Supplementary Material

Modeling the risk of spread and establishment for Asian Longhorned Beetle (*Anoplophora glabripennis*) in Massachusetts from 2008 – 2009

A.J. Shatz, J. Rogan, F. Sangermano, J. Miller and A. Elmes

S1. Host preferability map creation

An Asian Longhorned Beetle (ALB) host preferability map was developed for this study using a weighted linear combination of eight host species abundance models. Host species abundance models were derived from tree abundance data from the United States Department of Agriculture – Forest Service Forest Inventory Analysis (FIA) program (USDA Forest Service 2005) and 18 topo-climatic and remotely sensed environmental predictor variables (Table S1). Host abundance models were only created for host species with $\geq 10\%$ prevalence, hence excluding the *Aesculus*, *Salix*, and *Ulmus* genera. The host abundance models include red maple (*Acer rubrum*), sugar maple (*Acer saccharum*), yellow birch (*Betula alleghaniensis*), sweet birch (*Betula lenta*), paper birch (*Betula papyrifera*), green ash (*Fraxinus americana*), white oak (*Quercus alba*), red oak (*Quercus rubra*). Host species were chosen based upon observed ALB host preference in field and laboratory settings and the availability of FIA data in the state of Massachusetts (Bancroft et al. 2002, Haack et al. 2006, Ludwig et al. 2002, Morewood et al. 2004a, Morewood et al. 2004b, Wang 2012,)

Table S1. Topo-climatic and remotely sensed variables used for host species abundance model creation.

Variable	Description	Units	Spatial Resolution	Temporal Range
TEMP	Average mean annual temperature	kelvin	30 m	2000-2009
PRECIP	Average mean precipitation	centimeters	250 m	2007
DEM	Elevation	meters	30 m	n/a
SLOPE	Slope	degrees	30 m	n/a
ASPECT	Aspect	n/a	30 m	n/a
NDVI_LOFF and NDVI_LON	Leaf-off and leaf-on NDVI	undefined	30 m	Sept.-Oct. 2000
EVI_LOFF and EVI_LON	Leaf-off and leaf-on EVI	n/a	30 m	Sept.-Oct. 2000
LOFFB and LONB	Leaf-off and leaf-on Kauth-Thomas brightness	n/a	30 m	Sept.-Oct. 2000
LOFFG and LONG	Leaf-off and leaf-on Kauth-Thomas greenness	n/a	30 m	Sept.-Oct. 2000
LOFFW and LONW	Leaf-off and leaf-on Kauth-Thomas wetness	n/a	30 m	Sept.-Oct. 2000
LC2000	16 category land cover map of MA for 2000	categorical	30 m	2000
LU1999	21 category land use map of MA for 1999	categorical	30 m	1999
SGEO	Surficial geology map	categorical	30 m	n/a

S2. Abundance Model Creation

Variable Importance

The host tree abundance models were created using a species distribution modeling framework employing Random Forest classification and regression trees as implemented in R (Liaw and Wiener 2002). Prior to model creation, the importance of environmental predictor variables was determined for each species by calculating the Mean Square Error (MSE) generated after permuting each predictor variable from the model at each node of the classification tree (Brieman 2001, Ghimire et al. 2012). This process was performed using $k = 100$ bootstrap samples of host presence locations with replacement to generate a distribution of MSE. Only variables with a MSE significantly greater than zero were retained to create the abundance model. Variables important to host species abundance model creation are shown in Table S2

Table S2. Predictor variables used for presence/absence and abundance model construction with variable importance scores normalized from 0-100.

Variable	<i>A. rubrum</i>	<i>A. saccharum</i>	<i>B. alleghaniensis</i>	<i>B. lenta</i>	<i>B. papyrifera</i>	<i>F. americana</i>	<i>Q. alba</i>	<i>Q. rubra</i>
TEMP	38 (7)	100 (1)	73 (2)	100 (1)	100 (1)	100 (1)	70 (3)	92 (2)
PRECIP	26 (14)	21 (8)	71 (3)	---	---	28 (6)	27 (10)	66 (6)
DEM	100 (1)	56 (2)	100 (1)	64 (4)	40 (2)	---	73 (2)	92 (2)
SLOPE	---	16 (10)	11 (7)	99 (2)	26 (4)	32 (4)	---	---
ASPECT	---	---	---	8 (12)	---	---	---	---
NDVI_LOFF	31 (10)	---	11 (7)	40 (5)	---	---	66 (4)	30 (12)
NDVI_LON	48 (4)	36 (3)	---	77 (3)	21 (5)	23 (9)	52 (5)	67 (5)
EVI_LOFF	43 (6)	---	26 (5)	30 (8)	---	---	41 (7)	38 (11)
EVI_LON	72 (2)	36 (3)	---	---	---	28 (6)	---	100 (1)
LOFFB	---	---	---	26 (10)	---	---	51 (6)	54 (7)
LONB	32 (9)	16 (10)	---	16 (11)	---	---	---	---
LOFFG	31 (10)	26 (7)	13 (6)	40 (5)	---	34 (3)	100 (1)	79 (4)
LONG	44 (5)	21 (8)	---	31 (7)	---	16 (10)	36 (8)	43 (9)
LOFFW	29(13)	---	---	28 (9)	---	39 (2)	---	49 (8)
LONW	38 (7)	---	---	---	16 (6)	30 (5)	---	43 (9)
LC2000	30(12)	32 (5)	66 (4)	---	---	27 (8)	30 (9)	22 (15)
LU1999	61 (3)	28 (6)	---	---	---	---	---	29 (13)
SGEO	24 (15)	---	---	---	27 (3)	---	23 (11)	28 (14)

Abundance Model Workflow

The host tree abundance models were developed using a two-part classification and regression tree approach employed in Prasad et al. (2006). Host species locations were converted into binary presence/absence data and classified into presence/absence distribution models using $k = 2000$ bootstrapped samples. Abundance models were created using the original host species abundance data using identical Random Forest model parameters. The final abundance maps were created by masking the abundance model with the classification model such that locations with a classified value of 0 were given an abundance of 0, and locations with a classified value of 1 retained the abundance model value. A flowchart of the model creation workflow is shown in Figure S2b.

Figure S2. A flowchart outlining the workflow to create both presence/absence and abundance models for the final host preferability map.

S3. Abundance Model Validation

Host species locations were bootstrapped into 50/50 calibration/validation datasets to assess the discrimination capacity and the reliability of both presence/absence and abundance models respectively (Fielding and Bell 1997; Pierce and Ferrier 2000). The discrimination capacity of the presence/absence models was assessed using a True Skill Statistic measurement, and the reliability of models were assessed using a Root Mean Square Error (RMSE) and a Pearson's Correlation Coefficient measurement (Liu et al. 2009). Validation measurements for all eight host species are shown below in Figure S3.

Figure 4. Box and whisker plots of (a) TSS, (b) RMSE, and (c) Correlation for evaluating model accuracy. Boxes represent 95% confidence intervals. Whiskers represent minimum and maximum. Labels above each plot show the number of bootstrap iterations performed.

S4. Host Preferability Index Creation

The host preferability index was created through a weighted linear combination scheme using a Multi-Criteria Evaluation (MCE) method. All eight host abundance models were linearly transformed into 8-bit integers (0 – 255) and were multiplied by a weights of importance from zero to one. Weights of importance were calculated based on the relative attractiveness of host species to ALB with respect to each other using an Analytical Hierarchy Process algorithm (Saaty 1977). The weighted abundance models were first added, and the summed model was then divided by 255 to create a host preferability index ranging from zero to one, with values closer to zero containing low abundance of attractive hosts and values closer to one containing high abundances of attractive hosts. The weights of importance assigned to host abundance models are shown below in Table S4.

Table S4. Weights of importance assigned to each host species abundance model as part of the MCE method for creating the host preferability map. Host species are listed in alphabetical order.

Species	Weight
<i>A. rubrum</i>	0.3998
<i>A. saccharum</i>	0.2049
<i>B. alleghaniensis</i>	0.0859
<i>B. lenta</i>	0.1298
<i>B. papyrifera</i>	0.0859
<i>F. americana</i>	0.0282
<i>Q. alba</i>	0.0286
<i>Q. rubra</i>	0.0370

References

- Bancroft, J.S., Smith, M.T., Chaput, E.K., Tropp, J., 2002. Rapid test of the suitability of host-trees and the effects of larval history on *Anoplophora glabripennis* (Coleoptera: Cerambycidae). *Journal of the Kansas Entomological Society* 75, 308–316.
- Breiman, L., 2001. Random Forest. *Machine Learning* 45, 5–32.
- Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24, 38–49.
- Ghimire, B., Rogan, J., Galiano, V.R., Panday, P., Neeti, N., 2012. An evaluation of bagging, boosting, and random forest for land-cover classification in Cape Cod, Massachusetts, USA. *GIScience and Remote Sensing* 49, 623–643.
- Haack, R.A., Bauer, L.S., Gao, R., McCarthy, J.J., Miller, D.L., Petrice, T.R., Poland, T.M., 2006. *Anoplophora glabripennis* within-tree distribution, seasonal development, and host suitability in China and Chicago. *The Great Lakes Entomologist* 9, 169–183.
- Liaw, A., Wiener, M., 2002. Classification and regression by RandomForest. *R News* 2, 18–22.
- Liu, C., White, M., Newell, G., 2009. Measuring the accuracy of species distribution models: a review. Presented at the 18th World IMACS/MODSIM Congress, Cairns, Australia.
- Ludwig, S.W., Lazarus, L., McCullough, D.G., Hoover, K., Montero, S., Sellmer, J.C., 2002. Methods to evaluate host tree suitability to the Asian Longhorned Beetle, *Anoplophora glabripennis*. *Journal of Environmental Horticulture* 20, 175–180.
- Morewood, W.D., Hoover, K., Neiner, P.R., McNeil, J.R., Sellmer, J.C., 2004a. Host tree resistance against the polyphagous wood-boring beetle *Anoplophora glabripennis*. *Entomologia Experimentales et Applicata* 110, 79–86.
- Morewood, W.D., Neiner, P.R., Sellmer, J.C., Hoover, K., 2004b. Behavior of adult *Anoplophora glabripennis* on different tree species under greenhouse conditions. *Journal of Insect Behavior* 17, 215–226.

- Pearce, J., Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133, 225–245.
- Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and Random Forest for ecological prediction. *Ecosystems* 9, 181–199.
- Saaty, T.L., 1977. A scaling method for properties in hierarchical structures. *Journal of Mathematical Psychology* 15, 234–281.
- USDA Forest Service, 2005. Forest Inventory & Analysis Data Collection and Analysis.
- Wang, B. Asian Longhorned Beetle: Annotated Host List. USDA-APHIS-PPQ, Center for Plant Health Science and Technology, Otis Laboratory. Available online: http://www.aphis.usda.gov/plant_health/plant_pest_info/asian_lhb/downloads/hostlist.pdf (accessed on 30 June 2015).