

RESEARCH ARTICLE

Mapping dengue risk in Singapore using Random Forest

Janet Ong¹, Xu Liu¹, Jayanthi Rajarethinam¹, Suet Yheng Kok¹, Shaohong Liang¹, Choon Siang Tang², Alex R. Cook³, Lee Ching Ng^{1,4*}, Grace Yap¹

1 Environmental Health Institute, National Environment Agency, Singapore, **2** Environmental Public Health Operations, National Environment Agency, Singapore, **3** Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, **4** School of Biological Sciences, Nanyang Technological University, Singapore

* NG_lee_ching@nea.gov.sg



Abstract

Background

Singapore experiences endemic dengue, with 2013 being the largest outbreak year known to date, culminating in 22,170 cases. Given the limited resources available, and that vector control is the key approach for prevention in Singapore, it is important that public health professionals know where resources should be invested in. This study aims to stratify the spatial risk of dengue transmission in Singapore for effective deployment of resources.

Methodology/principal findings

Random Forest was used to predict the risk rank of dengue transmission in 1km² grids, with dengue, population, entomological and environmental data. The predicted risk ranks are categorized and mapped to four color-coded risk groups for easy operation application. The risk maps were evaluated with dengue case and cluster data. Risk maps produced by Random Forest have high accuracy. More than 80% of the observed risk ranks fell within the 80% prediction interval. The observed and predicted risk ranks were highly correlated ($\rho \geq 0.86$, $P < 0.01$). Furthermore, the predicted risk levels were in excellent agreement with case density, a weighted Kappa coefficient of more than 0.80 ($P < 0.01$). Close to 90% of the dengue clusters occur in high risk areas, and the odds of cluster forming in high risk areas were higher than in low risk areas.

Conclusions

This study demonstrates the potential of Random Forest and its strong predictive capability in stratifying the spatial risk of dengue transmission in Singapore. Dengue risk map produced using Random Forest has high accuracy, and is a good surveillance tool to guide vector control operations.

OPEN ACCESS

Citation: Ong J, Liu X, Rajarethinam J, Kok SY, Liang S, Tang CS, et al. (2018) Mapping dengue risk in Singapore using Random Forest. *PLoS Negl Trop Dis* 12(6): e0006587. <https://doi.org/10.1371/journal.pntd.0006587>

Editor: Benjamin Althouse, Institute for Disease Modeling, UNITED STATES

Received: November 9, 2017

Accepted: June 6, 2018

Published: June 18, 2018

Copyright: © 2018 Ong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The dengue risk map described in the manuscript uses geo-referenced dengue case data which reveal patient's locality and hence compromise patient privacy. Thus, the data could not be made available due to patient confidentiality. However, qualified researchers could contact Ministry of Health, Singapore for access to the case data at Vernon_LEE@moh.gov.sg

Funding: This study was funded by the Ministry of Finance, Singapore under the Reinvestment Fund. The funders had no role in study design, data

collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Dengue fever, the most prevalent mosquito-borne viral disease today, is caused by Dengue virus (DENV) and transmitted to human by *Aedes* mosquitoes, primarily the *Ae. aegypti* and *Ae. albopictus*. The key approach to mitigating dengue transmission is to control the *Aedes* population, and this often involve vector control strategies such as larval source reduction and preventive surveillance that are labour-intensive and require effective deployment of valuable resources. Spatial risk profiling of dengue transmission is therefore necessary to ensure the optimal utilization of limited resources, and achieving maximum impact of dengue vector control. Here, we developed a dengue risk map by stratifying the spatial risk of dengue transmission in Singapore. Random Forest was used to predict the risk rank of dengue transmission in 1km² grids, and the predicted risk ranks are then categorized and mapped to color-coded risk groups. The dengue risk map is a good surveillance tool to guide vector control operations. Valuable resources can be deployed in a strategic manner, mitigating the spread of dengue transmission.

Introduction

Dengue is a viral infection caused by one of the four closely related yet antigenically distinct virus serotypes (DENV-1, DENV-2, DENV-3 and DENV-4), and transmitted by *Aedes* mosquitoes, primarily the *Ae. aegypti* and *Ae. albopictus* [1,2]. Infection confers lifelong immunity to the infecting serotype [3]. However, it increases risk for dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS), a deadly form that present with severe complications, in subsequent infections [4]. Since the publication of the GBD 2010, it was estimated that 390 million dengue infections occur each year globally, of which 500,000 develop into DHF [5,6]. Dengue poses a substantial public health threat globally, especially throughout the tropical and subtropical regions [7,8].

Located one and a half degrees north of the equator and lying in the dengue belt, Singapore is prone to dengue transmission, with all four dengue serotypes co-circulating and frequent introduction of new genotype virus [9]. Though intensive vector control efforts have successfully suppressed the *Aedes* population, from an *Aedes* house index of over 50% in the 1960's to the present 1–2%, Singapore remains susceptible to dengue outbreaks [10–12]. The increased in human population density and the low herd immunity resulting from sustained period of low dengue transmission are factors that may have contributed to the resurgence of dengue in Singapore [13,14]. A significant amount of funding and resources has been allocated for dengue every year [15]. The estimated economic and disease burden of dengue were 9–14 disability-adjusted life years (DALYs) per 100,000 population and US\$41.5 million per annum [16].

A dengue temporal model was developed in 2013 by the Environmental Health Institute, a research institute of the Singapore's National Environment Agency (NEA) in collaboration with the National University of Singapore (NUS) to aid vector control measures. The model predicts trends and incidence up to 12 weeks ahead, providing early warnings of outbreak and facilitating public health response to moderate impending outbreak [17]. This model was able to accurately project an upward trend of dengue cases in 2013 and 2014, predicting the two major outbreaks [18]. NEA has been using the model in planning vector control and public communication [19]. However, a limitation of the model is the missing spatial resolution as it does not highlight areas with high risk of dengue transmission. Given that NEA's key strategy in dengue control is preventive surveillance and larval source reduction, a labour-intensive activity that requires effective deployment of a limited pool of skilled vector control officers,

spatial risk profiling of dengue transmission is thus necessary for effective deployment of resources, and achieving maximum impact.

In this paper, we describe a new approach for spatial risk stratification of dengue transmission in Singapore. Using Random Forest, we quantify the risk of dengue transmission in different areas and categorize them into different risk groups to guide the pre-emptive source reduction exercise conducted by NEA vector control officers. Predictive performance of the model is evaluated with both dengue cases and clusters.

Materials and methods

Statistical analysis

Proposed by Leo Breiman, Random Forest is an ensemble machine learning method that uses an ensemble of decision trees [20]. In Random Forest, several ($N = 1000$) bootstrap samples are drawn from the training set data, and an unpruned decision tree $f_n(x)$, is fitted to each bootstrap sample. At each node of the decision tree, variable selection is carried out on a small random subset of the predictor variables, so as to avoid the “small n large p ” problem. The best split on these predictors is used to split the node. The predicted response is obtained by averaging the predictions of all trees, i.e. $\frac{1}{N} \sum_{n=1}^N f_n(x)$ (Fig 1). Random Forest was used to predict the percentile rank of dengue case count in 1km^2 grids, with past dengue exposure (total number of cases in previous year, total number of cases in neighbouring grids in previous year and number of non-resident cases in previous year), human population (estimated population density), vector population (estimated ratio of *Aedes aegypti* mosquitoes out of all *Aedes* mosquitoes—breeding percentage) and environmental data (vegetation index, connectivity index and ratio of residential area). The predicted percentile ranks are then categorized and mapped to four color-coded risk groups (RG1-4, lowest risk as RG1 and highest risk of dengue transmission as RG4) for easy operation application. Although administrative boundaries are

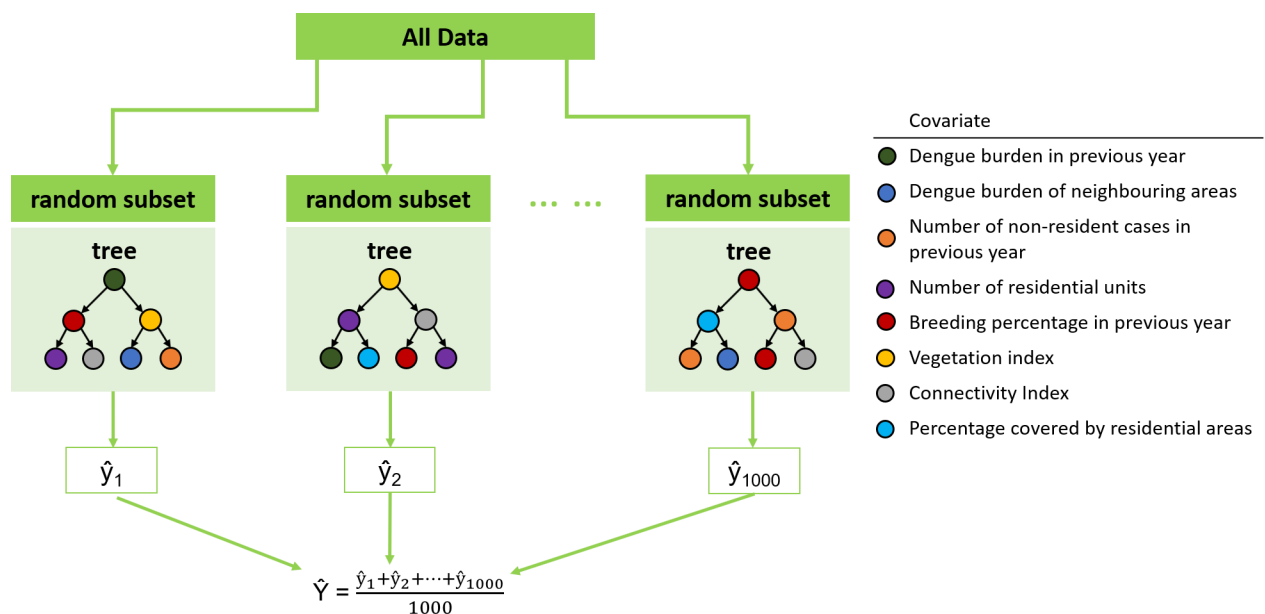


Fig 1. Framework of random forest algorithm. 1000 random bootstrap samples were drawn from the data, and an unpruned decision tree is fitted to each bootstrap sample. At each node, a small subset of the covariates was chosen at random to optimize the split. The predicted risk rank is obtained by averaging the prediction of all trees.

<https://doi.org/10.1371/journal.pntd.0006587.g001>

more compatible with ground operation, 1km² grids were used as study units as they are more consistent in area size and do not change over time. We use residential grids exclusively for the analysis and risk mapping. Random Forest analyses were performed using the randomForest package implemented in the R statistical language [21].

Model evaluation

Data from 2006 to 2013 were used to parameterize the model, and performance of the model is evaluated with new dengue case data from 2014 to 2016. Apart from visually comparing the risk map and distribution of dengue cases, we applied the following quantitative metrics to evaluate the model: 1. correlation between predicted and observed percentile ranks, 2. coverage of prediction intervals, 3. summary statistics of the number of cases within each risk group, and 4. weighted (square) Kappa agreement coefficients of risk grouping.

In addition to using dengue case data, data on dengue cluster, which indicates possible transmission within the locality, were considered for model evaluation as well. We investigated the odds of clusters forming in high (RG 3 and 4) and low (RG 1 and 2) risk areas, and examined if transmission intensity, comprising of cluster’s growth rate, transmission duration and cluster size differ between high and low risk areas. Differences were analysed using Kruskal-Wallis tests.

Data

Table 1 shows the various risk factors considered for the risk mapping. The risk factors were identified from literature review and examined with historical data [11,22,23]. All data (Dengue, Population and Entomological) were aggregated to the 1km² grids. The time period used for all variables was January 2006 to December 2016, and their sources are:

Dengue cases. Dengue is a notifiable disease in Singapore, where medical practitioners are required to notify all clinically diagnoses and laboratory confirmed dengue cases to the Ministry of Health (MOH), Singapore [24]. Residential and workplace address and onset date of each dengue cases are recorded and shared with NEA on a daily basis. Dengue cases were tagged to the address, either residential or workplace address, after epidemiological investigation has been carried out by officers to determine and confirm the location where the cases acquired dengue. The addresses were then geocoded using the Geographic Information System (GIS). Geo-referenced data on dengue cases was extracted from the GIS database of NEA, and anonymized prior to analysis.

Population density. The number of residential units were provided by Housing Development Board (HDB) for public housing and sourced online from the Real Estate Information

Table 1. Overview of the risk factors used for dengue risk mapping.

Category	Type	Covariates
Dengue exposure	Spatio-temporal	Dengue burden in previous year
		Dengue burden of neighbouring grids in previous year
		Number of non-resident cases in previous year
Population	Spatio-temporal	Number of residential units
Entomological	Spatio-temporal	Breeding percentage in previous year
Environmental	Spatial	Vegetation index
	Spatial	Connectivity index
	Spatial	Percentage covered by residential areas

<https://doi.org/10.1371/journal.pntd.0006587.t001>

System (REALIS), an online database managed by Urban Redevelopment Authority, for private housing.

Entomological. Breeding Percentage (BP) is an in-house index developed by NEA to estimate the proportion of *Ae. aegypti* relative to *Ae. albopictus*, which is ubiquitous in Singapore [25]. BP is calculated from the number of *Aedes* mosquito breeding sites recorded during ground inspections carried out by NEA using the formula:

$$BP = \frac{N_{aegypti}(t)}{N_{aegypti}(t) + N_{albopictus}(t) - N_{mixed}(t)}$$

NEA carried out routine inspection surveillance across Singapore throughout the year. These inspections include those scheduled for regular preventive surveillance, and those conducted in response to dengue transmission in a location. To estimate the yearly BP for each grid, geo-referenced data on *Aedes spp.* larval counts from the routine surveillance was extracted from the GIS database of NEA and mapped the location of *Aedes* breeding sites onto each grid to extract the number of *Ae. aegypti* and/or *Ae. albopictus* breeding sites found within each grid for each year. BP value for grids with inspections was calculated by definition. For grids that were not inspected, their BP values were estimated using ordinary Kriging with a spherical variogram model.

Environmental factors. Vegetation index, also known as the Normalized Difference Vegetation Index (NDVI), is an index of plant “greenness” or photosynthetic activity. NDVI data was provided by Centre for Remote Imaging, Sensing and Processing in NUS after processing satellite image. Connectivity index measures the total connectivity (accessibility) of the grid relative to all other grids, and is derived from public transport data from Future City Lab ETH-NUS.

Dengue clusters. Dengue cases are clustered for vector operations purposes based on their geographical and temporal proximity. A dengue cluster is formed when two or more cases are located within a 150-meter radius and with the onsets of illness within a 14-day period. Dengue clusters are generated using the Geographical Information System (GIS), and information such as transmission duration, serotypes detected and the number of dengue cases is recorded for every cluster [12].

Results

Associations between covariates and dengue burden were examined through partial dependence plot. Consistent with our prior knowledge, all covariates are associated with dengue burden, as contrasted by the flat line partial effect of random noise (Fig 2). Among the covariates, the number of residential units, dengue burden in previous year and the breeding percentage in previous year are top-ranked in terms of variable importance (Fig 3), and impose a larger influence on model accuracy, relative to the other covariates. This, therefore, suggests that population density, dengue burden and abundance of *Ae. aegypti* are significant risk factors for dengue transmission.

The predicted percentile ranks were categorized and mapped to four color-coded risk groups based on the three quartiles so that the number of grids in each risk group is approximately the same. The distribution of risk groups is comparable in all three years, with high risk groups (RG 3 and 4) congregating in the eastern part of Singapore. When dengue cases were overlaid onto the risk maps, we observed good agreement between the cases and risk groups (Fig 4). Majority of the cases fell in risk group 3 and 4. There was strong positive correlation between the observed and predicted risk ranks, a correlation of 0.86 ($P < 0.01$), 0.87 ($P < 0.01$) and 0.88 ($P < 0.01$) in 2014, 2015 and 2016 respectively. In addition, the risk level

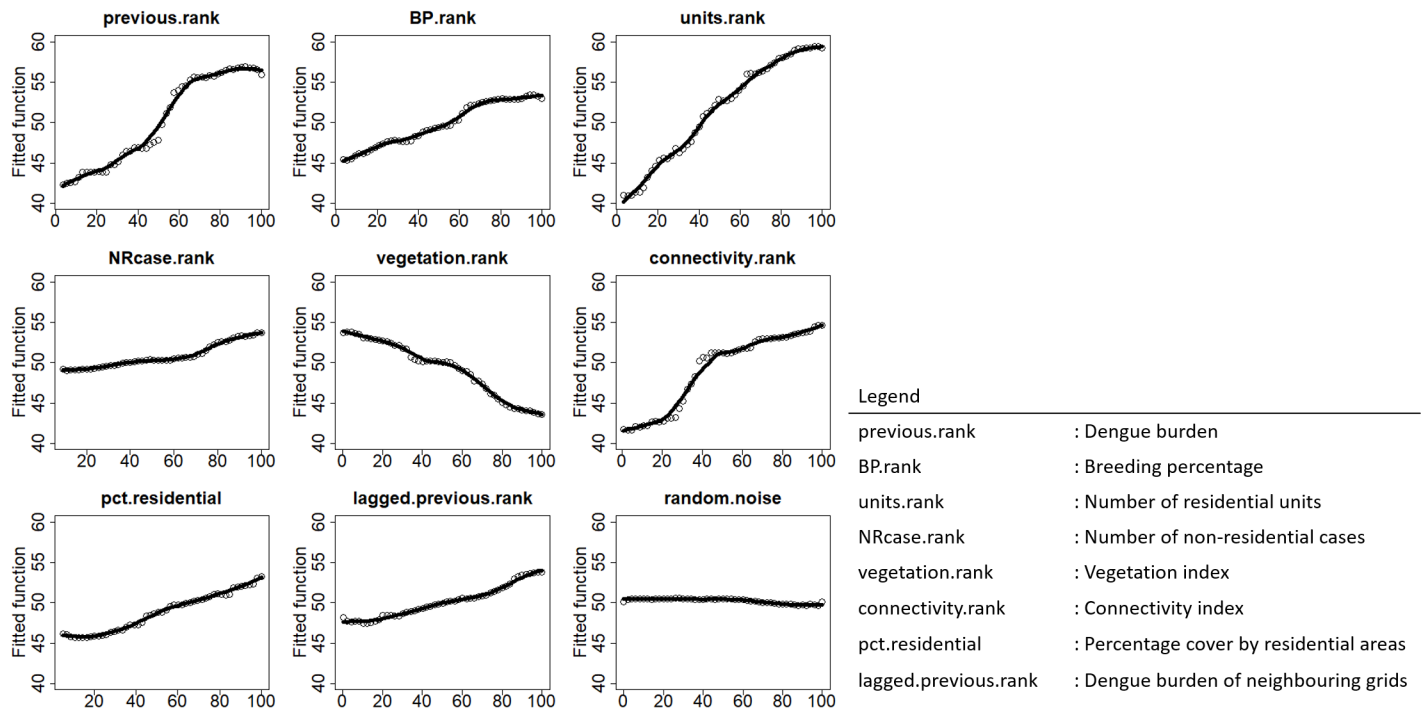


Fig 2. Partial dependence plot of the risk factors showing how dengue burden varies with one variable when all other variables are held constant at their average values.

<https://doi.org/10.1371/journal.pntd.0006587.g002>

commensurate with case density. The predicted risk levels were in excellent agreement with the case density, a weighted Kappa coefficient of 0.814 ($P < 0.01$) in 2014, 0.839 ($P < 0.01$) in 2015 and 0.821 ($P < 0.01$) in 2016. This is further supported by the increasing trend of dengue case count from risk group 1 to 4 (Table 2). Fig 5 shows the predicted percentile ranks and its 80% prediction interval. 82% and 83% of the observed percentile ranks fell within the 80% prediction interval in 2014 and 2015 respectively. In 2016, 81% of the observed percentile ranks

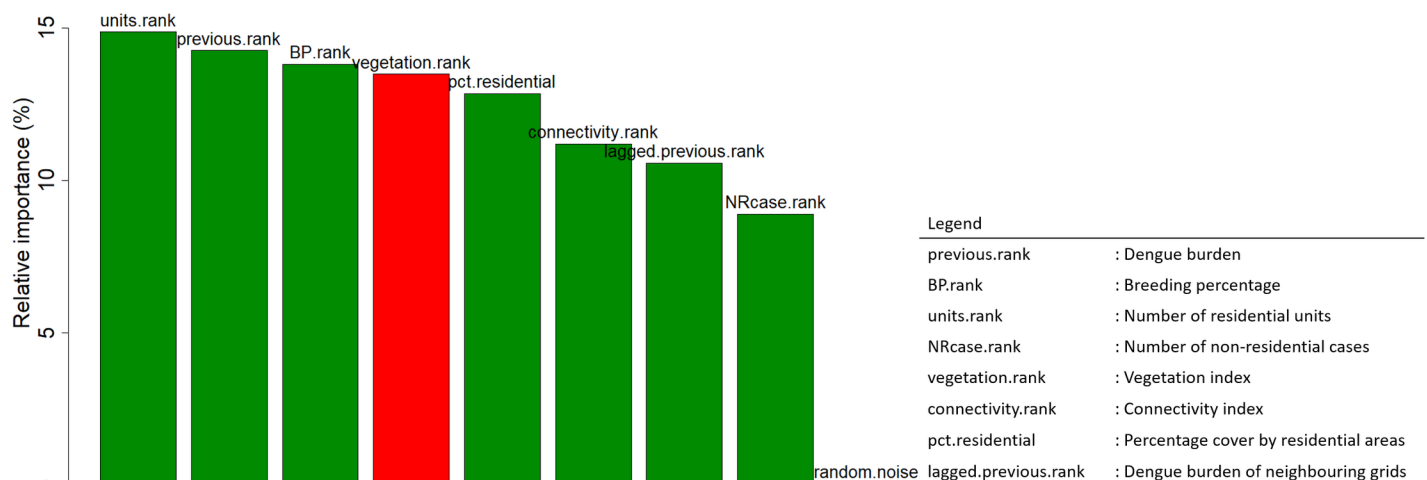


Fig 3. Variable importance plot showing population density, dengue burden and breeding percentage having stronger predictive power than other variables.

<https://doi.org/10.1371/journal.pntd.0006587.g003>

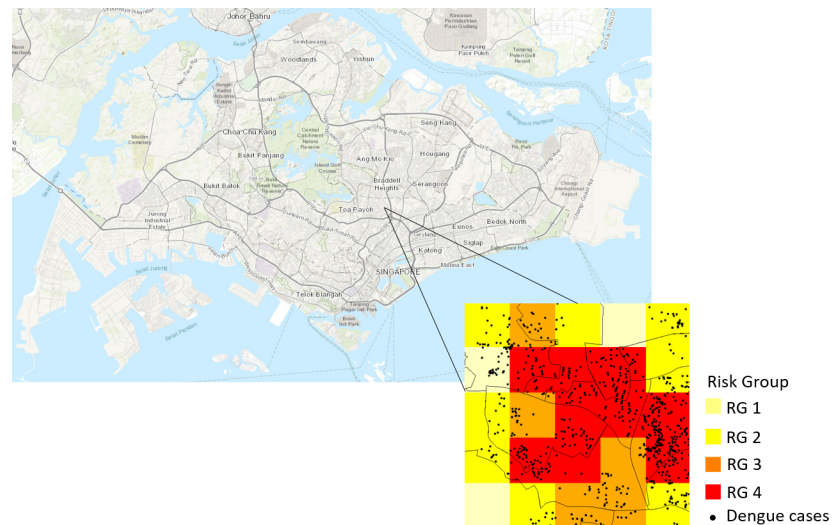


Fig 4. Risk grouping of a small area with 2016 dengue cases (black circles) overlaid. The risk groups are color-coded, with RG 4 (highest risk) as red and RG 1 (lowest risk) in light yellow. The figure was created using R software with base layer obtained from <https://landsatlook.usgs.gov/>.

<https://doi.org/10.1371/journal.pntd.0006587.g004>

fell within the 80% prediction interval. Overall, cases in 2015 have slightly better agreement with the risk map than in 2014 and 2016.

Evaluation of risk maps with 2014 to 2016 clusters data shows that the number of dengue clusters in high risk areas was almost 8 times the low risk areas (Fig 6). Each year, close to 90% of the dengue clusters were found in high risk areas, which represent 22% of Singapore land area and 50% of residential areas. The odds of cluster forming in high risk areas was higher than in low risk areas for all three years. The odds ratios were 11.1 ($P < 0.01$), 14.6 ($P < 0.01$) and 12.1 ($P < 0.01$) for 2014, 2015 and 2016 respectively. Clusters were further stratified by the number of serotypes into single serotype and multiple serotypes clusters. High risk areas have a larger proportion of multiple serotypes clusters than low risk areas, and interestingly, 3-serotypes clusters were only present in high risk areas, especially in RG4 (Fig 6). Transmission intensity, comprising of cluster’s growth rate, transmission duration and cluster size were significantly different between single serotype and multiple serotypes clusters ($P < 0.01$). Clusters

Table 2. Summary statistics of dengue case count in grids for all risk groups in 2014–2016.

Year	Risk Group	Percentile Rank	Count	Min	1 st Quartile	Median	Mean	3 rd Quartile	Max
2014	1	[0, 25)	79	0.0	2.0	4.0	8.5	9.0	157.0
	2	[25, 50)	78	3.0	10.0	16.0	21.9	22.0	151.0
	3	[50, 75)	79	13.0	27.5	43.0	59.8	63.0	512.0
	4	[75, 100]	79	27.0	53.5	75.0	112.6	163.5	390.0
2015	1	[0, 25)	78	0.0	1.0	3.0	4.5	6.0	29.0
	2	[25, 50)	79	2.0	9.0	14.0	16.0	19.0	83.0
	3	[50, 75)	79	7.0	20.0	29.0	38.3	41.5	176.0
	4	[75, 100]	79	23.0	38.5	56.0	64.3	78.5	304.0
2016	1	[0, 25)	79	2.0	6.0	12.0	16.9	22.0	67.0
	2	[25, 50)	78	12.0	28.0	37.0	39.5	48.0	91.0
	3	[50, 75)	79	26.0	51.0	60.0	61.7	73.0	92.0
	4	[75, 100]	79	41.0	73.0	85.0	82.2	94.0	100.0

<https://doi.org/10.1371/journal.pntd.0006587.t002>

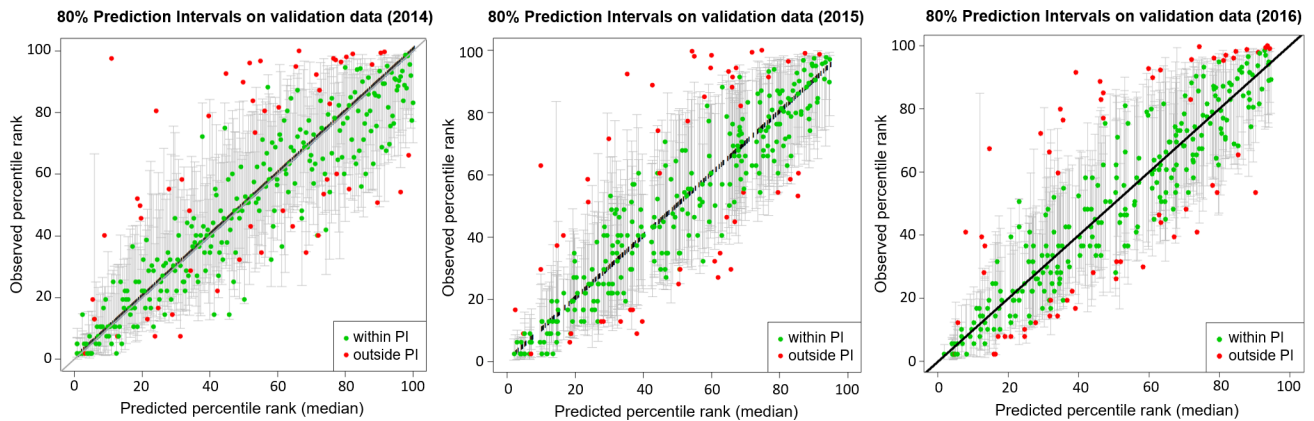


Fig 5. The predicted percentile ranks and its 80% prediction interval for 2014 to 2016 (left to right). In each panel, the green circles indicate the predicted percentile ranks that fall within the prediction intervals and the red circles indicate the predicted percentile ranks that fall outside the prediction intervals.

<https://doi.org/10.1371/journal.pntd.0006587.g005>

with more serotypes present have a faster growth rate, longer transmission duration and larger cluster size (Table 3). The same characteristics were seen when we grouped the clusters by high and low risk areas. Though there were less clusters in low risk areas, the transmission intensity of clusters in these low risk areas was of no significant difference ($P > 0.1$) when compared with those in high risk areas (Table 3).

Discussion

Dengue has been endemic in Singapore since its first reported outbreak in 1901 [26]. Though the dengue temporal model is capable of predicting impending outbreaks, it does not indicate where the outbreak will be [17]. As a result, source reduction inspections are conducted on a frequency based on the risk level of the premises types (e.g. construction sites are of higher risk than apartment homes). Spatial risk mapping of dengue transmission is therefore essential for the prioritization and allocation of scarce resources especially manpower need to inspect premises.

Dengue risk map has been developed in many countries as a surveillance tool to enhance public health preparedness for dengue outbreak [27]. Statistical approaches such as logistic

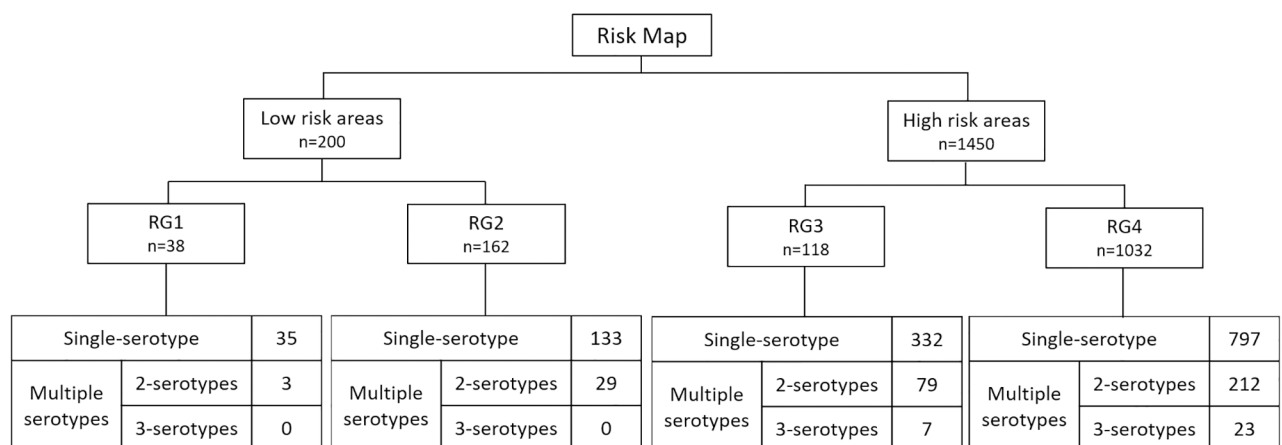


Fig 6. Stratification of clusters (2014–2016) by risk groups and the number of serotypes present in cluster.

<https://doi.org/10.1371/journal.pntd.0006587.g006>

Table 3. The median transmission duration, cluster size and growth rate of clusters with different number of serotypes present in Singapore (2014–2016), and separated by low^a and high^b risk areas.

No. of serotypes in cluster	Median transmission duration (days)			Median cluster size (no. of cases)			Median growth rate (no. of new cases per day)			No. of clusters		
	Combined*	Low risk areas	High risk areas	Combined*	Low risk areas	High risk areas	Combined*	Low risk areas	High risk areas	Combined	Low risk areas	High risk areas
1	13	13	13	4	4	4	0.28	0.28	0.28	1319	168	1129
2	21	18	21	7	7	7	0.35	0.34	0.35	328	32	291
3	53	n/a	93	25	n/a	70	0.52	n/a	0.68	31	0	30
4	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0	0	0

^a Low risk areas are RG 1 and RG 2 grids.

^b High risk areas are RG 3 and RG 4 grids.

n/a, data not available

Kruskal-Wallis test (P < 0.01 are highlighted by *)

<https://doi.org/10.1371/journal.pntd.0006587.t003>

regression models, generalized linear models and general additive models were most commonly used to compute risk level and create dengue risk map [28–32]. Although very good predictive accuracy can be achieved from Random Forest, it has yet to be reported in the development of dengue risk map [27]. In this paper, we demonstrated the use of Random Forest, an ensemble learning method that has garnered much interest in the machine-learning community, to develop a dengue risk map with high accuracy and robustness. Studies have shown that Random Forest has excellent performance in classification tasks, and even outperforms its counterparts such as discriminant analysis, neural networks and support vector machines [33,34]. The methodology has several advantages over the traditional approaches, with the utmost advantage being highly tolerant to interactions among the input covariates. Dengue transmission is a multi-factorial stochastic process where often one risk factor is correlated with other risk factors, making it difficult to quantify the effect of a particular risk factor as well as to construct a risk map using classical modelling method such as regression.

The model ranked the overall risk of dengue transmission of different areas in a year and mapped the ranks as color-coded risk groups. By comparing the risk groupings of the grids over the years, NEA could identify recurring risk areas (i.e. grids that are persistently risk group 4 over the years) that are of concerns, fluctuating risk areas (i.e. grids that have fluctuating risk grouping over the years) and even potential risk areas that were not seen in the previous years (i.e. grids whose risk group change from 1 to 4). Evaluation using latest dengue case data showed the model had strong predictive capability. Strong positive correlation between the observed and predicted risk ranks, and an almost perfect agreement between the predicted risk levels and case density were observed. High risk areas are where clusters, in particular multiple serotypes clusters are most likely to occur. However, surprisingly, despite the difference in risk levels, there was no difference in the transmission intensity of clusters in high and low risk areas, and this may be attributed to the presence of small pockets of high *Ae. aegypti* population within the low risk areas. For instance, construction sites along Flora Road and Belgravia Drive had led to large dengue cluster of size 46 and 35 in traditionally low risk areas in 2014 and 2016 respectively. This, therefore, highlights the importance of ground inspections in identifying high risk sites in low risk areas.

The dengue risk map complements the dengue temporal model in allowing the operation department of NEA to prioritise vector control efforts. While the dengue temporal model provides the time component of when the next outbreak will be, it is thus now possible for NEA to deploy limited resources ahead of time, targeting at the places with high risk of transmission.

There are, however, some limitations to the use of Random Forest, the key on being the model not amenable to interpretation. The Random Forest is an ensemble method—it constructs many “weak” models and then combines them to achieve a “strong” model. There is no explicit formulae-form relationship between risk of dengue transmission and risk factors, making it virtually impossible to decompose a particular prediction output into contribution of risk factors. Understanding that the primary objective is to accurately stratify the risk of dengue transmission liberated us from concerns over interpretability. Nevertheless, the Random Forest model is able to offer some insights about dengue transmission by estimating importance and partial effects of variable at a macro level.

The dengue risk map has become an integral part of Singapore’s dengue control program. The dengue risk map would be generated at the start of each year, and NEA operations would use the risk map as a guide to prioritize resource allocation for dengue control and plan the preventive surveillance activities for the year. Dengue risk map has been used since 2015 by the operational division of NEA to guide targeted preventive interventions. Future work will include incorporating real time data to develop a spatio-temporal risk map.

Conclusions

This study demonstrates the potential of Random Forest and its strong predictive capability in stratifying the spatial risk of dengue transmission in Singapore. Dengue risk map produced using Random Forest has high accuracy, and is a good tool to guide vector control operations, allowing targeted preventive measures before and in times of dengue outbreak. Valuable resources can then be deployed in a strategic manner, mitigating the spread of dengue transmission.

Acknowledgments

The authors would like to thank the Communicable Disease Division of Ministry of Health, Singapore; Housing Development Board, Singapore; Urban Redevelopment Authority, Singapore; Centre for Remote Imaging, Sensing and Processing, National University of Singapore, Singapore and Future Cities Laboratory, Singapore for providing the data used in this project. This project is also made possible from the effort put in by both the Environmental Health Institute and the Environmental Public Health Operations of the National Environmental Agency, Singapore.

Author Contributions

Conceptualization: Alex R. Cook, Lee Ching Ng, Grace Yap.

Data curation: Janet Ong, Jayanthi Rajarethinam.

Formal analysis: Janet Ong, Xu Liu.

Funding acquisition: Lee Ching Ng.

Investigation: Janet Ong, Xu Liu, Jayanthi Rajarethinam, Suet Yheng Kok, Shaohong Liang, Choon Siang Tang, Alex R. Cook.

Methodology: Janet Ong, Xu Liu, Alex R. Cook, Lee Ching Ng, Grace Yap.

Project administration: Lee Ching Ng, Grace Yap.

Resources: Lee Ching Ng.

Supervision: Lee Ching Ng, Grace Yap.

Visualization: Janet Ong.

Writing – original draft: Janet Ong.

Writing – review & editing: Janet Ong, Lee Ching Ng, Grace Yap.

References

1. Recker M, Blyuss KB, Simmons CP, Hien TT, Wills B, Farrar J, et al. Immunological serotype interactions and their effect on the epidemiological pattern of dengue. *Proceedings of the Royal Society of London B: Biological Sciences*. 2009 Jan 1; rspb-2009.
2. Rosen L, Shroyer DA, Tesh RB, Freier JE, Lien JC. Transovarial transmission of dengue viruses by mosquitoes: *Aedes albopictus* and *Aedes aegypti*. *The American journal of tropical medicine and hygiene*. 1983 Sep 1; 32(5):1108–19. PMID: [6625066](#)
3. Simmons CP, Farrar JJ, van Vinh Chau N, Wills B. Dengue. *New England Journal of Medicine*. 2012 Apr 12; 366(15):1423–32. <https://doi.org/10.1056/NEJMra1110265> PMID: [22494122](#)
4. Ooi EE, Goh KT, Gubler DJ. Dengue prevention and 35 years of vector control in Singapore. *Emerging infectious diseases*. 2006 Jun; 12(6):887. <https://doi.org/10.3201/10.3201/eid1206.051210> PMID: [16707042](#)
5. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, Myers MF. The global distribution and burden of dengue. *Nature*. 2013 Apr; 496(7446):504. <https://doi.org/10.1038/nature12060> PMID: [23563266](#)
6. Chaturvedi UC, Shrivastava R. Dengue haemorrhagic fever: A global challenge. *Indian journal of medical microbiology*. 2004 Jan 1; 22(1):5. PMID: [17642678](#)
7. Gibbons RV, Vaughn DW. Dengue: an escalating problem. *BMJ: British Medical Journal*. 2002 Jun 29; 324(7353):1563. PMID: [12089096](#)
8. Gubler DJ. Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. *Trends in microbiology*. 2002 Feb 1; 10(2):100–3. PMID: [11827812](#)
9. Lee KS, Lo S, Tan SS, Chua R, Tan LK, Xu H, et al. Dengue virus surveillance in Singapore reveals high viral diversity through multiple introductions and in situ evolution. *Infection, Genetics and Evolution*. 2012 Jan 31; 12(1):77–85. <https://doi.org/10.1016/j.meegid.2011.10.012> PMID: [22036707](#)
10. Hapuarachchi HC, Koo C, Rajarethinam J, Chong CS, Lin C, Yap G, et al. Epidemic resurgence of dengue fever in Singapore in 2013–2014: A virological and entomological perspective. *BMC infectious diseases*. 2016 Jun 17; 16(1):300.
11. Koh BK, Ng LC, Kita Y, Tang CS, Ang LW, Wong KY, et al. The 2005 dengue epidemic in Singapore: epidemiology, prevention and control. *Annals Academy of Medicine Singapore*. 2008 Jul 1; 37(7):538.
12. Ler TS, Ang LW, Yap GS, Ng LC, Tai JC, James L, et al. Epidemiological characteristics of the 2005 and 2007 dengue epidemics in Singapore—similarities and distinctions. *Western Pacific surveillance and response journal: WPSAR*. 2011 Apr; 2(2):24. <https://doi.org/10.5365/WPSAR.2010.1.1.011> PMID: [23908885](#)
13. Low SL, Lam S, Wong WY, Teo D, Ng LC, Tan LK. Dengue seroprevalence of healthy adults in Singapore: serosurvey among blood donors, 2009. *The American journal of tropical medicine and hygiene*. 2015 Jul 8; 93(1):40–5. <https://doi.org/10.4269/ajtmh.14-0671> PMID: [26013376](#)
14. Struchiner CJ, Rocklöv J, Wilder-Smith A, Massad E. Increasing dengue incidence in Singapore over the past 40 years: population growth, climate and mobility. *PLoS One*. 2015 Aug 31; 10(8):e0136286. <https://doi.org/10.1371/journal.pone.0136286> PMID: [26322517](#)
15. MOF (Ministry of Finance). Singapore Budget. 2016. Available: http://www.singaporebudget.gov.sg/budget_2013/expenditure_overview/mewr.html
16. Carrasco LR, Lee LK, Lee VJ, Ooi EE, Shepard DS, Thein TL, et al. Economic impact of dengue illness and the cost-effectiveness of future vaccination programs in Singapore. *PLoS neglected tropical diseases*. 2011 Dec 20; 5(12):e1426. <https://doi.org/10.1371/journal.pntd.0001426> PMID: [22206028](#)
17. Shi Y, Liu X, Kok SY, Rajarethinam J, Liang S, Yap G, et al. Three-month real-time dengue forecast models: an early warning system for outbreak alerts and policy decision support in Singapore. *Environmental health perspectives*. 2016 Sep; 124(9):1369. <https://doi.org/10.1289/ehp.1509981> PMID: [26662617](#)
18. Seltenrich N. Singapore success: new model helps forecast dengue outbreaks. *Environmental health perspectives*. 2016 Sep; 124(9):A167. <https://doi.org/10.1289/ehp.124-A167> PMID: [27581656](#)

19. NEA (National Environmental Agency). Newsroom: Dengue Cases May Exceed 30,000 in 2016. 2016. Available: <http://www.nea.gov.sg/corporate-functions/newsroom/news-releases/category/public-health/dengue-cases-may-exceed-30000-in-2016>
20. Breiman L. Random forests. *Machine learning*. 2001 Oct 1; 45(1):5–32.
21. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
22. Cheong YL, Leitao PJ, Lakes T. Assessment of land use factors associated with dengue cases in Malaysia using Boosted Regression Trees. *Spatial and spatio-temporal epidemiology*. 2014 Jul 31; 10:75–84. <https://doi.org/10.1016/j.sste.2014.05.002> PMID: 25113593
23. Stoddard ST, Morrison AC, Vazquez-Prokopec GM, Soldan VP, Kochel TJ, Kitron U, et al. The role of human movement in the transmission of vector-borne pathogens. *PLoS neglected tropical diseases*. 2009 Jul 21; 3(7):e481. <https://doi.org/10.1371/journal.pntd.0000481> PMID: 19621090
24. Infectious Diseases Act: Agency SNE, ed. Chapter 137, 2002.
25. Ho ZJ, Hapuarachchi HC, Barkham T, Chow A, Ng LC, Lee JM, Leo YS, Prem K, Lim YH, de Sessions PF, Rabaa MA. Outbreak of Zika virus infection in Singapore: an epidemiological, entomological, virological, and clinical analysis. *The Lancet Infectious Diseases*. 2017 Aug 1; 17(8):813–21. [https://doi.org/10.1016/S1473-3099\(17\)30249-9](https://doi.org/10.1016/S1473-3099(17)30249-9) PMID: 28527892
26. More FW. Observations on dengue fever in Singapore. *J. Malaya Branch Br. Med. Assoc.* 1904; 1:24–9.
27. Louis VR, Phalkey R, Horstick O, Ratanawong P, Wilder-Smith A, Tozan Y, et al. Modeling tools for dengue risk mapping—a systematic review. *International journal of health geographics*. 2014 Dec 9; 13(1):50.
28. Cordeiro R, Donalisio MR, Andrade VR, Mafra AC, Nucci LB, Brown JC, et al. Spatial distribution of the risk of dengue fever in southeast Brazil, 2006–2007. *BMC Public Health*. 2011 May 20; 11(1):355.
29. Honorio NA, Nogueira RM, Codeco CT, Carvalho MS, Cruz OG, Magalhaes MD, et al. Spatial evaluation and modeling of dengue seroprevalence and vector density in Rio de Janeiro, Brazil. *PLoS neglected tropical diseases*. 2009 Nov 10; 3(11):e545. <https://doi.org/10.1371/journal.pntd.0000545> PMID: 19901983
30. Shafie A. Evaluation of the spatial risk factors for high incidence of dengue fever and dengue hemorrhagic fever using GIS application. *Sains Malaysiana*. 2011 Aug 1; 40(8):937–43.
31. Sriprom M, Chalvet-Monfray K, Chaimane T, Vongsawat K, Bicout DJ. Monthly district level risk of dengue occurrences in Sakon Nakhon Province, Thailand. *Science of the total environment*. 2010 Oct 15; 408(22):5521–8. <https://doi.org/10.1016/j.scitotenv.2010.08.024> PMID: 20817262
32. Wu PC, Lay JG, Guo HR, Lin CY, Lung SC, Su HJ. Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. *Science of the total Environment*. 2009 Mar 15; 407(7):2224–33. <https://doi.org/10.1016/j.scitotenv.2008.11.034> PMID: 19157509
33. Li X, Wang Y. Applying various algorithms for species distribution modelling. *Integrative Zoology*. 2013 Jun 1; 8(2):124–35. <https://doi.org/10.1111/1749-4877.12000> PMID: 23731809
34. Liaw A, Wiener M. Classification and regression by randomForest. *R news*. 2002 Dec 3; 2(3):18–22.