Supporting Information

Lauer et al. 10.1073/pnas.1714457115

SI Text

As displayed in Table S4, the smallest MAE by any model for any region was for Bangkok (MOPH region 0) using the incidence-only model (MAE = 0.286). However, because the baseline MAE for Bangkok was only slightly higher (MAE = 0.289), the incidence-only model rMAE was the second largest of any region (rMAE = 0.99). Thus, although the incidence-only model accurately forecasted DHF incidence in Bangkok, it did not add much value over a 10-y median due, in part, to there being no outbreaks in Bangkok during the testing phase.

Conversely, the incidence-only model had about twice as much error in MOPH region 12 (MAE = 0.59) as in Bangkok. However, the baseline model had nearly three times as much error as in Bangkok (MAE = 0.86), and therefore, the incidence-only rMAE for MOPH region 12 was the lowest of any model for any region (rMAE = 0.69). Thus, despite greater absolute error from the incidence-only model forecasts, there was more added benefit for that region over the baseline forecasts than for Bangkok. These examples show how MAE and rMAE can be used in tandem to give a more complete evaluation of model performance.

The receiver operating characteristic (ROC) curves for the incidence-only and WIP model outbreak forecasts for the testing phase are both significantly above the line of no discrimination but are not significantly different from each other (Fig. 5). The incidence-only model area under the receiver operating characteristic curve (AUC; estimate: 84.2%, 95% CI: 78.5–89.9%) was slightly larger than the WIP model AUC (82.9%, 95% CI: 76.3–89.6%). The sensitivity of the WIP model is marginally larger than that of the incidence-only model when specificity is large, suggesting that the WIP model showed very slightly better performance than the incidence-only model at larger outbreak thresholds.

The predictive distributions samples used to make the outbreak forecasts could have been obtained by estimating parameters in a Bayesian framework, including drawing posterior samples of the dispersion parameter, which may have changed the predictive performance of the models. However, due to the coverage rates observed by our model (80% of forecasts covered by the 80% prediction interval), we did not believe that it would be worth the additional computational complexity to use these methods.

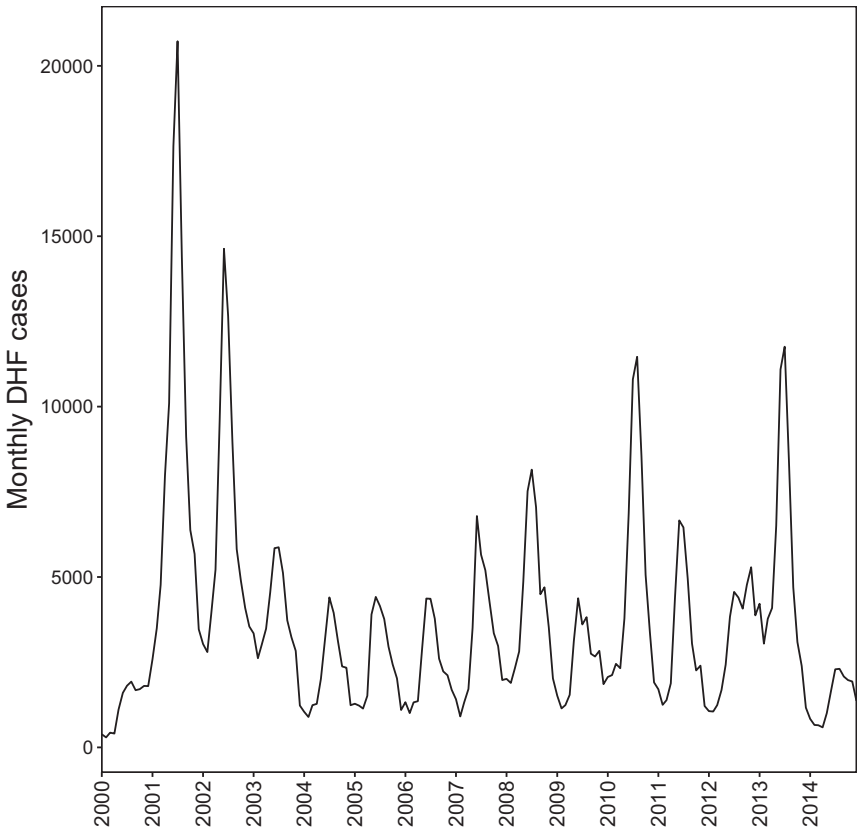

Fig. S1. Aggregated time series of DHF cases from 2000 to 2014.

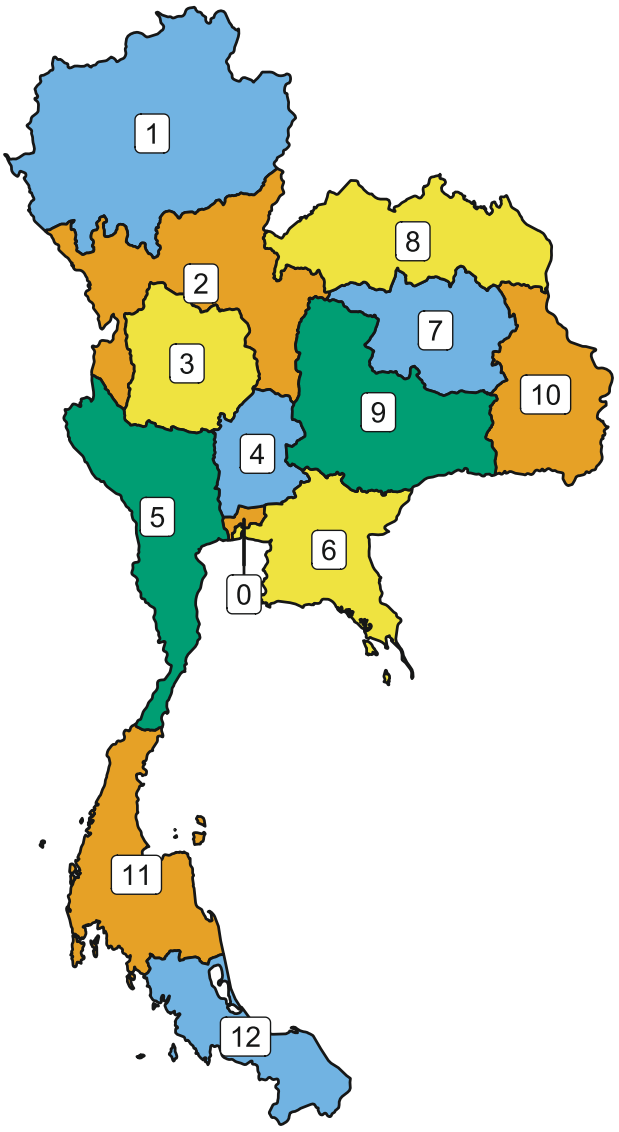

Fig. S2. Map of the Thailand MOPH administrative regions. These 13 MOPH regions are geographically clustered sets of four to eight provinces [with the exception of Bangkok (region 0), which is its own region] co-operatively managed by a regional health office.

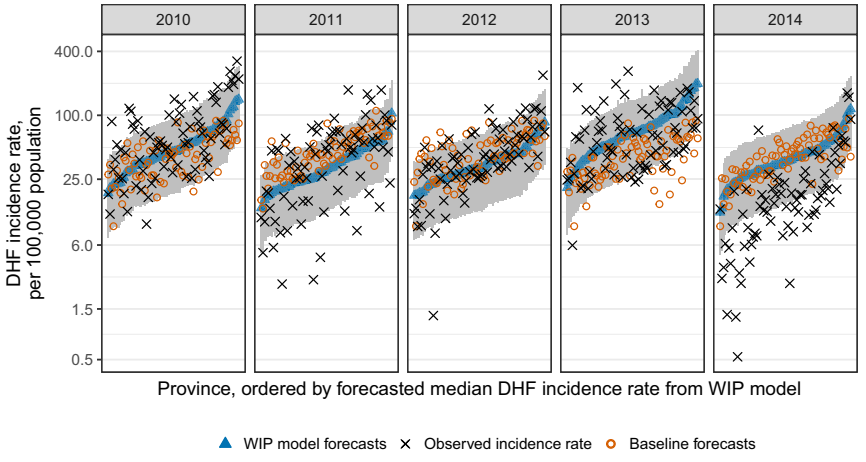

Fig. S3. WIP model forecasts for each year of the testing phase compared with the baseline forecasts and the observed values. Forecasts for the annual DHF incidence rate per 100,000 population from the WIP model (blue triangles with gray 80% prediction intervals), baseline forecasts (red circles), and observed values (black x) for each province and year in the testing phase.

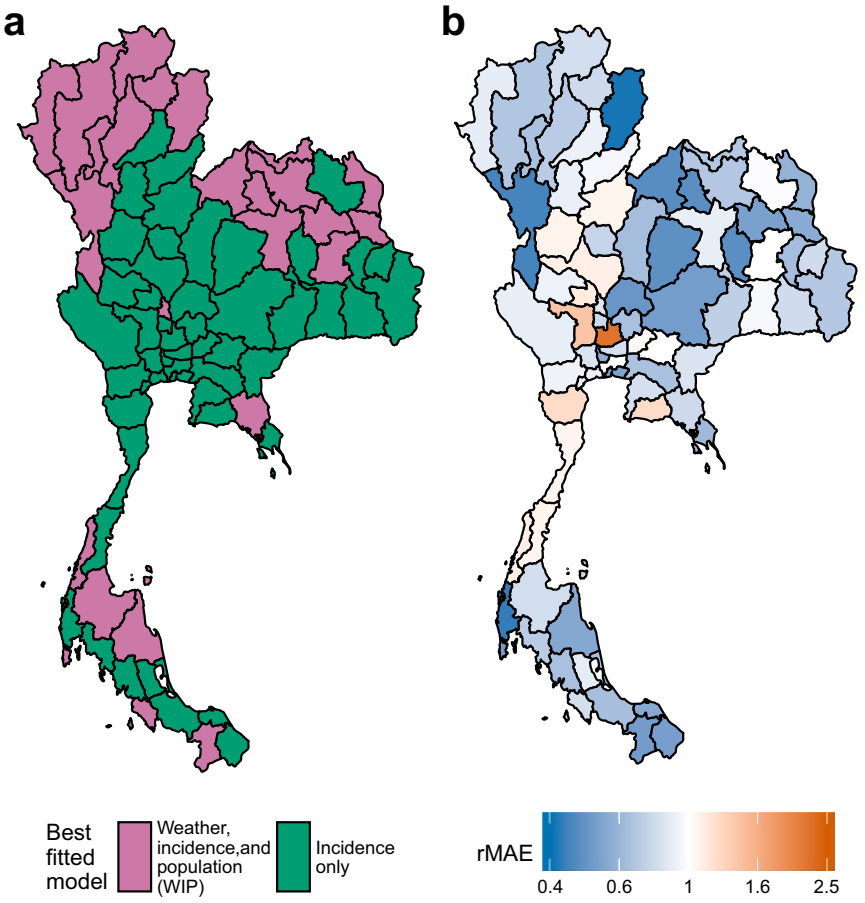

Fig. S4. Geographic variation in model and performance by province. (*A*) The best fitted model in the testing phase for each Thai province, which shows spatial patterns of performance. (*B*) The rMAE of the forecasts for each province from the models in *A* over the baseline forecasts. Provinces with less error than the baseline are blue, provinces with more error than the baseline are red, and provinces equal to the baseline are white.

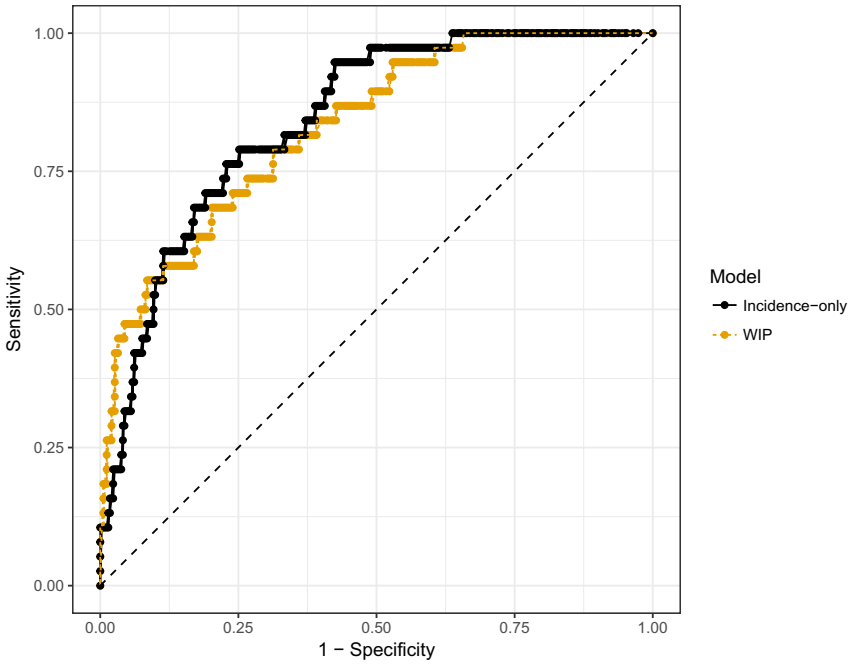

Fig. S5. Comparison of ROC curves by model. The ROC curve based on the incidence-only model and WIP models' sensitivity and specificity on outbreak forecasts during the testing phase. Both curves are comfortably above the line of no discrimination (dashed), indicating that their outbreak forecasts are better than random. The AUC for the WIP model (82.9%) is a bit lower than that of the incidence-only model (84.2%).

Table S1. Covariates considered for inclusion before model selection

Covariate type and covariate name	Incidence only	WIP
Incidence		
Estimated relative susceptibility rate		
Last high-season incidence rate		
Last postseason incidence rate		
Preseason incidence rate	✓	✓
Demographics		
Population per square kilometer		
Provincial population		✓
Humidity		
Maximum low-season humidity		
Minimum low-season humidity		
Mean January humidity		
Mean February humidity		
Mean March humidity		
Rainfall		
Maximum low-season rainfall (NOAA)		
Total low-season rainfall (ESRL)		
Total low-season rainfall (NOAA)		
Maximum January rainfall (NOAA)		
Total January rainfall (ESRL)		✓
Total January rainfall (NOAA)		
Maximum February rainfall (NOAA)		
Total February rainfall (ESRL)		
Total February rainfall (NOAA)		
Maximum March rainfall (NOAA)		
Total March rainfall (ESRL)		
Total March rainfall (NOAA)		
Temperature		
Maximum low-season temperature (NCDC)		
Mean low-season temperature (ESRL)		✓
Minimum low-season temperature (NCDC)		
Mean January temperature (ESRL)		✓
Mean January temperature (NCDC)		
Mean January temperature (NOAA)		
Mean February temperature (ESRL)		
Mean February temperature (NCDC)		
Mean February temperature (NOAA)		
Mean March temperature (ESRL)		
Mean March temperature (NCDC)		
Mean March temperature (NOAA)		

Incidence only indicates the covariates that were included in the incidence-only model. WIP indicates the covariates that were included in the WIP model.

Table S2. Results for each model across all regions and years in the testing phase

Model	MAE	rMAE	Forecasts better than baseline, %	80% PI coverage	AIC
WIP	0.64	0.87	56.3	69.7	**9,909**
Incidence only	**0.59**	**0.81**	**64.7**	**80.0**	10,055
Baseline	0.73	1.00			

Numbers in bold highlight the model that performed best for each metric. AIC, Akaike information criterion; PI, predictive interval.

Table S3. Annual results for each model across all regions in the testing phase

Model	Year	Mean provincial incidence	Outbreaks	MAE	rMAE	Forecasts better than baseline, %	80% PI coverage
WIP	2010	78	12	0.53	0.73	60.5	76.3
Incidence only				**0.53**	**0.73**	**68.4**	**80.3**
Baseline				0.72	1.00		
WIP	2011	47	2	0.65	1.05	47.4	61.8
Incidence only				**0.59**	**0.95**	**55.3**	**78.9**
Baseline				0.61	1.00		
WIP	2012	48	1	0.43	0.90	60.5	**86.8**
Incidence only				**0.43**	**0.90**	**61.8**	89.5
Baseline				0.48	1.00		
WIP	2013	74	23	0.57	0.79	51.3	**75.0**
Incidence only				**0.56**	**0.77**	**55.3**	85.5
Baseline				0.73	1.00		
WIP	2014	24	0	1.00	0.89	61.8	48.7
Incidence only				**0.85**	**0.75**	**82.9**	**65.8**
Baseline				1.13	1.00		

Numbers in bold highlight the model that performed best for each metric in each year. PI, predictive interval.

Table S4. Regional results for each model across all years in the testing phase

Model	MOPH region	Provinces	Mean provincial incidence	Outbreaks	MAE	rMAE	Forecasts better than baseline	80% PI coverage
WIP	12	7	83	3	0.62	0.72	68.6	65.7
Incidence only					**0.59**	**0.69**	**77.1**	**74.3**
Baseline					0.86	1.00		
WIP	9	4	74	3	0.58	0.84	70.0	70.0
Incidence only					**0.48**	**0.70**	**75.0**	**85.0**
Baseline					0.69	1.00		
WIP	8	6	37	5	**0.78**	**0.71**	70.0	66.7
Incidence only					0.83	0.75	70.0	66.7
Baseline					1.10	1.00		
WIP	10	5	43	4	0.56	0.82	60.0	**76.0**
Incidence only					**0.49**	**0.71**	**72.0**	88.0
Baseline					0.69	1.00		
WIP	1	8	45	12	**0.82**	**0.74**	77.5	50.0
Incidence only					0.94	0.84	77.5	**55.0**
Baseline					1.11	1.00		
WIP	7	4	51	4	0.56	0.76	60.0	**75.0**
Incidence only					**0.55**	**0.75**	**75.0**	70.0
Baseline					0.74	1.00		
WIP	11	7	72	3	0.71	0.79	62.9	62.9
Incidence only					**0.69**	**0.77**	**65.7**	**77.1**
Baseline					0.90	1.00		
WIP	6	8	65	1	0.51	1.01	42.5	**75.0**
Incidence only					**0.39**	**0.79**	**55.0**	95.0
Baseline					0.50	1.00		
WIP	2	5	52	2	0.62	0.93	48.0	72.0
Incidence only					**0.56**	**0.84**	**68.0**	**76.0**
Baseline					0.66	1.00		
WIP	4	8	33	0	0.60	1.04	42.5	**80.0**
Incidence only					**0.52**	**0.88**	**57.5**	92.5
Baseline					0.58	1.00		
WIP	3	5	51	0	0.68	1.17	52.0	64.0
Incidence only					**0.57**	**0.98**	**56.0**	**80.0**
Baseline					0.58	1.00		
WIP	0	1	63	0	0.58	2.00	20.0	100.0
Incidence only					**0.29**	**0.99**	**40.0**	100.0
Baseline					0.29	1.00		
WIP	5	8	47	1	0.56	1.24	37.5	**77.5**
Incidence only					0.47	1.03	**45.0**	92.5
Baseline					**0.46**	**1.00**		

Numbers in bold highlight the model that performed best for each metric in each region. The regions are sorted by best model performance using rMAE from lowest to highest. PI, predictive interval.

Other Supporting Information Files

Table S5 (XLSX)
Table S6 (XLSX)
Table S7 (XLSX)
Table S8 (XLSX)
Table S9 (XLSX)