http://dx.doi.org/10.1016/j.phrp.2014.04.004
pISSN 2210-9099 eISSN 2233-6052

ORIGINAL ARTICLE

Prediction Forecast for *Culex tritaeniorhynchus* Populations in Korea

Nam-Hyun Kim [a,1], Wook-Gyo Lee [b,1], E-Hyun Shin [b], Jong Yul Roh [b], Hae-Chun Rhee [a], Mi Yeoun Park [b,*]

[a] *School of Economics, Sungkyunkwan University, Seoul, Korea.*
[b] *Division of Medical Entomology, Korea National Institute of Health, Osong, Korea.*

Received: April 1, 2014
Revised: April 17, 2014
Accepted: April 27, 2014

KEYWORDS:
Culex tritaeniorhynchus,
Japanese encephalitis,
out-of-sample prediction

Abstract

Objectives: Japanese encephalitis is considered as a secondary legal infectious disease in Korea and is transmitted by mosquitoes in the summer season. The purpose of this study was to predict the ratio of *Culex tritaeniorhynchus* to all the species of mosquitoes present in the study regions.

Methods: From 1999 to 2012, black light traps were installed in 10 regions in Korea (Busan, Gyeonggi, Gangwon, Chungbuk, Chungnam, Jeonbuk, Jeonnam, Gyeongbuk, Gyeongnam, and Jeju) to capture mosquitoes for identification and classification under a dissecting microscope. The number of mosquitoes captured/week was used to calculate its daily occurrence (mosquitoes/trap/night). To predict the characteristics of the mosquito population, an autoregressive model of order p (AR(p)) was used to execute the out-of-sample prediction and the in-sample estimation after presumption.

Results: Compared with the out-of-sample method, the sample-weighted regression method's case was relatively superior for prediction, and this method predicted a decrease in the frequency of *Cx. tritaeniorhynchus* for 2013. However, the actual frequency of this species showed an increase in frequency. By contrast, the frequency rate of all the mosquitoes including *Cx. tritaeniorhynchus* gradually decreased.

Conclusion: The number of patients with Japanese encephalitis has been strongly associated with the occurrence and density of vector mosquitoes, and the importance of this infectious disease has been highlighted since 2010. The 2013 prediction indicated an increase after an initial decrease, although the ratio of the two mosquito species decreased. The increase in vector density may be due to changes in temperature and the environment. Thus, continuous prevalence prediction is warranted.

---

*Corresponding author.
E-mail: miyeoun@korea.kr

[1] N.-H.K. and W.-G.L are co-first authors of the paper.

1. Introduction

Japanese encephalitis is considered as a secondary legal infectious disease in Korea and is one of the main mosquito-borne infectious diseases of the summer season. *Culex tritaeniorhynchus*, which transmits Japanese encephalitis, is dispersed not only in Korea, but also in other areas such as Japan, China, Southeast Asia, India, and Pakistan. This major mosquito species infects approximately 68,000 individuals each year, resulting in approximately 20,000 deaths annually [1—3].

Since the first reported incidence of the disease in Korea from the U.S. forces stationed in the Incheon area in 1946 [4,5], the incidence of Japanese encephalitis has significantly increased since 1949, affecting at least 5616 people, and resulting in 2797 deaths [6,7]. Moreover, 1000—3000 individuals were infected with the disease each year until the 1960s. The incidence of Japanese encephalitis significantly decreased in the 1970s compared with the 1960s. From 1984 to 2009, this infectious disease was almost eradicated, with less than 10 cases reported every year. However, 28 cases were reported in 2010, with a possibility of an increase in vector mosquito density due to changes in temperature and the environment. Thus, continuous prevalence prediction is warranted.

Seasonal identification is very important in managing mosquitoes [8]. It has been reported that the rapid decrease in the incidence of Japanese encephalitis between 1960 and 1970 was due to the decrease in the density of the vector mosquito [9—11], which is similar to that observed in Japan [12,13].

In this research, black light traps were installed in 10 regions of Korea (Busan, Gyeonggi, Gangwon, Chungbuk, Chungnam, Jeonbuk, Jeonnam, Gyeongbuk, Gyeongnam, and Jeju) for the last 12 years from 1999 to 2012 and mosquito data were collected. Using the data collected, a simple AR(p) model was used to estimate and predict the ratio of Japanese encephalitis vector mosquitoes. Thus, this research was conducted to predict mosquito occurrence in order to control the incidence of Japanese encephalitis.

2. Materials and methods

2.1. Data

Data for this investigation were directly acquired by the National Institutes of Health from the Public Health and Environment Research Institute of 10 regions in Korea (Busan, Gyeonggi, Gangwon, Chungbuk, Chungnam, Jeonbuk, Jeonnam, Gyeongbuk, Gyeongnam, and Jeju), two times a week from May to October (data collection period: 1999—2012). In addition, this investigation used the mosquito occurrence density data of Japanese encephalitis prediction programs of the last 14 years using the mosquito classification key of the regional health centers.

2.2. Collection region and equipment

Cowsheds have been identified as the main region of vector mosquito occurrence in all the 10 Korean regions. A black light trap, which is commonly used for mosquito density studies, was installed at a height of 1.5—1.8 m within the cowshed. The light traps were operated two times a week from 19:00 PM to 06:00 AM the following day [14]. The mosquitoes collected in the trap were carefully transported to the laboratory. Then, the mosquitoes were placed in a plastic bag with a cotton ball of ether or chloroform. Next, the plastic bag was completely sealed or kept in the freezer for at least 2 hours. After killing, the mosquitoes were identified and classified by observing them under a dissection microscope. Based on the number of mosquitoes collected, the daily average density of mosquitoes was calculated (i.e., mosquitoes/trap/night).

2.3. Preliminary data analysis

Figure 1 shows the distribution of *Cx. tritaeniorhynchus* and all other mosquitoes in Korea by week from 1999 to 2012. The population of all mosquitoes including *Cx. tritaeniorhynchus* changed at 2—3-year intervals. Moreover, after 2010, the *Cx. tritaeniorhynchus* population decreased compared with the population of all other species of mosquitoes. These changes might have been caused by an increase in its natural enemies or temperature, although this analysis did not include the analysis of factors affecting mosquito density and mainly focused on predicting the population dynamics of mosquitoes.

2.3.1. Unit root test

The time series using the unit root test is relatively unstable. Therefore, this may cause problems of spurious regression, especially when using general regression analysis. Thus, the variables and mosquito data used in this study were executed using a unit root test to verify whether these could be considered as stable time-series data. The augmented Dickey—Fuller test [15] and Phillips—Perron test [16] were used for the unit root test.

Table 1 shows the results of the unit root test using the mosquitoes collected/week. In the table, "none" indicates the absence of a constant term and trend; "intercept" indicates a constant term; and both constant term and trend are considered as "trend." The null hypothesis that the unit root exists in the rates of all mosquitoes and *Cx. tritaeniorhynchus* was rejected and the unstable-level variables were used in the analysis.

2.3.2. Summary statistics

Table 2 shows the results of the statistical analysis. The ratio of all mosquitoes to *Cx. tritaeniorhynchus* was positive (+). The density of all mosquitoes was highest in 2003 (87,194 mosquitoes). The density of

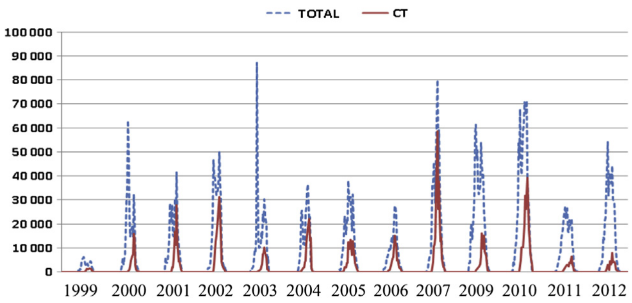

Figure 1. Distribution of mosquitoes and *Culex tritaeniorhynchus* (CT) per year.

Table 1. Unit root test (level variables).

	ADF			PP		
Weekly data	None	Intercept	Trend	None	Intercept	Trend
All mosquitoes	−5.423*	−6.190*	−6.246*	−5.655*	−6.672*	−6.751*
Culex tritaeniorhynchus	−8.786*	−9.529*	−9.549*	−6.862*	−7.708*	−7.720*
Ratio	−6.177*	−7.340*	−7.336*	−5.771*	−6.708*	−6.705*

*Significant at the 1% level. ADF = augmented Dickey−Fuller test; PP = Phillips−Perron test.

Cx. tritaeniorhynchus was highest in 2007 (58,769 mosquitoes), which was the highest value ever recorded.

2.3.3. Lag selection

Before the prediction using the AR(*p*) model, the Akaike information criterion and the Schwarz information criterion tests were used to determine the proper time deviation *p*. Table 3 shows the density of all mosquitoes 1 week before the analysis, as well as the *Cx. tritaeniorhynchus* density and ratio variable 4 weeks before the analysis. The autocorrelation function was used to validate the time difference, which was high in the 1-year unit (Figure 2) and was maintained for 4 years. Thus, the data from the same period of last year to the same period 4 years ago were used to predict the AR figure.

2.4. Methods

The AR(*p*) model used in this research is presented in Eq. (1). Each *p* predicts the same period of last year's AR(*p*) model.

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + u_t,\ u_t \sim N\left(0,\ \sigma_u^2\right) \tag{1}$$

AR(*p*) model was used to execute the in-sample and out-of-sample analyses. For the in-sample prediction, *β*, which is the estimation gained from the prediction of the total period, was used for the in-sample estimation, as shown in Eq. (2).

$$\tilde{y}_t = \beta_1 \hat{y}_{t-1} + \beta_2 \hat{y}_{t-2} + \cdots + \beta_p \hat{y}_{t-p} \tag{2}$$

For the out-of-sample prediction, two methods were used, namely, rolling regression (RO) and adding regression (AD). First, the RO is a prediction method while moving a certain number of samples. In cases of insufficient data, it is not advisable to use the prediction method. Second, the AD method executes the out-of-sample prediction by accumulating the samples. The out-of-sample prediction in this study initially predicts mosquito density until 2007. Then, the sample was moved by 1-year units for prediction analysis. The *t* data were used for the out-of-sample prediction analysis and the $\tilde{y}_{t+1}$ of $t + 1$ is the same as Eq. (3).

Table 2. Basic statistical data on mosquito density.

Variables	Mean	Maximum value	Minimum value	Standard deviation
All mosquitoes	7846.3(562.9)*	87,194.0	0.0	14,635.1
Culex tritaeniorhynchus	1991.9(217.0)*	58,769.0	0.0	5641.4
Ratio	10.2(0.7)*	73.4	0.0	17.7

The number in parenthesis represent standard deviation. * Significant at the 1% level.

Table 3. Time lag test.

Variables	All mosquitoes		*Culex tritaeniorhynchus*		Ratio of *Culex tritaeniorhynchus* to all mosquitoes	
Test method	AIC	SIC	AIC	SIC	AIC	SIC
Time lag	10	1	4	4	9	4

AIC = Akaike information criterion; SIC = Schwarz information criteria.

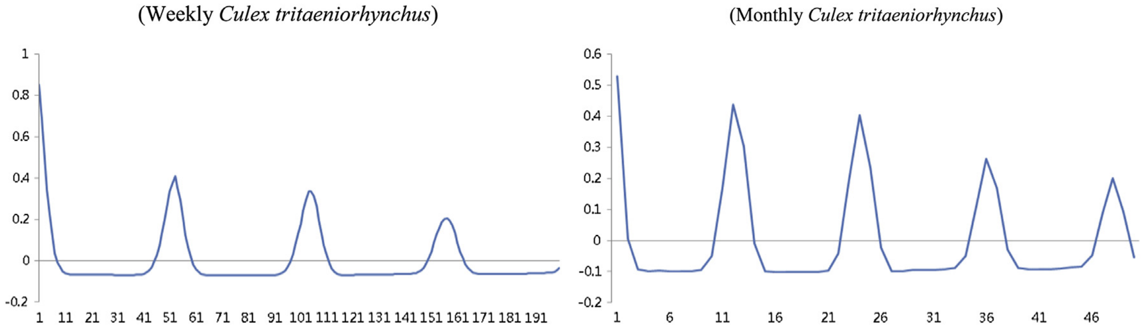
(Weekly *Culex tritaeniorhynchus*) (Monthly *Culex tritaeniorhynchus*)

Figure 2. Autocorrelation.

Table 4. Autoregressive model estimation (weekly data).

Variable	Last year	2 years ago	3 years ago	4 years ago	R^2
All mosquitoes	0.435(0.044)**	0.269(0.054)**	−0.024(0.056)	0.310(0.055)**	0.637
Japanese encephalitis	0.138(0.043)**	0.353(0.043)**	0.189(0.050)**	0.068(0.049)	0.394
Ratio of Japanese encephalitis	0.162(0.044)**	0.381(0.041)**	0.381(0.044)**	−0.085(0.045)***	0.717

The results of the autoregressive model showed no differences in superiority with that of autoregressive−moving-average model. The results of the weekly data were similar to that of the monthly data. Thus, the weekly data, which consists of more data, is described. * Significant at $p < 0.05$. ** Significant at $p < 0.01$. *** Significant at $p < 0.01$.

$$\tilde{y}_{t+1} = \beta_1\hat{y}_t + \beta_2\hat{y}_{t-1} + \cdots + \beta_p\hat{y}_{t-p-1} \quad (3)$$

After executing an in-sample estimation and an out-of-sample prediction, the mean-square prediction errors (MSPEs) were calculated to select the model that shows superior prediction results. MSPE pertains to the average of the square of the difference between the actual value and the estimation. A smaller MSPE value indicates a relatively superior prediction.

3. Results

Table 4 shows the prediction results of the AR model using the same period of last year. Similar to most cases, the AR model generates a significant value.

Mean square error (MSE) was calculated to compare the estimation of the model according to the variable using the predicted results. Table 5 shows the MSE value when predicting the in-sample model described in Table 4.

Table 6 shows the MSPE calculated during the out-of-sample prediction of the model as shown in Table 4. Two methods can be used for the out-of-sample prediction as described earlier. First, the RO is a method for prediction while moving the interval number of the sample. When data are insufficient, an inaccurate prediction might be generated. Second, the AD method executes the out-of-sample prediction by accumulating the sample. The out-of-sample prediction here initially predicts up to 2007. The sample was then moved by 1-year units for prediction. In this research, two cases were analyzed. However, Table 6 shows that the MSPE generated using the AD method was smaller when the sample size was smaller than the population of mosquitoes.

Figure 3 shows the out-of-sample results used in the RO analysis. For 2013−2016, past estimations were used to identify the predicted values. First, the ratio of all mosquitoes and *Cx. tritaeniorhynchus* gradually increased from 2013 with regular changes. For *Cx. tritaeniorhynchus*, the term of the change was shorter

Table 5. MSE of in-sample.

	In-sample prediction		
	All mosquitoes	*Culex tritaeniorhynchus*	Ratio
MSE	8.881×10^7	1.851×10^7	84.823

MSE = mean square error.

Table 6. MSPE of out-of-sample prediction.

	Out-of-sample prediction		
	All mosquitoes	*Culex tritaeniorhynchus*	Ratio of *Culex tritaeniorhynchus*
MSPE (RO)	2.045×10^8	8.852×10^7	86.069
MSPE (AD)	5.229×10^7	3.476×10^7	78.351

AD = adding regression; MSPE = mean-square prediction error; RO = rolling regression.

compared with all mosquitoes. However, using the RO method, the accuracy of the value that barely appeared in *Cx. tritaeniorhynchus* in 2013 was lower compared with that using the AD method.

Figure 4 shows the results of the out-of-sample prediction using the AD method. The difference in the estimation and the actual value between 2009 and 2012 was low compared with that shown in Figure 3. In addition, when looking at the estimations from 2013 to 2016, 2013 showed a decrease in the density of *Cx. tritaeniorhynchus*, and an increase was detected a year later. By contrast, the ratio of all mosquitoes to *Cx. tritaeniorhynchus* gradually decreased.

4. Discussion

In this research, the data on mosquito density for the Japanese encephalitis prediction program acquired from the Public Health and Environment Research Institute of 10 regions in Korea from May to October of 1999 to 2012 were used in the AR(p) model. The MSPEs of the in-sample and out-of-sample predictions were compared. The relatively superior model was used to predict the future mosquito populations.

The MSPE value was low when the AR method was used. When prediction was executed using the AR method, the mosquito population again showed an increase and a decrease in a certain term and interval. For the estimations of 2013–2016 using estimations of up to 2012, the density of *Cx. tritaeniorhynchus* initially decreased in 2013, and then increased. By contrast, the ratio of all mosquitoes to *Cx. tritaeniorhynchus* showed a gradual decrease.

Not only *Cx. tritaeniorhynchus* but all mosquitoes are influenced by factors related to their habitat such as number of disinfection events, temperature, and rainfall as related to humidity. If these factors are appropriately controlled, more superior results could be acquired for the prediction of *Cx. tritaeniorhynchus* and all other

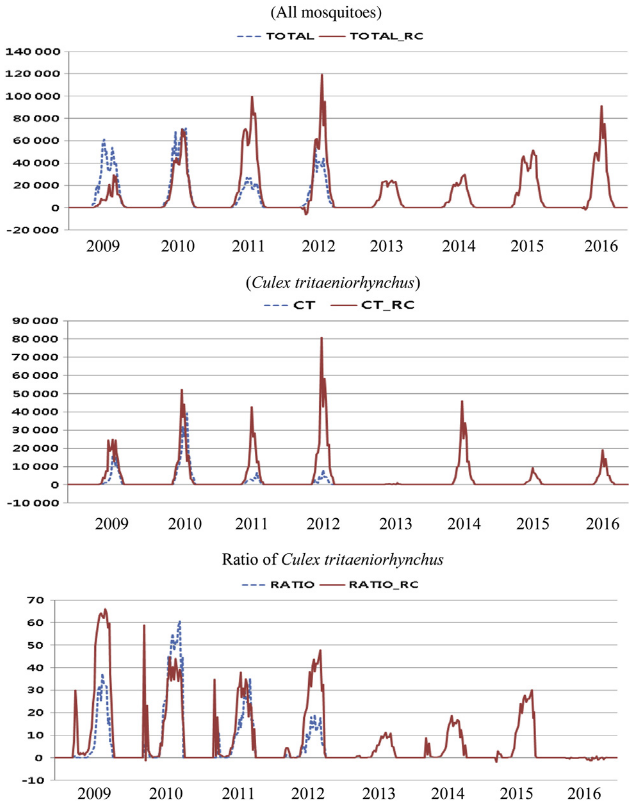

Figure 3. Out-of-sample results used in rolling window regression (weekly data). CT, *Culex tritaeniorhynchus*. RC: Rolling window regression of CT.

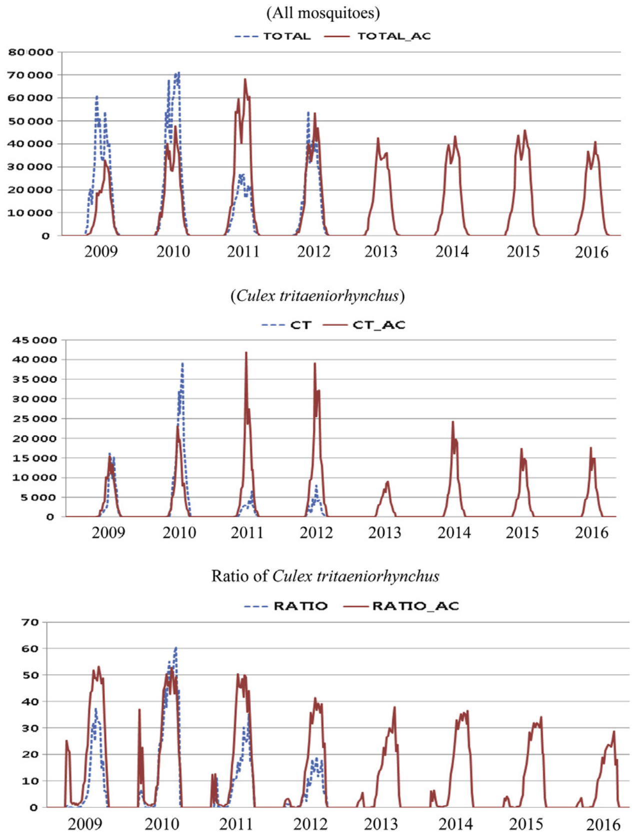

Figure 4. Out-of-sample results used in adding window regression (weekly data). CT, *Culex tritaeniorhynchus*. AC: Adding window regression of CT.

mosquitoes. However, when predicting with methodologies such as the Kalman filter to control these factors, this methodology could also be confusing. In addition, there is a possibility that the result will not be exact due to the lack of data or other factors that cannot be observed. Thus, based on the properties of the data on *Cx. tritaeniorhynchus* and all mosquitoes, the ratios of all mosquitoes to *Cx. tritaeniorhynchus* were predicted through a relatively simple AR(p) model. Furthermore, the *Cx. tritaeniorhynchus* population initially decreased and subsequently increased, although the density of *Cx. tritaeniorhynchus* decreased compared with that of all mosquitoes.

Conflicts of interest

All contributing authors declare no conflicts of interest.

Acknowledgments

This work was supported by grants from the National Vector Control and Surveillance work carried out by the Korean National Institute of Health (No. 4836-303-210-13) and a grant from the National Research Foundation of Korea funded by the Korean Government (Grant No. NRF-2011-358-B00007).

References

1. Rosen L. The natural history of Japanese encephalitis virus. Annu Rev Microbiol 1986;40:395−414.
2. Igarashi A. Japanese encephalitis virus infection, and control. In: Kurstak E, editor. Control of virus disease. 2nd ed. New York: Marcel Dekker, Inc; 1992. p. 309−42.
3. Centers for Disease Control and Prevention (CDC). Japanese encephalitis surveillance and immunization—Asia and the Western Pacific, 2012. MMWR Morb Mortal Wkly Rep 2013 Aug;62(33):658−62.
4. Sabin AB. Epidemiological studies on Japanese B encephalitis virus. Am J Hyg 1947;46:356.
5. Ueba NA, Meada KB, Mitsuda B, et al. Natural infections of swine by Japanese encephalitis virus and its protection with vaccine. Proc Osaka Prefect Inst Public Health Ed Environ Health 1968;6:56.
6. Lee JW, Kim KH, Kim ID. Mechanical and serological investigation research on Korea's epidemic encephalitis. Korean J Prev Med 1974;7(2):403−15.
7. Hong HK. The ecological investigation of *Culex tritaeniorhynchus*, Japanese encephalitis vector mosquitoes. Korean J Epidemiol 1983;5(1):29−40.
8. Jeong YS, Lee DK. Prevalence and seasonal abundance of the dominant mosquito species in a large marsh near Coast of Ulsan. Kor J Appl Entomol 2003;42(2):125−32.
9. Kim KH. Epidemiological features of Japanese encephalitis in the Republic of Korea. Korean J Virol 1979;9(1):43−55.
10. Shim JC, Kim CL, Lee WJ, et al. Population density investigation of *Culex tritaeniorhynchus* and forecasting the occurrence of Japanese encephalitis. Korean J Entomol 1990;20(4):213−22.
11. Shim JC, Yun YH, Lee WJ, et al. The population density investigation of Korean *Culex tritaeniorhynchus*. Rep NIH Korea 1990; 27(1):165−72.

12. Wada Y, Oda T, Mogi M, et al. Ecology of Japanese encephalitis virus in Japan. II. The population of vector mosquitoes and the epidemic of Japanese encephalitis. Trop Med 1975;17(3):111—27.
13. Matsuzaki S. Population dynamics of *Culex tritaeniorhynchus* in relation to the epidemics of Japanese encephalitis in Kochi Prefecture, Japan. Jpn J Sanit Zool 1990;41(3):247—55.
14. Saito K, Fujita L. A brief note on mosquitoes collected with a light-trap in the city of Tokyo, 1966. Jap J Sanit Zool 1967;18(4):304.
15. Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc 1979;74(366):427—31.
16. Phillips P, Perron P. Testing for a unit root in time series regression. Biometrika 1988;75(2):335—46.