Supplementary Information for
Paninvasion severity assessment of a US grape pest to disrupt the global wine market

Nicholas A. Huron*
Jocelyn E. Behm
Matthew R. Helmus

Integrative Ecology Lab, Center for Biodiversity, Department of Biology, Temple University, Philadelphia, PA 19122 USA.

* corresponding author, nahuron@temple.edu

This PDF file includes:

Supplementary Methods

Below, we provide additional details for terminology and methods for the analyses conducted in our study.

Term Definitions

- **alignment correlation**—multivariate relationship among invasion potentials.
- **establishment potential**—likelihood of a region to contain suitable habitat for transported individuals of a non-native species to form a spreading population.
- **impact potential**—likelihood of a region to experience negative economic effects from an established non-native species.
- **invasion potentials**—likelihoods of a species to move through stages in an invasion process across regions[1]. We focus on the main stages: transport, establishment, impact.
- **MaxEnt**—abbreviation for maximum entropy, a presence-only SDM methodology that uses machine-learning to estimate the probability distribution of maximum entropy based on environmental variables and species occurrence records[2,3].
- **paninvasion**—invasion of a species at the global scale that disturbs a global economic market.
- **paninvasion risk**—the likelihood of a regional invasive species to become a globally invasive species and cause economic repercussions.
- **paninvasive species**—globally invasive species that goes through the three main invasion stages and thus can disturb global economic markets.
- **phylloxera**—*Daktulosphaira vitifoliae* is grapevine root pest native to North America that was responsible for the Great Wine Blight of the late 1800's, which was the largest economic disturbance to the global wine market ever recorded. The disruption was mitigated by widespread planting of European vines that were grafted to North American grapevine root stocks. The paninvasion of phylloxera continues to this day[4,5].
- **species distribution model (SDM)**—spatial model used to predict the environmental niche, habitat suitability, and establishment potential of a species.
- **spotted lanternfly (SLF)**—*Lycorma delicatula* is a planthopper native to China, Vietnam, and India. It invaded South Korea and Japan in the early 2000's and the

northeastern US ca. 2014. It is known to feed on >100 different host species, including grapes[6,7].

- **transport potential**—likelihood of a region to have an introduction of a non-native species.
- **tree of heaven (TOH)**—*Ailanthus altissima* is a paninvasive deciduous tree that is native to China, Taiwan, and northern Korea, but has been spread globally. It is a highly preferred host for SLF and may determine SLF establishment potential.

Supplementary methods: Confirmation of relationship between import tonnage and SLF invasion status for transport potential

The prevailing hypothesis on SLF transport potential is that regions that import more tonnage of commodities from the invaded US region also import more total tonnage of goods and trade infrastructure (e.g., cargo containers, pallets, railcars) that inadvertently transport SLF egg masses long-distances[6,8–12]. To test this hypothesis, we fit two logistic regressions with our metric of transport potential as the covariate. This metric was the $\log_{10}$ of the average annual metric total tonnage imported between 2012 and 2017 from US states invaded by SLF (main text Fig. 3). We regressed the presence/absence of established populations and regulatory incidents (i.e., has a state experienced and reported any observations of SLF, dead, moribund, or alive, independent of the presence of established populations?). For both establishment and regulatory incidents, the relationship between SLF-status and our measure of transport potential was significant, thereby providing support for our estimate of SLF transport potential (Supplemental Table 1). These results suggest that total tonnage of imports is a suitable proxy for transport potential until new metrics are developed that include refined pathway analyses.

Supplementary methods: Modeling establishment potential and the influence of chilling periods for diapause

We estimated establishment potential as an ensemble from three global species distribution models (SDMs): a multivariate SDM of TOH (*sdm_toh*), a multivariate SDM of SLF (*sdm_slf1*), and a univariate SDM of SLF that modeled SLF presence on the predicted values from *sdm_toh (sdm_slf2)*. Models were constructed with MaxEnt ver. 3.4.1 by following best practices for estimating unbiased niche models[3,13,14]. We first queried GBIF for TOH and SLF presences on October 20, 2020. For TOH, 67,100 records were obtained and for SLF 3,180 records were obtained[15]. Records were checked for errors, duplicate records removed, and the remaining records were rarefied (spatially filtered) by omitting records <10 km from each other to reduce bias from spatial autocorrelation[16,17]. The result was 8,578 unique, error checked TOH presence records and 325 unique, error checked SLF presence records. Thus, *sdm_toh* was built on 8,578 TOH global presence records, and *sdm_slf1* and *sdm_slf2* were built on 325 SLF presence records (see our research compendium for the data, https://github.com/ieco-lab/slfrsk).

To find the best models that explained TOH and SLF presences, we started with 22 potential covariates hypothesized to influence SLF and TOH global distributions. The covariates included 20 topographic and bioclimatic variables from WorldClim, which is a standard database of covariates used in global SDMs[18,19]. WorldClim has also been used in two previous SDMs for SLF[20,21]. In addition to these 20 covariates, we added Global Forest Canopy Height[22] because SLF feeds on multiple tree species[23], and Global Access to Cities[24] because TOH and SLF are often established along transportation networks[9]. We analyzed these covariates to identify an uncorrelated subset to include in final best-fit SDMs with low model collinearity. To do this, we calculated pairwise Pearson correlations among the 22 covariates, and fit each covariate to SLF and TOH in univariate SDMs (i.e., 44 models in total). We then compared covariates that were highly correlated and retained only the covariates that fit best to the TOH and SLF presences. This reduction of potential covariates resulted six minimally correlated covariates (pairwise absolute Pearson correlations <0.70) that we fit in our models: annual mean temperature (BIO01), mean diurnal temperature range (BIO02), annual precipitation (BIO12), precipitation seasonality (BIO15), elevation (ELEV), and access to cities (ATC).

We fit *sdm_toh* and *sdm_slf1* with these six covariates; *sdm_slf2* was fit from the *sdm_toh* predicted values. The three models were fit under default settings of the MaxEnt program except for the following changes: (1) all features were enabled but still set to “Auto Features”, (2) response curves were created, (3) variable importance was measured via jackknifing (we did not do this for *sdm_slf2* because it was a univariate model), (4) the threshold rule was set to “Minimum Training Presence”, and (5) the number of replicates was set to five for SLF and ten for TOH. This last modification sets the number of *k*-fold cross-validation replicates and determines the test proportion from *k*, thus we validated the three models with *k*-fold cross-validation via evaluation of the receiver operating characteristic of the AUC (area under the curve) and omission error[2,25–27]. For AUC, the fraction of true positives relative to type I error (positive background points) is compared at all possible thresholds for each model[2,25]. The resultant AUCs were assessed relative to a random model where AUC = 0.50, such that values close to 1.00 indicate strong model performance and those ≤ 0.50 suggest poor performance[25]. Given presence only data, measured AUC cannot reach 1.00, but model AUCs that approach 1.00 are considered to perform well[2,28]. Given concerns with model evaluation with AUC[29–31], we also confirmed model performance with average omission error, which is the proportion of presence point(s) predicted with suitability less than the threshold averaged across replicates[26,27].

All three models performed well according to AUC and omission error. Models yielded test AUC values >0.75 while boasting average test omission error rates <0.01, indicating that each model performed better than random and identified areas of known species presence as suitable for the cross-validation partitions. *sdm_toh* had a slightly lower AUC (0.7779) and omission error (0.0003) than *sdm_slf1* (AUC = 0.9828, omission = 0.0064) and *sdm_slf2* (AUC = 0.9675, omission = 0.0032). For both multivariate SDMs, we compared the variable contributions for congruence. The top four contributing variables were the same for both models (ATC, BIO01, BIO12, and BIO15 in descending order). The remaining two variables (ELEV and BIO02) contributed <2% in each model, with ELEV contributing more in *sdm_toh* and BIO02 contributing more in *sdm_slf1*. For *sdm_toh*, two other variables, BIO12 and BIO15 also contributed <2% each but still contributed more than ELEV and BIO02 (for a more detailed comparison, see our research compendium, https://ieco-lab.github.io/slfrsk/).

We averaged our three best-fit models to produce one ensemble image at the 30 arcsecond resolution, and intersected this image with state and country polygons[13]. We then calculated summary statistics for the ensemble pixels within each state and country (mean, median, and maximum). The R function we wrote to perform this task, extract_enm2(), is available with the R companion package, slfrsk (see https://github.com/ieco-lab/slfrsk). Establishment potential for the 50 US states and 223 countries was estimated as the maximum pixel value for each state and country. Results and conclusions with mean and median pixel values instead of max were qualitatively similar (see https://ieco-lab.github.io/slfrsk/).

Although our work suggests widespread establishment potential, SDM-based establishment potential might overestimate suitability in warmer climates if SLF require a chilling period to initiate diapause to complete development[32]. However, recent work suggests that while SLF can diapause as eggs in the invaded US region, native populations across China include sub-tropical regions that do not provide the colder temperatures necessary for completing diapause[33], and SLF in the US do not require diapause to develop[34]. Indeed, under lab conditions, eggs in the US that do not undergo diapause exhibit higher survivorship than those that do undergo diapause[35]. This observation suggests that our global ensemble model does not overestimate SLF establishment potential and instead may be a conservative estimate, especially for warmer regions (main text Fig. 4).

In summary, our estimate of SLF global establishment potential was based on an ensemble of models for SLF and TOH environmental suitabilities. Two previous estimates of SLF global establishment potential have been published but did not include TOH, were not ensemble estimates, and were not built on as many presence records[20,21]. These other estimates also did not include an anthropogenic covariate like Global Access to Cities[24], which we found to be important in determining TOH and SLF environmental suitability. Finally, although our estimate of SLF establishment potential is broadly like these previous estimates (as observed by comparing our map to theirs), it differs in three key ways: we provide our estimate in a finer resolution, our estimate differs across globally important viticultural regions, and we provide the data as open access. To visualize and download our estimate please see our Google Earth Engine app (https://ieco.users.earthengine.app/view/ieco-slf-riskmap).

Supplementary Table 1 Logistic regression of spotted lanternfly (SLF) status on trade with established US states as average annual metric total tonnage demonstrates a significant relationship for all US states and Washington DC. Trade with established states predicts both presence or absence of established SLF populations and record of SLF regulatory incidents (identification of SLF, deceased, moribund, or alive). Logistic regression model coefficients are shown above with standard error below in parentheses.

	Establishment Status	Regulatory Status
Log_{10}(average annual metric tonnage)	5.64^{***}	3.10^{***}
	(2.03)	(0.90)
Constant	-42.74^{***}	-22.64^{***}
	(15.27)	(6.49)
Observations	51	51
Log likelihood	-9.80	-18.15
Akaike information criterion	23.61	40.30
Notes:	$^{***}P < .01$	

Supplemental References

1. Blackburn, T. M. *et al.* A proposed unified framework for biological invasions. *Trends Ecol. Evol.* **26**, 333–339 (2011).
2. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231–259 (2006).
3. Phillips Steven J., Anderson Robert P., Dudík Miroslav, Schapire Robert E., & Blair Mary E. Opening the black box: An open-source release of Maxent. *Ecography* **40**, 887–893 (2017).
4. Gale, G. *Dying on the vine*. (University of California Press, 2011).
5. Ordish, G. *The great wine blight*. (Charles Scribner's Sons, 1972).
6. Urban, J. M. Perspective: Shedding light on spotted lanternfly impacts in the USA. *Pest Manag. Sci.* **76**, 10–17 (2019).
7. Barringer, L. & Ciafré, C. M. Worldwide feeding host plants of spotted lanternfly, with significant additions from North America. *Environ. Entomol.* **49**, 999–1011 (2020).
8. Liu, H. Oviposition substrate selection, egg mass characteristics, host preference, and life history of the spotted lanternfly (Hemiptera: Fulgoridae) in North America. *Environ. Entomol.* **48**, 1452–1468 (2019).
9. Parra, G., Moylett, H. & Bulluck, R. USDA-APHIS-PPQ-CPHST Technical working group summary report spotted lanternfly, *Lycorma delicatula* (White, 1845). (2018).
10. Lee, D.-H., Park, Y.-L. & Leskey, T. C. A review of biology and management of *Lycorma delicatula* (Hemiptera: Fulgoridae), an emerging global invasive species. *J. Asia-Pac. Entomol.* **22**, 589–596 (2019).

11. Liu, H. Seasonal development, cumulative growing degree-days, and population density of spotted lanternfly (Hemiptera: Fulgoridae) on selected hosts and substrates. *Environ. Entomol.* **49**, 1171–1184 (2020).

12. Park, M., Kim, K.-S. & Lee, J.-H. Genetic structure of *Lycorma delicatula* (Hemiptera: Fulgoridae) populations in Korea: Implication for invasion processes in heterogeneous landscapes. *Bull. Entomol. Res.* **103**, 414–424 (2013).

13. Araújo, M. B. & New, M. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **22**, 42–47 (2007).

14. Araújo, M. B. *et al.* Standards for distribution models in biodiversity assessments. *Sci. Adv.* **5**, eaat4858 (2019).

15. GBIF. GBIF: The global biodiversity information facility. https://www.gbif.org/ (2019).

16. Zhu, G., Illan, J. G., Looney, C. & Crowder, D. W. Assessing the ecological niche and invasion potential of the Asian giant hornet. *Proc. Natl. Acad. Sci.* **117**, 24646–24648 (2020).

17. Boria, R. A., Olson, L. E., Goodman, S. M. & Anderson, R. P. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecol. Model.* **275**, 73–77 (2014).

18. Fick, S. E. & Hijmans, R. J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).

19. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).

20. Jung, J.-M., Jung, S., Byeon, D. & Lee, W.-H. Model-based prediction of potential distribution of the invasive insect pest, spotted lanternfly *Lycorma delicatula* (Hemiptera: Fulgoridae), by using CLIMEX. *J. Asia-Pac. Biodivers.* **10**, 532–538 (2017).

21. Wakie, T. T., Neven, L. G., Yee, W. L. & Lu, Z. The establishment risk of *Lycorma delicatula* (Hemiptera: Fulgoridae) in the United States and globally. *J. Econ. Entomol.* **113**, 306–314 (2020).

22. Simard, M., Pinto, N., Fisher, J. B. & Baccini, A. Mapping forest canopy height globally with spaceborne lidar. *J. Geophys. Res. Biogeosciences* **116**, (2011).

23. Barringer, L. E., Donovall, L. R., Spichiger, S.-E., Lynch, D. & Henry, D. The first New World record of *Lycorma delicatula* (Insecta: Hemiptera: Fulgoridae). *Entomol. News* **125**, 20–23 (2015).

24. Weiss, D. J. *et al.* A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **553**, 333–336 (2018).

25. Fielding, A. H. & Bell, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* **24**, 38–49 (1997).

26. Pearson, R. G., Raxworthy, C. J., Nakamura, M. & Peterson, A. T. Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *J. Biogeogr.* **34**, 102–117 (2007).

27. Anderson, R. P. & Gonzalez, I. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecol. Model.* **222**, 2796–2811 (2011).

28. Wiley, E. O., McNyset, K. M., Peterson, A. T., Robins, C. R. & Stewart, A. M. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. **16**, 8 (2003).

29. Peterson, A. T., Papeş, M. & Soberón, J. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecol. Model.* **213**, 63–72 (2008).

30. Lobo, J. M., Jiménez-Valverde, A. & Real, R. AUC: A misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* **17**, 145–151 (2008).

31. Jiménez-Valverde, A. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Glob. Ecol. Biogeogr.* **21**, 498–507 (2012).

32. Shim, J.-K. & Lee, K.-Y. Molecular characterization of heat shock protein 70 cognate cDNA and its upregulation after diapause termination in *Lycorma delicatula* eggs. *J. Asia-Pac. Entomol.* **18**, 709–714 (2015).

33. Du, Z. *et al.* Global phylogeography and invasion history of the spotted lanternfly revealed by mitochondrial phylogenomics. *Evol. Appl.* (2021) doi:https://doi.org/10.1111/eva.13170.

34. Xin, B. *et al.* Exploratory survey of spotted lanternfly (Hemiptera: Fulgoridae) and its natural enemies in China. *Environ. Entomol.* **50**, 36–45 (2020).

35. Keena, M. A. & Nielsen, A. L. Comparison of the hatch of newly laid *Lycorma delicatula* (Hemiptera: Fulgoridae) eggs from the United States after exposure to different temperatures and durations of low temperature. *Environ. Entomol.* 1–8 (2021) doi:10.1093/ee/nvaa177.