Modeling the distribution of *Culex tritaeniorhynchus* to predict Japanese encephalitis distribution in the Republic of Korea

Penny Masuoka[1], Terry A. Klein[2], Heung-Chul Kim[3], David M. Claborn[4], Nicole Achee[1], Richard Andre[1], Judith Chamberlin[1], Jennifer Small[5], Assaf Anyamba[5], Dong-Kyu Lee[6], Suk H. Yi[2], Michael Sardelis[7], Young-Ran Ju[8], John Grieco[1]

[1]*Department of Preventive Medicine and Biometrics, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814, USA;* [2]*65th Medical Brigade/US Army MEDDAC-Korea, Unit 15281, APO AP 96205-5281, USA;* [3]*5th Medical Detachment, 168th Multifunctional Medical Battalion, 65th Medical Brigade, Unit 15247, APO AP 96205-5247, USA;* [4]*Center for Homeland Security, Missouri State University, 901 South National Avenue, Springfield, MO 65897, USA;* [5]*Biospheric Sciences Branch, NASA Goddard Space Flight Center, Code 614.4, Greenbelt, MD 20771, USA;* [6]*Department of Health and Environment, Kosin University, Busan, 606-701, Republic of Korea;* [7]*National Center for Medical Intelligence, 1607 Porter Street, Fort Detrick, MD 21702, USA;* [8]*Department of Arboviruses, Center for Immunology and Pathology, Korea National Institute of Health, 122-701, Republic of Korea*

Abstract. Over 35,000 cases of Japanese encephalitis (JE) are reported worldwide each year. *Culex tritaeniorhynchus* is the primary vector of the JE virus, while wading birds are natural reservoirs and swine amplifying hosts. As part of a JE risk analysis, the ecological niche modeling programme, Maxent, was used to develop a predictive model for the distribution of *Cx. tritaeniorhynchus* in the Republic of Korea, using mosquito collection data, temperature, precipitation, elevation, land cover and the normalized difference vegetation index (NDVI). The resulting probability maps from the model were consistent with the known environmental limitations of the mosquito with low probabilities predicted for forest covered mountains. July minimum temperature and land cover were the most important variables in the model. Elevation, summer NDVI (July-September), precipitation in July, summer minimum temperature (May-August) and maximum temperature for fall and winter months also contributed to the model. Comparison of the *Cx. tritaeniorhynchus* model to the distribution of JE cases in the Republic of Korea from 2001 to 2009 showed that cases among a highly vaccinated Korean population were located in high-probability areas for *Cx. tritaeniorhynchus*. No recent JE cases were reported from the eastern coastline, where higher probabilities of mosquitoes were predicted, but where only small numbers of pigs are raised. The geographical distribution of reported JE cases corresponded closely with the predicted high-probability areas for *Cx. tritaeniorhynchus*, making the map a useful tool for health risk analysis that could be used for planning preventive public health measures.

Keywords: *Culex tritaeniorhynchus*, geographical distribution, ecological niche modeling, Japanese encephalitis virus, Republic of Korea.

Corresponding author:
Penny Masuoka
Department of Preventive Medicine and Biometrics
Uniformed Services University of the Health Sciences
4301 Jones Bridge Road
Bethesda, MD 20814, USA
Tel. +1 301 295 3732; Fax +1 301 295 1971
E-mail: pmasuoka@usuhs.mil

Introduction

Culex tritaeniorhynchus Giles is the primary vector of Japanese encephalitis (JE) virus, and JE is distributed in parts of Asia and the Pacific (Straus and Straus, 2002; Erlanger et al., 2009). Worldwide, more than 35,000 cases of JE, with approximately

10,000 deaths, are reported annually. Although only about 1 per 200 infections with JE are symptomatic, many of those who survive suffer mild to severe long-term neurological sequelae.

While JE virus has been detected in several species of mosquitoes, the primary vector is *Cx. tritaeniorhynchu*s, which prefers to feed on water birds, pigs and cattle. The primary natural reservoirs for JE are large water birds associated with streams, ditches, canals and rice paddies. Swine, when in proximity to large water birds, serve as amplifying hosts, which results in high infection rates in the mosquito populations.

A national vaccination programme, initiated in 1967, greatly reduced the incidence of JE cases in the Republic of Korea (ROK) over the last two decades from thousands of cases (e.g. 6,897 in 1958) to 0-7 cases annually (K-CDC, 2008). Additionally, the K-CDC vector surveillance programme alerts the public when high populations of *Cx. tritaeniorhynchus* occur and advises the use of personal protective measures against mosquito bites. As a result of the vaccination and vector surveillance programmes, transmission and disease risk among unvaccinated populations (e.g. US military and travelers) is unknown. While the incidence of JE remains low in the ROK, JE cases are on the increase in other countries (e.g. Bangladesh, Cambodia, India and Lao People's Democratic Republic) which lack comprehensive vaccination and vector control programmes (Erlanger et al., 2009).

Ecological niche modeling programmes have been used to predict the distribution of vector populations (e.g. Peterson and Shaw, 2003; Moffett et al., 2007; Foley et al., 2009). These modeling programs use known occurrence locations of the species and environmental data in a raster format to infer environmental requirements of the species, and produce output maps predicting species distributions. In our current research effort, we used the Maxent programme (Phillips et al., 2006) to model the distribution of *Cx. tritaeniorhynchus* in the ROK.

Initial *Cx. tritaeniorhynchus* modeling efforts in the ROK, used 27 adult mosquito collection sites that were part of a routine surveillance programme on US military bases (Masuoka et al., 2009). In this earlier study, several problems were noted in the use of these data for modeling. The sampling areas were restricted primarily to urban land cover at lower elevations (<310 m). To improve environmental sampling in the current study, additional mosquito collection sites were incorporated. The accuracy and contribution of the environmental input variables to the resulting model were examined. The output prediction map of *Cx. tritaeniorhynchus* distribution was compared with reported JE case locations from 2001 to 2009, swine density and human population density in the ROK.

Understanding the geographical distribution of *Cx. tritaeniorhynchus* can lead to targeted preventive measures against the transmission of JE virus to humans, e.g. vaccination in high risk areas, mosquito control and use of effective insect repellents. Because mosquito distributions are limited by environmental factors (e.g. temperature, precipitation, elevation, etc.), these variables can be used to develop mosquito distribution models.

Materials and methods

Mosquito data

Mosquito data were collected by the U.S. Army's 65th Medical Brigade (Yongsan Army Garrison, ROK) (Kim et al., 2004, 2005). Adult and larval mosquito data came from collections made from May through October in the period 2002-2009 at US military installations, training sites, and other non-military areas of the ROK. Although the mosquitoes were collected by a variety of methods (i.e. dipping and multiple light-trap types), the data were used only to demonstrate the presence of *Cx. tritaeniorhynchus* for the purpose of modeling. Each positive collection site was assigned a value of one mosquito, regardless of the number of mosquitoes collected at that site. Using this strategy, a total of 96 unique col-

lection locations (presence records) for *Cx. tritae-niorhynchus* were obtained.

To study the seasonal abundance of adult mosquitoes in the ROK, data from bimonthly adult mosquito collections, using black-light traps model "Black Hole" by BioTrap (http://www.bio-trap.com), were examined for seven locations in agricultural settings during the summer of 2009. The numbers of *Cx. tritaeniorhynchus* mosquitoes collected bimonthly were totaled and mapped to show their distribution. The seven collection locations were included as part of the 96 presence records for the model.

Environmental data

Environmental raster data for this project were obtained from several sources. Each raster layer was resampled to 1-km pixels, georeferenced and subset to the ROK study site.

Climate data for the model were obtained from WorldClim, version 1.4 (http://www.worldclim.org). Gridded WorldClim precipitation and temperature data are based on monthly ground weather station measurements and averaged from years 1950 to 2000 for each month (Hijmans, 2005).

Elevation data were also obtained from WorldClim. The programme processed the Shuttle Radar Topography Mission (SRTM) data obtained from NASA (http://www2.jpl.nasa.gov/srtm) to the same projection and scale as the other WorldClim layers.

Land cover data, derived from Moderate Resolution Imaging Spectroradiometer (MODIS), were obtained from Boston University (http://www.modis.bu.edu/landcover). The data are in the International Geosphere Biosphere Programme (IGBP) classification scheme consisting of 17 land cover classes.

Rice fields serve as an important larval habitat for *Cx. tritaeniorhynchus*. Unfortunately, the IGBP classification does not have a separate class for rice, which is the primary agricultural crop in the ROK. The ROK produced nearly twice as much rice by

weight as any other single crop in 2007 (http://fao-stat.fao.org/site/339/default.aspx). While cropland cannot be directly substituted for rice land cover, it does provide an indication of where rice fields are likely to be located.

The land cover was used in its original format and was also processed with a filter that counted the number of cropland pixels within each 5x5 block of pixels in the file (approximately 5x5 km). The filtering allows the model to take into account the amount of cropland surrounding the mosquito collection sites and not just the one pixel (1 km^2) at the collection site. In capture and release studies, *Cx. tritaeniorhynchus* mosquitoes have been collected up to 8.4 km from the release site (Wada et al., 1969) but are believed to have a more typical flight range of <1.0 km. Therefore, a 5x5 pixel filter includes pixels within 2.5 km of each collection site to account for this flight range.

The normalized difference vegetation index (NDVI) was used in the model as a measure of the amount of healthy green vegetation on the ground. NDVI was derived from SPOT vegetation sensor data. The NDVI data were processed to create 12 monthly mean composite NDVI images from data collected from May 1998 to April 2008.

Modeling

The ecological niche modeling programme Maxent (Phillips et al., 2004, 2006; Phillips and Dudik, 2008) was used to model the distribution of *Cx. tritaeniorhynchus*. Maxent takes raster environmental layers (e.g. temperature, elevation) and a text file of species' locations, infers the environmental requirements for the species and produces a map that predicts the distribution of a species. This programme can be downloaded from Princeton University (http:// www.cs.princeton.edu/~schapire /maxent). Maxent is one of several programmes available for ecological niche modeling and was selected for this project based on its performance compared to other modeling programmes (Elith et

al., 2006) and because it does not require species absence data, which were not always available for this study.

To test the accuracy of the model, 25% of the 96 *Cx. tritaeniorhynchus* presence records were randomly selected and reserved for testing, and the remaining 75% were used for building the model (training points). Maxent uses two methods to calculate the accuracy of the model. The first method is a calculation of the area under the curve (AUC) for the receiver operation characteristic (ROC), a graphical representation of the sensitivity *versus* 1 minus the specificity. The AUC method has been described by other authors and is widely used to evaluate models (Swets, 1988; Fielding and Bell, 1997), including ecological niche models (Phillips et al., 2006).

In the second method to test accuracy, a threshold (e.g. minimum training presence) is used to split the continuous range of probability values into two sets representing predicted presence or absence. The programme then calculates the p-values for the null hypothesis that test points are predicted to be no better than they would by a random prediction (Phillips et al., 2006).

In order to determine the importance of each environmental variable to predicting the species presence, the Maxent programme performs a jackknife test. The jackknife procedure runs the model numerous times: (i) a model is created with all variables; (ii) models are created with all variables excluding one variable each time the model is run; and (iii) models are created using only one variable at a time (Phillips et al., 2006). The importance of an environmental variable is determined based on having a large training gain when the variable is used alone in the model and a subsequent decrease in training gain when removed from the model.

Because many of the collection sites fell within urban land cover pixels, we wanted to compare the influence of land cover *versus* climate and elevation on the output probability maps. Therefore, the Maxent model was run using three different sets of environmental data: (i) all available layers (climate, elevation, and land cover related layers

such as NDVI); (ii) climate and elevation; and (iii) climate.

Comparison data sets

To help determine how useful the model is in predicting JE for the ROK, the model output was compared to three other geographic data sets. JE human case data were obtained for 2001-2009 from the Disease Web Statistics System of the Korea Centers for Disease Control (K-CDC) (http://stat.cdc.go.kr/). Case locations were reported at either the city or district level. An approximate latitude/longitude position, for the center of each city or district where cases were reported, was determined using a geographical information systems (GIS) approach.

As swine serve as an amplifying host of JE, comparison of the distribution and density of pig populations with the model is useful in providing a better understanding of potential human disease. A gridded estimate of pig density (number of pigs per km^2) was obtained from the Food and Agricultural Organization (FAO) (http://www.fao.org/geonetwork).

Human population density, in combination with other factors, can affect potential human JE disease rates. A LandScan™ 2008 High Resolution Global Population Data Set was obtained from UT-Battelle, Oak Ridge National Laboratory, to compare with the locations of the JE cases (http://www.ornl.gov /sci/landscan/). LandScan is worldwide population data stored in a 30- by 30-second resolution latitude/longitude grid. Each cell within the grid represents an estimate of the number of people located within that grid cell during daylight hours (Dobson et al., 2000).

Results

Maxent modeling results

Of the 96 collection sites for *Cx. tritaeniorhynchus*, 72 points were used to build the model

(training points) and 24 points were withheld to test the model (Fig. 1a). The model was run three times to observe the effect of using: (i) all layers (land cover, NDVI, climate and elevation); (ii) only the climate and elevation; and (iii) only the climate layers (Fig. 1b-1d). The climate only and the climate with elevation models were similar. The model using land cover showed higher probabilities in the same areas as the other two models, but locally, the land cover layers made small differences to the probability values. The final model using all variables is available for download at http://www.mosquitomap.org.

Maxent's statistical evaluation of the models indicated that all three models provided useful predictions (Table 1). For all of the models, the AUC was above 0.9, indicating very high accuracy (Swets, 1988; Manel et al., 2001). The model that included all variables had the highest AUC. Relatively high
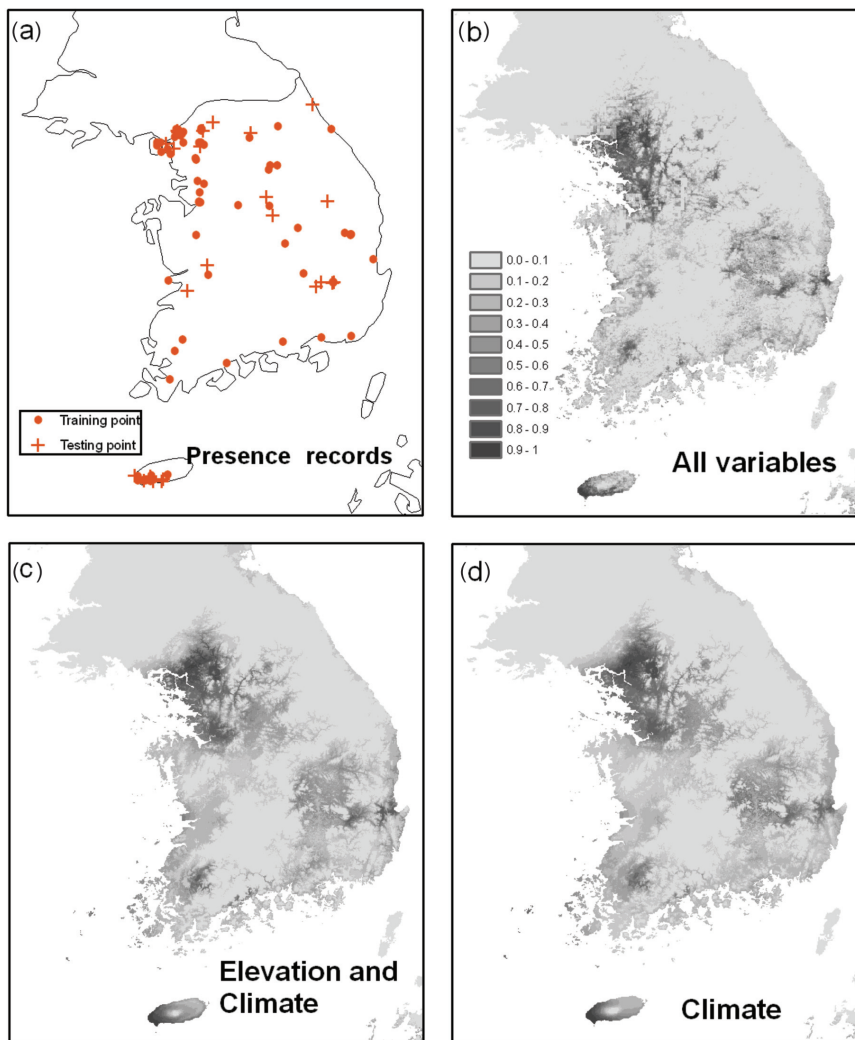

Fig. 1. Modeling results. (a) presence records for *Cx. tritaeniorhynchus* used in the model showing the randomly selected training and testing points; (b) Maxent output map showing the predicted probability (0 to 1) of occurrence for *Cx. tritaeniorhynchus* based on all variables; (c) Maxent output map using elevation and climate; and (d) Maxent output map using only climate.

Table 1. Statistical evaluation of models built using same presence records but different sets of environmental variables.

Input data to model	AUC training points	AUC test points	P-value minimum training presence
All environmental data	0.940	0.884	<0.001
Climate and elevation data only	0.921	0.898	<0.001
Climate data only	0.915	0.889	<0.001

AUC values (>0.8) for the testing points, were another indication of the predictive power of the model. Finally, Maxent tests the null hypothesis that the test points are predicting no better than a random prediction using various thresholds (Phillips et al., 2006). All models demonstrated highly significant, better-than-random performance (P <0.001).

Contributions of the variables to the model

The jackknife test (Fig. 2) showed minimum temperature in July (tmin07) as the environmental variable with the highest training gain when used alone in the model, which indicates that it had the most predictive ability of any variable. The variable, which decreased the gain the most when excluded from the model, was land cover (BU_Landcover_ealaea.sds01), indicating that land cover had the most unique contribution to the model.

Comparison of training gains of models built with single variables illustrates some interesting variations during different seasons. For example, the training gain increased in the summer months for both NDVI (spot07mean-spot09mean) and minimum temperature (tmin05-tmin08) variables. However, for maximum temperature, the training gain was higher in the fall and winter months, indicating that winter maximum temperatures help determine the probability of mosquito occurrence. The training gain for precipitation was highest during the month of July before the peak *Cx. tritaeniorhynchus* populations occur.

Response of Cx. tritaeniorhynchus *to environmental variables*

An examination of the environmental conditions at the collections sites (Tables 2, 3 and 4) and
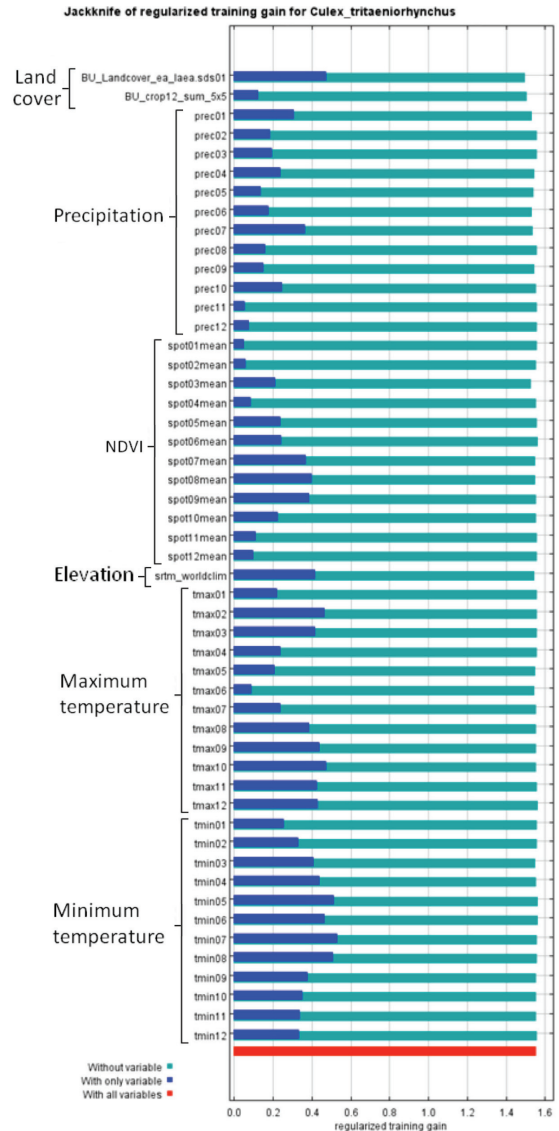

Fig. 2. Jackknife of regularized training gain. Red bar represents the training gain achieved by a model using all variables. Each dark blue bar represents the training gain achieved in the model using a single variable. Each aqua bars represents the training gain achieved when that particular variable is dropped from the model. See abbreviations of variable names in Table 2.

Table 2. Mean, minimum and maximum values for selected environmental variables calculated for collection site locations.

Environmental variable (unit)	Abbreviation	Mean	Minimum	Maximum
Elevation (m)	srtm_worldclim	82	3	517
Precipitation June (mm)	prec06	151	103	250
Precipitation July (mm)	prec07	308	134	455
Precipitation August (mm)	prec08	239	143	299
Precipitation September (mm)	prec09	155	123	204
Precipitation October (mm)	prec10	51	41	90
Minimum temperature June (°C)	tmin06	16.5	13.9	18.1
Minimum temperature July (°C)	tmin07	21.3	18.6	23.4
Minimum temperature August (°C)	tmin08	21.5	18.4	24.1
Minimum temperature September (°C)	tmin09	15.8	12.4	20.2
Minimum temperature October (°C)	tmin10	8.5	4.7	14.0
Maximum temperature June (°C)	tmax06	25.8	21.4	27.9
Maximum temperature July (°C)	tmax07	28.4	25.3	30.0
Maximum temperature August (°C)	tmax08	29.5	26.9	30.9
Maximum temperature September (°C)	tmax09	25.4	23.3	26.8
Maximum temperature October (°C)	tmax10	20.1	17.8	22.3
NDVI June	spot06mean	0.561	0.200	0.828
NDVI July	spot07mean	0.620	0.216	0.796
NDVI August	spot08mean	0.658	0.240	0.828
NDVI September	spot09mean	0.634	0.236	0.832
NDVI October	spot10mean	0.562	0.176	0.812

graphs of environmental variables *versus* the probability of occurrence (Fig. 3), illustrate the environmental limits of *Cx. tritaeniorhynchus* in the ROK.

Elevation: The maximum elevation at the collection sites is 517 m with a mean elevation of 82 m (Table 2). The elevation response curve (Fig. 3) shows that as elevation increases, the probability of JE occurrence decreases.

Land cover: As determined from the land cover, urban/built-up areas, and croplands are the dominant land cover types at the collection sites (Table 3). Many of the cropland areas are rice fields, which provide suitable habitat for *Cx. tritaeniorhynchus*. The urban land cover class may be over emphasised because many of the collection sites are located on US military bases. Many of these urban military bases are adjacent to rice cropland as demonstrated by the 5x5 filtered cropland images; 60 of the 96

collection sites (62.5%) were within or near cropland pixels (Table 4). The probability of *Cx. tritaeniorhynchus* occurrence increased with increasing numbers of cropland pixels within a 5x5 pixel window up to 15 cropland pixels but subsequently decreased (Fig. 3).

Climate: The highest mean summer climate values at the collection sites occurred during July for precipitation, August for minimum temperature and August for maximum temperature (Table 2). *Cx. tritaeniorhynchus* probability of occurrence generally increases as minimum temperature in July increases (Fig. 3).

NDVI: The highest mean NDVI at the collection sites occurred during August (Table 2). For the SPOT mean NDVI image in August, the probability of *Cx. tritaeniorhynchus* presence initially increases with increasing NDVI, and then decreases at the highest NDVI values (Fig. 3).

Table 3. Land cover at the 96 *Cx. tritaeniorhynchus* collection sites.

Land cover class	Number of sites
Water	7
Evergreen broadleaf forest	1
Mixed forests	9
Closed shrublands	2
Woody savannas	3
Savannas	4
Croplands	24
Urban and built-up	37
Croplands/natural vegetation	7
Unclassified	2

Table 4. Number of cropland pixels occurring at the collection sites.

Number of crop pixels	Number of sites
0	36
1	3
2	9
3	8
4	4
5	3
6	2
7	2
8	1
9	3
10	8
11	1
13	1
14	1
15	8
16	1
17	1
18	1
21	2
23	1

Comparison of Cx. tritaeniorhynchus *models to other geographic data*

JE cases reported from 2001-2009 were diagnosed in patients residing in areas of predicted high probabilities for *Cx. tritaeniorhynchus* (Fig. 4). Although the site of transmission was not determined and case locations are approximate due to reporting by city or district rather than patient address, the general pattern of cases matched the general trend of the model. Similarly, JE cases were usually reported near or in predicted areas of high densities of swine (Fig. 5a). Although cases were reported in both urban and rural areas, more cases were reported in urban areas (Fig. 5b).

Development of Cx. tritaeniorhynchus *populations over one summer*

The seasonal numbers of *Cx. tritaeniorhynchus* mosquitoes collected bimonthly at seven pig farms during the summer of 2009 showed a high degree of variability (Fig. 6). In early June, low numbers of *Cx. tritaeniorhynchus* mosquitoes were collected at only one site on the southern coast of the ROK. As the season progressed, the numbers increased and mosquitoes were collected further north. *Cx. tritaeniorhynchus* was found at six of the seven collection sites in late August with numbers decreasing in September as cooler weather returned and rice fields were drained marking the end of the growing season.

Discussion

The present model predicted lower probabilities of *Cx. tritaeniorhynchus* for higher elevations, which is consistent with literature reports that *Cx. tritaeniorhynchus* is infrequently collected above 1,000 m in elevation (Peiris et al., 1993; Pandey et al., 2003). The highest collection site in the ROK occurred at 517 m and the model approaches zero probability of occurrence at approximately 600 m and above (Fig. 3). At higher elevations in the ROK, the forested, steeply sloped hillsides and mountains have very
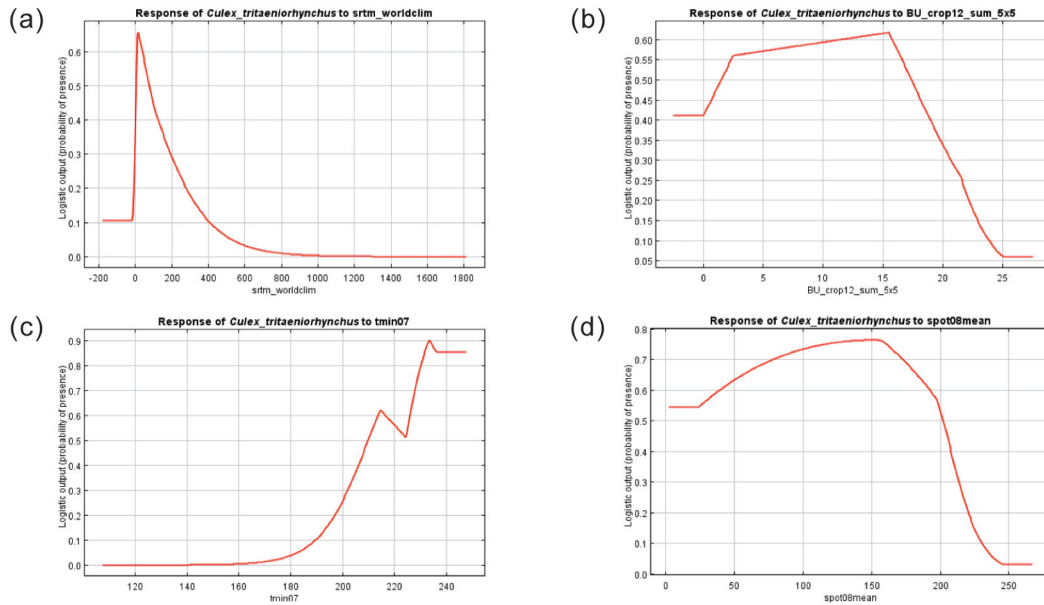
Fig. 3. Graphs of probability of *Cx. tritaeniorhynchus* occurrence *versus* selected variables: (a) elevation (srtm_worldclim); (b) count of cropland pixels within a 5x5 pixel area (BU_crop12_sum_5x5); (c) minimum temperature in July (tmin07); and (d) NDVI in August (spot08mean).

limited larval habitat (e.g. intermittent to permanent streams and rock and stream pools). However, additional sampling would be needed to determine the actual elevation limit of *Cx. tritaeniorhynchus* in the ROK.

In our initial modeling efforts (Masuoka et al., 2009), the data were collected primarily on US military bases, resulting in an over-sampling of urban land cover and a predictive bias towards this variable in the model. This was a concern since *Cx. tritaeniorhynchus* primarily breeds in rice fields and is not an urban container breeder. However, adult mosquitoes can fly from rice fields that are adjacent to peri-urban villages and outlying perimeters of larger urban areas. While additional collection sites were established and the new data contain more sites located in croplands, urban pixels still account for 37 of the 96 collection sites (Table 3).

Removing land cover as a variable in the model addresses the issue of how the urban collection sites affect the model. NDVI is a type of land cover since it maps the amount of vegetation on the ground.
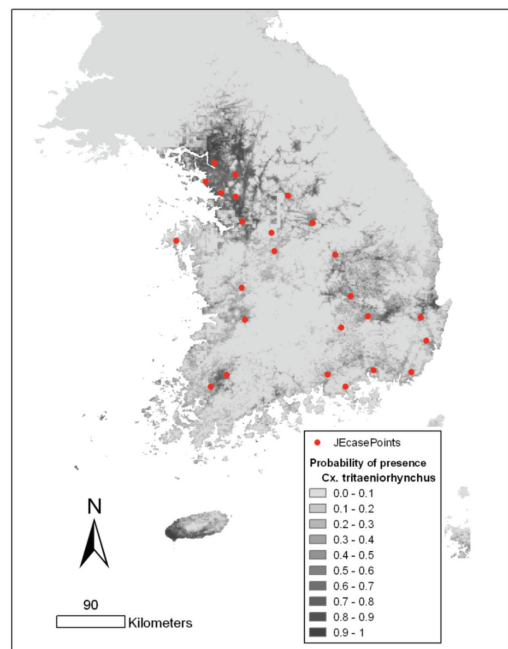

Fig. 4. Location (red dots) of reported cases of JE in the Republic of Korea from 2001-2009. Background image is the Maxent probability of presence map for *Cx. tritaeniorhynchus* created using all environmental input variables.

When land cover, the filtered cropland image, and NDVI were removed from the model (Fig. 1), the prediction was similar to the model using all variables. While land cover in the model does affect the probability locally, the overall prediction pattern is similar to the other models.

From the jackknife test, NDVI contributed more training gain to the model in the summer months. As NDVI is a measurement of the amount of healthy green vegetation on the ground, the NDVI difference between vegetated and non-vegetated areas occurs during the growing season, which may account for the increase in training gain during the summer.

The NDVI response curve (Fig. 3d) plots the initial increase of *Cx. tritaeniorhynchus* probability of occurrence as NDVI increases but then shows a decreasing probability at higher NDVI values. The drop in probability with the highest NDVI values may reflect the under-sampling of densely vegetated sites (many sample sites contain urban areas with lower NDVI). In addition, mixed pixels containing water (low NDVI) and vegetation have lower NDVI than pure vegetation pixels containing no water or mosquito habitat. Finally, previous studies have shown initial planting and tillering rice stages to have the highest immature abundance and decreasing larval abundance with increasing rice height (Mutero et al., 2000; Sunish and Ruben, 2001; Jacob et al., 2007). Increased rice height is thought to decrease larval abundance by blocking oviposition and reducing the temperature of the water by blocking sunlight.

Precipitation in July, minimum temperature in the summer months and maximum temperature in the fall and winter months all strongly contributed to the training gain in the jackknife test. *Cx. tritaeniorhynchus* mosquitoes overwinter as adults on the Korean peninsula and first appear at its southern tip (Hong et al., 1993, 1994). The adults increase in number with peak populations occurring in August and September. Heavy rainfall in July and warmer minimum summer temperatures would be conducive for habitat formation and mosquito development, respectively. In the model, more training gain
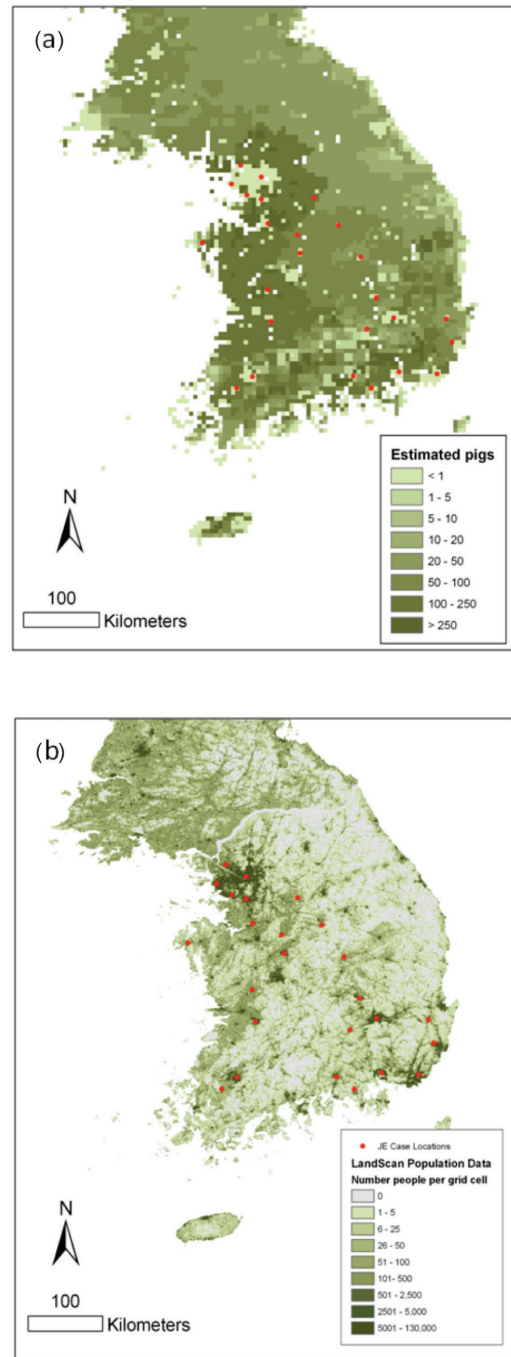

Fig. 5. (a) Locations of reported cases of JE (red dots) compared to predicted pig density (pigs per km²) in the Republic of Korea for 2005, (b) comparison of human population to locations of reported JE cases (2001-2009) shown in red. Source of background image: LandScan 2008™, ORNL, UT-Battelle, LLC.

Number *Cx. tritaeniorhynchus* collected

□ 0   • 1 - 10   • 11 - 46   ● 47 - 51   ● 52 - 300   ● 301 - 500   ● 501 - 7710
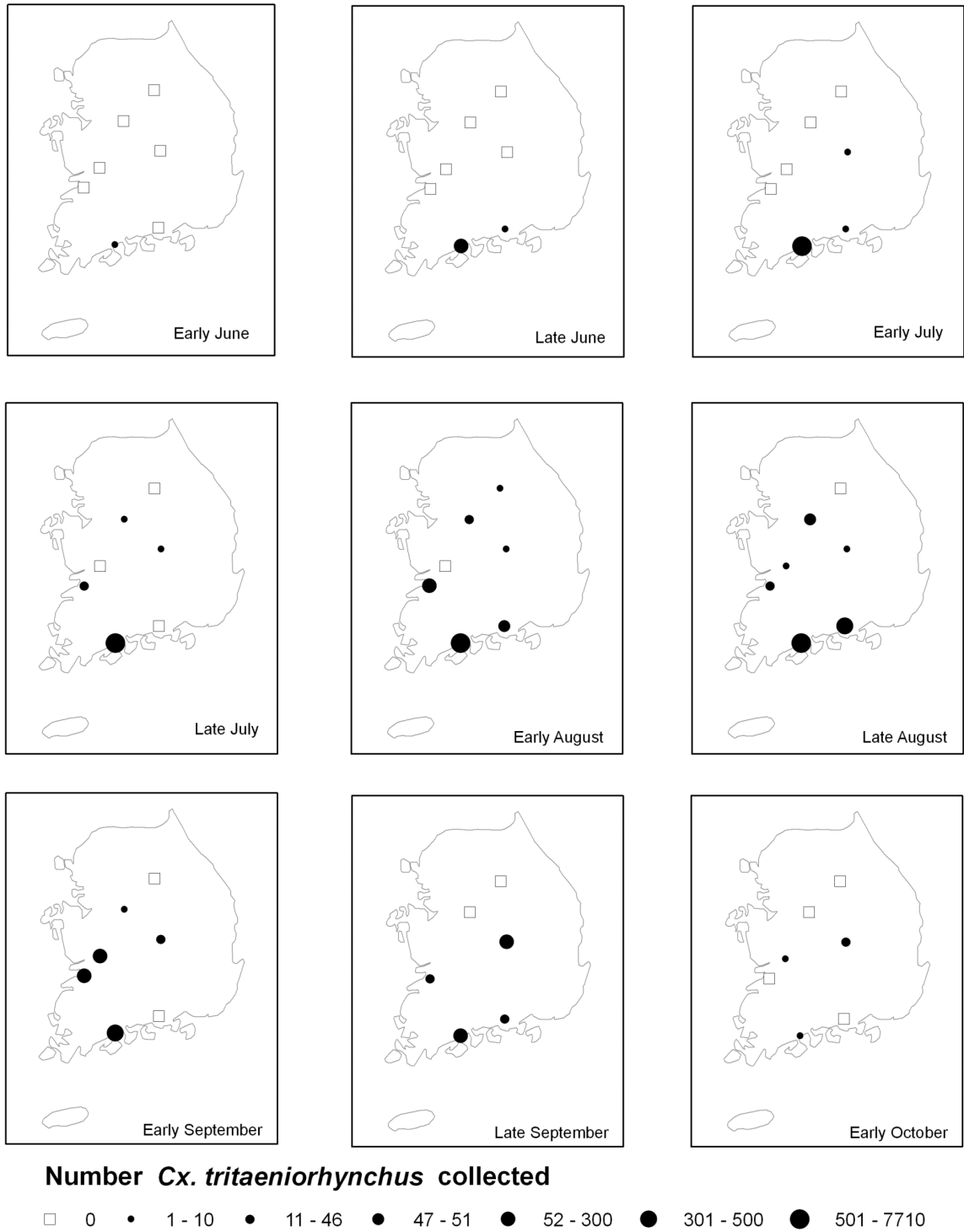
Fig. 6. Number of adult *Culex tritaeniorhynchus* mosquitoes collected over the summer of 2009 from seven locations.

in winter months for the maximum temperature may be a reflection of warmer temperatures that are important for the survival of overwintering adult mosquitoes. Because *Cx. tritaeniorhynchus* use the rice fields as breeding sites, we incorrectly expected that the sum of cropland pixels in a 5x5 pixel area would be a more predictive variable in the model than the land cover. The jackknife test showed that the sum of the cropland pixels did not produce much gain in the model when used by itself. In the response curves (e.g. Fig. 3b), the sum of the cropland initially increases the probability and then drops after 15 pixels of cropland. This is a reflection of small cropland areas often associated with narrow stream and river valleys surrounded by forested hills and mountains.

There is an increased probability for the presence of *Cx. tritaeniorhynchus* on the northeast coastline (Fig. 4), with adults collected at two locations (Fig. 1a), but no recent cases of JE. The population image (Fig. 5b) depicts some cities and smaller towns in the area of predicted occurrence, so the lack of JE cases is not due to low population numbers. The lack of JE cases from these regions may be the result of low estimated pig (amplifying hosts) populations along the east coast (Fig. 5a), reducing the risk of JE infections in mosquitoes and transmission to humans. Human behaviours and practices in this region such as housing structure or mosquito control practices, may also have an effect.

In summary, the Maxent output probability distribution maps were consistent with known ecological characteristics for the distribution of *Cx. tritaeniorhynchus*. The jackknife analysis showed that elevation, land cover, minimum summer temperatures, maximum fall and winter temperatures, and summer NDVI are important in modeling the species. An overlay of recent JE cases on the mosquito probability map showed that these maps are a useful prediction of areas that are at increased risk for JE. The prediction maps are useful for public health officials to develop JE control strategies by focusing limited resources on vaccine, vector control and disease prevention programmes.

The model developed here estimates the probability of species occurrence. In disease control, the abundance of the vector is also an important factor. In future work, we will examine the use of satellite weather data and NDVI to predict mosquito abundance within the geographic high-risk areas and attempt to expand the model to include other parts of Southeast Asia.

Acknowledgements

References

Dobson JE, Bright EA, Coleman PR, Durfee RC, Worley BA, 2000. LandScan: a global population database for estimating populations at risk. Photogram Eng Rem Sensing 66, 849-857.

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijamans RJ, Huettmann F, Lethwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton J, Sobéron J, Williams S, Wisz MS, Zimmermann NE, 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129-151.

Erlanger TE, Weiss S, Keiser J, Utzinger J, Wiedenmayer K, 2009. Past, present and future of Japanese encephalitis. Emerg Infect Dis 15, 1-7.

Fielding AH, Bell JF, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Cons 24, 38-49.

Foley DH, Klein TA, Kim HC, Sames WJ, Wilkerson RC, Rueda LM, 2009. Geographic distribution and ecology of potential malaria vectors in the Republic of Korea. J Med Entomol 46, 680-692.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A, 2005. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25, 1965-1978.

Hong HK, Kim CM, Lee YJ, Ko YG, 1993. Hibernation study of Korean mosquitoes. Annual Meeting for Entomological Society of Korea, Dongguk University,

Seoul, Korea, p. 20.

Hong HK, Kim CM, Shim JC, Shin EH, 1994. Hibernation study of Korean mosquitoes. Annual Meeting for Entomological Society of Korea, Korea University, Seoul, Korea, p. 34.

Jacob BG, Muturi EJ, Mwangangi JM, Funes J, Caamano EX, Muriu S, Shililu J, Githure J, Novak RJ, 2007. Remote and field level quantification of vegetation covariates for malaria mapping in three rice agro-village complexes in Central Kenya. Int J Health Geogr 6, 21.

Kim HC, Chong ST, Pike JG, O'Guinn ML, Pacha LA, Lee HS, Klein TA, 2004. Seasonal prevalence of mosquitoes collected from light traps in the Republic of Korea, 2002. Entomol Res 34, 177-186.

Kim HC, Wilkerson RC, Pecor JE, Lee WJ, Lee JS, O'Guinn ML, Klein TA, 2005. New records and reference collection of mosquitoes (Diptera: Culicidae) on Jeju Island, Republic of Korea. Entomol Res 35, 56-65.

Korea Centers for Disease Control and Prevention (K-CDC), 2008. Japanese encephalitis in Korea since 2001. Public Health Weekly Report (PHWR) 1, 289-292.

Manel S, Williams HC, Ormerod SJ, 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. J Appl Ecol 38, 921-931.

Masuoka P, Klein TA, Kim HC, Claborn DM, Achee N, Andre R, Chamberlin J, Taylor K, Small J, Anyamba A, Sardelis M, Grieco J, 2009. Modeling and analysis of mosquito and environmental data to predict the risk of Japanese encephalitis. ASPRS Annual Conference, Baltimore, Maryland, USA, 63-68.

Moffett A, Shackelford N, Sarkar S, 2007. Malaria in Africa: vector species' niche models and relative risk maps. PLoS One 2, e824.

Mutero CM, Blank H, Konradsen F, van der Hoek, 2000. Water management for controlling the breeding of *Anopheles* mosquitoes in rice irrigation schemes in Kenya. Acta Trop 76, 253-263.

Pandey B, Yamamoto A, Morita K, Kurosawa Y, Rai S, Adhikari S, Kandel P, Kurane I, 2003. Serodiagnosis of Japanese encephalitis among Nepalese patients by particle agglutination assay. Epidemiol Infect 131, 881-885.

Peiris JSM, Amerasinghe FP, Arunagiri CK, Perera LP, Karunaratne PP, Ratnayake CB, Kulatilaka TA, Abeysinghe MRN, 1993. Japanese encephalitis in Sri Lanka: comparison of vector and virus ecology in different agro-climatic areas. Trans R Soc Trop Med Hyg 87, 541-548.

Peterson AT, Shaw JJ, 2003. *Lutzomyia* vectors for cutaneous leishmaniasis in southern Brazil: ecological niche models, predicted geographic distributions, and climate change effects. Int J Parasitol 33, 19-31.

Phillips SJ, Anderson RP, Schapire RE, 2006. Maximum entropy modeling of species geographic distributions. Ecol Model 190, 231-259.

Phillips SJ, Dudik M, 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161-175.

Phillips SJ, Dudík M, Schapire RE, 2004. A maximum entropy approach to species distribution modeling. Proceedings of the 21st International Conference on Machine Learning, 655-662.

Straus JH, Straus EG, 2002. Viruses and human disease. Academic Press, 383 pp.

Sunish IP, Ruben R, 2001. Factors influencing the abundance of Japanese encephalitis vectors in ricefields in India - I. Abiotic. Med Vet Entomol 15, 381-392.

Swets JA, 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285-1293.

Wada Y, Kawai S, Oda T, Miyagi I, Suenaga O, Nishigaki J, Omori N, 1969. Dispersal experiment of *Culex tritaeniorhynchus* in Nagasaki area. Trop Med 11, 37-44.