony known for prisons



E4WARNING

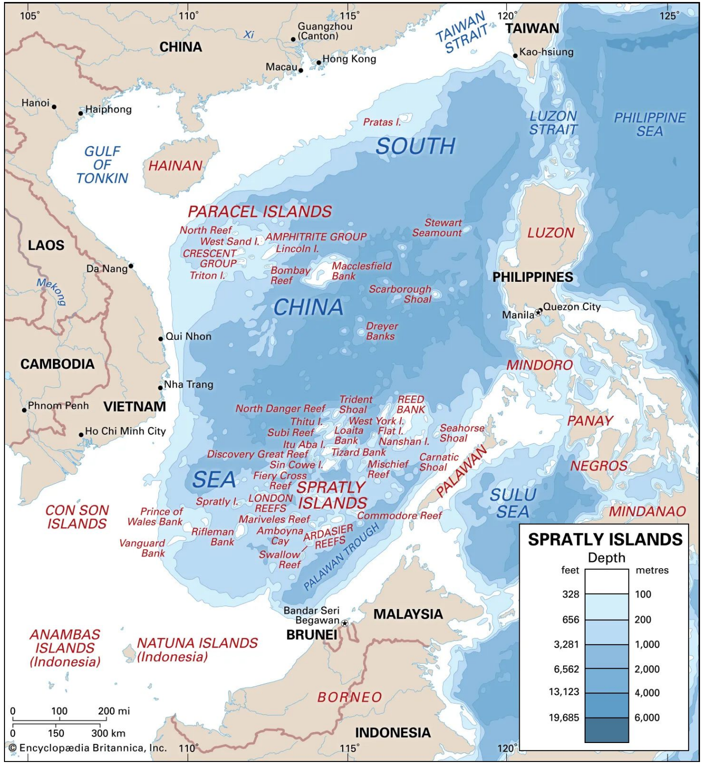
Disputed Islands

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

1. Paracel Islands (70355): occupied by China since 1974 despite being claimed by Vietnam

2. Spratly Islands (70698): ongoing dispute in South China Sea claimed by China, Taiwan, Vietnam, The Philippines, Malaysia, and Brunei

3. Con Son islands in Con Dao District in Ba Ria - Vung Tau Province (70711): ex-french colony known for prisons



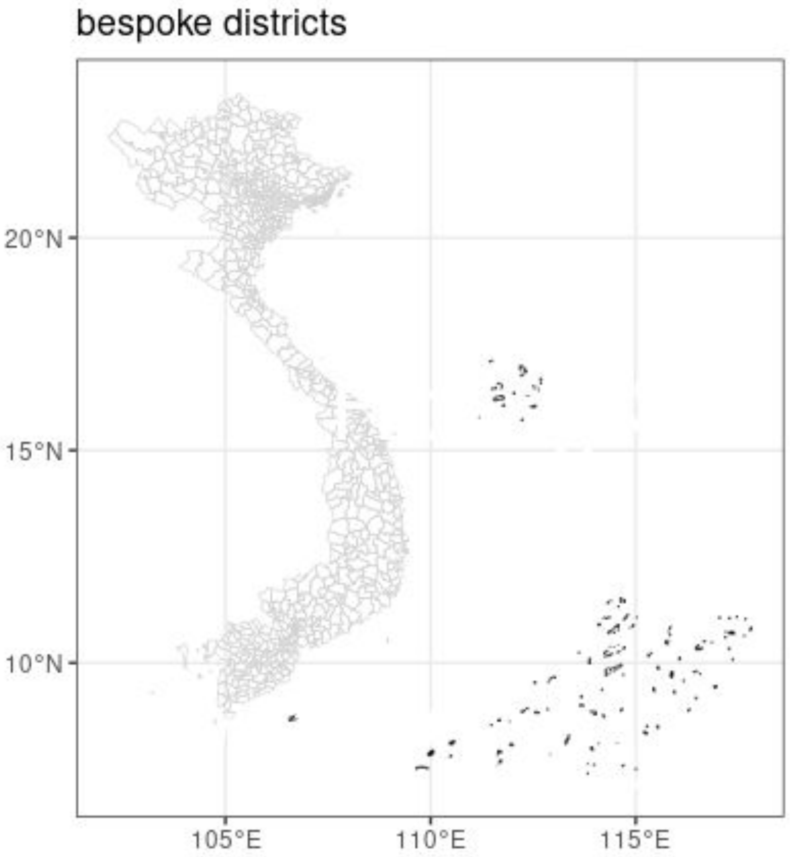
E4WARNING

Dynamic district designations

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

GADM: 710 Districts
Bespoke regions: 672 districts

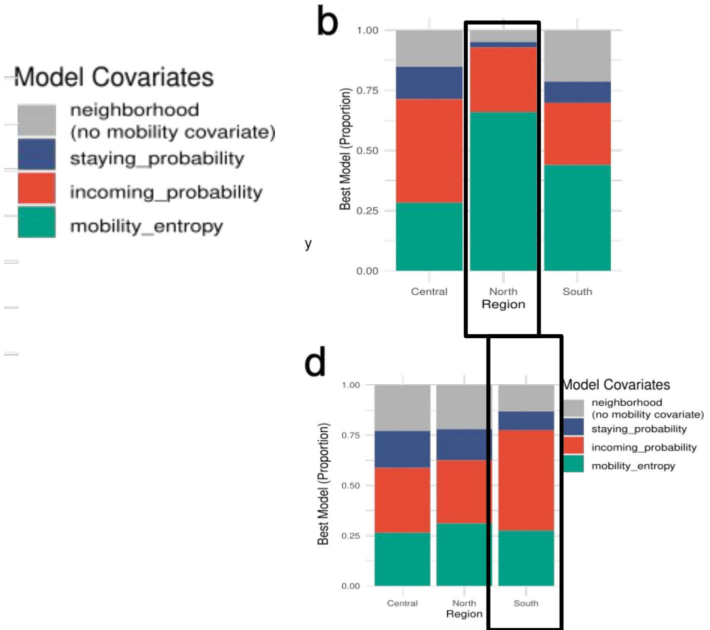
July 1, 2025, Vietnam eliminated the district-level (huyện/quận) administration, transitioning to a two-tier system.
Districts are replaced by a direct, grassroots-level structure consisting of wards (phường) in urban areas, communes (xã) in rural areas, and special zones on islands, directly under provincial or city control



E4WARNING

Some interpretation..

Top: In the north mobility covariates fulfill need of neighborhood matrix implying it mostly encodes network connectivity
Bottom: In the South models with neighborhood matrix and mobility perform best implying climate, vector habitat, sociodemographic factors may be more important



<u>Including spatial covariates</u>

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + staying\_probability$$

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + incoming\_probability$$

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + Shannon\ mobility\ entropy$$

<u>Including spatial covariates and neighborhood matrix</u>

$$log(\rho_{s,t}) = \alpha + \delta_{m(t)} + \gamma_{r,a(t)} + u_s + v_s + staying\_probability$$

Conclusions

Potentially using point estimates is better for dynamic administrative regions