Chiang Mai J. Sci. 2018; 45(6) : 2509-2514
http://epg.science.cmu.ac.th/ejournal/
Contributed Paper

Influenza Activity and Province-level Weather Variations in Thailand, 2009 to 2014, Using Random Forest Time-series Approach

Romrawin Chumpu, Nirattaya Khamsemanan* and Cholwich Nattee*
Sirindhorn International Institute of Technology, Thammasat University, Pathum Thani 12120, Thailand.
* Author for correspondence; e-mail: nirattaya@siit.tu.ac.th; cholwich@siit.tu.ac.th

Received: 25 April 2017
Accepted: 6 September 2017

ABSTRACT

Influenza activity in Thailand, like many tropical regions, is a burden on public health and economy. The goal of this study is to find links between influenza activity and weather variations of the province-level in Thailand. We use a Random Forest time-series model to analyze the data and evaluate the importance of each variable. Data from 2009 to 2013, in weekly units, is used as a training set to create prediction models. Data from 2014 is used as a test set to validate prediction models. Five high populated provinces are selected to represent different geological regions in Thailand for this study. Although, results indicate that the number of influenza cases from the previous week yields the highest importance for influenza activity, weather variations are not without their impacts. Influenza activity in different provinces is associated with different sets of weather variations and their importance. In all selected provinces, the change in temperatures plays a significant role in influenza activity but its impacts are location-dependent. Correlation coefficients between predicted and observed influenza cases in 2014 are from 0.55 to 0.91.

Keywords: influenza activity, infectious disease, random forest prediction

1. INTRODUCTION

The World Health Organization estimates that annual seasonal influenza results in 3 to 5 million cases of severe illness and 250,000 to 500,000 deaths worldwide. In different parts of the world, influenza outbreaks occur in different periods of time. This suggests the association between seasonal influenza incidences and weather and environmental variations. In temperate regions, influenza activity is at the highest during the winter time. In tropical regions, influenza incidences peak during raining seasons when temperature and humidity are high [1, 2]. Even though the core factors that cause influenza outbreaks are still unknown, recent studies focus on weather conditions that promote the transmissibility of the influenza virus. In this study, we investigate the relation between weather parameters and reported influenza incidences at the province-level in Thailand. Five high populated provinces are selected to represent different parts of Thailand; Bangkok (capital), Chiang Mai (north), Ubon Ratchathani (northeast), Chonburi (coastal) and Phuket (south). These provinces are chosen, in part, due to the completion of both weather data and influenza data.

Based loosely on the classical susceptible-infected-recovered (SIR) framework [9], we also take the number of influenza cases up to the prior two weeks into our consideration. We employ a Random Forest time-series model. A recent study suggests that this technique provide more advantages than other existing regression techniques [7]. Among these advantages, the Random Forest technique provides how "important" an input variable is to the outcome.

We believe that this study helps fill the gap in knowledge of the link between weather conditions and influenza cases in tropical regions. Even though Thailand is not a large country, weather conditions are quite various in different parts. In addition, influenza pandemics of each province occur at different times throughout the year. Due to these reasons, this study is conducted at the province-level to reveal key factors that control influenza outbreaks locally that otherwise could be misleading in a larger scale study, such as at the national level. The ability to accurately predict future outbreaks in the province-level can also be helpful for government agencies and policy makers. Healthcare personnel and additional budget can be allocated in advance to prepare and quarantine the outbreaks.

2. MATERIALS AND METHODS

2.1 Influenza Data

We collect weekly reported influenza cases from January 2009 to December 2014 of each province of Thailand from the Ministry of Public Health.

2.2 Weather Data

The weather parameters used in this study are relative humidity, maximum temperature, minimum temperature, and precipitation. The daily weather data of each province is obtained from the Thai Meteorological Department. We average seven daily weather parameters, starting from Monday of each week, to obtain weekly climate data.

2.3 Statistical Analysis

We employ the Random Forest technique in analyzing the link between four meteorological parameters and influenza reported cases. The Random Forest is an ensemble learning technique where a prediction model is a large collection of decision trees. Each decision tree is constructed from a subset of the training set and a randomly selected set of variables. A prediction is conducted by averaging the predicted values from all decision trees. Random Forest models also provide the "importance" of each input variable. The importance of each variable illustrates the percent increase in the mean square error when the value of the variable is randomly changed. This importance indicates how the variables affect the outcome. The higher the magnitude of the importance, the more influence a variable has.

To predict the number of influenza cases of a considered week, we use 14 potential inputs of the Random Forest model which are relative humidity (Rel. Humid) , maximum temperature (TMax), minimum temperature (TMin), precipitation of the considered week (Prcp), relative humidity (Rel. Humid lag 1), maximum temperature (TMax lag 1), minimum temperature (TMin lag 1), precipitation of the prior week to the considered week (Prcp lag 1), relative humidity

(Rel. Humid lag 2), maximum temperature (TMax lag 2), minimum temperature (TMin lag 2), precipitation of the prior two weeks to the considered week (Prcp lag 2) and the number of reported influenza cases one (#Cases lag 1) and two weeks (#Cases lag 2) prior to the considered week. Since not all 14 potential variables are relevant to the number of influenza cases, we conduct an exhaustive search to find the subset of the potential variables that results in the best-fit prediction model. A Random Forest model is built for each possible subset of the 14 potential variables. Therefore, 2^{14} (16,384) models are constructed. We select the model with the highest correlation coefficient as a prediction model for a province. The Random Forest models in this work are constructed using R, a software environment for statistical computing, with the randomForest package [10].

3. Results and Discussion

The main results of this study are shown in Table 1, and Table 2. Table 1 shows the performance of the best prediction model and the importance of each variable as defined in [8]. Table 2 shows the percentage change in the number of considered week's influenza cases when a variable increases from its median value by 10% of its range.

The results show that, in all provinces, the key factor that dictates influenza cases in the considered week is the number of influenza cases (#Cases lag 1) in the previous week with the highest importance. This finding supports the traditional hypothesis of the SIR dynamic framework [9] that the number of infected cases depends on the previous number of infected cases. However, all best-fit prediction models of selected provinces contain at least one weather variable. This implies that the SIR model alone cannot describe influenza cases in Thailand. Our results support the popular hypothesis [1-6] that influenza activity is associated with weather variations.

In all selected provinces, temperatures, both maximum and minimum, are considered as factors in the change of influenza cases. However, the impacts of temperature variations are location-dependent. For example, the number of influenza cases increases by 6.25% (95% confidence interval (CI) 2.09,10.41) if the minimum temperature of the considered week increases by 10% in Ubon Rathcathani (non-coastal). Whereas in Phuket (coastal), the number of influenza cases decreases by 7.77% (or increases by -7.77, 95% confidence interval (CI): -12.25, -3.29). Unlike temperate regions, our findings suggest that the change in temperature does play a role in the change of influenza activity, however its impacts are location-dependent.

Surprisingly, the change in relative humidity does not play a significant role in the change of influenza activity. Relative humidity variables only appear in Ubon Rathchathani and Phuket. In both provinces, if the relative humidity increases by 10%, then number of predicted influenza cases increases. The findings support previous studies that the relative humidity has a positive association with influenza activity in tropical regions [6].

The statistical analysis reveals that at the province-level, different sets of weather variables have impacts on influenza cases. This suggests that a larger scale study on the association of weather conditions and influenza activity, such as at the national level, may overlook and oversimplify impacts of local weather variations. We also discovered that weather conditions up to the prior two weeks may have some influences on the influenza activity.

Table 1. Random forest variable importance of best-fit models of selected provinces.

Variable	Bangkok R^2=0.89 MSE=69.59	Chiang Mai R^2=0.91 MSE=32.29	Ubon Ratchathani R^2=0.65 MSE=11.07	Chonburi R^2=0.55 MSE=6.36	Phuket R^2=0.76 MSE=8.95
Rel. Humid		-0.94	11.84		
TMax					
TMin			6.26		2.65
Prcp				-3.55	
#Cases lag 1	70.61	71	55.96	76.29	47.11
Rel. Humid lag 1					1.09
TMax lag 1	3.01	6.47		0.94	
TMin lag 1	12.69	9.83			
Prcp lag 1					
#Cases lag 2	13.92	17.6	4.11	4.49	2.91
Rel. Humid lag 2					2.73
TMax lag 2		8.16	4.07		-0.73
TMin lag 2		7.44			3.52
Prcp lag 2					

Table 2. Percentage change in the number of influenza cases with 95% confidence interval (CI) when a variable increases from its median value by 10% of its range.

Variable	Bangkok %Chg (95%CI)	Chiang Mai %Chg (95%CI)	Ubon Ratchathani %Chg (95%CI)	Chonburi %Chg (95%CI)	Phuket %Chg (95%CI)
Rel. Humid		1.15 (-2.37,4.66)	22.06 (16.25,27.86)		
TMax					
TMin			6.25 (2.09,10.41)		-7.77 (-12.25,-3.29)
Prcp				-30.44 (-32.86,-28.03)	
#Cases lag 1	122.95 (117.66,128.24)	252.03 (229,81,274.24)	251.19 (230.74,271.64)	136.86 (130.47,143.25)	115.25 (109,15,121.35)
Rel. Humid lag 1					4.88 (0.42,9.35)
TMax lag 1	-2.06 (-4.27,0.15)	-6.67 (-10.56,-2.79)		-6.6 (-10.52,-2.86)	
TMin lag 1	2.61 (0.8,4.42)	17.08 (13, 21.17)			
Prcp lag 1					
#Cases lag 2	-26.85 (-28.75,24.95)	12.48 (8.82,16.15)	79.81 (74.13,85.48)	-5.57 (-10,-1.13)	18.95 (14.65,23.26)
Rel. Humid lag 2					7.18 (2.23,12.13)
TMax lag 2		-2.18 (-5.9,1.54)	2.6 (-0.67,5.88)		-11.28 (-14.97,-7.59)
TMin lag 2		-3.76 (-7.32,-0.2)			1.17 (-3.37,5.71)
Prcp lag 2					

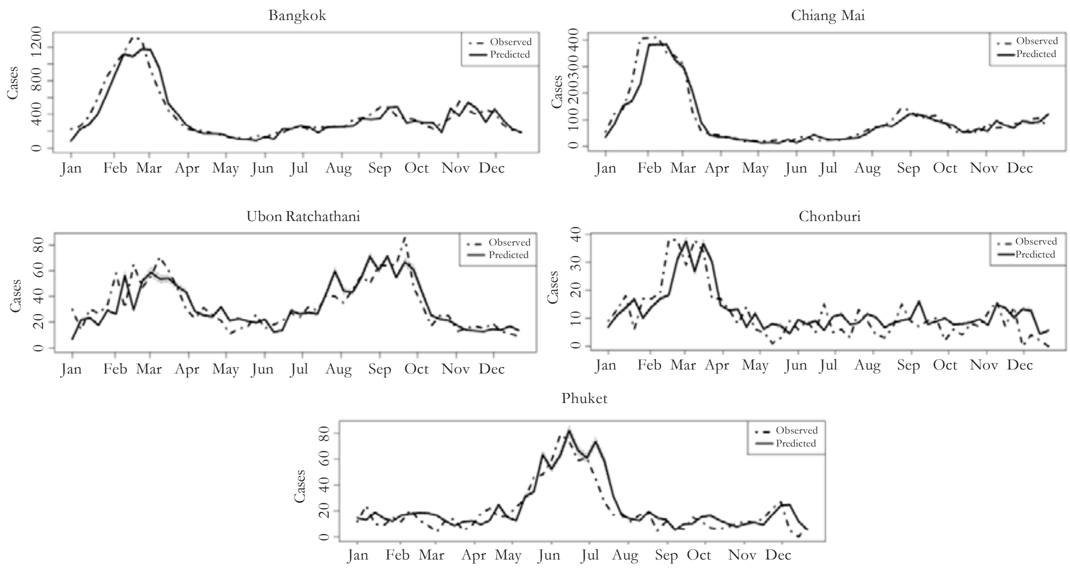

Figure 1. Predicted and observed influenza cases in 2014 of provinces in Thailand. The dotted line is the observation and the black line is the predicted cases with the shaded areas are the 95%CI.

4. CONCLUSION

This work studies the link between influenza activity and weather variations in provincial level in Thailand using random forest technique. The results suggest that a small-scale study of links between influenza activity and weather location, like province-level, is more suitable in tropical region. Different sets of weather variations are associated with influenza activity differently in different provinces. Weather variations and previous cases up to two previous weeks may have impacted influenza cases. The most important factor is the number of cases from the previous week. But weather conditions also have influences on all locations. A change in temperatures plays a notable role in influenza activity, however, its impacts are location-dependent. Relative humidity plays a less significant role and has a positive association with influenza activity.

ACKNOWLEDGEMENT

The corresponding authors gracefully acknowledge the financial support provided by Thammasat University Research Fund under the TU Research Scholar, Contract No. 2/22/2559. The first author is supported by the Junior Science Talent Project, NSTDA, Thailand.

REFERENCES

[1] Tamerius J., Nelson M.I., Zhou S.Z., Viboud C., Miller M.A. and Alonso W.J., *Environ. Health Persp.*, 2010; 439-445. DOI 10.1289/ehp.1002383. PMID: 21097384.

[2] Yaari R., Katriel G., Huppert A., Axelsen J.B. and Stone L., *J. R. Soc. Interface*, 2013; **8**: 856-867. DOI 10.1098/rsif.2010.0515.

[3] Chittaganpitch M., Supawat K., Olsen S.J., Waicharoen S., Patthamadilok S., Yingyong T., Brammer L., Epperson S.P., Akrasewi P. and Sawanpanyalert P.,

Influenza Other Resp. Viruses., 2012; **6**: 27683.

[4] Moura F.E., *Curr. Opin. Infect. Dis*., 2010; **23**: 41520.

[5] Yang L., Ma S., Chen P.Y., HeJ.F., Chan K.P., Chow A., Ou C.Q., Deng A.P., Hedley A.J., Wong C.M. and Peiris J.S., *Vaccine*, 2011; **29(48)**: 8909-8914.

[6] Imai C., Brooks W.A., Chung Y., Goswami D., Anjali B.A., Dewan A., Kim H. and Hashizume M., *Glob. Health Action*, 2014; **7(1)**: 24413.

[7] Kane M., Price N., Scotch M. and Robinowitz P., *BMC Bioinformatics*, 2014; **15**: 276.

[8] Breiman L., *Mach. Learn*., 2001; **45**: 5-32.

[9] Kermack W.O. and McKendrick J.D., *P. R. Soc. London*, 1927; **A115**: 700-721.

[10] Liaw A. and Wiener M., *R News*, 2002; **2(3)**: 18-22.