Geographic potential of the world's largest hornet, *Vespa mandarinia* Smith (Hymenoptera: Vespidae), worldwide and particularly in North America

Claudia Nuñez-Penichet [1, 2], **Luis Osorio-Olvera** [2, 3], **Victor H. Gonzalez** [1, 4], **Marlon E. Cobos** [1, 2], **Laura Jiménez** [1, 2], **Devon A. DeRaad** [1, 2], **Abdelghafar Alkishe** [1, 2], **Rusby G. Contreras-Díaz** [3, 5], **Angela Nava-Bolaños** [2], **Kaera Utsumi** [1], **Uzma Ashraf** [6], **Adeola Adeboje** [1], **A. Townsend Peterson** [1, 2], **Jorge Soberon** [Corresp. 1, 2]

[1] Department of Ecology & Evolutionary Biology, University of Kansas, Lawrence, KS, United States

[2] Biodiversity Institute, University of Kansas, Lawrence, Kansas, United States

[3] Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad de México, Ciudad de México, Mexico

[4] Undergraduate Biology Program, University of Kansas, Lawrence, KS, United States

[5] Posgrado en Ciencias Biológicas. Unidad de Posgrado, Universidad Nacional Autónoma de México, Ciudad de México, Ciudad de México, México

[6] Department of Environmental Sciences and Policy, Lahore School of Economics, Lahore, Pakistan

Corresponding Author: Jorge Soberon
Email address: jsoberon@ku.edu

The Asian giant hornet (AGH, *Vespa mandarinia*) is the world's largest hornet, occurring naturally in the Indomalayan region, where it is a voracious predator of pollinating insects including honey bees. In September 2019, a nest of Asian giant hornets was detected outside of Vancouver, British Columbia, multiple individuals were detected in British Columbia and Washington state in 2020, and another nest was found and eradicated in Washington state in November 2020, indicating that the AGH could have successfully wintered in North America. Because hornets tend to spread rapidly and become pests, reliable estimates of the potential invasive range of *V. mandarinia* in North America are needed to assess likely human and economic impacts, and to guide future eradication attempts. Here, we assess climatic suitability for AGH in North America, and suggest that, without control, this species could establish populations across the Pacific Northwest and much of eastern North America. Predicted suitable areas for AGH in North America overlap broadly with areas where honey production is highest, as well as with species-rich areas for native bumble bees and stingless bees of the genus *Melipona* in Mexico, highlighting the economic and environmental necessity of controlling this nascent invasion.

Geographic potential of the world's largest hornet, *Vespa mandarinia* Smith (Hymenoptera: Vespidae), worldwide and particularly in North America

Claudia Nuñez-Penichet[1, 2], Luis Osorio-Olvera[2, 3], Victor H. Gonzalez[1, 4], Marlon E. Cobos[1, 2], Laura Jiménez[1, 2], Devon A. DeRaad[1, 2], Abdelghafar Alkishe[1, 2], Rusby G. Contreras-Díaz[3, 5], Angela Nava-Bolaños[2], Kaera Utsumi[1], Uzma Ashraf[6], Adeola Adeboje[1], A. Townsend Peterson[1, 2], Jorge Soberon[1, 2*]

[1]Department of Ecology & Evolutionary Biology, University of Kansas, Lawrence, KS, USA.

[2]Biodiversity Institute, University of Kansas, Lawrence, KS, USA.

[3]Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad de México, Ciudad de México, México.

[4]Undergraduate Biology Program, University of Kansas, Lawrence, KS, USA.

[5]Posgrado en Ciencias Biológicas. Unidad de Posgrado, Universidad Nacional Autónoma de México, Ciudad de México, Ciudad de México, México.

[6]Department of Environmental Sciences and Policy, Lahore School of Economics, Lahore, Pakistan.

[*]Corresponding Author: Jorge Soberon

Department of Ecology & Evolutionary Biology and Biodiversity Institute, University of Kansas, 1345 Jayhawk Blvd., Lawrence, Kansas 66045, USA. jsoberon@ku.edu

ORCID:

CNP: https://orcid.org/0000-0001-7442-8593

LOO: https://orcid.org/0000-0003-0701-5398

VHG: https://orcid.org/0000-0002-4146-1634

MEC: https://orcid.org/0000-0002-2611-1767

LJ: https://orcid.org/0000-0002-6683-9576

DAD: https://orcid.org/0000-0003-3105-985X

AbA: https://orcid.org/0000-0003-2927-514X

RGCD: https://orcid.org/0000-0002-0569-8984

ANB: https://orcid.org/0000-0002-4371-5415

KU: https://orcid.org/0000-0001-5935-7299

UA: https://orcid.org/0000-0003-4319-9315

AdA: https://orcid.org/0000-0001-8513-7804

ATP: http://orcid.org/0000-0003-0243-2379

JS: https://orcid.org/0000-0003-2160-4148

ABSTRACT

The Asian giant hornet (AGH, *Vespa mandarinia*) is the world's largest hornet, occurring naturally in the Indomalayan region, where it is a voracious predator of pollinating insects including honey bees. In September 2019, a nest of Asian giant hornets was detected outside of Vancouver, British Columbia, multiple individuals were detected in British Columbia and Washington state in 2020, and another nest was found and eradicated in Washington state in November 2020, indicating that the AGH could have successfully wintered in North America. Because hornets tend to spread rapidly and become pests, reliable estimates of the potential invasive range of *V. mandarinia* in North America are needed to assess likely human and economic impacts, and to guide future eradication attempts. Here, we assess climatic suitability for AGH in North America, and suggest that, without control, this species could establish populations across the Pacific Northwest and much of eastern North America. Predicted suitable areas for AGH in North America overlap broadly with areas where honey production is highest, as well as with species-rich areas for native bumble bees and stingless bees of the genus *Melipona* in Mexico, highlighting the economic and environmental necessity of controlling this nascent invasion.

Keywords: Asian giant hornet, dispersal simulation, ecological niche modeling, invasive species, pollinator threats

Introduction

Invasive species represent major threats to biodiversity, as they can alter ecosystem processes and functions (Pyšek & Richardson, 2010; Vilà et al., 2011), and often contribute to the decline of imperiled species (e.g., Wilcove et al., 1998; Dueñas et al., 2018). The economic damage to agriculture, forestry, and public health, resulting from invasive species totals nearly US $120 billion annually in the United States alone (Pimentel, Zuniga & Morrison, 2005), and more than US $172 million for Canada (Colautti et al., 2006).

Even in the midst of the global uncertainty and socio-economic distress resulting from the COVID-19 pandemic, the recent detection of the Asian Giant Hornet (AGH, *Vespa mandarinia* Smith, Hymenoptera: Vespidae), in North America (Bérubé, 2020; Wilson et al., 2020), received significant public attention. This social insect is the world's largest hornet (2.5–4.5 cm body length), and occurs naturally across Asia, including in India, Nepal, Sri Lanka, Vietnam, Taiwan, and Japan, at elevations ranging between 850 and 1900 (Matsuura & Sakagami, 1973; Archer, 2008; Smith-Pardo, Carpenter & Kimsey, 2020). As in other temperate-zone social species, annual colonies of the AGH, which may contain up to 500 workers, die at the onset of winter and mated queens overwinter in underground cavities. After emerging in the spring, each queen starts a new colony in a pre-existing cavity, typically in tree roots or an abandoned rodent nest (Archer, 2008). Like other species of *Vespa*, AGH is a voracious predator of insects, and is notable for preying on bees and other social Hymenoptera. Attacks on honey bee hives occur late in the development of the hornet colony and prior to the emergence of reproductive individuals (males and new queens), the timing of which depends on location (e.g., Matsuura & Sakagami, 1973; Matsuura, 1988; Archer, 2008).

In its native range, AGH attacks several species of bees, some of which have developed sophisticated defense mechanisms against attacks (Ono et al., 1995; Kastberger, Schmelzer & Kranner, 2008; Fujiwara, Sasaki & Washitani, 2016). The best documented, colony-level defense mechanism is in the Asiatic honey bee, *Apis cerana* Fabricius, which can detect site-marking pheromones released by AGH scouts, and responds by engulfing a single hornet in a ball consisting of up to 500 bees. The heat generated by the vibration of the bees' flight muscles, and the resulting high levels of CO_2 from respiration effectively kill the hornet (Ono et al., 1995; Sugahara, Nishimura & Sakamoto, 2012). In contrast, European honey bees (*A. mellifera* L.) does not detect and respond to AGH marking pheromones, and colonies are mostly defenseless against AGH attacks (McClenaghan et al., 2019). In Japan, as few as a dozen AGH can destroy a European honey bee colony of up to 30,000 individuals, and extirpate thousands of beehives annually (Matsuura & Sakagami, 1973).

In addition to the potential threat to the beekeeping industry (Alanis et al. 2020), the introduction of AGH in North America is also concerning for public health. Their powerful stings can induce severe allergic reactions or even death in hypersensitive individuals and sometimes have long-term health efects in people who receive multiple stings (Schmidt et al., 1986; Yanagawa et al., 2007). Annually, as many as 30-40 people may have die from AGH stings in Japan, most as a result of anaphylaxis or sudden cardiac arrest (Matsuura & Sakagami, 1973); similar deadly cases have been reported from China during outbreak events (Li et al., 2015). Although invasive species are typically limited by dispersal ability and suitability of novel environments, vespid hornets are well known for their invasive success and excellent dispersal capacity (Beggs et al., 2011; Monceau, Bonnard & Thiéry, 2014). As such, the introduction of AGH in the Pacific Northwest is already presenting a potentially serious ecological and socio-

economic risk in North America. Here, we use ecological niche modeling (ENM) to detect areas of suitable environments for this species worldwide, with particular emphasis on North America. We also use a dispersal simulation approach to detect potential invasion paths of this species within North America. A similar methodology for projecting AGH invasion potential has been implemented by Zhu et al. (2020); we build upon this framework by introducing several modifications to the modelling approach, and investigating further the potential ecological impact on the species richness of the genus *Bombus* and *Melipona* as well as the possible economic effects of an AGH invasion in North America.

Methods

Occurrence and environmental data

We downloaded occurrence data for *V. mandarinia* from the Global Biodiversity Information Facility database (GBIF; https://www.gbif.org/). We kept records from the species' native range (Fig. 1) separate from non-native occurrences facilitated by human introduction. We cleaned occurrences from the native distribution following Cobos et al. (2018) by removing duplicates and records with inconsistent georeferencing (coordinates outside country limits, on the sea, or missing, as recommended in the literature of data cleaning; Chapman, 2005). To avoid model fitting derived from spatial autocorrelation and overdominance of specific regions due to sampling bias, we thinned these records spatially in two ways: by geographic distance and by density of records per country (Fig. 2). In the first case (distance-based thinning; Anderson, 2012), we excluded occurrences that were <50 km away from another occurrence record. In the second thinning approach (country-density thinning), we accounted for potential differences in survey and reporting activities among countries and randomly reduced numbers of occurrences in countries with the densest sampling, namely Japan, Taiwan, and South Korea (from 30, 6, and

5, to 6, 2, and 2 occurrences, respectively), to match the approximate reference density of India, Nepal, and China. We used the package ellipsenm (Cobos et al., 2020; available at https://github.com/marlonecobos/ellipsenm) in R 3.6.2 (R Core Team, 2019) to clean and thin the data. We retained 172 occurrence records for *V. mandarinia* after initial data cleaning, 49 records after the distance-based thinning approach, and 18 records after the country-density thinning approach (Fig. 1). We then treated both data sets independently in all subsequent analysis steps.

For environmental predictors, we used bioclimatic variables at 10' resolution (~18 km at the Equator) from the MERRAclim database (Vega, Pertierra & Olalla-Tárraga, 2018). We excluded four variables because they are known to contain spatial artifacts as a result of combining temperature and humidity information (Escobar et al., 2014): mean temperature of most humid quarter, mean temperature of least humid quarter, specific humidity mean of warmest quarter, and specific humidity mean of coldest quarter. The 15 variables remaining were masked to an area for model calibration (**M**, see Ecological niche modeling).

These 15 variables were further processed into two subsets, each used separately in all subsequent analyses. One set consisted of submitting the variables to a principal component analysis (PCA) to reduce dimensionality and multicollinearity. The other set was created with the raw environmental variables that had a Pearson's correlation coefficient ≤0.85 measured in the calibration area, choosing the most biologically relevant or interpretable variables based on our knowledge of AGH natural history (Simões et al., 2020). As a result of this selection process, we obtained six raw variables: isothermality (BIO3), maximum temperature of warmest month (BIO5), minimum temperature of coldest month (BIO6), temperature annual range (BIO7), specific humidity of most humid month (BIO13), and specific humidity of least humid month

(BIO14). The PCA was performed with the raw variables masked to the M area, and principal components for the entire world were obtained by transforming raw variables (at world extent) using the scaling and rotations from the PCA obtained for M. Projecting the PCA results from the accessible area to the entire world ensures that values are comparable and prevents further problems when models are projected. For further analyses, we kept the first four PC axes, as they explained 97.9% of the cumulative variance (Figure S1). All analyses were done in R; specifically, raster processing was completed using the packages raster (Hijmans et al., 2020), rgeos (Bivand et al., 2020b), and rgdal (Bivand et al., 2020a); PCA was performed using the ntbox package (Osorio-Olvera et al., 2020).

Ecological niche modeling

To identify a calibration area (ostensibly equivalent to **M**; Owens et al., 2013) for our models, we considered a region contained within a buffer of 500 km around the known occurrence records after the 50 km thinning process (Fig. 1). This distance was selected considering that little is known about the species' dispersal ability (Matsuura & Sakagami, 1973; APHIS, 2020a), as no information exists about queens' dispersal capacities; however, we consider that given the species body size, long-distance dispersal events should not be discarded. We used all pixels in **M** (15,411) as the background across which to calibrate the models.

Given uncertainty deriving from specific treatments of occurrence records and environmental predictors in ecological niche modeling (Alkishe et al., 2020), we calibrated models via four distinct schemes: (1) using raw variables and distance-based thinned occurrences, (2) using PCs and distance-based thinned occurrences, (3) using raw environmental

variables and country-density thinned occurrences, and (4) using PCs and country-density thinned occurrences (Fig. 2). For each scheme, we calibrated models five times, each time randomly selecting 50% of the occurrences for calibrating models (random k-fold evaluation, where k = 5), and using the remaining records for testing (Cobos et al., 2019a).

Each process of model calibration consisted of creating and evaluating candidate models using Maxent (Phillips, Anderson & Schapire, 2006; Phillips et al., 2017). Since many choices are needed to parameterize Maxent (Merow, Smith & Silander, 2013), we chose distinct parameter settings: 10 regularization multiplier values (0.10, 0.25, 0.50, 0.75, 1, 2, 3, 4, 5, 6), eight feature classes (lq, lp, lqp, qp, q, lqpt, lqpth, lqph, where l is linear, q is quadratic, p is product, t is threshold, and h is hinge), and all combinations of more than two predictor variables (Cobos et al., 2019b; Table S1-S2). This resulted in a total of 4560 models using raw variables and 880 using PCs that were tested, in tandem with the two subsets of occurrence data described above. We assessed model performance using partial ROC (for statistical significance; Peterson, Papeş & Soberón, 2008), omission rates (E = 5%, for predictive ability; Anderson, Lew & Peterson, 2003), and Akaike Information Criterion corrected for small sample sizes (AICc, for model complexity; Warren & Seifert, 2011). We selected models with delta AICc $\leq$2 (Cobos et al., 2019a) from those that were statistically significant and had omission rates below 5%.

After model calibration, we created models with the selected parameter values, using all occurrences after the corresponding thinning process, with 10 bootstrap replicates, cloglog output (Phillips et al., 2017), and model transfers using three types of extrapolation (free extrapolation, extrapolation with clamping, no extrapolation; Owens et al., 2013). Not all calibration processes identified models that met all three criteria of model selection; we did not consider those models in further analyses (Fig. 2; Table 1). As a final evaluation step, we

191 binarized all model replicates given a modified least presence (5% of omission) and then tested
192 whether each replicate of the selected models was able to anticipate the known invasive records
193 of the species in British Columbia, Canada and Washington, USA. For each scheme, using only
194 those model replicates that met the selection criteria and correctly predicted independent
195 occurrences (known invaded localities in North America; independent testing), we created a
196 consensus per sample and two types of final consensus: (1) a median of the medians obtained for
197 each parameterization (continuous), and (2) the sum of all suitable areas derived from binarizing
198 each replicate using a modified least presence (5% omission) threshold (this represents the
199 number of coincidences; Pearson et at. 2007; Fig. 2). Our exploration of different modeling
200 pipelines allowed us to highlight how different methodologies produce different results.

201 As we transferred models to the entire world, we used the mobility-oriented parity metric
202 (MOP; Owens et al., 2013) to detect areas where strict or combinational extrapolation risks could
203 be expected, given the presence of non-analogous conditions with respect to the environments
204 manifested across the calibration area. The areas where extrapolation risks were detected using
205 MOP were deleted from our binary results (suitable areas) to avoid potentially problematic
206 interpretations based on extrapolative situations. Model calibration, production of selected
207 models with replicates, and MOP analyses were done in R using the package kuenm (Cobos et
208 al., 2019a); raster processing and independent testing of models were done using the package
209 raster in R.

210

Dispersal simulations

We used the binary outputs from the final consensus models (suitable and unsuitable areas, without areas of strict extrapolation) to simulate invasion dynamics of the AGH. All simulations were started from the Pacific Northwest, from sites already known to be occupied by the AGH. The simulations were performed using the cellular automaton dynamic model included in the bam R package (Osorio-Olvera & Soberón, 2020; available at https://github.com/luismurao/bam). Under this discrete model, given an occupied area at time t, two layers of information are needed to obtain the occupied area at time $t+1$: (i) the binary layer of suitability for the species, and (ii) a connectivity matrix determined by the species' ability to reach neighboring cells in one time unit (known as "Moore's neighborhood"; Gray et al., 2003, that defines patches that are connected by dispersal). At each step, each of the suitable cells can be either occupied or not by the species. If a cell is occupied, adjacent cells can be visited by the species, and if suitable, they become occupied. This method is similar to the one implemented in the MigClim R package (Engler, Hordijk & Guisan, 2012), but uses a simpler dispersal kernel and parameterization.

With each of the final consensus models for *V. mandarinia*, we performed a set of simulations in which we explored different degrees of connectivity (1, 2, 4, 8, 10, and 12 neighbor cells pixels per unit time) and different suitability thresholds (10 equidistant levels from 3–10% of the presence points to explore variability in sensitivity to the amount of area classified as suitable) to create the binary maps. Since no information is available about dispersal capacities of queens of *V. mandarinia,* all simulations were done with 200 arbitrary time steps that ensure reaching a steady state. In the end, we visualized the simulation results by summing the occupied distribution layers obtained from each set of simulations. A value of 100 in these

final layers means that the species reached that cell in 100% of the simulations, whereas a value of 0 means that the species never reached that cell. Further details regarding the simulation processes can be found in the Supplementary Information.

Honey production and native bee richness in North America

To explore potential ecological and economic impacts of the invasion of the AGH in North America, we explored annual, state-level production of honey (for Mexico, Untied States, and Canada) as well as species richness of bumble bees (*Bombus* Latreille) and stingless bees of the genus *Melipona* Illiger in Mexico and the United States. We extracted data on 2016 honey production (in US dollars) for the United States from the U.S. Department of Agriculture (USDA; available at https://quickstats.nass.usda.gov/#4A0314DA-F3E5-3B06-ADD1-CA8032FBD937), from the Instituto Nacional de Estadística, Geografía e Informática (INEGI) for Mexico (https://atlasapi2019.github.io/cap4.html), and from the government of Canada website (https://www.agr.gc.ca/eng/horticulture/horticulture-sector-reports/statistical-overview-of-the-canadian-honey-and-bee-industry-2018/?id=1571143699779) for Canada. For native species richness, we obtained a list of species of bumble bees and stingless bees of the genus *Melipona* that occur in Mexico and the United States from Discover Life (https://www.discoverlife.org/) and downloaded their occurrence data from GBIF. We chose these bee taxa as likely targets of AGH because the species in these groups are of similar body size and behavior to the typical prey of these hornets: they are social insects that form annual or perennial colonies that can have a few hundreds to as many as 10,000 individuals (Cueva del Castillo, Sanabria-Urbán & Serrano-Meneses, 2015; Viana et al., 2015), and store honey and pollen inside their nests (Michener, 2000). To summarize species richness of these two genera,

257 we created a presence absence matrix (PAM; Arita et al., 2008) for North America, based on
258 geographic coordinates of occurrence data, with a pixel size of one degree. The PAM was
259 created in R with the package biosurvey (Nuñez-Penichet et al., 2020; available at
260 https://github.com/claununez/biosurvey).

261 To assure transparency and reproducibility of our work, we include an Overview, Data,
262 Model, Assessment, and Prediction protocol (ODMAP; Zurell et al., 2020) in our supplementary
263 materials. This metadata summary provides a detailed key to the steps of our analyses. The data
264 and R code used in this research are openly available at http://hdl.handle.net/1808/30602 and
265 https://github.com/townpeterson/vespa repositories, respectively.

266

267 **Results**

268 *Model calibration*

269 The number of models that met the selection criteria was considerably smaller than the
270 total number of models tested (Table 1, see Supplemental information for more details of the
271 calibration results). The calibration schemes including raw variables had fewer models selected
272 than those using PCs (11, 19, 6, 15 models selected for raw/distance-thinned, PC/distance-
273 thinned, raw/country-density, and PC/country-density, respectively). Not all replicates of
274 selected models anticipated the *V. mandarinia* invaded areas in North America successfully, so
275 we kept only those that predicted all known invasive records. The number of replicates retained
276 varied among distinct calibration schemes and types of extrapolation used (Table 1).

277

Ecological niche model predictions

In our models, areas predicted as suitable for the AGH varied among calibration schemes, in both extension and geographic pattern (Fig. 3, Figures S2-S4). The differences are conspicuous between the two types of thinning approaches, which resulted in models created with different numbers of occurrence records. Models with country-density thinning (18 records) resulted in broad predicted suitable areas worldwide, with areas of higher values of suitability concentrated in tropical regions (Fig. 3, Figures S2-S4). In contrast, models created with the greater number of occurrences (49 records) from the geographic distance thinning predicted more patches of suitable areas across large extensions of Southeast Asia, Europe, West Africa, Central America, northern South America, and the Pacific Northwest and southeastern United States (Fig. 3, Figures S2-S4). In the calibration area, the areas detected with high levels of suitability were larger in the scheme with geographic distance thinned occurrences and the raw variables and smaller in the predictions obtained with the country-density thinned occurrences and the PCs as environmental predictors (Fig. 3). In all schemes, the two northernmost occurrence points of this species in China were accorded relatively low levels of suitability (Fig. 3). Predicted suitable areas for this hornet worldwide were also different among types of extrapolation considered in this study, especially as regards its distribution size rather than location (Figures S2-S4).

In North America, across multiple model calibration schemes, our various models agreed in predicting suitable areas for AGH in the Pacific region of southwestern Canada, the Pacific Northwest, the southeastern United States, and from central Mexico south to southernmost Panama (Fig. 4). Our model calibration schemes also agreed in identifying the Rocky Mountains and Great Plains as unsuitable for this species (Fig. 4).

301 The proportion of area identified as suitable varied among the data thinning schemes. In
302 the case of models created with raw variables, the proportion was 0.171 and 0.164 for spatially
303 thinned and country-density thinned records, respectively. When PCs were used, suitable
304 proportions were 0.248 and 0.239, for spatially thinned and country-density thinned records,
305 respectively (Table S3).

306 *Extrapolation risks in model projections*

307 The pattern of areas detected with risk of extrapolation was similar worldwide between
308 thinning methods, but different between raw variables and PCs (Fig. 5, Figure S8). Most tropical
309 areas predicted as suitable were identified as regions with high extrapolation risk (Figure S8).
310 For raw variables, the areas with extrapolation risk in North America included most of Canada
311 and Alaska, whereas for PCs areas with extrapolation risk included large portions of Mexico and,
312 the central-southwestern United States, as well as the islands north of Hudson Bay in Canada
313 (Fig. 5).

314

315 *Simulations of potential invasion*

316 The simulations of potential sequences of colonization and dispersal of AGH in North
317 America, starting from the known invaded localities, showed agreement among calibration
318 schemes in predicting an invasion across the Pacific Northwest from southernmost Alaska to
319 southernmost California in the United States (Fig. 6). In contrast, we found that the dispersal
320 distance required to invade all the way to the East Coast of North America varied among
321 calibration schemes. In the schemes using raw variables, the route of invasion to reach the East
322 Coast goes from the Pacific Northwest down to California and Mexico, and then up the East

Coast of North America. A dispersal distance of 10 cells (where each cell represents ~18 km) was enough to reach the East Coast (see A and C panels in Fig. 6). For the scheme using the 50 km spatially-thinned occurrences and PCs, the invasion follows a more direct route from the Pacific Northwest to the East Coast that goes through the United States, and the required dispersal distance to reach the East Coast was only 4 cells (Fig. 6B). Finally, in the case of country-density thinned occurrences and PCs, the invasion goes from the Pacific Northwest through Canada to the Atlantic, and then down the East Coast to the United States. A distance of 8 cells was needed to make this invasion route possible (Fig. 6D).

Honey production and native bee richness

The areas in North America that our models identified as highly suitable for AGH overlapped broadly with the states where honey production is highest. This overlap was particularly noticeable in southern Mexico and in some states of the Pacific Northwest and eastern US (Fig. 7). We found a similar pattern with the species richness of *Bombus* and *Melipona*.

Discussion

The patterns of suitability that we found in North America across multiple input data processing schemes are broadly concurrent with the results obtained by Zhu et al. (2020) and Alaniz et al. (2020) (Fig. 6), who used an ensemble modeling approach for the potential invasion of AGH. This concordance with the results of these works (both among our selected models, and

between our models and the ensemble models), gives us confidence that the Pacific Northwest and southeastern United States represent suitable areas for AGH. In contrast with the results of Zhu et al. (2020), however, our dispersal simulations indicate a larger potential invasion area in the United States, with the AGH potentially crossing to eastern North America via a southern invasion route, through Mexico and Texas; a southeast-ward route crossing Idaho, Wyoming, and Colorado; or a northern route across Canada and the Great Lakes region (Fig. 6).

Quantifying the probability of the AGH following any one of the individual dispersal routes presented would require precise quantification of dispersal ability, and discerning the real-world validity of each of the four modeling outcomes. Instead of attempting to guess, we present several models that offer multiple plausible invasion scenarios. Across all scenarios presented, the AGH is expected to establish populations along the coastal Pacific Northwest via short-distance dispersal, and it is likely to invade the southeastern United States if it has even moderate dispersal potential (Fig. 6). It is important to note that these potential invasion routes consider only the natural dispersal ability of this hornet, and do not take into account the effect of potential accidental human-aided dispersal through the transport of soil and wood, where fertilized queen AGHs overwinter (Archer, 1995). Such unwitting human-aided dispersal is a serious concern, as it could potentiate a rapid invasion of this hornet to environmentally suitable, yet currently isolated places across North America. Our simulations allowing AGH to disperse to larger numbers of neighbor cells are perhaps a good illustration of what could be expected if dispersal events to very long distances occur.

Contrasts between our prediction of extensive invasion potential, and Zhu et al.'s (2020) more conservative predictions, arise from Zhu et al.'s (2020) use of MigClim (Engler, Hordijk & Guisan, 2012) to model dispersal of the AGH in western North America. MigClim is a cellular

automaton platform that models the state of grid cells as occupied or unoccupied. Although we used the same modelling technique, our dispersal kernel is a much simpler "Moore Neighborhood" (Gray et al., 2003) approach, in which cells surrounding an occupied focal cell (to 1,2...*d* neighbors) may become occupied, depending on their suitability. MigClim instead assumes a probabilistic contagion model that requires parameter estimates for number of propagules, and short- and long-distance-decay rates. Given the lack of empirical data to inform values for those parameters, we prefer a simpler algorithm to explore how connected clusters of suitable cells are across different values of the single parameter *d*. Another factor resulting in these differences is the number of simulation steps used in our approach (200). From a biological perspective, this implies that 200 dispersal events resulting in colonization of suitable cells happened. Although this number may appear excessive, it gives a view of a scenario in which no action is taken to prevent AGH invasion in North America and the species builds to large local populations. For a more conservative view of the expected invasion, one could concentrate in areas with high values of suitability on the layers obtained from our simulations.

The areas in North America that our models identified as highly suitable for this hornet overlap broadly with the states where honey production is highest, and species richness of *Bombus* and *Melipona* are highest (Fig. 7). These results give credence to public concerns that, if established, the AGH could pose a serious economic threat to the beekeeping industry in Oregon, northern California, Georgia, Alabama, and Florida. In the United States alone, the European honey bee provides at least $15 billion worth of pollination services and generates between $300 and 500 million in harvestable honey and other products each year (Calderone, 2012). Indeed, Alaniz et al. (2020) estimate that if spread across the US, the AGH could threaten between 11 and 100 million dollars for hive derived products and honey bee-pollinated crops production. In

Mexico, impacts on the honey bee industry are also expected in tropical areas of the country that have suitable areas for the AGH, particularly in the states of Yucatán, Campeche, and Quintana Roo. Beekeepers in the United States and Mexico may have to adopt mitigation practices to avoid serious losses, such as those developed by Japanese beekeepers including the use of protective screens or traps at the hive entrance that can exclude AGHs based on body size (Matsuura & Sakagami, 1973; Mahdi, Glaiim & Ibrahim, 2008). Potential establishment of the AGH in North America adds an additional layer of environmental and economic stress to a beekeeping industry already suffering from high annual hive mortality rates resulting from the combined effects of pesticides, diseases, and poor nutrition (Goulson et al., 2015).

The ecological impact of AGH on the local bee fauna is more challenging to predict than the economic impact on honey production, because it is not clear which native bee species would be particularly targeted by AGH in North America. We explore *Bombus* and *Melipona* species as potential prey candidates of AGH because, among the >4000 bee species occurring in this region (Ascher & Pickering, 2020), these two groups of bees are social, locally abundant, and make annual or perennial colonies (Michener, 2007; Cueva del Castillo, Sanabria-Urbán & Serrano-Meneses, 2015; Viana et al., 2015). Thus, they may represent predictable food sources for the AGH, particularly in areas where bee colonies remain active year-round. It is crucial to consider this potential threat because both *Bombus* and *Melipona* bees are important pollinators that have already experienced population losses and local extirpations, reflecting changes in landscape and agricultural intensification (Brown & Albrecht, 2001; Cameron et al., 2011). Furthermore, these North American bee species are predicted to lack the specialized behavioral responses exhibited by the Asian honey bee, which makes them vulnerable to the attacks of the AGH. The economic and cultural importance of *Melipona* species in America is well-

413 documented, particularly in the Yucatan Peninsula in Mexico, where these bees have been
414 traditionally raised for honey and were even considered gods outright in Mayan times (Ayala,
415 Gonzalez & Engel, 2013; Quezada-Euán et al., 2018). It is important to mention, however, that
416 the risk to *Melipona* species may be lower than that to *Bombus* species because entrances to the
417 hives of some species of *Melipona* are narrow, allowing a single bee to pass at a time (Couvillon
418 et al., 2007), unlike the entrances to the hives of honey bees and many bumble bees, which are
419 wider.

420 The AGH is not the first Hymenoptera to invade North America, and species of *Vespa* are
421 well-known for their invasive success and excellent dispersal capacity (Beggs et al., 2011;
422 Monceau, Bonnard & Thiéry, 2014). The European hornet, *Vespa crabro* L., a Eurasian species
423 that was accidentally introduced to North America in the 1800s, occupies a range in the United
424 States that encompasses most state east of the Great Plains (Smith-Pardo et al. 2020). The
425 solitary giant resin bee, *Challomegachile sculpturalis* is an Asian taxon which was recently
426 introduced in the United States. Only 15 years after its initial detection near Baltimore,
427 Maryland, this species had invaded most of the southeastern United States (Hinojosa-Díaz et al.,
428 2005). These examples indicate considerable precedent for hornet invasion and establishment in
429 the southeastern United States, but the AGH poses a unique biodiversity risk as a direct predator
430 of bees. Because the Pacific Northwest is consistently predicted as suitable for the AGH,
431 preventing further establishment and spread of recently detected introduced populations near
432 Seattle and Vancouver is essential. If these introduced individuals are not eradicated, they may
433 flourish under the suitable climatic conditions, establishing many more colonies that will be
434 difficult to control. Preventing establishment of the AGH in the Pacific Northwest is especially
435 critical because an established AGH population in the Pacific Northwest would provide a source

436 population for potential long-range dispersers that could use multiple potential invasion routes
437 (Fig. 6) to reach suitable habitat in the eastern United States, facilitating full-scale invasion. In
438 light of this, we recommend official monitoring protocols for the vulnerable Pacific Northwest
439 region including on-going citizen science and outreach efforts (https://agr.wa.gov/hornets),,
440 which may be the fastest and most effective way to detect potential range expansions.

441 Although AGH is primarily found in temperate areas in its native range, some of its
442 populations reach subtropical regions like Taiwan (Archer, 2008), which indicates a broad
443 temperature tolerance. This southern part of the species' native range might explain why our
444 models predicted suitable areas in South America, Africa, and elsewhere (Figure 2S-S7).
445 Although temperature is a critical factor that determines the abundance and distribution of
446 organisms (Sunday, Bates & Dulvy, 2012), factors such as desiccation resistance may be equally
447 important for some species. For example, for ants and some bees, desiccation tolerance is a good
448 predictor of species' distributions (Bujan, Yanoviak & Kaspari, 2016; Burdine & McCluney,
449 2019). For example, humidity is important for the regulation of temperature in nests of the
450 European hornet (Klingner et al., 2005) and, in some species of stingless bees, regulation of
451 humidity appears to be more important than regulation of temperature to maintain colony health
452 (Ayton et al., 2016). Unfortunately, heat and desiccation tolerances, factors that might improve
453 predictions of this species' distributional potential, are unknown for the AGH. In other hornets,
454 subtropical populations tend to have longer population cycles than temperate populations
455 (Archer, 2008), so negative impacts of an AGH invasion may be stronger in tropical or
456 subtropical areas.

457 In summary, our modeling approach allowed us to recognize how predicted suitable
458 areas can be depending on distinct schemes of data treatment. We showed that this variability

459 can derive from crucial decisions made during the initial steps of ecological niche modeling
460 exercises. These results highlight the importance of such initial decisions, as well as the need to
461 recognize sources of variability in predictions of suitability. This point is of special importance
462 in predicting the potential for expansion of invasive species, as uncertainty increases when
463 models are transferred to areas where environmental conditions are different. Our analyses and
464 simulations revealed the potential of the AGH to invade large areas in North America and the
465 likely paths of such an invasion. We also showed that predicted suitable areas for the AGH
466 overlap broadly with those where honey production is highest in the United States and Mexico,
467 as well as with species-rich areas for bumble bees and stingless bees. These results bring light to
468 the potential implications of uncontrolled dispersal of the AGH to suitable environments in
469 North America, and highlight the need for rapid eradication actions to mitigate potential
470 biodiversity and economic losses.

471

472 **Acknowledgments**

473 We would like to thank the members of the KUENM group for their support in the development
474 of this manuscript. We also thank Allan Smith-Pardo for letting us use the photograph of AGH in
475 lateral view (Fig. 1B). ANB would like to thank Secretaría de Educación, Ciencia, Tecnología e
476 Innovación de la Ciudad de México.

477

478 **References**

479 APHIS. 2020a. New pest response guidelines: *Vespa mandarinia*. U.S. Department of
480 Agriculture, Animal and Plant Health Inspection Service, Plant Protection and

Quarantine.

APHIS. 2020b. Asian Giant Hornet. U.S. Department of Agriculture, National Invasive Species Information Center. https://www.invasivespeciesinfo.gov/terrestrial/invertebrates/asian-giant-hornet.

Alaniz AJ, Carvajal MA, Vergara PM, 2020. Giants are coming? Predicting the potential spread and impacts of the giant Asian hornet (*Vespa mandarinia*, Hymenoptera: Vespidae) in the United States. *Pest Management Science*. DOI: 10.1002/ps.6063.

Alkishe A, Cobos ME, Peterson AT, Samy AM. 2020. Recognizing sources of uncertainty in disease vector ecological niche models: an example with the tick *Rhipicephalus sanguineus* sensu lato. *Perspectives in Ecology and Conservation*. DOI: 10.1016/j.pecon.2020.03.002.

Anderson RP. 2012. Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. *Annals of the New York Academy of Sciences* 1260:66–80. DOI: 10.1111/j.1749-6632.2011.06440.x.

Anderson RP, Lew D, Peterson AT. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* 162:211–232. DOI: 10.1016/S0304-3800(02)00349-6.

Archer ME. 1995. Taxonomy, distribution and nesting biology of the *Vespa mandarinia* group (Hym., Vespinae). *Entomologist's Monthly Magazine* 131:47–53.

Archer ME. 2008. Taxonomy, distribution and nesting biology of species of the genera *Provespa* Ashmead and *Vespa* Linneaus (Hymenoptera, Vespidae). *Entomologist's Monthly Magazine* 144:69–101.

Arita HT, Christen JA, Rodríguez P, Soberón J. 2008. Species diversity and distribution in

presence-absence matrices: mathematical relationships and biological implications. *American Naturalist* 172:519–532. DOI: 10.1086/590954.

Ascher JS, Pickering J. 2020.Discover life bee species guide and world checklist (Hymenoptera: Apoidea: *Anthophila*). *Available at https://www.discoverlife.org/mp/20q?guide=Apoidea_species* (accessed July 1, 2020).

Ayala R, Gonzalez VH, Engel MS. 2013. Mexican stingless bees (Hymenoptera: Apidae): diversity, distribution, and indigenous knowledge. In: Vit P, Pedro SRM, Roubik D eds. *Pot-Honey: A Legacy of Stingless Bees*. New York, NY: Springer, 135–152. DOI: 10.1007/978-1-4614-4960-7_9.

Ayton S, Tomlinson S, Phillips RD, Dixon KW, Withers PC. 2016. Phenophysiological variation of a bee that regulates hive humidity, but not hive temperature. *Journal of Experimental Biology* 219:1552–1562. DOI: 10.1242/jeb.137588.

Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, Soberón J, Villalobos F. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling* 222:1810–1819. DOI: 10.1016/j.ecolmodel.2011.02.011.

Beggs JR, Brockerhoff EG, Corley JC, Kenis M, Masciocchi M, Muller F, Rome Q, Villemant C. 2011. Ecological effects and management of invasive alien Vespidae. *BioControl* 56:505–526. DOI: 10.1007/s10526-011-9389-z.

Bérubé C. 2020. Giant alien insect invasion averted Canadian beekeepers thwart apicultural disaster (... or at least the zorn-bee apocalypse). *American Bee Journal*:209–214.

Bivand R, Keitt T, Rowlingson B, Pebesma E, Sumner M, Hijmans R, Rouault E, Warmerdam F, Ooms J, Rundel C. 2020a. *rgdal: Bindings for the "geospatial" data abstraction library*.

https://cran.r-project.org/web/packages/rgdal/index.html.

Bivand R, Rundel C, Pebesma E, Stuetz R, Hufthammer KO, Giraudoux P, Davis M, Santilli S. 2020b. *rgeos: Interface to geometry engine - open source ('GEOS')*. https://cran.r-project.org/web/packages/rgeos/index.html.

Brown JC, Albrecht C. 2001. The effect of tropical deforestation on stingless bees of the genus *Melipona* (Insecta: Hymenoptera: Apidae: Meliponini) in central Rondonia, Brazil. *Journal of Biogeography* 28:623–634. DOI: 10.1046/j.1365-2699.2001.00583.x.

Bujan J, Yanoviak SP, Kaspari M. 2016. Desiccation resistance in tropical insects: causes and mechanisms underlying variability in a Panama ant community. *Ecology and Evolution* 6:6282–6291. DOI: 10.1002/ece3.2355.

Burdine JD, McCluney KE. 2019. Differential sensitivity of bees to urbanization-driven changes in body temperature and water content. *Scientific Reports* 9:1643. DOI: 10.1038/s41598-018-38338-0.

Cameron SA, Lozier JD, Strange JP, Koch JB, Cordes N, Solter LF, Griswold TL. 2011. Patterns of widespread decline in North American bumble bees. *Proceedings of the National Academy of Sciences USA* 108:662–667. DOI: 10.1073/pnas.1014743108.

Chapman AD. 2005. Principles and methods of data cleaning. Copenhagen, GBIF.

Cobos ME, Jiménez L, Nuñez-Penichet C, Romero-Alvarez D, Simões M. 2018. Sample data and training modules for cleaning biodiversity information. *Biodiversity Informatics* 13:49–50. DOI: 10.17161/bi.v13i0.7600.

Cobos ME, Osorio-Olvera L, Soberón J, Peterson AT, Barve V, Barve N. 2020. *ellipsenm: Ecological niche characterizations using ellipsoids*. https://github.com/marlonecobos/ellipsenm.

550 Cobos ME, Peterson AT, Barve N, Osorio-Olvera L. 2019a. kuenm: an R package for detailed
551 development of ecological niche models using Maxent. *PeerJ* 7:e6281. DOI:
552 10.7717/peerj.6281.
553 Cobos ME, Peterson AT, Osorio-Olvera L, Jiménez-García D. 2019b. An exhaustive analysis of
554 heuristic methods for variable selection in ecological niche modeling and species
555 distribution modeling. *Ecological Informatics* 53:100983. DOI:
556 10.1016/j.ecoinf.2019.100983.
557 Colautti RI, Bailey SA, Van Overdijk CD, Amundsen K, MacIsaac H J. 2006. Characterized and
558 projected costs of nonindigenous species in Canada. *Biological invasions* 8(1):45–59.
559 Couvillon MJ, Wenseleers T, Imperatriz-Fonseca VL, Nogueira-Neto P, Ratnieks FLW. 2007.
560 Comparative study in stingless bees (Meliponini) demonstrates that nest entrance size
561 predicts traffic and defensivity. *Journal of Evolutionary Biology* 21:194–201. DOI:
562 10.1111/j.1420-9101.2007.01457.x.
563 Cueva del Castillo R, Sanabria-Urbán S, Serrano-Meneses MA. 2015. Trade-offs in the evolution
564 of bumblebee colony and body size: a comparative analysis. *Ecology and Evolution*
565 5:3914–3926. DOI: 10.1002/ece3.1659.
566 Dueñas M-A, Ruffhead HJ, Wakefield NH, Roberts PD, Hemming DJ, Diaz-Soltero H. 2018.
567 The role played by invasive species in interactions with endangered and threatened
568 species in the United States: a systematic review. *Biodiversity and Conservation*
569 27:3171–3183. DOI: 10.1007/s10531-018-1595-x.
570 Engler R, Hordijk W, Guisan A. 2012. The MIGCLIM R package – seamless integration of
571 dispersal constraints into projections of species distribution models. *Ecography* 35:872–
572 878. DOI: 10.1111/j.1600-0587.2012.07608.x.

573 Escobar LE, Lira-Noriega A, Medina-Vogel G, Peterson AT. 2014. Potential for spread of the
574 white-nose fungus (*Pseudogymnoascus destructans*) in the Americas: use of Maxent and
575 NicheA to assure strict model transference. *Geospatial Health* 9:221–229. DOI:
576 10.4081/gh.2014.19.
577 Fujiwara A, Sasaki M, Washitani I. 2016. A scientific note on hive entrance smearing in
578 Japanese *Apis cerana* induced by pre-mass attack scouting by the Asian giant hornet
579 *Vespa mandarinia*. *Apidologie* 47:789–791. DOI: 10.1007/s13592-016-0432-z.
580 GBIF.org (07 May 2020) GBIF occurrence download. https://doi.org/10.15468/dl.kzcgc2**.**
581 Gray L, New A, Science K, Wolfram S. 2003. A mathematician looks at Wolfram's new kind of
582 science. *Notices of the American Mathematical Society 50 (2) (2003) 200–211, URL*
583 *http://www.ams.org/notices/200302/fea-gray.pdf. URL*
584 *http://www.ams.org/notices/200302/fea-gray.pdf* 50:200–211.
585 Hijmans RJ, Etten J van, Sumner M, Cheng J, Bevan A, Bivand R, Busetto L, Canty M, Forrest
586 D, Ghosh A, Golicher D, Gray J, Greenberg JA, Hiemstra P, Hingee K, Geosciences I for
587 MA, Karney C, Mattiuzzi M, Mosher S, Nowosad J, Pebesma E, Lamigueiro OP, Racine
588 EB, Rowlingson B, Shortridge A, Venables B, Wueest R. 2020. *raster: Geographic data*
589 *analysis and modeling*. https://cran.r-project.org/web/packages/raster/index.html.
590 Hinojosa-Díaz IA, Yáñez-Ordóñez O, Chen G, Peterson AT, Engel MS. 2005. The North
591 American invasion of the giant resin bee (Hymenoptera: Megachilidae). *Journal of*
592 *Hymenoptera Research* 14:69–77.
593 Kastberger G, Schmelzer E, Kranner I. 2008. Social waves in Giant Honeybees repel hornets.
594 *PLoS ONE* 3:e3141. DOI: 10.1371/journal.pone.0003141.
595 Klingner R, Richter K, Schmolz E, Keller B. 2005. The role of moisture in the nest

596 thermoregulation of social wasps. *Naturwissenschaften* 92:427–430. DOI:
597 10.1007/s00114-005-0012-y.
598 Li X-D, Liu Z, Zhai Y, Zhao M, Shen H-Y, Li Y, Zhang B, Liu T. 2015. Acute interstitial
599 nephritis following multiple Asian Giant Hornet stings. *American Journal of Case*
600 *Reports* 16:371–373. DOI: 10.12659/AJCR.893734.
601 Matsuura M. 1988. Ecological study on Vespine wasps (Hymenoptera: Vespidae) attacking
602 honeybee colonies: 1. Seasonal changes in the frequency of visits to apiaries by Vespine
603 wasps and damage inflicted, especially in the absence of artificial protection. *Applied*
604 *Entomology and Zoology* 23:428–440.
605 Matsuura M, Sakagami SF. 1973. A bionomic sketch of the Giant Hornet, *Vespa mandarinia*, a
606 serious pest for Japanese apiculture. *Journal of the Faculty of Science, Hokkaido*
607 *University: Series 6. Zoology* 19:125–162.
608 McClenaghan B, Schlaf M, Geddes M, Mazza J, Pitman G, McCallum K, Rawluk S, Hand K,
609 Otis GW. 2019. Behavioral responses of honey bees, *Apis cerana* and *Apis mellifera*, to
610 *Vespa mandarinia* marking and alarm pheromones. *Journal of Apicultural Research*
611 58:141–148. DOI: 10.1080/00218839.2018.1494917.
612 Merow C, Smith MJ, Silander JA. 2013. A practical guide to MaxEnt for modeling species'
613 distributions: what it does, and why inputs and settings matter. *Ecography* 36:1058–1069.
614 DOI: 10.1111/j.1600-0587.2013.07872.x.
615 Michener CD. 2000. *The Bees of the World*. Baltimore: Johns Hopkins University Press.
616 Michener CD. 2007. *The Bees of the World*. Baltimore: Johns Hopkins University Press.
617 Monceau K, Bonnard O, Thiéry D. 2014. *Vespa velutina*: a new invasive predator of honeybees
618 in Europe. *Journal of Pest Science* 87:1–16. DOI: 10.1007/s10340-013-0537-3.

Nuñez-Penichet C, Cobos ME, Peterson AT, Soberón J, Barve N, Barve V, Gueta T. 2020. *biosurvey: Tools for biological survey planning*. https://github.com/claununez/biosurvey.

Ono M, Igarashi T, Ohno E, Sasaki M. 1995. Unusual thermal defense by a honeybee against mass attack by hornets. *Nature* 377:334–336. DOI: 10.1038/377334a0.

Osorio-Olvera L, Soberón J. 2020. *bam: Species distribution models in the light of the BAM theory*. https://github.com/luismurao/bam.

Osorio-Olvera L, Lira-Noriega A, Soberón J, Peterson AT, Falconi M, Contreras-Díaz RG, Martínez-Meyer E, Barve N, Barve V. 2020. NTBOX: an R package with a graphical user interface for modeling and evaluating multidimensional ecological niches. *Methods in Ecology and Evolution* 11(10): 1199–11206. DOI: 10.1111/2041-210X.13452.

Owens HL, Campbell LP, Dornak LL, Saupe EE, Barve N, Soberón J, Ingenloff K, Lira-Noriega A, Hensz CM, Myers CE, Peterson AT. 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling* 263:10–18. DOI: 10.1016/j.ecolmodel.2013.04.011.

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. *Journal of Biogeography* 34(1):102–117.

Peterson AT, Papeş M, Soberón J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* 213:63–72. DOI: 10.1016/j.ecolmodel.2007.11.008.

Phillips SJ, Anderson RP, Dudík M, Schapire RE, Blair ME. 2017. Opening the black box: an open-source release of Maxent. *Ecography* 40:887–893. DOI: 10.1111/ecog.03049.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species

642 geographic distributions. *Ecological Modelling* 190:231–259. DOI:
643 10.1016/j.ecolmodel.2005.03.026.
644 Pimentel D, Zuniga R, Morrison D. 2005. Update on the environmental and economic costs
645 associated with alien-invasive species in the United States. *Ecological Economics*
646 52:273–288. DOI: 10.1016/j.ecolecon.2004.10.002.
647 Pyšek P, Richardson DM. 2010. Invasive species, environmental change and management, and
648 health. *Annual Review of Environment and Resources* 35:25–55. DOI: 10.1146/annurev-
649 environ-033009-095548.
650 Quezada-Euán JJG, Nates-Parra G, Maués MM, Roubik DW, Imperatriz-Fonseca VL. 2018. The
651 economic and cultural values of stingless bees (Hymenoptera: Meliponini) among ethnic
652 groups of tropical America. *Sociobiology* 65:534–557. DOI:
653 10.13102/sociobiology.v65i4.3447.
654 R Core Team. 2019. *R: A language and environment for statistical computing*. Vienna, Austria:
655 R Foundation for Statistical Computing.
656 Schmidt JO, Yamane S, Matsuura M, Starr CK. 1986. Hornet venoms: lethalities and lethal
657 capacities. *Toxicon* 24:950–954. DOI: 10.1016/0041-0101(86)90096-6.
658 Simões M, Romero-Alvarez D, Nuñez-Penichet C, Jiménez L, Cobos ME. 2020. General theory
659 and good practices in ecological niche modeling: a basic guide. *Biodiversity Informatics*
660 15:67–68.
661 Smith-Pardo AH, Carpenter JM, Kimsey L. 2020. The diversity of hornets in the genus *Vespa*
662 (Hymenoptera: Vespidae; Vespinae), their importance and interceptions in the United
663 States. *Insect Systematics and Diversity* 4:1–27.
664 Sugahara M, Nishimura Y, Sakamoto F. 2012. Differences in heat sensitivity between Japanese

honeybees and hornets under high carbon dioxide and humidity conditions inside bee balls. *Zoological Science* 29:30–36. DOI: 10.2108/zsj.29.30.

Sunday JM, Bates AE, Dulvy NK. 2012. Thermal tolerance and the global redistribution of animals. *Nature Climate Change* 2:686–690. DOI: 10.1038/nclimate1539.

Vega GC, Pertierra LR, Olalla-Tárraga MÁ. 2018. MERRAclim, a high-resolution global dataset of remotely sensed bioclimatic variables for ecological modelling. *Scientific Data* 4:170078. DOI: 10.1038/sdata.2017.78.

Viana JL, Sousa H de AC, Alves RM de O, Pereira DG, Silva Jr. JC, Paixão JF da, Waldschmidt AM, Viana JL, Sousa H de AC, Alves RM de O, Pereira DG, Silva Jr. JC, Paixão JF da, Waldschmidt AM. 2015. Bionomics of *Melipona mondury* Smith 1863 (Hymenoptera: Apidae, Meliponini) in relation to its nesting behavior. *Biota Neotropica* 15. DOI: 10.1590/1676-06032015009714.

Vilà M, Espinar JL, Hejda M, Hulme PE, Jarošík V, Maron JL, Pergl J, Schaffner U, Sun Y, Pyšek P. 2011. Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities and ecosystems. *Ecology Letters* 14:702–708. DOI: 10.1111/j.1461-0248.2011.01628.x.

Warren DL, Seifert SN. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* 21:335–342. DOI: 10.1890/10-1171.1.

Wilcove DS, Rothstein D, Dubow J, Phillips A, Losos E. 1998. Quantifying threats to imperiled species in the United States. *BioScience* 48:607–615. DOI: 10.2307/1313420.

Wilson TM, Takahashi J, Spichiger S-E, Kim I, van Westendorp P. 2020. First reports of *Vespa mandarinia* (Hymenoptera: Vespidae) in North America represent two separate maternal

lineages in Washington State, United States, and British Columbia, Canada. Annals of the Entomological Society of America 113:468–472. DOI: 10.1093/aesa/saaa024.

Yanagawa Y, Morita K, Sugiura T, Okada Y. 2007. Cutaneous hemorrhage or necrosis findings after *Vespa mandarinia* (wasp) stings may predict the occurrence of multiple organ injury: a case report and review of literature. *Clinical Toxicology* 45:803–807. DOI: 10.1080/15563650701664871.

Zhu G, Illan JG, Looney C, Crowder DW. 2020. Assessing the ecological niche and invasion potential of the Asian giant hornet. *Proceedings of the National Academy of Sciences USA*. DOI: 10.1073/pnas.2011441117.

Zurell D, Franklin J, König C, Bouchet PJ, Dormann CF, Elith J, Fandos G, Feng X, Guillera-Arroita G, Guisan A, Lahoz-Monfort JJ, Leitão PJ, Park DS, Peterson AT, Rapacciuolo G, Schmatz DR, Schröder B, Serra-Diaz JM, Thuiller W, Yates KL, Zimmermann NE, Merow C. 2020. A standard protocol for reporting species distribution models. *Ecography*. DOI: 10.1111/ecog.04960.

Tables

Table 1. Summary of results of ecological niche modeling for *Vespa mandarinia*, including model calibration, model evaluation, and relevant Maxent settings for models selected after independent testing. Maxent settings are represented in the columns 'Regularization multiplier,' 'Feature classes,' and 'Variable sets'. The column 'Calibration processes' refers to each of the five calibrations processes that were done with different sets of random points, for every calibration scheme. The variables included in the sets mentioned on this table can be found in Table 1S-2S. E: free extrapolation; EC: extrapolation with clamping; NE: no extrapolation, PCs: principal components.

Figures

Figure 1. Hypothesis of accessible areas (**M**) and representation of the occurrence records of *Vespa mandarinia* across its native distribution. The three panels represent the occurrences left after cleaning (A) and after applying the two thinning approaches (B and C).

Figure 2. Schematic representation of methods used to obtain ecological niche models for *Vespa mandarinia*. The aim of the modeling process was to consider the variability resulting from different procedures and methodological decisions made during model calibration.

Figure 3. Median of potentially suitable areas for *Vespa mandarinia* predicted with free extrapolation for different calibration schemes in the calibration area (A, C, E, G) and in North America (B, D, F, H). Only models that anticipated the invaded areas of North America were included. The color pallet is standard for all figure panels.

Figure 4. Sum of all suitable areas for *Vespa mandarinia* in North America derived from binarizing each replicate of selected models (model transfers done with extrapolation) using a 5% threshold. Each replicate predicted the known invaded localities of this hornet.

Figure 5. Agreement of areas with extrapolation risk for models of *Vespa mandarinia* in North America, separated by calibration schemes.

Figure 6. Results from simulations of the potential dynamics of invasion of *Vespa mandarinia* in North America. Dark shades of green show areas that the species reached in a high percentage of scenarios, while light shades of green represent areas reached only rarely by the species. Arrows represent the general path of potential invasion.

Figure 7. Representation of potential ecological and economic impacts of an invasion of *Vespa*

732 *mandarinia*. A: honey production (in US dollars) in Mexico and the United States in 2016. B:
733 species richness of the genera *Bombus* (bumble bees) and *Melipona* (stingless bees) in North
734 America. The area shaded in gray represents the simulated potential invaded area of *Vespa*
735 *mandarinia* in North America obtained with the 50 km spatial thinning occurrences and PCs as
736 environmental predictors. We used this scenario because is the one that best connects the known
737 invaded areas with the eastern United States.

Figure 1

Hypothesis of accessible areas (M) and representation of the occurrence records of *Vespa mandarinia* across its native distribution.

The three panels represent the occurrences left after cleaning (A) and after applying the two thinning approaches (B and C).

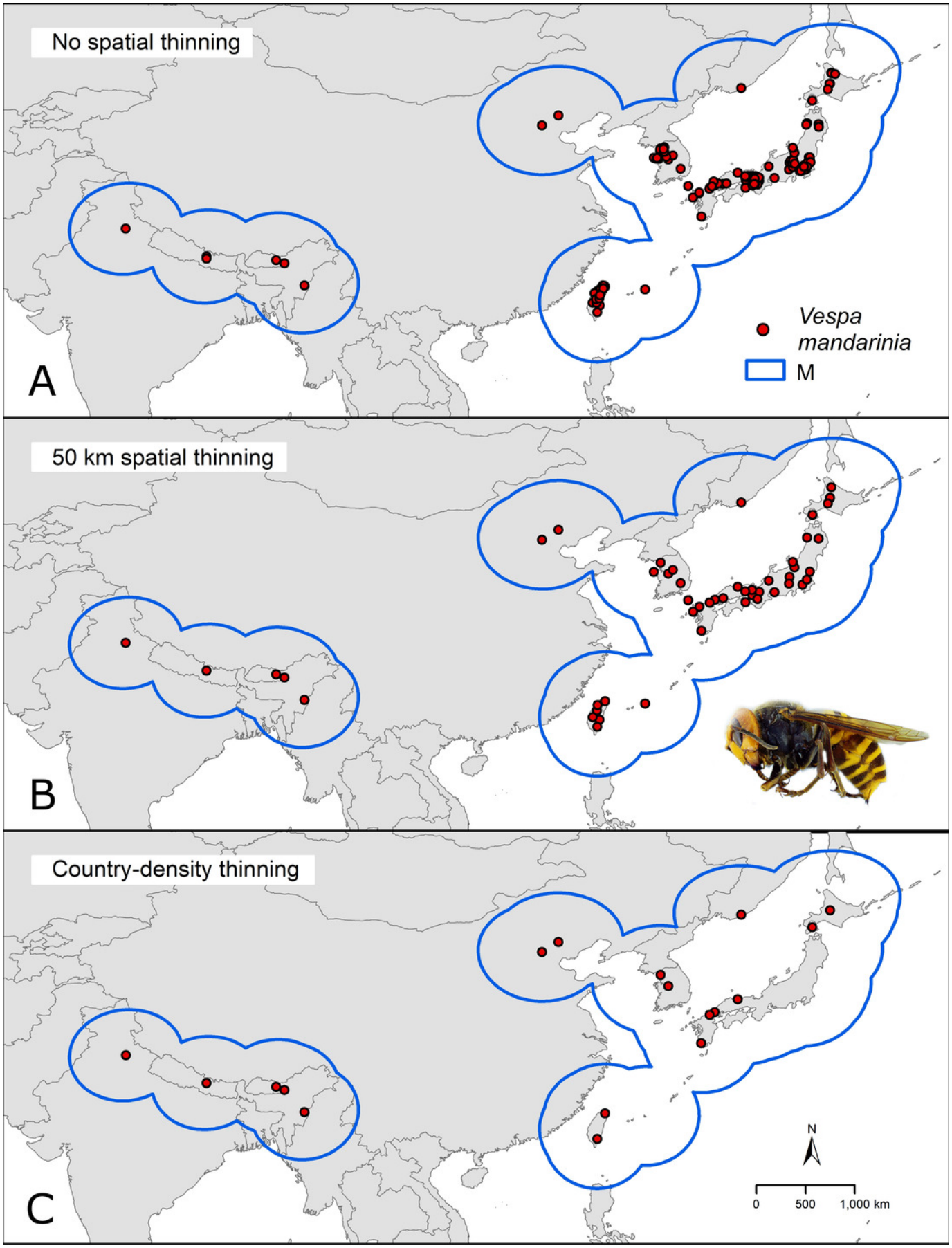
No spatial thinning
Vespa mandarinia
M
A
50 km spatial thinning
B
Country-density thinning
N
0
500
1,000 km
C

Figure 2

Schematic representation of methods used to obtain ecological niche models for *Vespa mandarinia.*

The aim of the modeling process was to consider the variability resulting from different procedures and methodological decisions made during model calibration.

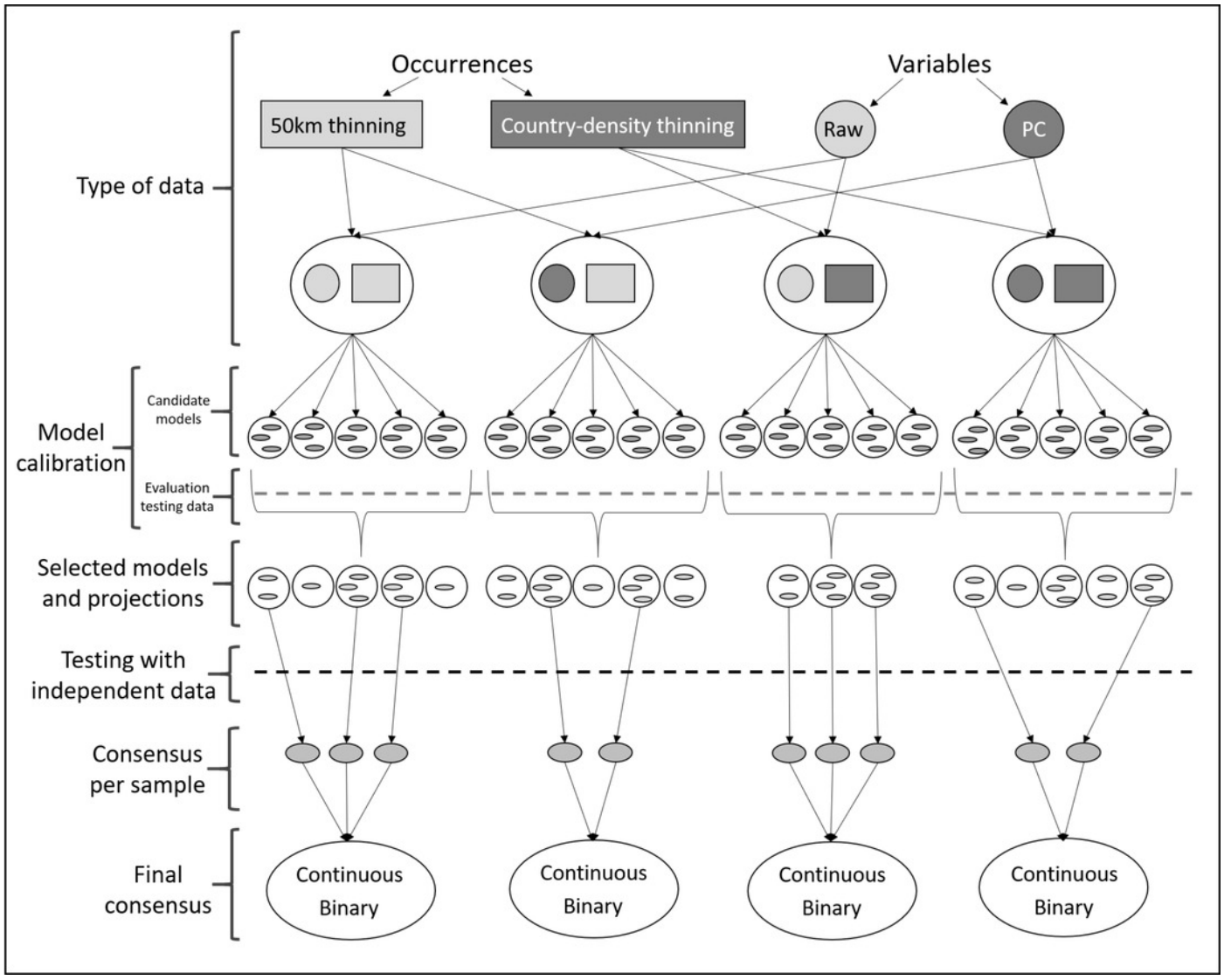

Figure 3

Median of potentially suitable areas for *Vespa mandarinia* predicted with free extrapolation for different calibration schemes in the calibration area (A, C, E, G) and in North America (B, D, F, H).

Only models that anticipated the invaded areas of North America were included. The color pallet is standard for all figure panels.

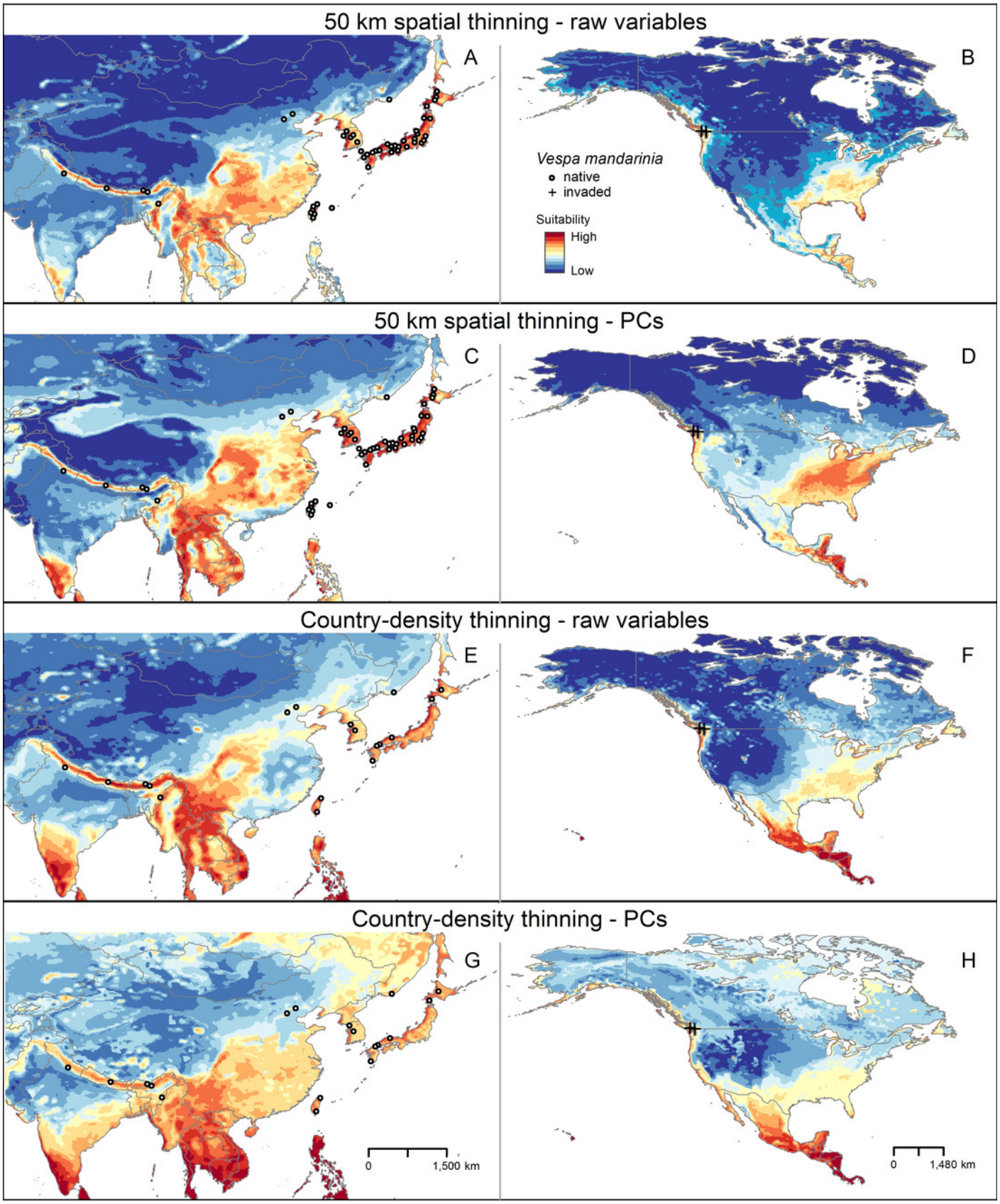
50 km spatial thinning - raw variables
A
B
Vespa mandarinia
native
invaded
Suitability
High
Low
50 km spatial thinning - PCs
C
D
Country-density thinning - raw variables
E
F
Country-density thinning - PCs
G
H
0
1,500 km
0
1,480 km

Figure 4

Sum of all suitable areas for *Vespa mandarinia* in North America derived from binarizing each replicate of selected models (model transfers done with extrapolation), using a 5% threshold.

Each replicate predicted the known invaded localities of this hornet.

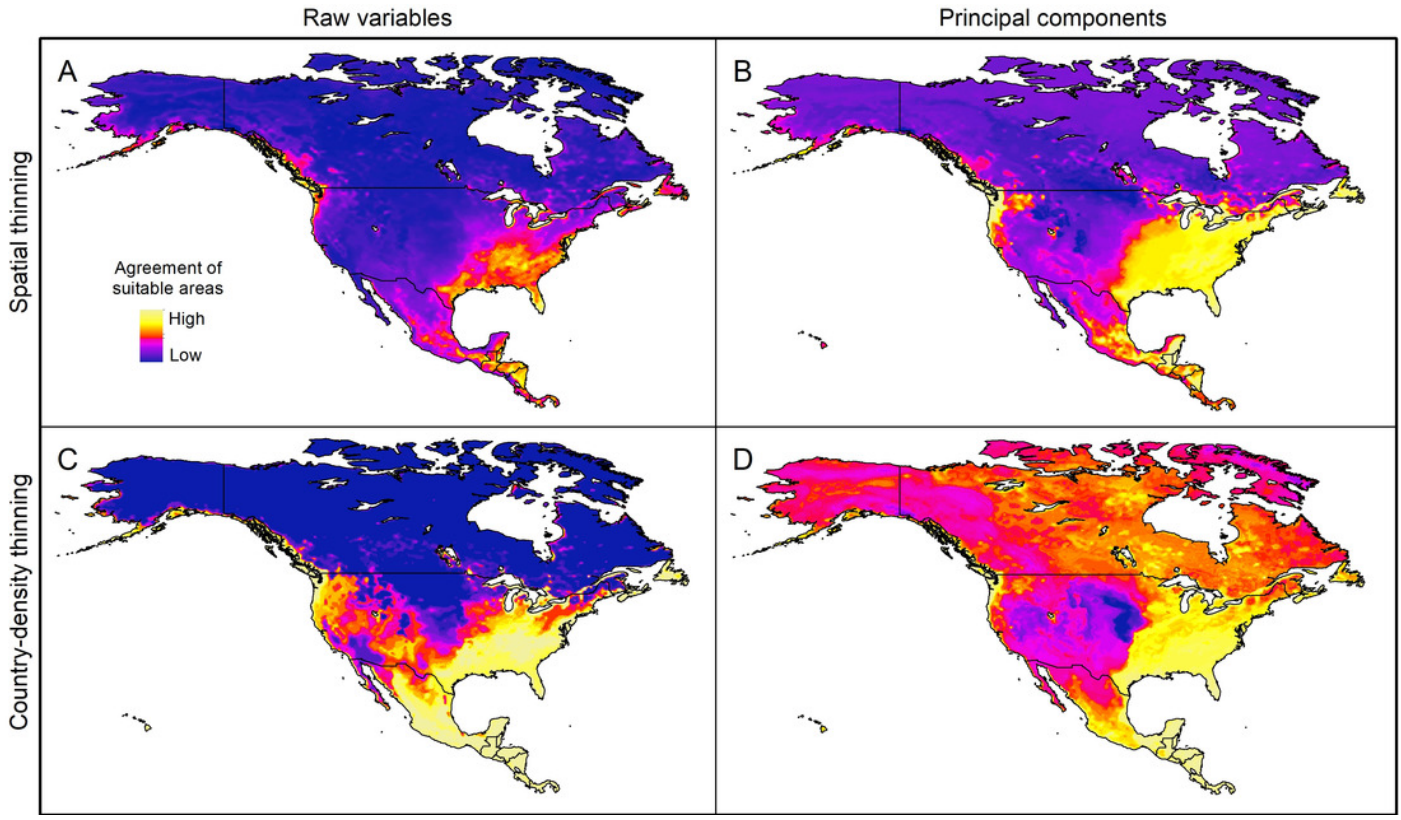

Figure 5

Agreement of areas with extrapolation risk for models of *Vespa mandarinia* in North America, separated by calibration schemes.

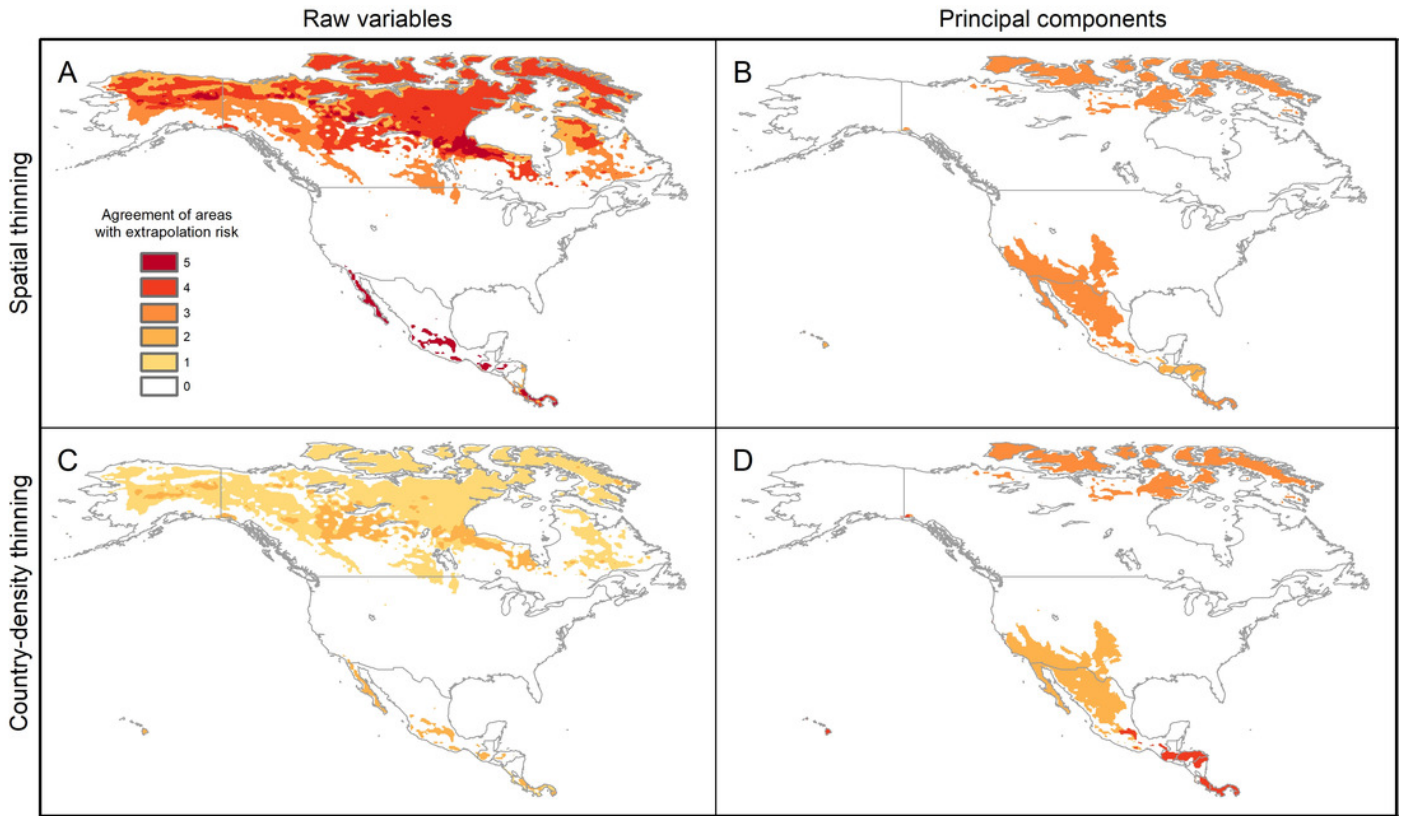

Figure 6

Results from simulations of the potential dynamics of invasion of *Vespa mandarinia* in North America.

Dark shades of green show areas that the species reached in a high percentage of scenarios, while light shades of green represent areas reached only rarely by the species. Arrows represent the general path of potential invasion.

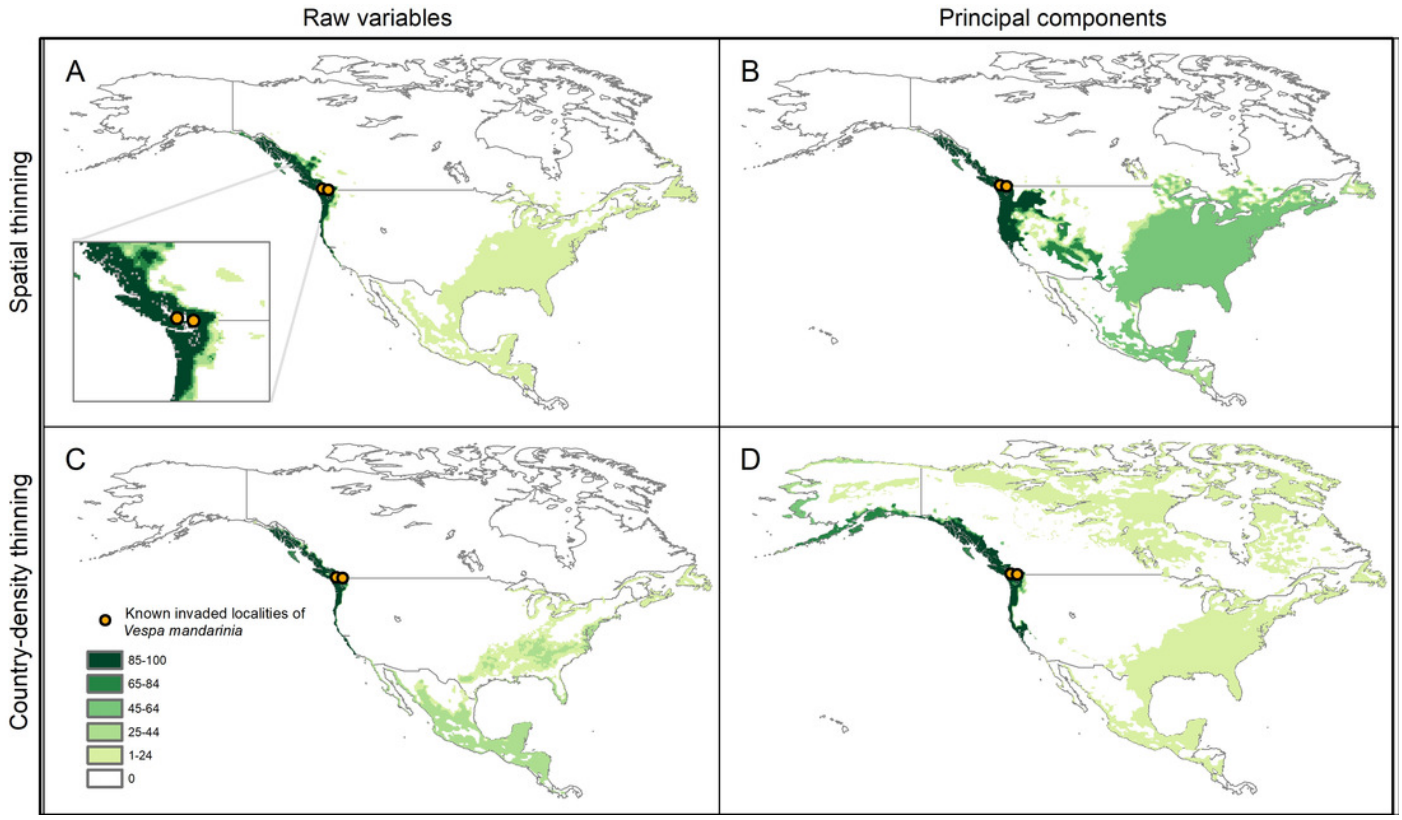

Figure 7

Representation of potential ecological and economic impacts of an invasion of *Vespa mandarinia*.

A: honey production (in US dollars) in Mexico and the United States in 2016. B: species richness of the genera *Bombus* (bumble bees) and *Melipona* (stingless bees) in North America. The area shaded in gray represents the simulated potential invaded area of *Vespa mandarinia* in North America obtained with the 50 km spatial thinning occurrences and PCs as environmental predictors. We used this scenario because is the one that best connects the known invaded areas with the eastern United States.

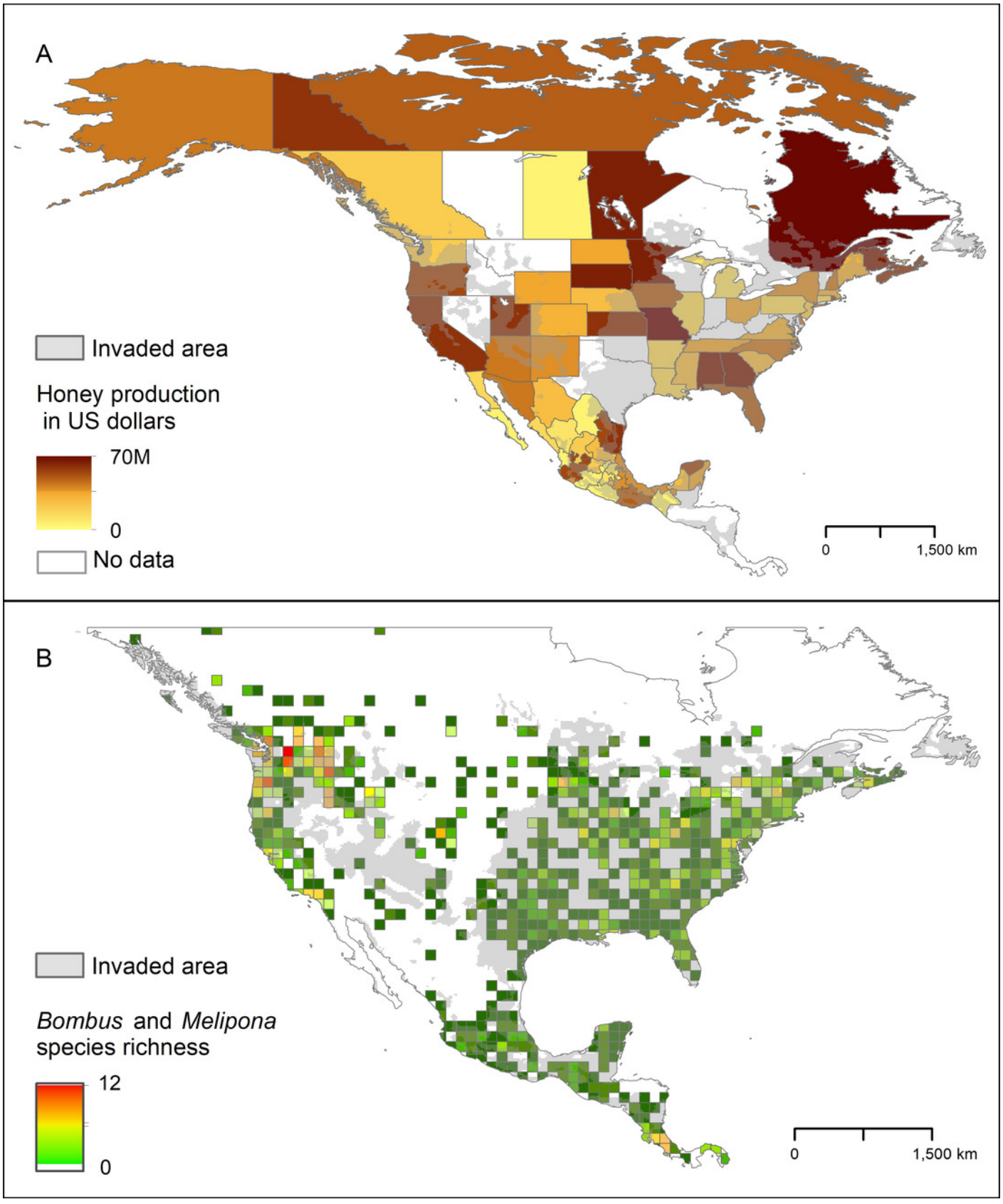
A
Invaded area
Honey production
in US dollars
70M
0
No data
0
1,500 km
B
Invaded area
Bombus and Melipona
species richness
12
0
0
1,500 km

Table 1(on next page)

Summary of results of ecological niche modeling for *Vespa mandarinia*, including model calibration, model evaluation, and relevant Maxent settings for models selected after independent testing.

Maxent settings are represented in the columns 'Regularization multiplier,' 'Feature classes,' and 'Variable sets'. The column 'Calibration processes' refers to each of the five calibrations processes that were done with different sets of random points, for every calibration scheme. The variables included in the sets mentioned on this table can be found in Table 1S-2S. E: free extrapolation; EC: extrapolation with clamping; NE: no extrapolation, PCs: principal components.

Calibration scheme	Calibration processes	Models meeting selection criteria	Models predicting independent records (E; EC; NE)	Regularization multiplier	Feature classes	Variable sets
Raw variables and distance thinned occurrences	1	6	8; 2; 10	0.25; 0.5; 0.75	lq; lqpt	42; 43; 50; 51; 57
	2	1	-	-	-	-
	3	1	6; 4; 4	0.75	lqpth	12
	4	1	2; -; 1	0.25	lq	21
	5	2	7; 7; 13	0.1; 0.25	lq	26
PCs and distance thinned occurrences	1	4	24; 18; 20	5	lqph; lqpth	7; 11
	2	2	10; 5; 5	0.25; 0.5	qp	11
	3	4	22; 23; 19	0.1; 0.25; 0.5; 0.75	lp	11
	4	3	9; 8; 11	0.25; 0.5; 0.75	qp	7
	5	6	21; 21; 26	0.1; .25; 0.5; 0.75	lqp	2; 9
Raw variables and country-density thinned occurrences	1	1	4; 4; 6	0.1	lqp	22
	2	2	4; 11; 8	0.1	lq; lqp	5; 22
	3	-	-	-	-	-
	4	3	15; 13; 16	0.1; 2	lq; lqph; lqpth	13; 32
	5	-	-	-	-	-
PCs and country-density thinned occurrences	1	7	32; 31; 32	0.25; 0.5; 0.7; 1	lp; lqpt; lqpth	2; 4; 5; 6; 8
	2	1	1; 3; 1	1	lqpth	8
	3	1	5; 6; 7	0.1	q	4
	4	4	24; 22; 18	0.1; 0.25; 0.5; 0.74	lp	1
	5	2	5; 7; 5	1	lqp	1; 6

1

2