Predicting Dengue Outbreaks in Cambodia

Anthony Cousien, Julia Ledien, Kimsan Souv, Rithea Leang, Rekol Huy, Didier Fontenille, Sowath Ly, Veasna Duong, Philippe Dussart, Patrice Piola,[1] Simon Cauchemez,[1] Arnaud Tarantola[1]

In Cambodia, dengue outbreaks occur each rainy season (May–October) but vary in magnitude. Using national surveillance data, we designed a tool that can predict 90% of the variance in peak magnitude by April, when typically <10% of dengue cases have been reported. This prediction may help hospitals anticipate excess patients.

Dengue is endemic to Cambodia; outbreaks are seasonal, occurring during the rainy season (May–October). However, the magnitude of outbreaks varies from year to year. When the epidemic is particularly large, the influx of patients with severe dengue in pediatric hospitals may saturate the healthcare system and negatively affect quality of care. However, adequate supportive care is crucial for patients with severe dengue and can decrease the fatality rate to <1% (*1*). Early prediction of the size of nascent dengue epidemics may improve healthcare planning and optimize allocation of healthcare resources. We used surveillance data to build a simple early warning tool based on the reported number of cases early in the season. Compared with other approaches used to predict dengue epidemics (*2–6*), this one is characterized by its simplicity because it relies only on the number of cases reported early in the season to predict the magnitude of the epidemic.

The Study

We used the monthly number of probable dengue cases reported by the National Dengue Surveillance System (NDSS) in Cambodia during 2004–2016. The NDSS includes passive surveillance of probable dengue pediatric inpatients reported by public hospitals to the Communicable Diseases Center of the Ministry of Health and a sentinel, pediatric hospital–based active surveillance system managed by the National Dengue Control Program of the National Center for Parasitology, Entomology and Malaria Control, Ministry of Health. A probable dengue case was defined as an acute febrile illness with ≥2 of the following: headache, retro-orbital pain, myalgia, arthralgia, rash, hemorrhage, and leukopenia, combined with either 1) a posteriori virologic confirmation, serologic confirmation, or both or 2) presence of ≥1 laboratory-confirmed case at the same location and time (*7*).

From January 2004 through December 2016, NDSS reported 215,574 probable dengue cases (Figure 1). During this period, we observed 2 outbreaks of particularly high magnitude, in 2007 (dengue virus serotype 3) and 2012 (dengue virus serotype 1). The magnitude of these outbreaks reached ≈10,000 cases versus the usual number of <5,000 cases. Incidence was always lowest during the dry season (i.e., November–April); the nadirs usually occurred in February and the peaks in July (8 times), August (4 times), and June (1 time, in 2007). On average, only 6.1% of the cases reported during a season (i.e., from February through January of the following year) are observed before the end of April (range 2.7%–9.0% of cases). We wanted to ascertain whether the small number of cases reported at the season's onset (i.e., up to April) could be used as an early warning tool for predicting the magnitude of that season's epidemic.

We observed a strong linear correlation between the magnitude of the peak and the number of cases reported at the beginning of the season, in February (Pearson correlation coefficient r = 0.78), March (r = 0.88), April (r = 0.95), February–March (r = 0.86), March–April (r = 0.95), and February–April (r = 0.94). Fitting a simple linear regression model to the data, we estimated that the number of cases reported explained the following parts of the variance in the peak magnitude for February (61%), March (78%), April (91%), February–March (73%), March–April (90%), and February–April (88%). The magnitude was therefore best predicted by the number of dengue cases reported in April. This simple model offered excellent accuracy for predicting the magnitude of the peak; mean absolute percentage error for 2007 was 2.5% and for 2012 was 1.9% (Figure 2, panel A). Predictions relying on data from March were also acceptably accurate; the error was larger, but the model was able to predict a larger than usual magnitude (Appendix Figure 1, https://wwwnc.cdc.gov/EID/article/25/12/18-1193-App1.pdf).

Author affiliations: Institut Pasteur du Cambodge, Phnom Penh, Cambodia (A. Cousien, J. Ledien, K. Souv, D. Fontenille, S. Ly, V. Duong, P. Dussart, P. Piola); Institut Pasteur, Paris, France (A. Cousien, S. Cauchemez); CNRS, Paris (A. Cousien, S. Cauchemez); National Center for Entomology, Parasitology and Malaria Control, Phnom Penh (R. Leang, R. Huy); Institut Pasteur, Noumea, New Caledonia (A. Tarantola)

DOI: https://doi.org/10.3201/eid2512.181193

[1]These senior authors contributed equally to this article.

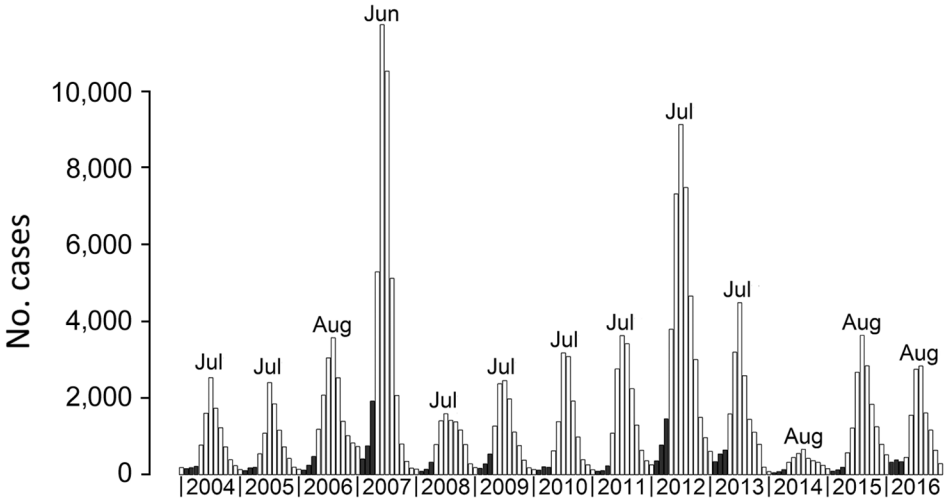


Figure 1. Monthly number of probable dengue cases reported to the National Dengue Surveillance System in Cambodia, 2004–2016. Dark gray bars represent the 3 months (February, March, and April) used as predictors for the magnitude of the following peak. For each year, the month corresponding to the peak of the epidemic is indicated.

To evaluate the performance of our model in a real-life situation, when the outcome of the ongoing epidemic remains unknown, we used a leave-one-out cross-validation procedure (*8–10*). We obtained the predicted value for season *s* by fitting our regression model to the 12 other seasons (i.e., excluding season *s* from the set of observations used to fit the parameters of the model [the training dataset]). The predictive power of our best fitting model remained very high; it was able to explain 90% of the variance of the magnitude of epidemics (Figure 2, panel B).

Our dataset contains information for only 2 large epidemics (2007 and 2012). If we trained the model on these 2 large epidemics only, performance would remain very good (98% variance explained). In contrast, when both epidemics were excluded from the training dataset, their magnitude was underestimated by 35% (2007) and 32% (2012). As expected, to be properly calibrated, the model needs to be trained on a mix of small and large epidemics; if 1 category is excluded from the training dataset, performance may be substantially degraded.

Of note, this loss of accuracy is mostly an issue for large epidemics. Given the small number of such epidemics in our dataset, robustly demonstrating predictability from this dataset alone remains difficult. We therefore explored whether similar patterns could be observed in 4 other countries in South Asia: Thailand, Vietnam, Laos, and the Philippines (*11–14*). To be comparable with our analysis for Cambodia, we used the month at which >5% of cases have been observed on average (Appendix Tables 4, 5). The results were promising for Vietnam (variance explained in the leave-one-out procedure was 64.3%), the Philippines (45.8%), and Thailand (33.4%) but bad for Laos (–53.5%) (Appendix Figures 3–6). This variability could be explained by several factors: national surveillance system characteristics, demographics, land cover, healthcare systems, or climate; all of these factors can affect dengue epidemiology and reporting. This analysis confirms the observation made for Cambodia that the number of dengue cases reported early in the epidemic year may provide early insight into the probable scale of the forthcoming epidemic.

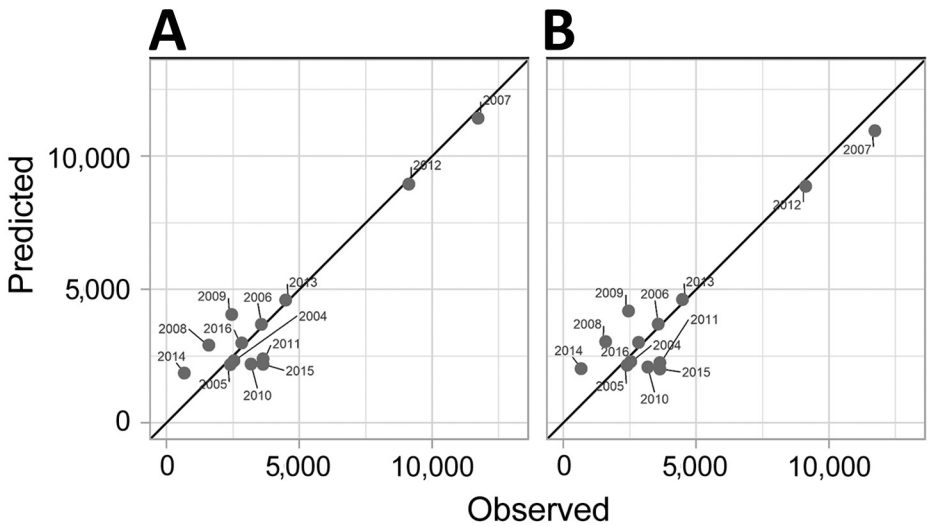


Figure 2. Dengue cases in Cambodia, 2004–2016. A) Observed versus predicted magnitude of the peak for each dengue season. We used a simple linear regression model, $M = \alpha + \beta N$, in which M indicates the magnitude of the peak and N the number of reported dengue-like cases in April. The black line represents the expected results with perfect prediction. B) Results for the leave-one-out cross-validation procedure.

Conclusions

The correlation between the number of patients hospitalized with probable dengue during the interepidemic period (i.e., the dry season) and the magnitude of the next outbreak peak during the rainy season was strong, even from February, which corresponds to the nadir of the incidence curves. Using dengue surveillance data for the end of the dry season (April), we were able to predict the magnitude of the peak for the next dengue outbreak, when typically <10% of cases have been observed and the peak is 2–3 months away. These results suggest that the intensity of rainfalls during the rainy season is not a major determinant of the occurrence of major outbreaks in Cambodia and that the outbreaks could be explained by conditions already present during the early stages of the outbreak (i.e., the part of the population immune to the circulating strains or weather conditions during the dry season). Our analysis is limited by the small number of epidemic seasons that are available to train our model for Cambodia (in particular, the small number of large epidemics), but similar patterns were observed in some other countries in South Asia.

In a setting where resources are limited and where pediatric hospitals face several other health issues (diarrheal diseases, other infectious diseases), the amount of available beds, medical supplies, and medical staff are usually appropriate for an average dengue outbreak. This simple and easy tool can help hospitals to plan in accordance with the predicted magnitude of the seasonal outbreak.

This study was supported by the Agence Française de Développement (ECOMORE Project), the Investissement d'Avenir program, the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases program (grant ANR-10-LABX-62-IBEID), the Models of Infectious Disease Agent Study of the National Institute of General Medical Sciences, and the AXA Research Fund.

About the Author

Dr. Cousien is a postdoctoral researcher at the Institut Pasteur. His research interests include the modeling of arbovirus epidemics.

References

1. World Health Organization. Dengue and severe dengue [cited 2017 Sep 7]. http://www.who.int/mediacentre/factsheets/fs117
2. Lowe R, Gasparrini A, Van Meerbeeck CJ, Lippi CA, Mahon R, Trotman AR, et al. Nonlinear and delayed impacts of climate on dengue risk in Barbados: a modelling study. PLoS Med. 2018;15:e1002613. https://doi.org/10.1371/journal.pmed.1002613
3. Lowe R, Barcellos C, Coelho CAS, Bailey TC, Coelho GE, Graham R, et al. Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts. Lancet Infect Dis. 2014;14:619–26. https://doi.org/10.1016/S1473-3099(14)70781-9
4. Lauer SA, Sakrejda K, Ray EL, Keegan LT, Bi Q, Suangtho P, et al. Prospective forecasts of annual dengue hemorrhagic fever incidence in Thailand, 2010–2014. Proc Natl Acad Sci U S A. 2018;115:E2175–82. https://doi.org/10.1073/pnas.1714457115
5. Johansson MA, Reich NG, Hota A, Brownstein JS, Santillana M. Evaluating the performance of infectious disease forecasts: a comparison of climate-driven and seasonal dengue forecasts for Mexico. Sci Rep. 2016;6:33707. https://doi.org/10.1038/srep33707
6. Reich NG, Lauer SA, Sakrejda K, Iamsirithaworn S, Hinjoy S, Suangtho P, et al. Challenges in real-time prediction of infectious disease: a case study of dengue in Thailand. PLoS Negl Trop Dis. 2016;10:e0004761. https://doi.org/10.1371/journal.pntd.0004761
7. Cambodian Ministry of Health. Standard Operating Procedures of the National Dengue Sentinel Surveillance System, National Dengue Control Program. Phnom Penh (Cambodia): The Ministry; 2010.
8. Stone M. Cross-validatory choice and assessment of statistical predictions. J R Stat Soc B. 1974;36:111–33. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x
9. Allen DM. The relationship between variable selection and data augmentation and a method of prediction. Technometrics. 1974;16:125–7. https://doi.org/10.1080/00401706.1974.10489157
10. Geisser S. The predictive sample reuse method with applications. J Am Stat Assoc. 1975;70:320–8. https://doi.org/10.1080/01621459.1975.10479865
11. Van Panhuis W, Cross A, Burke D, Choisy M. Project Tycho: Philippines (dengue) dataset: 1955–2010: Project Tycho [cited 2019 Sep 13]. https://www.tycho.pitt.edu/dataset/PH.38362002
12. Van Panhuis W, Cross A, Burke D, Choisy M. Project Tycho: Thailand (dengue) dataset: 1958–2010 [cited 2019 Sep 13]. https://www.tycho.pitt.edu/dataset/TH.38362002
13. Van Panhuis W, Cross A, Burke D, Choisy M. Project Tycho: Lao People's Democratic Republic (dengue) dataset 1979–2010: Project Tycho [cited 2019 Sep 13]. https://www.tycho.pitt.edu/dataset/LA.38362002
14. Van Panhuis W, Cross A, Burke D, Choisy M. Project Tycho: Vietnam (dengue) dataset: 1960–2010 [cited 2019 Sep 13]. https://www.tycho.pitt.edu/dataset/VN.38362002

Address for correspondence: Patrice Piola, Institut Pasteur du Cambodge, Epidemiology and Public Health Unit, 5 Blvd Monivong, BP 983, Phnom Penh, Cambodia; email: ppiola@pasteur-kh.org

Article DOI: https://doi.org/10.3201/eid2512.181193

Predicting Dengue Outbreaks in Cambodia

Appendix

Appendix Table 1. Magnitudes, timing of the peak and number of probable dengue cases reported in February, March, April, February-March, March-April and February to April to the National Dengue Surveillance System in Cambodia for seasonal dengue outbreaks between 2004 and 2016.

Year	Magnitude	Timing of the peak	Reported cases in February	Reported cases in March	Reported cases in April	Reported cases in February-March	Reported cases in March-April	Reported cases in February-April
2004	2,530	July	162	191	220	353	411	573
2005	2,403	July	113	180	195	293	375	488
2006	3,566	August	122	254	475	376	729	851
2007	11,726	June	416	756	1,917	1,172	2,673	3,089
2008	1,594	July	89	145	330	234	475	564
2009	2,451	July	172	284	544	456	828	1,000
2010	3,176	July	123	209	198	332	407	530
2011	3,628	July	90	112	234	202	346	436
2012	9,125	July	362	773	1,455	1,135	2,228	2,590
2013	4,483	July	340	545	646	885	1,191	1,531
2014	667	August	54	85	134	139	219	273
2015	3,631	August	97	133	196	230	329	426
2016	2,835	August	333	389	346	722	735	1,068

Appendix Table 2. Predicted magnitudes using the linear model $M = \alpha + \beta N$, with M the magnitude of the peak and N the number of reported probable dengue cases in either February, March, April, February-March, March-April and February to April. The last six columns present the results for the "leave-one-out" procedure, i.e., when the current season of the learning sample is excluded.

	Predicted magnitudes using reported cases						Predicted magnitudes in leave-one-out procedure using reported cases					
Year	in February	in March	in April	in February-March	in March-April	in February-April	in February	in March	in April	in February-March	in March-April	in February-April
2004	3,446	2,599	2,322	2,886	2,353	2,478	3,527	2,607	2,297	2,922	2,333	2,473
2005	2,509	2,473	2,187	2,444	2,217	2,199	2,522	2,481	2,161	2,449	2,194	2,174
2006	2,681	3,321	3,691	3,056	3,558	3,391	2,581	3,299	3,701	3,007	3,557	3,376
2007	8,303	9,074	11,432	8,920	10,921	10,737	6,465	7,503	10,952	7,264	9,926	9,597
2008	2,050	2,072	2,912	2,010	2,596	2,449	2,119	2,136	3,040	2,069	2,702	2,544
2009	3,637	3,665	4,061	3,645	3,933	3,880	3,738	3,768	4,195	3,747	4,056	3,999
2010	2,700	2,805	2,203	2,732	2,338	2,337	2,647	2,768	2,086	2,685	2,241	2,239
2011	2,069	1,694	2,397	1,774	2,107	2,029	1,835	1,387	2,257	1,482	1,915	1,820
2012	7,271	9,269	8,951	8,647	9,236	9,099	6,702	9,362	8,870	8,402	9,295	9,086
2013	6,850	6,656	4,608	6,806	5,308	5,623	7,430	7,064	4,620	7,291	5,393	5,755
2014	1,381	1,384	1,860	1,310	1,626	1,493	1,533	1,515	2,024	1,435	1,772	1,627
2015	2,203	1,935	2,193	1,980	2,043	1,996	2,002	1,692	2,018	1,743	1,838	1,779
2016	6,716	4,868	2,998	5,605	3,580	4,103	7,602	5,059	3,013	5,942	3,644	4,209

Appendix Table 3. Estimated parameters (and 95% confidence intervals) for the linear model $M = \alpha + \beta N$, with M the magnitude of the peak and N the number of probable dengue cases reported to NDSS in either February, March or April, and the corresponding proportion of the variance of M explained.

	Estimates for the model $M = \alpha + \beta N$, with M the magnitude and N the number of cases					
Description	February	March	April	February-March	March-April	February-April
α	348.0 [-1,948.8; 2,644.8]	410.4 [-1,153.7; 1,974.5]	1140.56 [288.8; 1992.3]	285.9 [-1,501.3; 2,073.0]	796.4 [-156.7; 1,749.6]	597.3 [-466.7; 1,661.2]
β	19.1 [8.9; 29.3]	11.5 [7.4; 15.5]	5.39 [4.2; 6.5]	7.4 [4.4; 10.3]	3.8 [2.9; 4.6]	3.3 [2.5; 4.1]
Variance explained by the model*	61%	78%	91%	73%	90%	88%
Variance explained, leave-one-out procedure†	42%	69%	90%	63%	87%	85%

*Corresponding to the squared correlation coefficient r^2.
†Ccorresponding to the predictive squared correlation coefficient q_{CV}^2 obtained in the "leave-one-out" cross validation procedure.

Appendix Table 4. Month with >5% of the total number of cases observed for 5 countries in Southeast Asia. Monthly data for the number of case reported in each dengue national surveillance system (excluding Cambodia) are publicly available in (*1–4*).

Country	Month with >5% of the total number of cases observed (%)
Cambodia	April (6.1%)
Thailand	February (7.3%)
Vietnam	March (6.0%)
Laos	March (5.6%)
Philippines	May (8.9%)

Appendix Table 5. Estimated parameters (and 95% confidence intervals) for the linear model $M = \alpha + \beta N$, with M the magnitude of the peak and N the number of probable dengue cases reported in national dengue surveillance systems (*1–4*), when 5% of the total number of cases have been observed, and the corresponding proportion of the variance of M explained.

	Estimates for the model $M = \alpha + \beta N$, with M the magnitude and N the number of cases				
Description	Cambodia	Thailand	Vietnam	Laos	Philippines
α	1140.6 [288.8; 1992.3]	5590.1 [1963.2; 9217.1]	–4875.4 [-10889.1; 1138.4]	3228.3 [-250; 6706.6]	–4327.1 [-10142.8; 1488.5]
β	5.4 [4.2; 6.5]	2 [0.8; 3.2]	8.1 [5.6; 10.7]	–3.1 [-28.5; 22.3]	10.2 [6.2; 14.2]
Variance explained by the model*	90.6	42.8	76.2	0.6	74.2
Variance explained, leave-one-out procedure†	89.8	33.4	64.3	–53.5	45.8

*Corresponding to the squared correlation coefficient r^2.
†Corresponding to the predictive squared correlation coefficient q_{CV}^2.

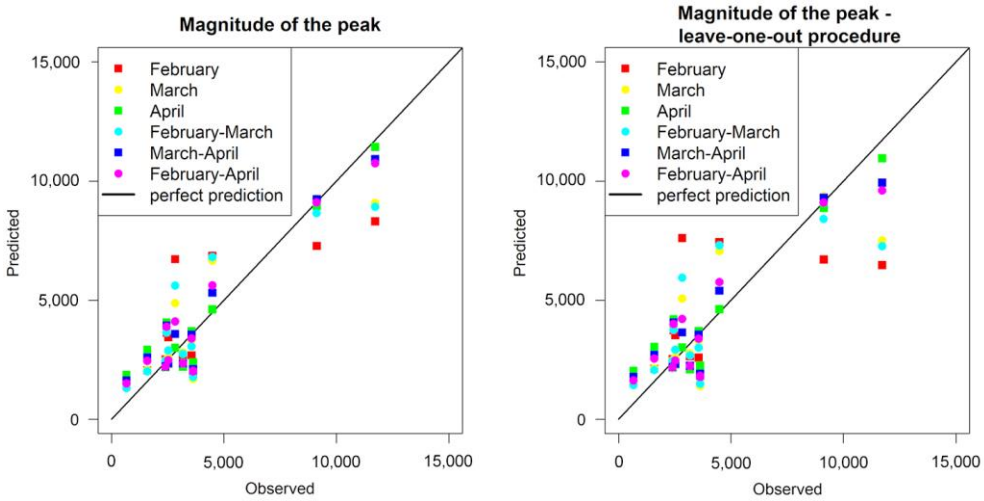
References

1. Van Panhuis W, Cross A, Burke D, Choisy M. Counts of Dengue reported in Philippines: 1955–2010: Project Tycho dataset. University of Pittsburgh; 2018 [cited 2019 Sep 13]. https://www.tycho.pitt.edu/dataset/PH.38362002

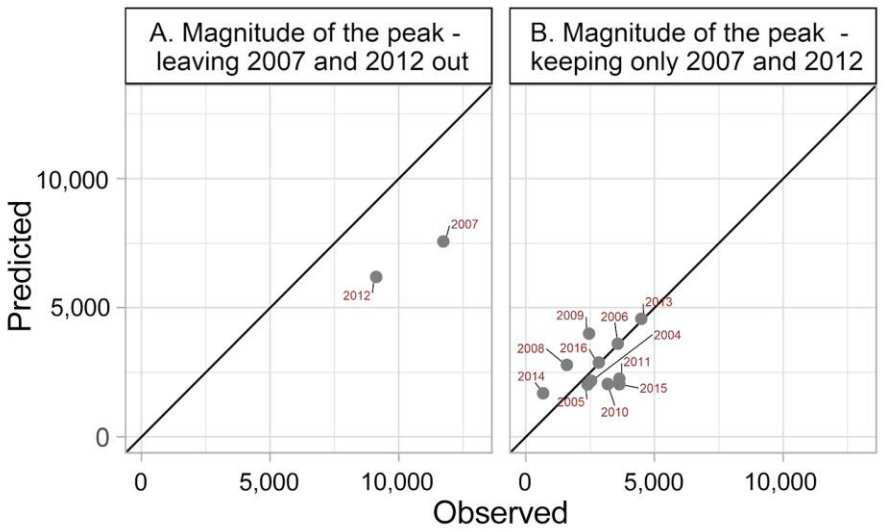
2. Van Panhuis W, Cross A, Burke D, Choisy M. Counts of Dengue reported in Thailand: 1958–2010: Project Tycho dataset. University of Pittsburgh; 2018 [cited 2019 Sep 13]. https://www.tycho.pitt.edu/dataset/TH.38362002

3. Van Panhuis W, Cross A, Burke D, Choisy M. Counts of Dengue reported in Lao People's Democratic Republic 1979–2010: Project Tycho dataset [Internet]. University of Pittsburgh; 2018 [cited 2019 Sep 13]. https://www.tycho.pitt.edu/dataset/LA.38362002

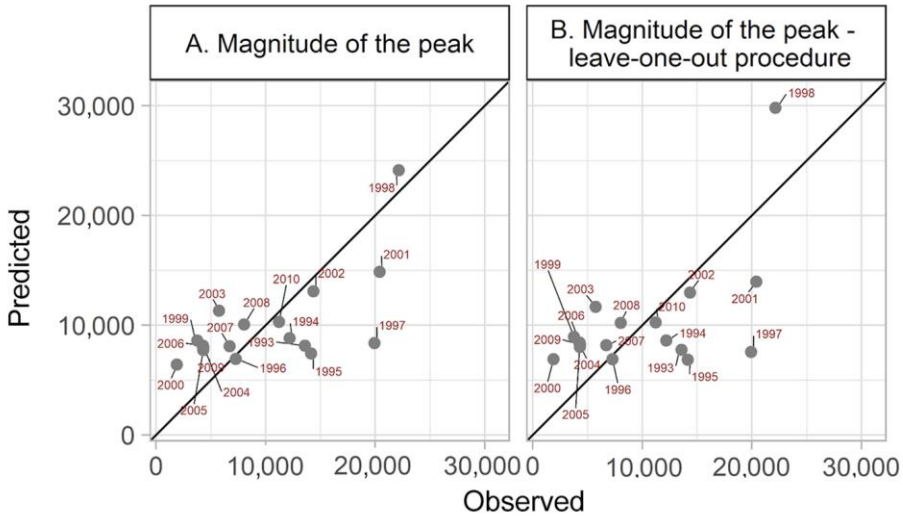
4. Van Panhuis W, Cross A, Burke D, Choisy M. Counts of Dengue reported in Vietnam: 1960–2010: Project Tycho dataset [Internet]. University of Pittsburgh; 2018 [cited 2019 Sep 13]. https://www.tycho.pitt.edu/dataset/VN.38362002



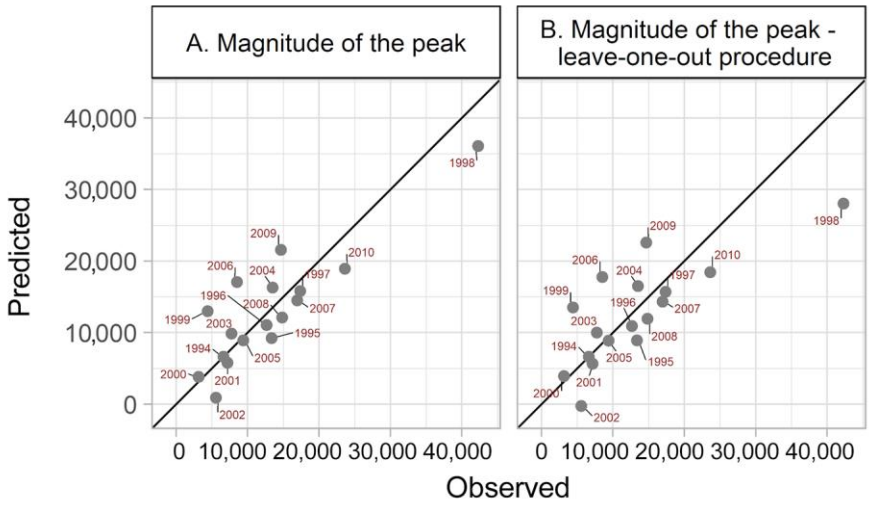
Appendix Figure 1. Observed versus predicted magnitude of the peak for each season using a simple linear regression model $M = \alpha + \beta N$, with M the magnitude of the peak and N the number of probable dengue cases reported to NDSS in April. The black line represents the expected results with perfect prediction. The figure on the right represents the results for the "leave-one-out" procedure.



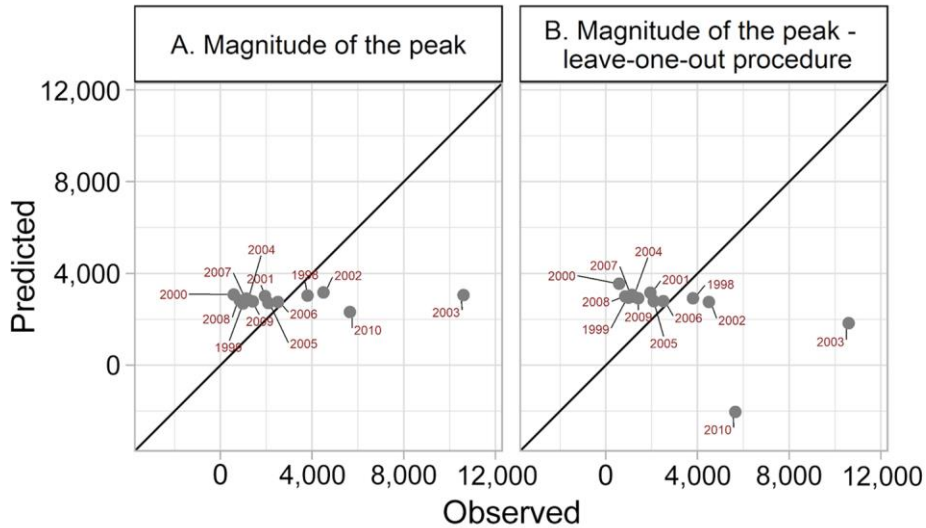
Appendix Figure 2. Observed versus predicted magnitude of the peak for each season using a simple linear regression model $M = \alpha + \beta N$, with M the magnitude of the peak and N the number of probable dengue cases reported to NDSS in April. The black line represents the expected results with perfect prediction. The figure on the left represents the results when we left 2007 and 2012 out of the learning sample and tried to predict the magnitude of the peak for these 2 years. Conversely, the figure on the right represent the results when we kept only 2007 and 2012 in the learning sample to predict the magnitude of the peak for the 11 other seasons.



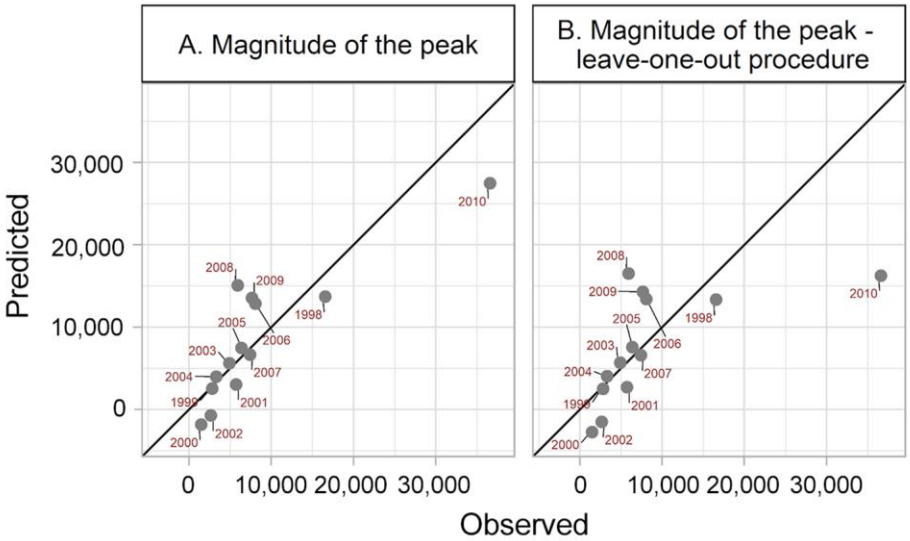
Appendix Figure 3. Observed *vs.* predicted magnitude of the peak for each season using a simple linear regression model $M = \alpha + \beta N$, with M the magnitude of the peak and N the number of reported dengue like cases in Thailand in February. The black line represents the expected results with perfect prediction. The figure on the right represents the results for the "leave-one-out" procedure.



Appendix Figure 4. Observed *vs.* predicted magnitude of the peak for each season using a simple linear regression model $M = \alpha + \beta N$, with M the magnitude of the peak and N the number of reported dengue like cases in Vietnam in March. The black line represents the expected results with perfect prediction. The figure on the right represents the results for the "leave-one-out" procedure.



Appendix Figure 5. Observed *vs.* predicted magnitude of the peak for each season using a simple linear regression model $M = \alpha + \beta N$, with M the magnitude of the peak and N the number of reported dengue like cases in Laos in March. The black line represents the expected results with perfect prediction. The figure on the right represents the results for the "leave-one-out" procedure.



Appendix Figure 6. Observed *vs.* predicted magnitude of the peak for each season using a simple linear regression model $M = \alpha + \beta N$, with M the magnitude of the peak and N the number of reported dengue like cases in Philippines in May. The black line represents the expected results with perfect prediction. The figure on the right represents the results for the "leave-one-out" procedure.