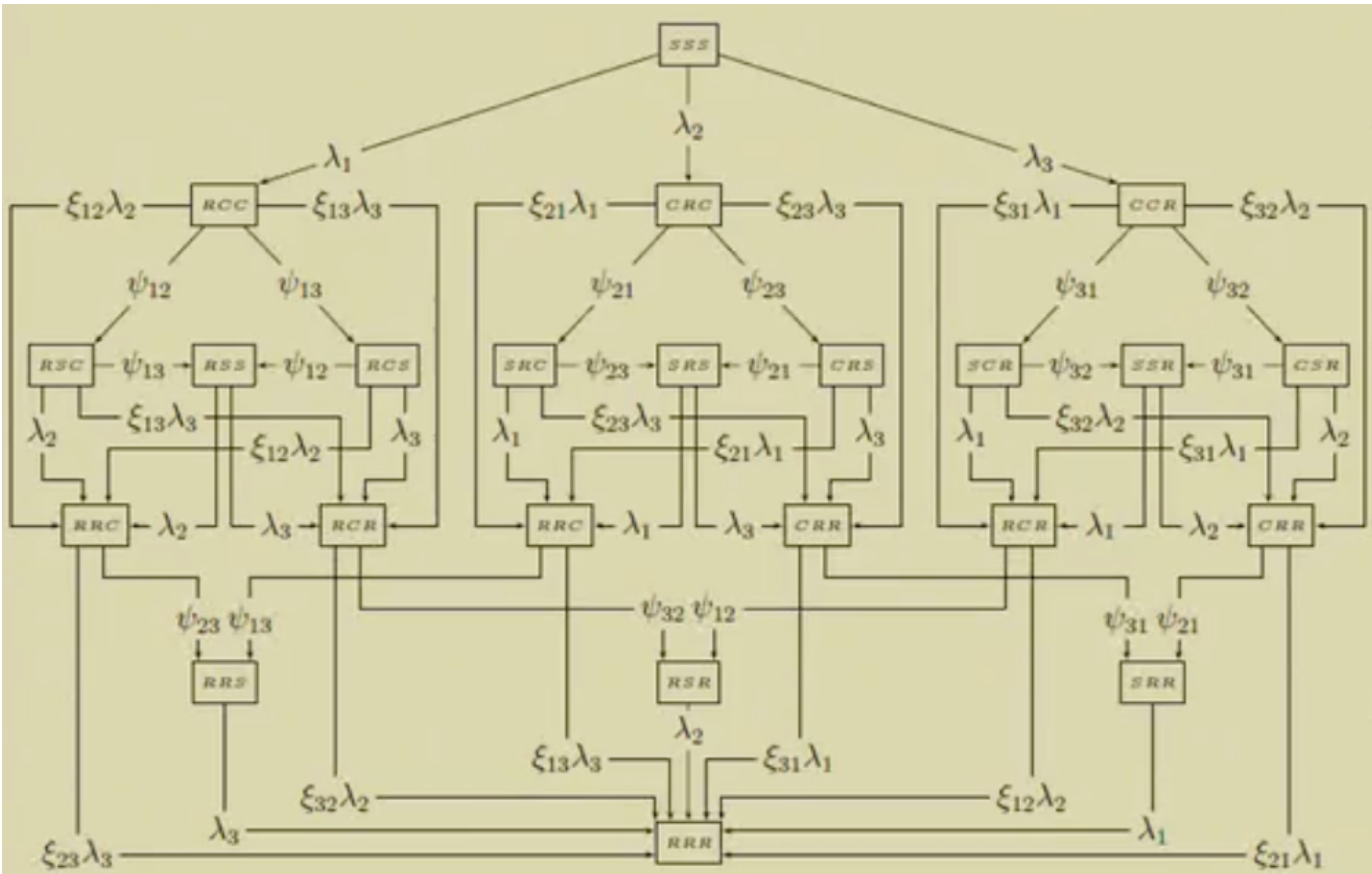
Predicting annual dengue hemorrhagic fever incidence in Thailand

Stephen Lauer, Reich Lab, UMass

Dengue

- Spread by *Aedes* mosquitoes
  - Transmit to humans via bites to mosquitoes via eggs
- Incubation period w/i humans 4-7 days, viremia 7-10 days
- Dengue hemorrhagic fever (DHF) is called break bone disease; can lead to organ failure and death
- Most infections, nearly all first infections asymptomatic
- Complex disease dynamics with four serotypes

Dengue

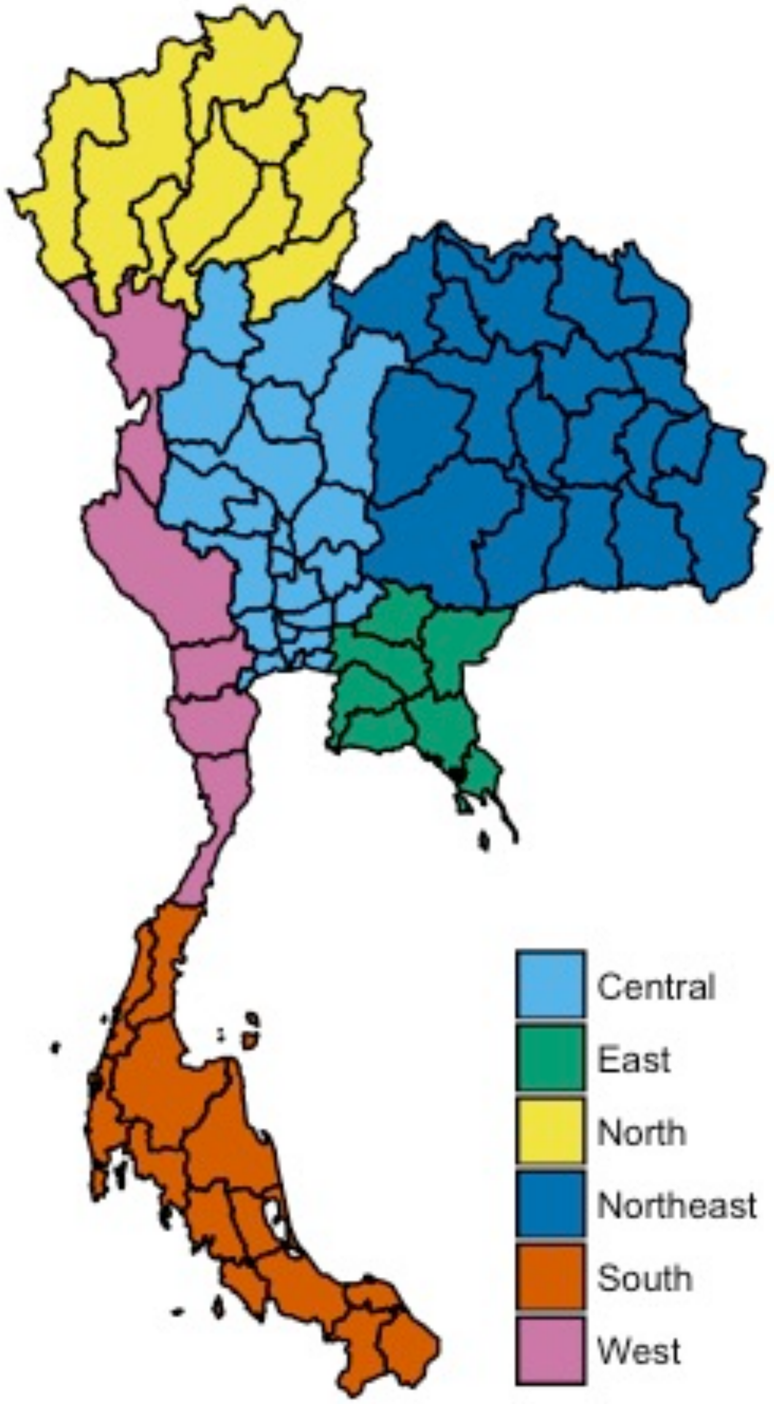


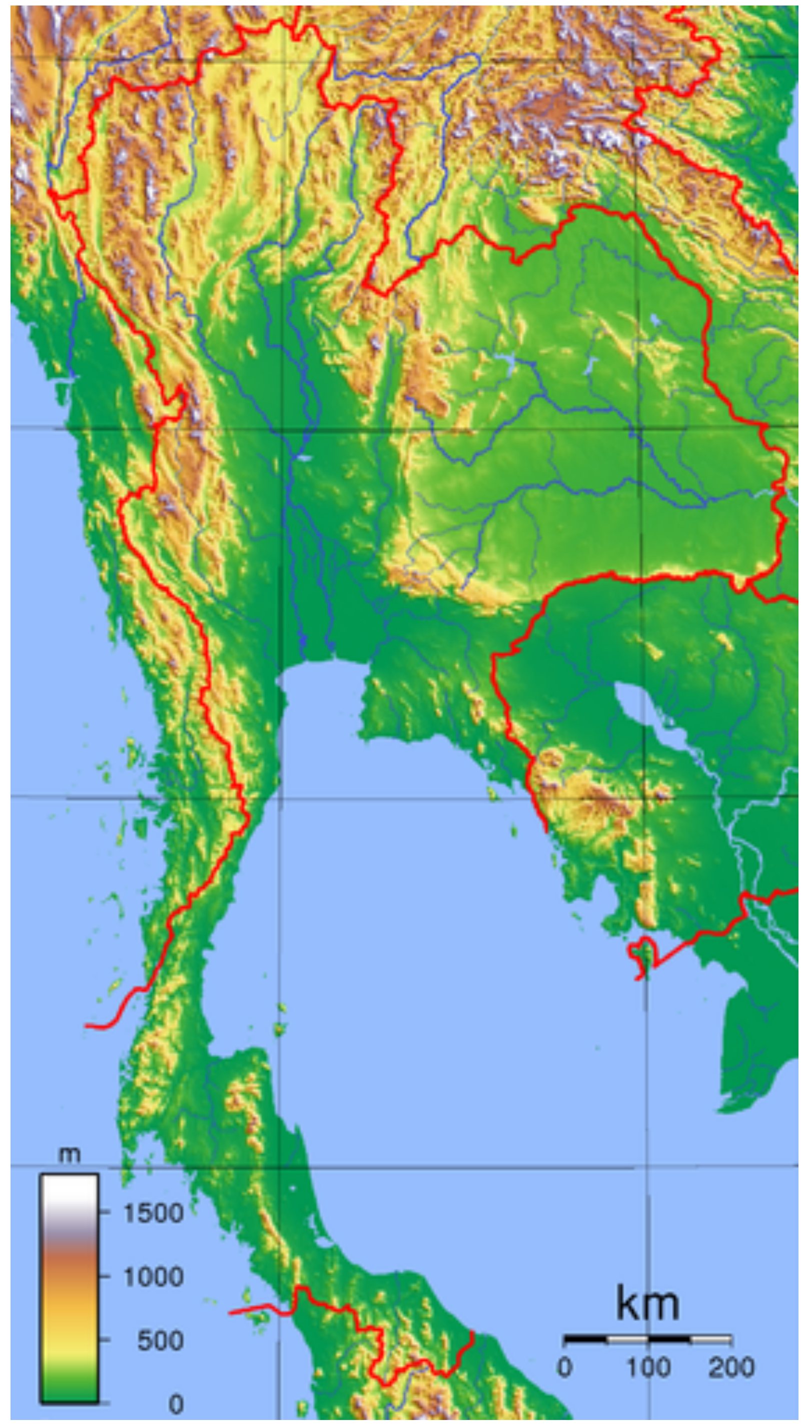
Source: Robert C Reiner Jr presentation at ASTMH 2015

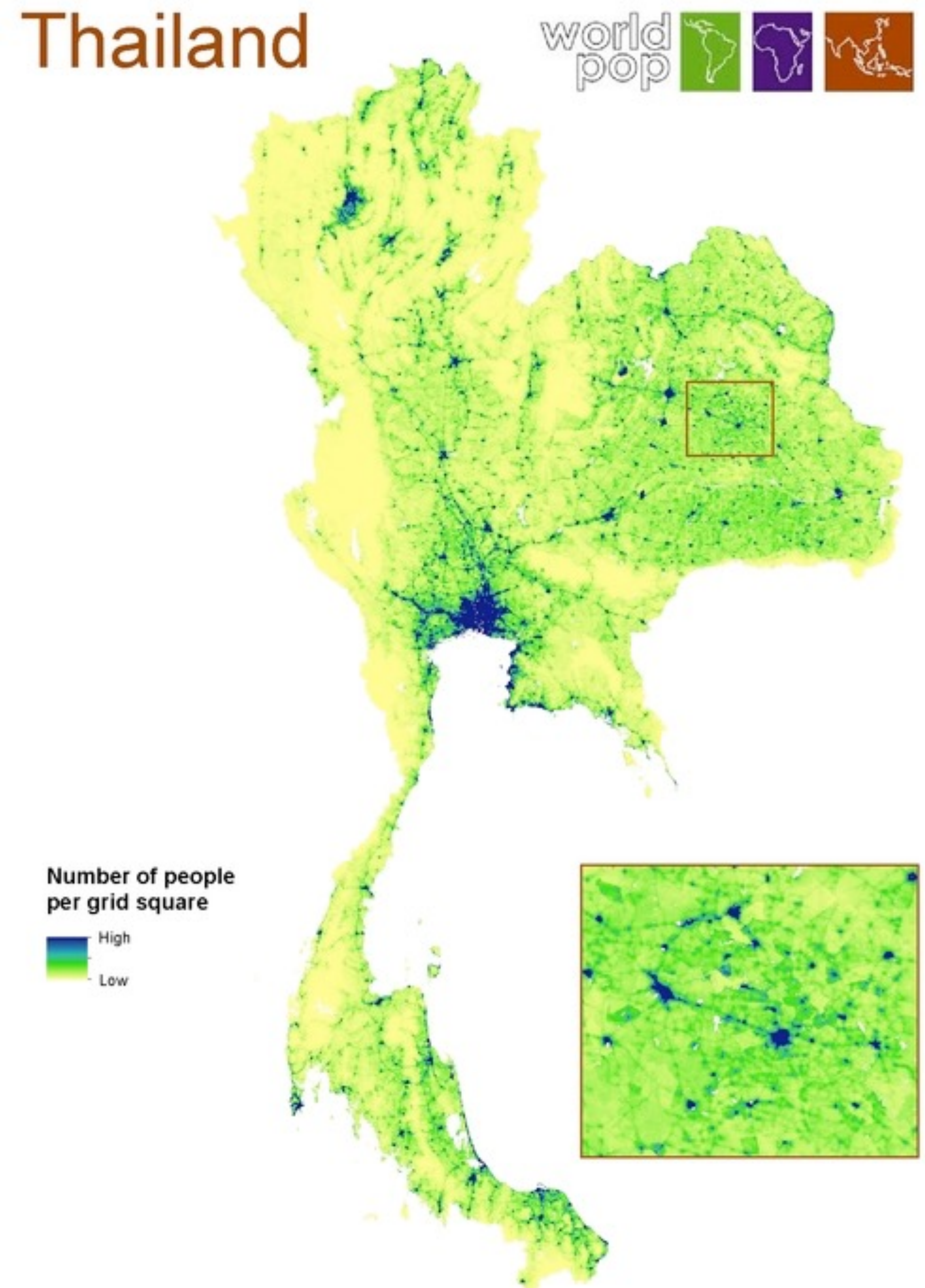
Aedes mosquitoes

- *Aedes aegypti* and *Aedes albopictus*
- Life cycles are dependent on weather
- Live longer at 20°-30°C; most live 40-60 days, but some live 100-120 (*Brady et al*)
- *A. albopictus* eggs require high humidity and temperature for first month, but may desiccate protecting them from harmful conditions
- *A. aegypti* eggs can survive most conditions in first two months, but are vulnerable to low humidity and high temperatures thereafter (Juliano et al)

Thailand geography



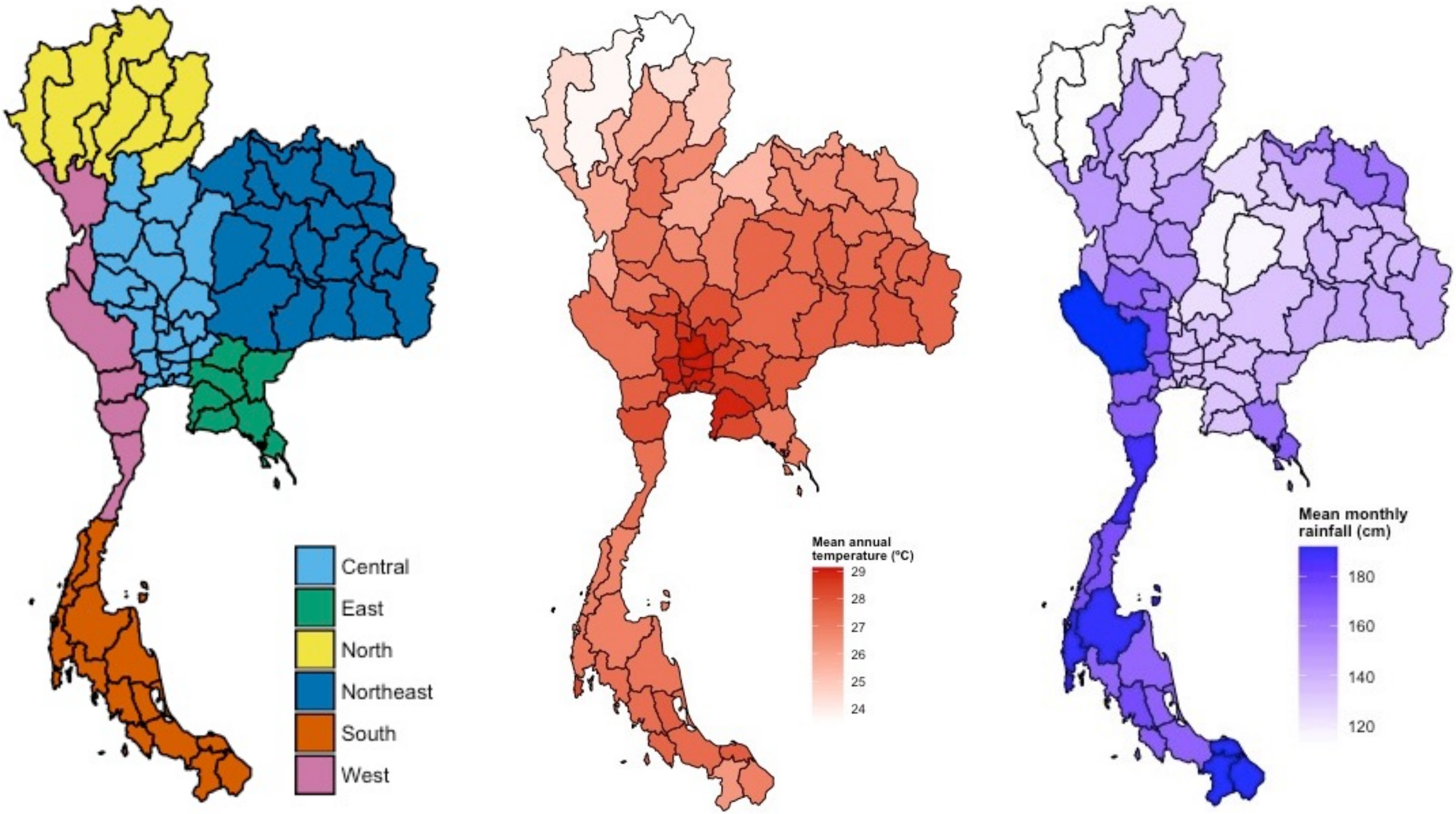




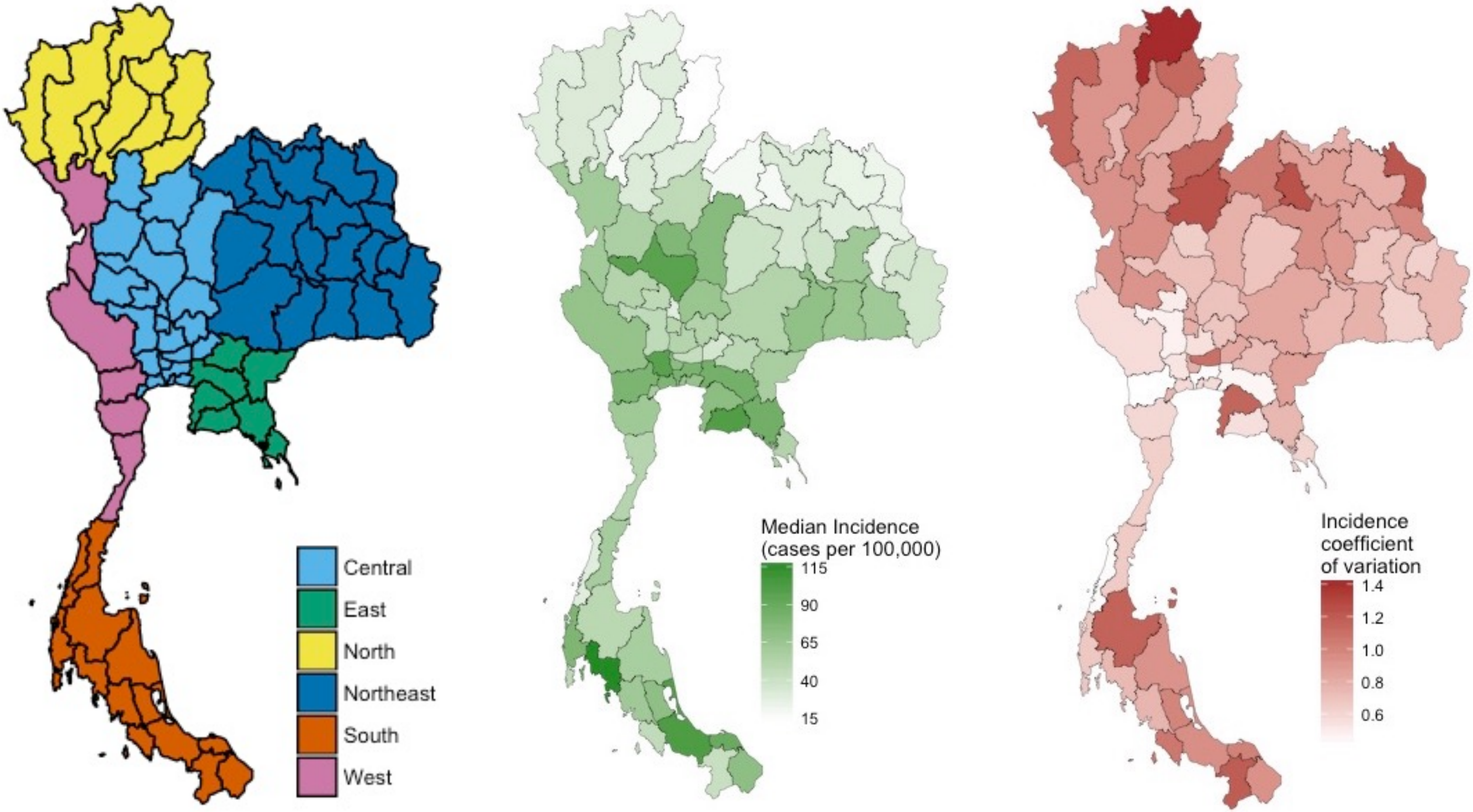
Source: Wikipedia

Source: WorldPop

Thailand weather



Thailand dengue

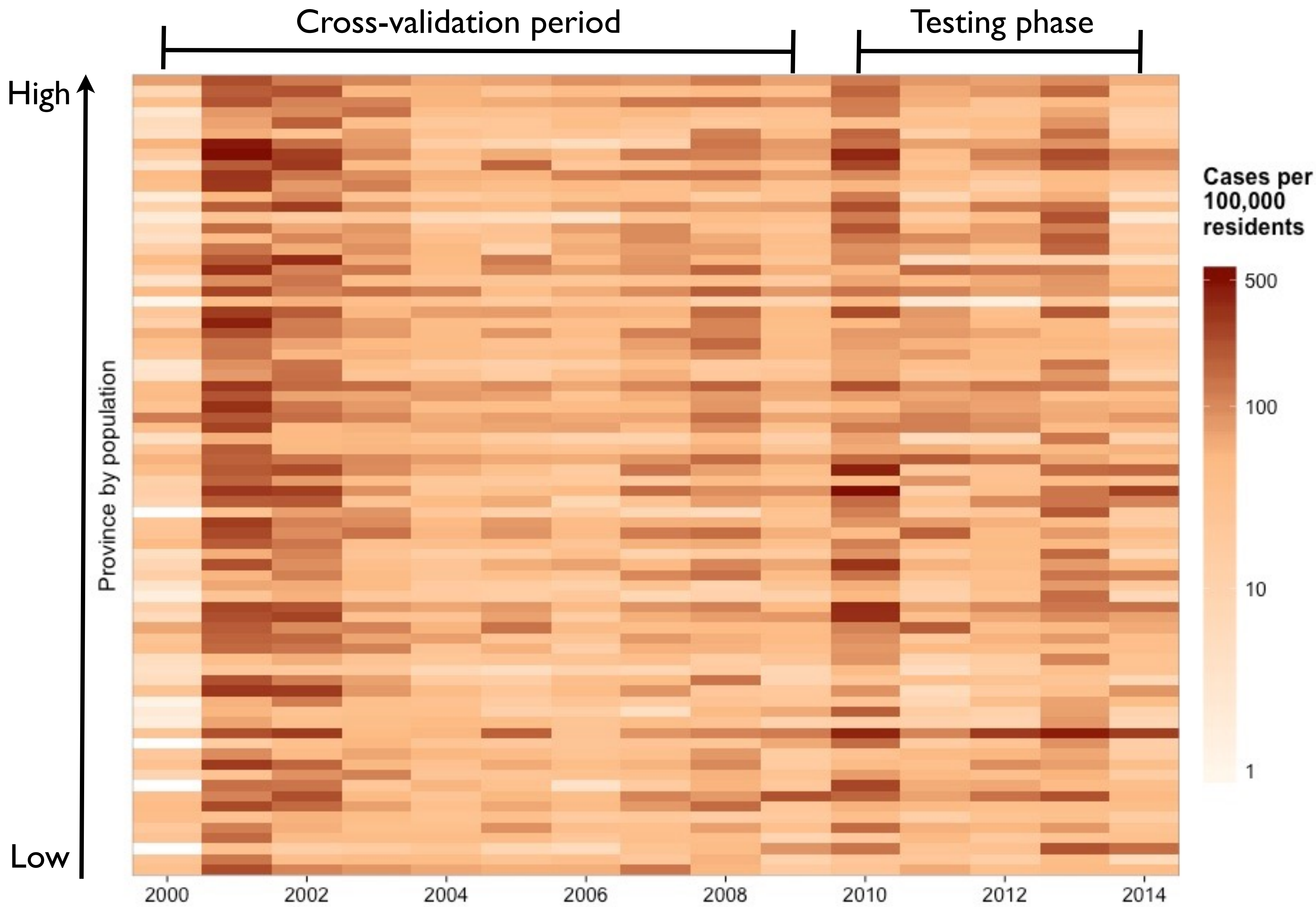


Objective and motivation

- Last April, Thai MOPH wanted to know magnitude and peak of coming season
- *Campbell and Scott* paper suggests that early season weather and incidence can predict Thailand dengue season timing and magnitude
- Objective: predict annual incidence based on early season information
- No vaccines or treatments are available and the effectiveness of vector control is questionable
- Accurate forecasts could play a critical role in assessing the efficacy of future interventions

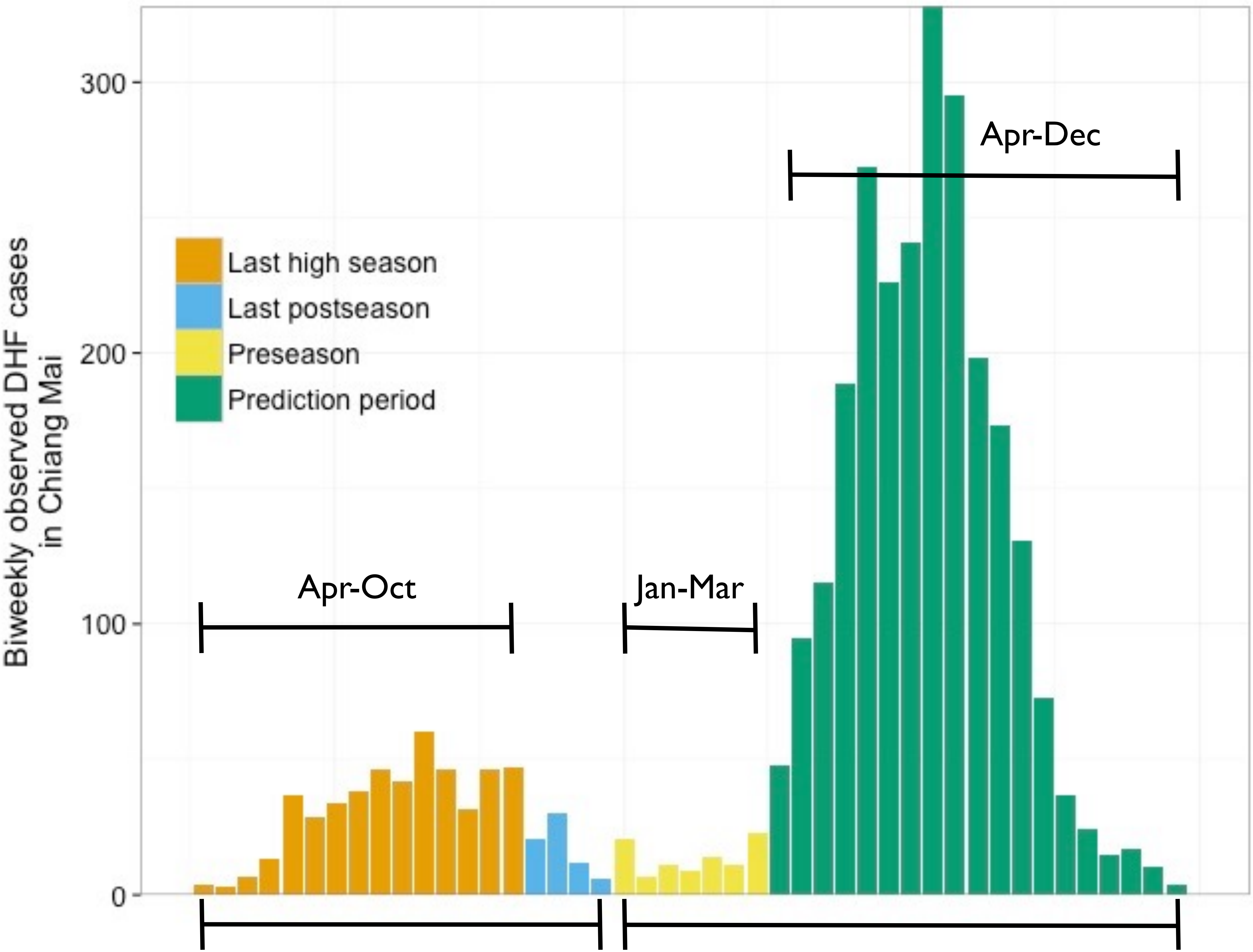
We want a method for building a predictive model that takes in a host of different variables and generates robust, validated, probabilistic forecasts

Dengue data



Cross-validation period
Testing phase
High
Low
Province by population
2000
2002
2004
2006
2008
2010
2012
2014
Cases per 100,000 residents
500
100
10
1

Seasonal timing variables



Weather covariates

- NOAA/NASA Grid data:
  - Monthly averages for temperature and rainfall for all 76 provinces
- NCDC station data:
  - Cross-sectional readings for temperature and humidity 4 times/day for 65 provinces
- NOAA station data:
  - Daily minimum and maximum temperatures and total precipitation for 35 provinces
- All data aggregated into min/max/mean temperature and humidity or max/total rainfall for Jan/Feb/Mar/low season

Prediction model

$$Y_{i,t} \sim NB(\lambda_{i,t}, r_{i,t})$$

$$log(E(Y_{i,t})) = \alpha_i + \sum_{j=1}^{J} f_j(X_{j,i,t})$$

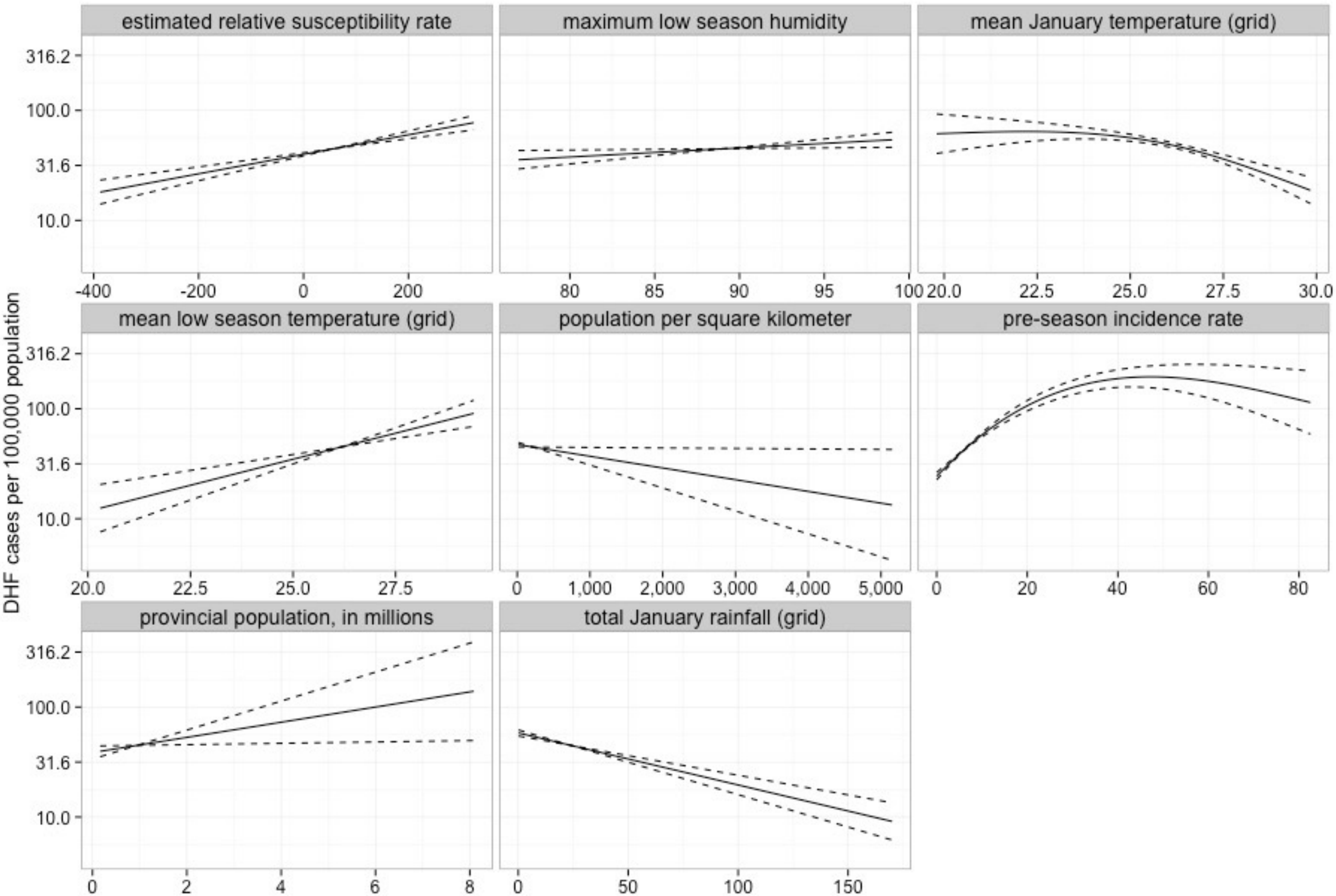
$$\alpha_i \sim Normal(0, \sigma^2)$$

- $Y_{i,t}$ DHF cases per 100,000 population in province *i* in year *t*
- α_i is the random effect for province i
- $f_j()$ is a cubic spline for covariate X_j if *J*, the number of covariates in the model, is greater than 0

Variable selection algorithm

1. Start with a "null" model (with only province random effects)
2. Build a new model for each covariate (35 new models), in the following manner:
    2.1. If the covariate is not currently in the model, add it to the model
    2.2. If the covariate is currently in the model, remove it from the model
3. Conduct leave-one-season-out cross validation for each new model on the years 2000-2009 and find the cross-validated mean absolute error (CV MAE) of each new model
4. If the lowest CV MAE for a new model is less than the best model from the last iteration, keep that new model and repeat steps 2-3. Otherwise, stop building.

Fitted model covariates



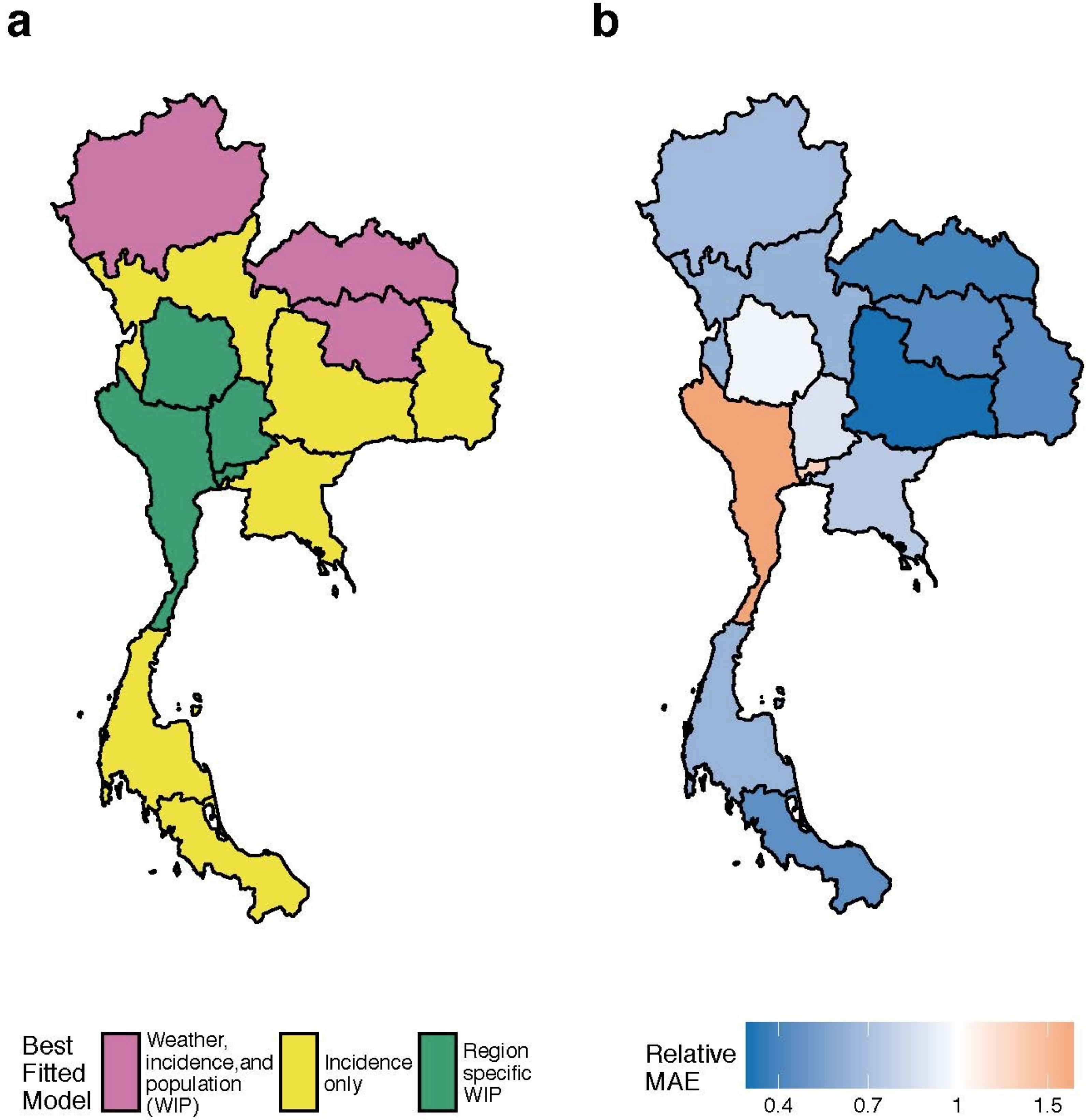
Testing phase

- Three models used in testing phase making predictions as if in real time:
    - Fitted model
    - Simple model: only covariate is pre-season rate. Least covariates of any model w/i one SD of fitted model
    - Baseline model: median incidence for each province over the past ten years

Overall Results

Model	# of Covariates	CV MAE	Test MAE
Fitted	8	0.509	0.599
Simple	1	0.568	0.571
Baseline	0	0.6	0.7

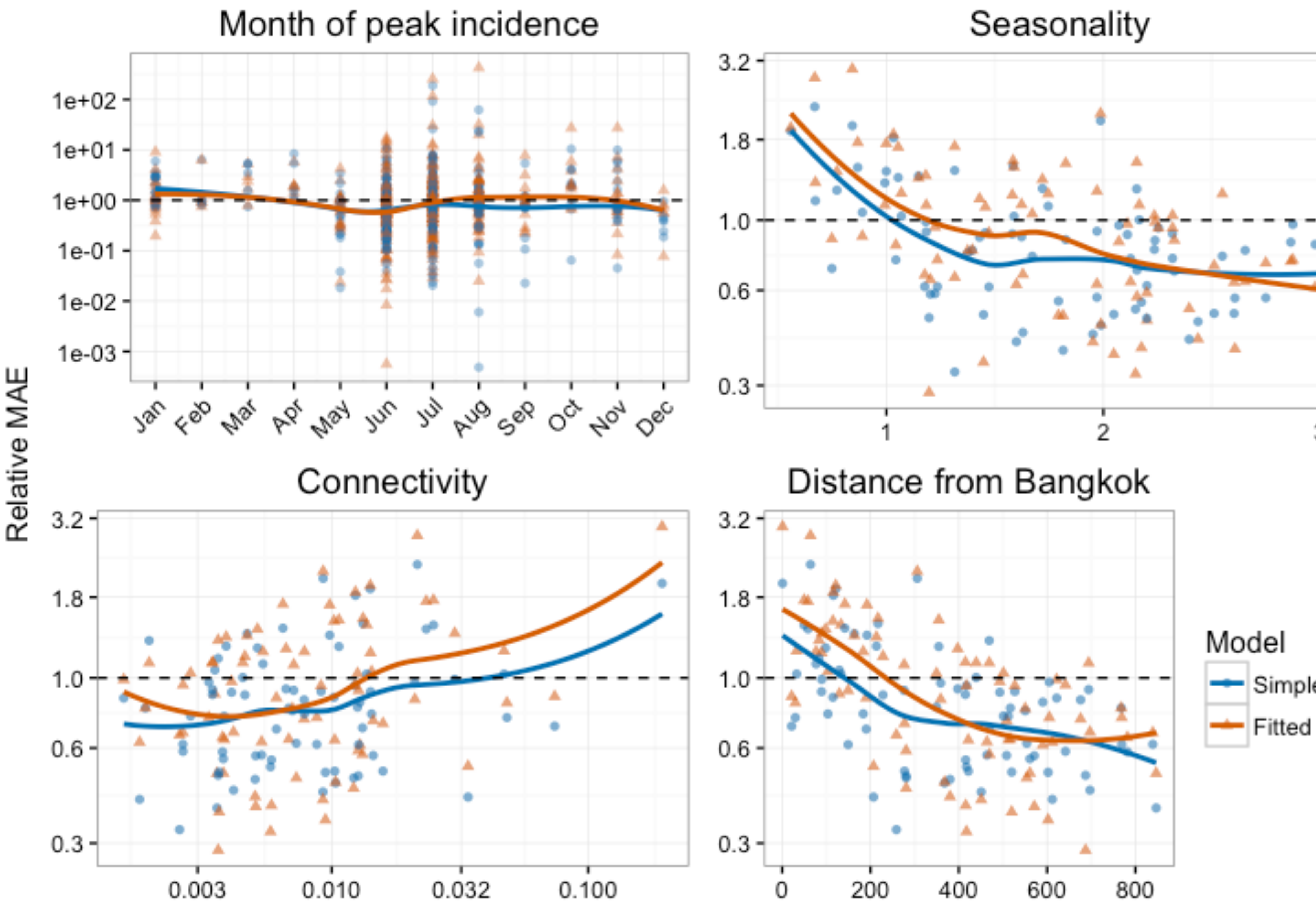
Results vary spatially



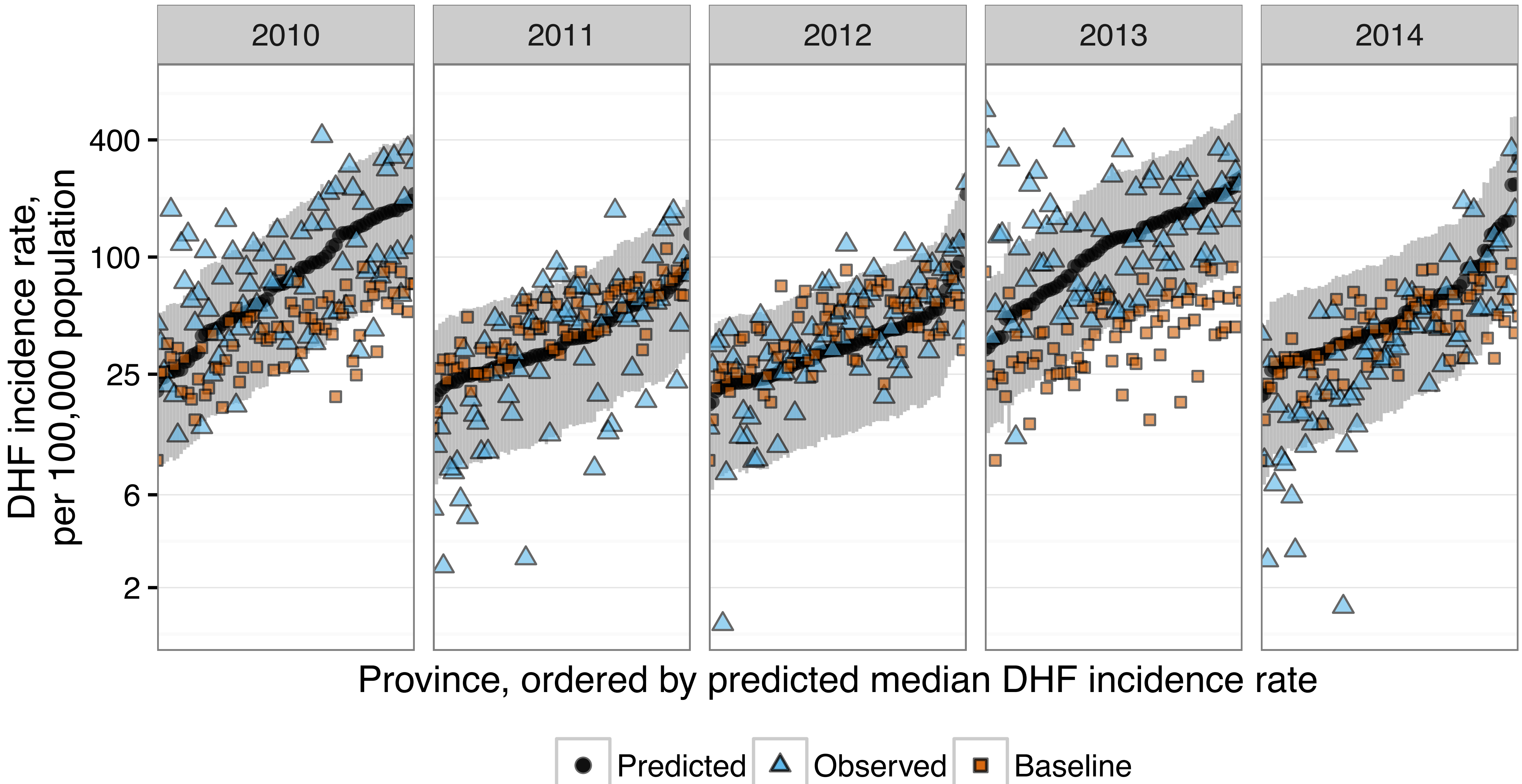
Results by year

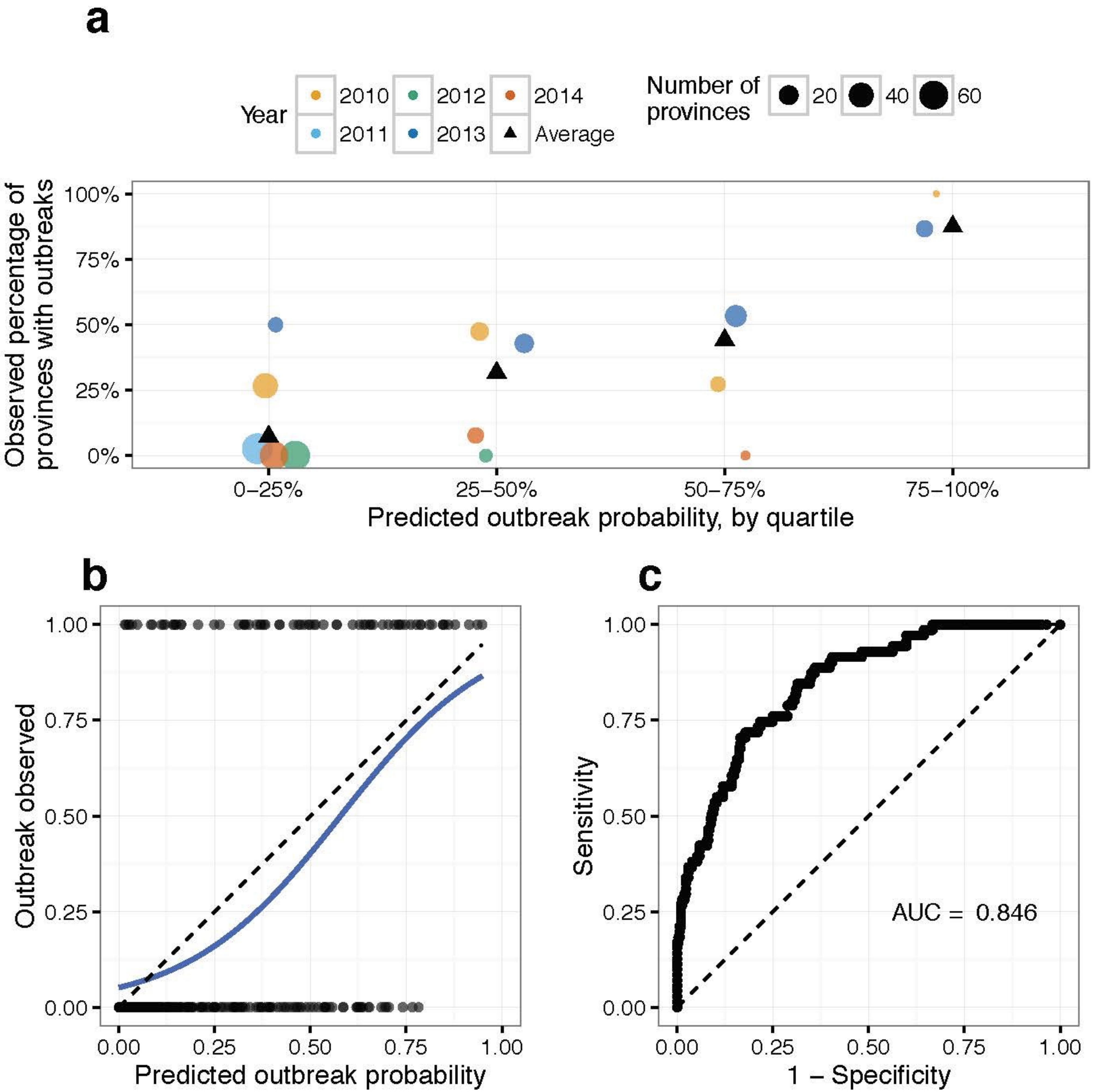
Year	Mean Incidence	Fitted MAE	Simple MAE	Baseline MAE
2010	104.7	0.56	0.587	0.89
2011	46	0.562	0.602	0.618
2012	47.5	0.45	0.418	0.483
2013	76.6	0.624	0.57	0.732
2014	38.8	0.797	0.678	0.777

MAE analysis



Probabilistic Forecasts





a
Year
2010
2011
2012
2013
2014
Average
Number of provinces
20
40
60
Observed percentage of provinces with outbreaks
100%
75%
50%
25%
0%
0–25%
25–50%
50–75%
75–100%
Predicted outbreak probability, by quartile
b
Outbreak observed
1.00
0.75
0.50
0.25
0.00
Predicted outbreak probability
c
Sensitivity
1 – Specificity
AUC = 0.846