

1 **Supplementary Information for**

2 Paninvasion severity assessment of a U.S. grape pest to disrupt the global wine market

3

4 Nicholas A. Huron*

5 Jocelyn E. Behm

6 Matthew R. Helmus

7

8 Integrative Ecology Lab, Center for Biodiversity, Department of Biology, Temple University,

9 Philadelphia, PA 19122 USA.

10

11 * corresponding author, nahuron@temple.edu

12

13 **This PDF file includes:**

14 Supplementary Methods

15 Supplementary Table 1

16 Supplementary References

17

18

19 **Supplementary Methods**

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

22

23 *Term Definitions*

- 24 • **alignment correlation**—multivariate relationship among invasion potentials.
- 25 • **establishment potential**—likelihood of a region to contain suitable habitat for
 26 transported individuals of a non-native species to form a spreading population.
- 27 • **impact potential**—likelihood of a region to experience negative economic effects from
 28 an established non-native species.
- 29 • **invasion potentials**—likelihoods of a species to move through stages in an invasion
 30 process across regions¹. We focus on the main stages: transport, establishment, impact.
- 31 • **MaxEnt**—abbreviation for maximum entropy, a presence-only SDM methodology that
 32 uses machine-learning to estimate the probability distribution of maximum entropy based
 33 on environmental variables and species occurrence records^{2,3}.
- 34 • **paninvasion**—invasion of a species at the global scale that disturbs a global economic
 35 market.
- 36 • **paninvasion risk**—the likelihood of a regional invasive species to become a globally
 37 invasive species and cause economic repercussions.
- 38 • **paninvasive species**—globally invasive species that goes through the three main
 39 invasion stages and thus can disturb global economic markets.
- 40 • **phylloxera**—*Daktulosphaira vitifoliae* is grapevine root pest native to North America
 41 that was responsible for the Great Wine Blight of the late 1800's, which was the largest
 42 economic disturbance to the global wine market ever recorded. The disruption was
 43 mitigated by widespread planting of European vines that were grafted to North American
 44 grapevine root stocks. The paninvasion of phylloxera continues to this day^{4,5}.
- 45 • **species distribution model (SDM)**—spatial model used to predict the environmental
 46 niche, habitat suitability, and establishment potential of a species.
- 47 • **spotted lanternfly (SLF)**—*Lycorma delicatula* is a planthopper native to China,
 48 Vietnam, and India. It invaded South Korea and Japan in the early 2000's and the

49 northeastern U.S. ca. 2014. It is known to feed on >100 different host species, including
 50 grapes^{6,7}.

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

56

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

59 The prevailing hypothesis on SLF transport potential is that regions that import more
 60 tonnage of commodities from the invaded U.S. region also import more total tonnage of goods
 61 and trade infrastructure (e.g., cargo containers, pallets, railcars) that inadvertently transport SLF
 62 egg masses long-distances. SLF propagules have been found hitchhiking on and in shipments of
 63 pharmaceutical containers, baking ingredients, paint shipments, building materials, boxes of
 64 pumpkins, pallets and many other commodities^{6,8–12}. To test if total tonnage can explain the
 65 current spread of SLF, we fit two logistic regressions with our metric of transport potential based
 66 on total tonnage as the covariate. This metric was the log₁₀ of the average annual metric total
 67 tonnage imported between 2012 and 2017 from U.S. states invaded by SLF (main text Fig. 3).
 68 We regressed the presence/absence of established populations and regulatory incidents (i.e., has
 69 a state experienced and reported any observations of SLF, dead, moribund, or alive, independent
 70 of the presence of established populations?). For both establishment and regulatory incidents, the
 71 relationship between SLF-status and our measure of transport potential was significant, thereby
 72 providing support for our estimate of SLF transport potential (Supplemental Table 1). These
 73 results suggest that total tonnage of imports is a suitable proxy for transport potential until new
 74 metrics are developed that include refined pathway analyses.

75

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

78 We estimated establishment potential as an ensemble from three global species
 79 distribution models (SDMs): a multivariate SDM of TOH (*sdm_toh*), a multivariate SDM of SLF

80 (*sdm_slf1*), and a univariate SDM of SLF that modeled SLF presence on the predicted values
81 from *sdm_toh* (*sdm_slf2*). Models were constructed with MaxEnt ver. 3.4.1 by following best
82 practices for estimating unbiased niche models^{3,13,14}. We first queried GBIF for TOH and SLF
83 presences on October 20, 2020. For TOH, 67,100 records were obtained and for SLF 3,180
84 records were obtained¹⁵. Records were checked for errors, duplicate records removed, and the
85 remaining records were rarefied (spatially filtered) by omitting records <10 km from each other
86 to reduce bias from spatial autocorrelation^{16,17}. The result was 8,022 unique, error checked TOH
87 presence records and 325 unique, error checked SLF presence records. Thus, *sdm_toh* was built
88 on 8,022 TOH global presence records, and *sdm_slf1* and *sdm_slf2* were built on 325 SLF
89 presence records (see our research compendium and Dryad repository for the data,
90 <https://github.com/ieco-lab/slfrsk> and <https://doi.org/10.5061/dryad.msbcc2g1b>).

91 To find the best models that explained TOH and SLF presences, we started with 22
92 potential covariates hypothesized to influence SLF and TOH global distributions. The covariates
93 included 20 topographic and bioclimatic variables from WorldClim, which is a standard database
94 of covariates used in global SDMs^{18,19}. WorldClim has also been used in two previous SDMs for
95 SLF^{20,21}. In addition to these 20 covariates, we added Global Forest Canopy Height²² because
96 SLF feeds on multiple tree species²³, and Global Access to Cities²⁴ because TOH and SLF are
97 often established along transportation networks⁹. We analyzed these covariates to identify an
98 uncorrelated subset to include in final best-fit SDMs with low model collinearity. To do this, we
99 calculated pairwise Pearson correlations among the 22 covariates, and fit each covariate to SLF
100 and TOH in univariate SDMs (i.e., 44 models in total). We then compared covariates that were
101 highly correlated and retained only the covariates that fit best to the TOH and SLF presences.
102 This reduction of potential covariates resulted six minimally correlated covariates (pairwise
103 absolute Pearson correlations <0.70) that we fit in our models: annual mean temperature
104 (BIO01), mean diurnal temperature range (BIO02), annual precipitation (BIO12), precipitation
105 seasonality (BIO15), elevation (ELEV), and access to cities (ATC).

106 We fit *sdm_toh* and *sdm_slf1* with these six covariates; *sdm_slf2* was fit from the
107 *sdm_toh* predicted values. The three models were fit under default settings of the MaxEnt
108 program except for the following changes: (1) all features were enabled but still set to “Auto
109 Features”, (2) response curves were created, (3) variable importance was measured via
110 jackknifing (we did not do this for *sdm_slf2* because it was a univariate model), (4) the threshold

111 rule was set to “Minimum Training Presence”, and (5) the number of replicates was set to five
112 for SLF and ten for TOH. This last modification sets the number of k -fold cross-validation
113 replicates and determines the test proportion from k , thus we validated the three models with k -
114 fold cross-validation via evaluation of the receiver operating characteristic of the AUC (area
115 under the curve) and omission error^{2,25-27}. For AUC, the fraction of true positives relative to type
116 I error (positive background points) is compared at all possible thresholds for each model^{2,25}. The
117 resultant AUCs were assessed relative to a random model where $AUC = 0.50$, such that values
118 close to 1.00 indicate strong model performance and those ≤ 0.50 suggest poor performance²⁵.
119 Given presence only data, measured AUC cannot reach 1.00, but model AUCs that approach
120 1.00 are considered to perform well^{2,28}. Given concerns with model evaluation with AUC²⁹⁻³¹,
121 we also confirmed model performance with average omission error, which is the proportion of
122 presence point(s) predicted with suitability less than the threshold averaged across replicates^{26,27}.

123 All three models performed well according to AUC and omission error. Models yielded
124 test AUC values >0.75 while boasting average test omission error rates <0.01 , indicating that
125 each model performed better than random and identified areas of known species presence as
126 suitable for the cross-validation partitions. *sdm_toh* had a slightly lower AUC (0.7779) and
127 omission error (0.0003) than *sdm_slf1* (AUC = 0.9828, omission = 0.0064) and *sdm_slf2* (AUC
128 = 0.9675, omission = 0.0032). For both multivariate SDMs, we compared the variable
129 contributions for congruence. The top four contributing variables were the same for both models
130 (ATC, BIO01, BIO12, and BIO15 in descending order). The remaining two variables (ELEV and
131 BIO02) contributed $<2\%$ in each model, with ELEV contributing more in *sdm_toh* and BIO02
132 contributing more in *sdm_slf1*. For *sdm_toh*, two other variables, BIO12 and BIO15 also
133 contributed $<2\%$ each but still contributed more than ELEV and BIO02 (for a more detailed
134 comparison, see our research compendium, <https://ieco-lab.github.io/slfrsk/>).

135 We averaged our three best-fit models to produce one ensemble image at the 30
136 arcsecond resolution, and intersected this image with state and country polygons¹³. We then
137 calculated summary statistics for the ensemble pixels within each state and country (mean,
138 median, and maximum). The R function we wrote to perform this task, `extract_enm2()`, is
139 available with the R companion package, `slfrsk` (see <https://github.com/ieco-lab/slfrsk>).
140 Establishment potential for the 50 U.S. states and 223 countries was estimated as the maximum

141 pixel value for each state and country. Results and conclusions with mean and median pixel
142 values instead of max were qualitatively similar (see <https://ieco-lab.github.io/slfrsk/>).

143 Although our work suggests widespread establishment potential, SDM-based
144 establishment potential might overestimate suitability in warmer climates if SLF require a
145 chilling period to initiate diapause to complete development³². However, recent work suggests
146 that while SLF can diapause as eggs in the invaded U.S. region, native populations across China
147 include sub-tropical regions that do not provide the colder temperatures necessary for completing
148 diapause³³, and SLF in the U.S. do not require diapause to develop³⁴. Indeed, under lab
149 conditions, eggs in the U.S. that do not undergo diapause exhibit higher survivorship than those
150 that do undergo diapause³⁵. This observation suggests that our global ensemble model does not
151 overestimate SLF establishment potential and instead may be a conservative estimate, especially
152 for warmer regions (main text Fig. 4).

153 In summary, our estimate of SLF global establishment potential was based on an
154 ensemble of models for SLF and TOH environmental suitabilities. Two previous estimates of
155 SLF global establishment potential have been published but did not include TOH, were not
156 ensemble estimates, and were not built on as many presence records^{20,21}. These other estimates
157 also did not include an anthropogenic covariate like Global Access to Cities²⁴, which we found to
158 be important in determining TOH and SLF environmental suitability. Finally, although our
159 estimate of SLF establishment potential is broadly like these previous estimates (as observed by
160 comparing our map to theirs), it differs in three key ways: we provide our estimate in a finer
161 resolution, our estimate differs across globally important viticultural regions, and we provide the
162 data as open access. To visualize and download our estimate please see our Google Earth Engine
163 app (<https://ieco.users.earthengine.app/view/ieco-slf-riskmap>).

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

	Establishment Status	Regulatory Status
Log ₁₀ (average annual metric tonnage)	5.64*** (2.03)	3.10*** (0.90)
Constant	-42.74*** (15.27)	-22.64*** (6.49)
<i>Observations</i>	51	51
<i>Log likelihood</i>	-9.80	-18.15
<i>Akaike information criterion</i>	23.61	40.30

Notes: *** $P < .01$

171 **Supplementary References**

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

- 192 11. Liu, H. Seasonal development, cumulative growing degree-days, and population density of
193 spotted lanternfly (Hemiptera: Fulgoridae) on selected hosts and substrates. *Environ.*
194 *Entomol.* **49**, 1171–1184 (2020).
- 195 12. Park, M., Kim, K.-S. & Lee, J.-H. Genetic structure of *Lycorma delicatula* (Hemiptera:
196 Fulgoridae) populations in Korea: Implication for invasion processes in heterogeneous
197 landscapes. *Bull. Entomol. Res.* **103**, 414–424 (2013).
- 198 13. Araújo, M. B. & New, M. Ensemble forecasting of species distributions. *Trends Ecol. Evol.*
199 **22**, 42–47 (2007).
- 200 14. Araújo, M. B. *et al.* Standards for distribution models in biodiversity assessments. *Sci. Adv.*
201 **5**, eaat4858 (2019).
- 202 15. Derived dataset GBIF.org. Filtered export of GBIF occurrence data. (2021)
203 doi:10.15468/DD.KS6ACS.
- 204 16. Zhu, G., Illan, J. G., Looney, C. & Crowder, D. W. Assessing the ecological niche and
205 invasion potential of the Asian giant hornet. *Proc. Natl. Acad. Sci.* **117**, 24646–24648
206 (2020).
- 207 17. Boria, R. A., Olson, L. E., Goodman, S. M. & Anderson, R. P. Spatial filtering to reduce
208 sampling bias can improve the performance of ecological niche models. *Ecol. Model.* **275**,
209 73–77 (2014).
- 210 18. Fick, S. E. & Hijmans, R. J. WorldClim 2: New 1-km spatial resolution climate surfaces for
211 global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
- 212 19. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution
213 interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).

- 214 20. Jung, J.-M., Jung, S., Byeon, D. & Lee, W.-H. Model-based prediction of potential
215 distribution of the invasive insect pest, spotted lanternfly *Lycorma delicatula* (Hemiptera:
216 Fulgoridae), by using CLIMEX. *J. Asia-Pac. Biodivers.* **10**, 532–538 (2017).
- 217 21. Wakie, T. T., Neven, L. G., Yee, W. L. & Lu, Z. The establishment risk of *Lycorma*
218 *delicatula* (Hemiptera: Fulgoridae) in the United States and globally. *J. Econ. Entomol.* **113**,
219 306–314 (2020).
- 220 22. Simard, M., Pinto, N., Fisher, J. B. & Baccini, A. Mapping forest canopy height globally
221 with spaceborne lidar. *J. Geophys. Res. Biogeosciences* **116**, (2011).
- 222 23. Barringer, L. E., Donovall, L. R., Spichiger, S.-E., Lynch, D. & Henry, D. The first New
223 World record of *Lycorma delicatula* (Insecta: Hemiptera: Fulgoridae). *Entomol. News* **125**,
224 20–23 (2015).
- 225 24. Weiss, D. J. *et al.* A global map of travel time to cities to assess inequalities in accessibility
226 in 2015. *Nature* **553**, 333–336 (2018).
- 227 25. Fielding, A. H. & Bell, J. F. A review of methods for the assessment of prediction errors in
228 conservation presence/absence models. *Environ. Conserv.* **24**, 38–49 (1997).
- 229 26. Pearson, R. G., Raxworthy, C. J., Nakamura, M. & Peterson, A. T. Predicting species
230 distributions from small numbers of occurrence records: A test case using cryptic geckos in
231 Madagascar. *J. Biogeogr.* **34**, 102–117 (2007).
- 232 27. Anderson, R. P. & Gonzalez, I. Species-specific tuning increases robustness to sampling bias
233 in models of species distributions: An implementation with Maxent. *Ecol. Model.* **222**, 2796–
234 2811 (2011).

- 235 28. Wiley, E. O., McNyset, K. M., Peterson, A. T., Robins, C. R. & Stewart, A. M. Niche
236 modeling and geographic range predictions in the marine environment using a machine-
237 learning algorithm. **16**, 8 (2003).
- 238 29. Peterson, A. T., Papeş, M. & Soberón, J. Rethinking receiver operating characteristic
239 analysis applications in ecological niche modeling. *Ecol. Model.* **213**, 63–72 (2008).
- 240 30. Lobo, J. M., Jiménez-Valverde, A. & Real, R. AUC: A misleading measure of the
241 performance of predictive distribution models. *Glob. Ecol. Biogeogr.* **17**, 145–151 (2008).
- 242 31. Jiménez-Valverde, A. Insights into the area under the receiver operating characteristic curve
243 (AUC) as a discrimination measure in species distribution modelling. *Glob. Ecol. Biogeogr.*
244 **21**, 498–507 (2012).
- 245 32. Shim, J.-K. & Lee, K.-Y. Molecular characterization of heat shock protein 70 cognate cDNA
246 and its upregulation after diapause termination in *Lycorma delicatula* eggs. *J. Asia-Pac.*
247 *Entomol.* **18**, 709–714 (2015).
- 248 33. Du, Z. *et al.* Global phylogeography and invasion history of the spotted lanternfly revealed
249 by mitochondrial phylogenomics. *Evol. Appl.* (2021) doi:<https://doi.org/10.1111/eva.13170>.
- 250 34. Xin, B. *et al.* Exploratory survey of spotted lanternfly (Hemiptera: Fulgoridae) and its natural
251 enemies in China. *Environ. Entomol.* **50**, 36–45 (2020).
- 252 35. Keena, M. A. & Nielsen, A. L. Comparison of the hatch of newly laid *Lycorma delicatula*
253 (Hemiptera: Fulgoridae) eggs from the United States after exposure to different temperatures
254 and durations of low temperature. *Environ. Entomol.* 1–8 (2021) doi:10.1093/ee/nvaa177.
255