SCIENTIFIC NOTE

VALIDATION OF ECOLOGICAL NICHE MODELS FOR POTENTIAL MALARIA VECTORS IN THE REPUBLIC OF KOREA

DESMOND H. FOLEY,[1] TERRY A. KLEIN,[2] HEUNG CHUL KIM,[3] TRACY BROWN,[1] RICHARD C. WILKERSON[1] AND LEOPOLDO M. RUEDA[1]

ABSTRACT. Data on molecularly identified adult and larval mosquitoes collected from 104 sites from the Republic of Korea (ROK) in 2007 were used to test the predictive ability of recently reported ecological niche models (ENMs) for 8 potential malaria vectors. The ENMs, based on the program Maxent and the least presence threshold criterion, predicted 100% of new collection locations for *Anopheles sinensis*, *An. belenrae*, *An. pullus*, and *An. sineroides*; 96% of locations for *An. kleini*; and 83% for *An. lesteri*, but were relatively unsuccessful for the infrequently collected non-Hyrcanus group species *An. koreicus* and *An. lindesayi japonicas*. The ENMs produced with the use of Maxent had fewer omission errors than those using the Genetic Algorithm for Rule-Set Prediction program. The results emphasize the importance of independent test data for validation and improvement of ENMs, and lend support for the further development of ENMs for predicting the distribution of malaria vectors in the ROK.

KEY WORDS   Ecological niche models, distribution, malaria, mosquito, Korea

Since 1993, *vivax* malaria has become an annual threat to military personnel and civilians in the Republic of Korea (ROK—South Korea), especially along the demilitarized zone (DMZ) separating the ROK from the Democratic People's Republic of Korea (DPRK—North Korea) (Kho et al. 1999, Park et al. 2003, Ciminera and Brundage 2006, Han et al. 2006, Yeom et al. 2007, Kim et al. 2009, Klein et al. 2009). To identify suitable areas for malaria transmission, we developed ecological niche models (ENMs) to predict the distribution of the 8 candidate malaria vector species reported from the ROK (Foley et al. 2009). These species are: *Anopheles sinensis* sensu stricto (s.s.) Wiedemann, *An. pullus* M. Yamada, *An. lesteri* Baisas and Hu (=*An. anthropophagus*), *An. sineroides* S. Yamada, *An. kleini* Rueda, *An. belenrae* Rueda, *An. lindesayi japonicus* S. Yamada, and *An. koreicus* S. Yamada and Watanabe. Because of the lack of morphological markers, a polymerase chain reaction (PCR) technique (Li et al. 2005) is required to separate 5 of the 6 members of the Hyrcanus group found in the ROK. Two of these species, *An. kleini* and *An. belenrae*, have only recently been recognized (Rueda 2005), necessitating a reassessment of information about malaria vectors in the ROK. Until the identity

of the malaria vectors are established, we consider all 8 species as candidates, although results from ENMs suggest that *An. lindesayi japonicus* and *An. koreicus* are unlikely to be malaria vectors (Foley et al. 2009).

Foley et al. (2009) used the Genetic Algorithm for Rule-Set Prediction (GARP) (Stockwell and Noble 1992) and a maximum entropy approach (Maxent) (Phillips et al. 2006) for ENMs of Korean mosquitoes. In certain comparisons, Maxent has been shown to outperform GARP (Elith et al. 2006, Phillips et al. 2006), and Hernandez et al. (2006) and Pearson et al. (2007) found that Maxent achieved better predictive success rates with small sample sizes. However, GARP may perform better at extrapolation for unsampled areas (i.e., transferability) than Maxent (Peterson et al. 2007, but see Phillips 2008). To reduce ENMs to presence and absence predictions, Foley et al. (2009) used the lowest presence threshold (LPT), which identifies pixels that are at least as suitable as the lowest value associated with a species' presence (Pearson et al. 2007).

We are interested in the performance of ENMs because quantification of the area of vector-borne disease risk (Mal-area), where humans, pathogens, and vectors potentially overlap (Foley et al. 2008), is dependent on realistic models of vector distribution. Model validation is often an integral part of ENM development. For example, the genetic algorithm in GARP allows rules applied to training data to evolve to maximize predictive accuracy based on test data. Extrinsic test data set aside from model development can also be overlaid onto predicted distributions to calculate omission error (percentage of independent test points not predicted by the model). Extrinsic test

[1] Division of Entomology, Walter Reed Army Institute of Research, 503 Robert Grant Avenue, Silver Spring, MD 20910.
[2] Force Health Protection and Preventive Medicine, 65th Medical Brigade/USAMEDDAC-Korea, Unit 15281, APO AP 96205-5281.
[3] 5th Medical Detachment, 168th Multifunctional Medical Battalion, 65th Medical Brigade, Unit 15247, APO AP 96205-5247.

Report Documentation Page

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **2010**	2. REPORT TYPE	3. DATES COVERED **00-00-2010 to 00-00-2010**

4. TITLE AND SUBTITLE **Validation of Ecological Niche Models for Potential Malaria Vectors in the Republic of Korea**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Walter Reed Army Institute of Research,Walter Reed Biosystematics Unit,503 Robert Grant Avenue,Silver Spring,MD,20910**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Data on molecularly identified adult and larval mosquitoes collected from 104 sites from the Republic of Korea (ROK) in 2007 were used to test the predictive ability of recently reported ecological niche models (ENMs) for 8 potential malaria vectors. The ENMs, based on the program Maxent and the least presence threshold criterion, predicted 100% of new collection locations for Anopheles sinensis, An. belenrae, An. pullus, and An. sineroides; 96% of locations for An. kleini; and 83% for An. lesteri, but were relatively unsuccessful for the infrequently collected non-Hyrcanus group species An. koreicus and An. lindesayi japonicas. The ENMs produced with the use of Maxent had fewer omission errors than those using the Genetic Algorithm for Rule-Set Prediction program. The results emphasize the importance of independent test data for validation and improvement of ENMs, and lend support for the further development of ENMs for predicting the distribution of malaria vectors in the ROK.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**4**	

Table 1. Validation of the ecological niche models in Foley et al. (2009) for *Anopheles belenrae*, *An. kleini*, *An. koreicus*, *An. lesteri*, *An. lindesayi japonicus*, *An. pullus*, *An. sinensis*, and *An. sineroides* based on comparison with new collection data. The number of collection grids that were correctly predicted positive by the Genetic Algorithm for Rule-Set Prediction (GARP) and Maxent models for each species are shown, along with the omission error rate (in parentheses).

Species	Specimens	Locations[1]	No. grids	Final no. grids	GARP	Maxent
An. belenrae	58	30	15	13	8 (0.38)	13 (0)
An. kleini	1,378	71	26	24	21 (0.13)	23 (0.04)
An. koreicus	10	10	2	2	1 (0.50)	0 (1.00)
An. lesteri	38	13	8	6	2 (0.67)	5 (0.17)
An. lindesayi	358	11	7	7	2 (0.71)	4 (0.43)
An. pullus	418	60	19	17	16 (0.06)	17 (0)
An. sinensis	6,898	84	37	30	27 (0.10)	30 (0)
An. sineroides	412	24	10	9	7 (0.22)	9 (0)

[1] Total numbers of new collection locations for each species were reduced to 0.00833° grids, then grids that occurred in "no data" areas of the models and that overlapped grid point positions used for model construction were removed, resulting in the final number of grids for each species that were available for comparison.

data can be a random subset of collection locations, or spatially stratified subsampling of collection points; the latter is seen as a more rigorous test of the transferability of the models (Peterson et al. 2007). With the use of a subsample of test data, Foley et al. (2009) found that *An. sinensis* had the best support among Maxent models for species with larger sample sizes, according to random expectations of omission error for the study area predicted present. For species with smaller sample sizes (<25), all except *An. lesteri*, gave better omission error than random. The GARP models for species with larger sample sizes indicated that all species gave better omission error than random but, for species with smaller sample sizes, only *An. lesteri* and *An. koreicus* gave better omission error than random.

Additional validation can be conducted by comparing an ENM with expert opinion about a species' distribution, or with observations about a species' distribution from the literature. A less common, but perhaps more convincing demonstration of the utility of ENMs, is to test predictions with newly collected data. Thus, we aimed to test the accuracy of the ENMs developed by Foley et al. (2009) by comparing predictions with independent mosquito collection data obtained from the ROK in 2007.

The numbers of adult and larval specimens and number of collection locations are shown in Table 1. Species were identified with the use of PCR as described by Li et al. (2005). Many of the collection sites for 2007 were close to one another, so points were converted to grids of 0.00833° resolution (1 km at the equator) that matched the grid size and position in the models of Foley et al. (2009). For example, for *An. sinensis*, this process resulted in 37 spatially independent grids (see Table 1), of which 35 did not coincide with any of the 80 grid sites for this species used in Foley et al. (2009). Five of these 35 grids occurred in areas of

the model that were classified as "no data," resulting in 30 grids that could be used for final comparison (see Table 1). Of these new *An. sinensis* sites, 90% were predicted positive by the GARP model, and 100% by the Maxent model of Foley et al. (2009). The Maxent model accurately predicted new locations for *An. sinensis*, even when these were geographically distant from model input locations.

The Maxent models of Foley et al. (2009) were updated with the addition of collection data from 2007 (Fig. 1). The addition of points for *An. belenrae*, *An. kleini*, and *An. pullus* resulted in a small increase in the areas predicted suitable for these species, although the overall distribution pattern did not change markedly. For *An. sinensis*, the addition of points reduced the area predicted suitable for this species from 54.6% to 46.3%, although the overall distribution pattern did not change markedly. For *An. sineroides*, the addition of points (mainly in the northwest of the ROK) increased the area predicted suitable for this species from 38.2% to 40.8%. For *An. lesteri*, the addition of points (mainly in the southeast) increased the area predicted suitable for this species from 4.4% to 10.5%, mainly due to a southward expansion. For *An. lindesayi japonicus*, the addition of points (mainly in the center and south of the ROK) resulted in a contraction in the south and west, and an extension to the east. For *An. koreicus*, the addition of points decreased the area predicted suitable for this species from 17.4% to 15.1%, and resulted in an expansion in the north and central part of the ROK, and a contraction in the south and east.

The ENMs produced with the use of Maxent had fewer omission errors than those using GARP under the conditions of this study. This finding is in agreement with Elith et al. (2006) and other authors, who observed that Maxent outperforms GARP, especially within more densely sampled landscapes. Whether Maxent models
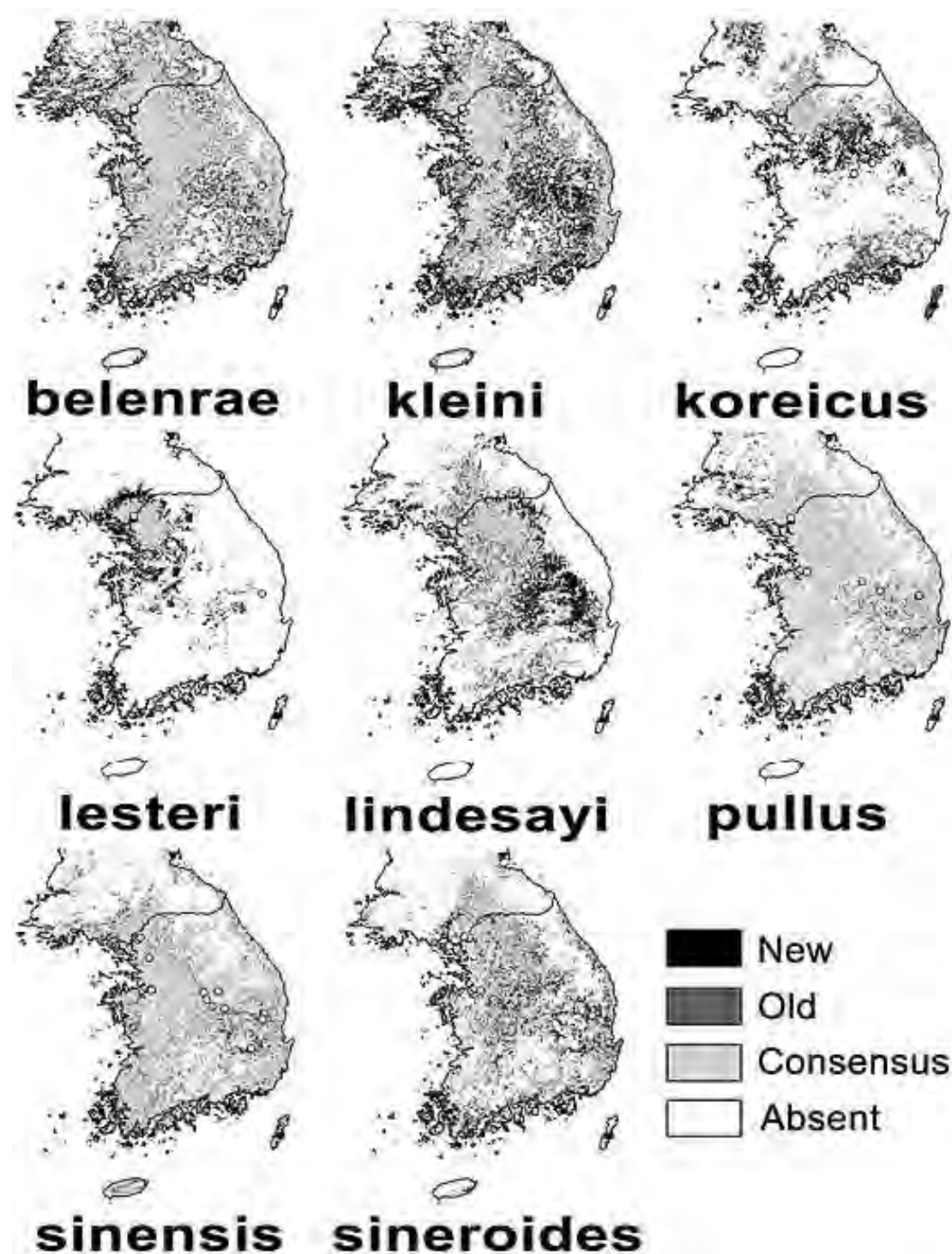
Fig. 1. Updated ecological niche models of *Anopheles belenrae*, *An. kleini*, *An. koreicus*, *An. lesteri*, *An. lindesayi japonicus*, *An. pullus*, *An. sinensis*, and *An. sineroides* in the Republic of Korea with the use of the program Maxent and new collection data from 2007. Figure shows the areas of agreement (Consensus) and disagreement between the current (New) models and those of Foley et al. (2009) (Old). The lowest presence threshold criterion was used, and new collection data for each species are shown by circles.

developed for the ROK would perform better than GARP models in unsampled areas of Southeast Asia is not known. We hope to validate the updated ENMs with additional sampling that was undertaken in the ROK in 2008.

The results for some Maxent models suggest that omission levels are lower than those reported by Foley et al. (2009). These authors suggested that among Maxent models for species with lower numbers of data points, only the *An. lesteri* model failed. However, we found that models for *An. koreicus* and *An. lindesayi japonicus* had higher levels of omission error than for *An. lesteri*. One possible reason for this discrepancy is that omission levels in Foley et al. (2009) were calculated from models built with a subset of points, whereas we tested their final models that were built on total available data points. For species with lower numbers of data points, additional collection data could improve the accuracy of ENMs.

Species habitat model accuracy can be important for incriminating the vector species, for understanding the ecological requirements of species based on satellite data, and for determining the Mal-area. We show that many of the ENMs of Foley et al. (2009) accurately predicts where those species occur. It is important to note that these models say nothing about the epidemiologically important parameters of abundance and longevity. However, we believe that our results lend support for the further development of ENMs for predicting the distribution of mosquito disease vectors in the ROK.

REFERENCES CITED

Ciminera P, Brundage J. 2006. Malaria in U.S. military forces: a description of deployment exposures from 2003 through 2005. *Am J Trop Med Hyg* 76:275–279.

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151.

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

Foley DH, Klein TA, Kim HC, Wilkerson RC, Rueda LM. 2008. Malaria risk assessment for the Republic of Korea based on models of mosquito distribution. *US Army Med Dep J* 46–53.

Han ET, Lee DH, Park KD, Seok WS, Kim YS, Tsuboi T, Shin EH, Chai JY. 2006. Reemerging vivax malaria: changing patterns of annual incidence and control programs in the Republic of Korea. *Korean J Parasitol* 44:285–294.

Hernandez PA, Graham CH, Master LL, Albert DL. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29:773–785.

Kho WG, Jang JY, Hong ST, Lee HW, Lee WJ, Lee JS. 1999. Border malaria characteristics of reemerging vivax malaria in the Republic of Korea. *Korean J Parasitol* 37:71–76.

Kim HC, Pacha LA, Lee WJ, Lee JK, Gaydos JC, Sames WJ, Lee HCS, Bradley K, Jeung GG, Tobler SK, Klein TA. 2009. Malaria in the Republic of Korea, 1993–2007. Variables related to re-emergence and persistence of *Plasmodium vivax* among Korean populations and U.S. Forces in Korea. *Mil Med* 174:762–769.

Klein TA, Pacha LA, Lee HCS, Kim HC, Lee WJ, Lee JK, Jeung GG, Sames WJ, Gaydos JC. 2009. *Plasmodium vivax* malaria among U.S. Forces Korea in the Republic of Korea. *Mil Med* 174:412–418.

Li C, Lee JS, Groebner JL, Kim HC, Klein TA, O'Guinn ML, Wilkerson RC. 2005. A newly recognized species in the *Anopheles Hyrcanus* group and molecular identification of related species from the Republic of South Korea (Diptera: Culicidae). *Zootaxa* 939:1–8.

Park JW, Klein TA, Lee HC, Pacha LA, Ryu SH, Yeom JS, Moon SH, Kim TS, Chai JY, Oh MD, Choe KW. 2003. Vivax malaria: a continuing health threat to the Republic of Korea. *Am J Trop Med Hyg* 69:159–167.

Pearson RG, Raxworthy CJ, Nakamura M, Peterson AT. 2007. Predicting species' distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *J Biogeog* 34:102–117.

Peterson AT, Papeş M, Eaton M. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* 30:550–560.

Phillips SJ. 2008. Transferability, sample selection bias and background data in presence-only modeling: a response to Peterson et al. *Ecography* 31:272–278.

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

Rueda LM. 2005. Two new species of Anopheles (Anopheles) Hyrcanus Group (Diptera: Culicidae) from the Republic of South Korea. *Zootaxa* 941:1–26.

Stockwell DRB, Noble IR. 1992. Induction of sets of rules from animal distribution data: a robust and informative method of analysis. *Math Comput Simul* 33:385–390.

Yeom JS, Kim TS, Oh S, Sim JB, Barn JS, Kim HJ, Kim YA, Ahn SY, Shin MY, Yoo JA, Park JW. 2007. *Plasmodium vivax* malaria in the Republic of Korea 2004–2005: changing patterns of infection. *Am J Trop Med Hyg* 76:865–868.