Diversity and Distributions, (Diversity Distrib.) (2013) 1–9

BIODIVERSITY RESEARCH

Regional distribution models with lack of proximate predictors: Africanized honeybees expanding north

Catherine S. Jarnevich[1]*, Wayne E. Esaias[2], Peter L. A. Ma[3], Jeffery T. Morisette[1], Jaime E. Nickeson[3], Thomas J. Stohlgren[1], Tracy R. Holcombe[1], Joanne M. Nightingale[3], Robert E. Wolfe[2] and Bin Tan[2]

[1]*U.S. Geological Survey, Fort Collins Science Center, 2150 Centre Ave Bldg. C, Fort Collins, CO 80526, USA,* [2]*NASA Goddard Space Flight Center, Greenbelt, MD, USA,* [3]*NASA Goddard Space Flight Center/Sigma Space, Greenbelt, MD, USA*

*Correspondence: Catherine S. Jarnevich, U.S. Geological Survey, Fort Collins Science Center, 2150 Centre Ave Bldg. C, Fort Collins, CO 80526, USA.
E-mail jarnevichc@usgs.gov

ABSTRACT

Aim Species distribution models have often been hampered by poor local species data, reliance on coarse-scale climate predictors and the assumption that species–environment relationships, even with non-proximate predictors, are consistent across geographical space. Yet locally accurate maps of invasive species, such as the Africanized honeybee (AHB) in North America, are needed to support conservation efforts. Current AHB range maps are relatively coarse and are inconsistent with observed data. Our aim was to improve distribution maps using more proximate predictors (phenology) and using regional models rather than one across the entire range of interest to explore potential differences in drivers.

Location United States of America.

Methods We provide a generalized framework for regional and local species distribution modelling with our more nuanced and spatially detailed forecast of potential AHB spread using multiple habitat modelling techniques and newly derived remotely sensed phenology layers.

Results Variable importance did differ between the two regions for which we modelled AHB. Phenology metrics were important, especially in the south-east.

Main conclusions Results demonstrate that incorporating a combination of both climate drivers and vegetation phenology information into models can be important for predicting the suitable habitat range of these pollinators. Regional models may provide evidence of differing drivers of distributions geographically. This framework may improve many local and regional species distribution modelling efforts.

Keywords
Africanized honeybee, *Apis mellifera*, habitat suitability, species distribution modelling, vegetation phenology.

INTRODUCTION

Species distribution modelling (SDM) has become a common tool over the last few years with applications to diverse disciplines and biological taxa including conservation biology (e.g. Urbina-Cardona & Flores-Villela, 2010), biological invasions (Measey *et al.*, 2012), risk assessments (Bolliger *et al.*, 2007), restoration (Fei *et al.*, 2012) and climate change impacts (Thomas *et al.*, 2004). While these models are often correlative in nature, physiological information about a species should inform environmental factors included in distribution models (Austin, 2002). However, it can be difficult to obtain spatially continuous information for relevant factors. Indirect predictors such as elevation are often used as surrogates for those thought to be causal due to their high correlation with direct predictors such as temperature (Guisan & Zimmermann, 2000). For plant species, direct predictors are often environmental or abiotic factors that are measured such as climate or soil data. For fauna species, however, direct predictors may be different, including factors such as food availability and competition. Creating spatially explicit continuous surfaces describing these factors may be difficult.

Published 2013.
This article is a U.S. Government work and is in the public domain in the USA.
DOI: 10.1111/ddi.12143
http://wileyonlinelibrary.com/journal/ddi

A Journal of Conservation Biogeography
Diversity and Distributions

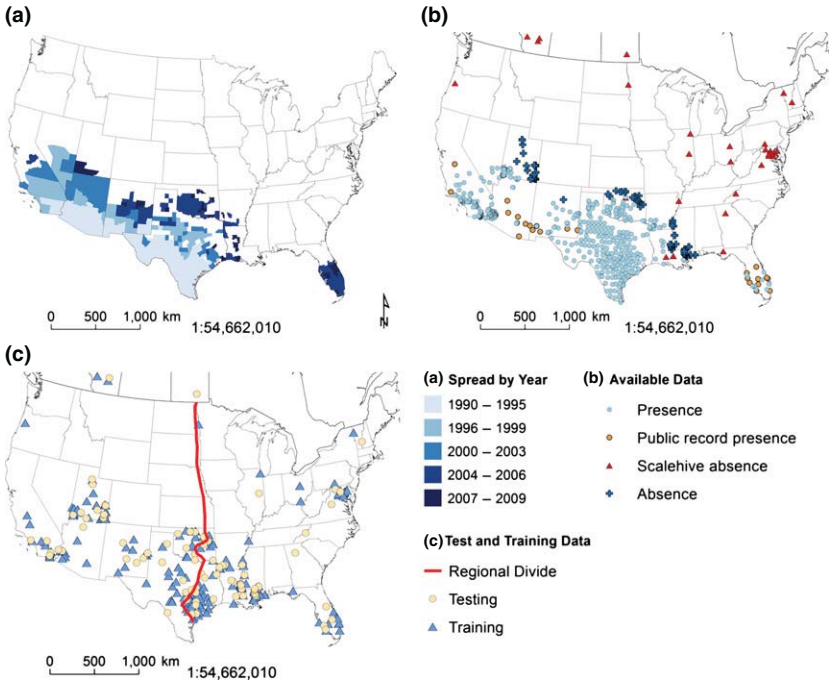

Figure 1 Africanized honeybee (AHB) location data (a) the spread of AHB by county from 1990 to 2009. (b) Presence points for AHB are from county and state apiculture records (blue) and derived from public records of AHB incidence (orange). Absence points were acquired from state apiculture records (blue crosses) or were generated from single nectar flow HoneyBeeNet scale hive records (red triangles). (c) Presence and absence points used in the model including training points (blue triangles) and testing points (yellow circles) with the regional divide between south-east and south-west (red line). (North America Albers Equal Area Conic projection, Datum NAD83).

Often species distribution models have been generated solely with climatic information, creating climatic envelopes that ignore other potentially important factors limiting the distribution of species (Heikkinen *et al.*, 2006). With remote sensing products continually becoming more easily accessible to ecologists, the products have increased as predictors (Zimmermann *et al.*, 2007; Bradley *et al.*, 2012). Most often these products have been land cover derivatives, coarse vegetation indices such as tree cover or leaf area index, or indices of greenness such as the normalized difference vegetation index (NDVI).

The conservation literature has recently recognized the importance of examining habitat relationships at regional levels to assess how relationships may change across biogeographical regions (McAlpine *et al.*, 2008). This need for distinctive models may be especially important when proximate predictors are unavailable, but the need for a model forces the use of the best available indirect predictor. For example, differences in distribution of Africanized honeybees (AHB) within the United States compared with European honeybees (EHB) are thought to derive from behavioural differences related to food storage and metabolism. While direct measure of these limiting factors may not be possible, climate or satellite imagery may act as surrogates. Over a large spatial extent covering multiple biogeographical provinces the relationship between the direct factor – food availability – and surrogates such as climate and satellite imagery – may differ.

Pollinator species have been modelled relatively infrequently compared with other taxonomic groups, but there could be important applications for both invasive pollinators and native pollinators in decline. We wished to explore using remote sensing-derived metrics related to the physiology of these species. For our example, we focused on AHB, a genetic hybrid cross of Tanzanian *Apis mellifera scutellata* and a variety of EHB strains such as *A. m. ligustica* (Harrison *et al.*, 2006), that have been spreading north in the Americas since their introduction to Brazil in 1957. These hybrids first reached the United States from Mexico in 1990 and have continued their northward spread (Fig. 1a), albeit at a slower rate across the south-eastern United States than the south-western United States (Villa *et al.*, 2002). AHB spread within the United States has been slower than the spread rate in the Neotropics and has been more erratic (Schneider *et al.*, 2004). Schneider *et al.* (2004) proposed several hypotheses for these observed differences including climatic differences (AHB may be more adapted to arid climates) and response to photoperiod (AHB tied to rainfall and floral abundance, which may make them less adapted to temperate conditions). Within the United States, AHB has had similar time and opportunity to expand in to the south-east as it has had to move north in the south-west. Predictions of the northern extent of the AHB potential habitat in North America can inform regional apiarists, safety officials and bee managers of the risks associated with

AHB. However, previous predictions of a northern limit for AHB spread were based on a simple temperature threshold derived from AHB distributions in South America (Taylor & Spivak, 1984) or observed physiology and behaviour (Southwick *et al.*, 1990). Interannual variation in these values may provide a fuzzy boundary for AHB extent (Harrison *et al.*, 2006). However, there have been reports of AHB presence in consecutive years considerably further north of areas previously predicted by temperature thresholds (Taylor & Spivak, 1984; Harrison *et al.*, 2006) and some of these were found during winter months (Fig. 1b). Conversely, the AHB extents derived using physiology, and behaviour show AHB presence much farther north than they are now presently found (Southwick *et al.*, 1990). These two existing methods for predicting the northern limit of AHB are inconsistent with each other and the observed presence data. Furthermore, previous prediction methods do not take advantage of the higher detailed environmental data, including physiologically relevant remote sensing products, and advanced modelling techniques that are currently available.

Our aim was to explore how phenology predictors may contribute to models of pollinator species and how the drivers of these models may differ between biologically defined regions. We focus on phenology predictors because, while climate data are commonly used in species distribution models, vegetation phenology is not and we believe that vegetation phenology, and thus forage availability, must influence the distribution of pollinator species. Phenology predictors act as a surrogate for seasonal availability of nectar related to blooming phenology (with respect to swarming, AHB show characteristics of a multivoltine population, whereas the EHB might be considered univoltine with respect to colony reproduction). We do this with an example of the potential northern limit of AHB, taking advantage of current location data, current species distribution modelling techniques, and concurrent environmental and climatic data. We hypothesized that drivers in the arid south-west would be related to temperature and precipitation, while those in the wetter south-east would be driven by vegetation phenology, defining the two regions using honeybee forage zones identified by Ayers & Harman (1992) based on natural floristic and land use patterns. These hypotheses are based on observed patterns of AHB distribution in the United States, including the differences in spread rates between the two regions. Hence, the models were developed for the continental United States and the south-west and the south-east regions.

METHODS

Species occurrence data

Presence data consist of both feral AHB and AHB/EHB hybrids, and we refer to the combination as AHB for simplicity in the following. We consulted with state apiculturists to compile field observations of AHB confirmed by DNA test across the United States (478 presence and 107 absence locations; Fig. 1a). Many of the counties in Arizona and Texas that were sites of initial United States invasion ceased collection of observations once AHB became common and the counties did not maintain historical records, resulting in regional data gaps in the well-established range. Thus, we supplemented the field observation presence locations with those from public safety and news records in the region (23 presence locations; Fig. 1a). We also added locations consisting of the centroids of small, heterogeneous counties where AHB are fairly ubiquitous in eastern Texas (140 presence locations; Fig. 1a). These data resulted in 641 presence and 107 absence locations.

As mentioned above we developed models for two subregions, the arid south-west and more humid south-east, along with the contiguous United States to allow potential differences in climate and vegetation drivers to be examined independently (Fig. 1c). Both regions contain areas currently invaded by AHB, and the south-east region encompasses the Atlantic and Gulf Coastal Plain and the Appalachian-Ozark Upland forage regions. The south-west region includes multiple forage regions. We subsampled our location data to a single location per environmental grid cell (30 arc second) to minimize pseudoreplication. From this subset, we randomly selected an equal number of presence and absence locations within the south-east and within the south-west.

Environmental data layers

We considered 40 environmental data layers consisting of climate, land cover and vegetation phenology variables to parameterize the models (see Table S1 in Supporting Information). Climate data included 19 bioclimatic layers from WorldClim (Hijmans *et al.*, 2005) that are derived by interpolation of average monthly climate data at 30-arc-second (approximately 1 km) resolution. Vegetation cover layers from the National Aeronautics and Space Administration's (NASA's) Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Continuous Fields (VCF) product, including percentage estimates for trees (for 2005), herbaceous vegetation and bare ground cover (for 2001; Hansen *et al.*, 2003), was included. The MODIS Land Surface Phenology product was the source of 15 metrics of seasonal variation in vegetation productivity from 2001 to 2007, excluding 2005 (Tan *et al.*, 2008). Similar to the long-term average climate used, we calculated the average of each phenology metric across all years available at the time of analysis (2001–2007 excluding 2005). All data were resampled to 30 arc seconds to match the lowest resolution data set using ENVI software (Exelis Visual Information Solutions, Boulder, CO, USA) with the nearest neighbour method for resampling.

To reduce multicollinearity issues and predictor redundancy, we examined Pearson's correlation coefficients between pairs of variables using SYSTAT 12 (Systat Software, Inc., Chicago, IL, USA) for each of the three regions. We retained the variable considered to be the more biologically meaningful from pairs of variables with Pearson's correlation

coefficient values >0.8 or <−0.8. Selected variables for each region are shown in Table S1. This selection process resulted in a reduction to 19 variables for the United States, 16 variables for the south-west and 18 variables for the south-east.

Species distribution modelling

We paired the selected environmental layers consisting of climate, land cover and vegetation phenology variables with a random partition of the presence/absence locations (random 70% of data points used to train the model, random 30% retained to test model) across the United States to build and evaluate habitat suitability models for AHB in North America. We used four statistical modelling techniques for binary data that have performed well in the past (Elith *et al.*, 2006). These techniques included generalized linear modelling (GLM; Hosmer & Lemeshow, 2000); boosted regression tree (BRT; Elith *et al.*, 2008); multivariate adaptive regression splines (MARS; Leathwick *et al.*, 2006); and random forest (RF; Prasad *et al.*, 2006). All techniques except RF included variable selection within the modelling algorithm, so although each technique began with the same set of predictor variables (from Table S1), each final model depended on a unique set of variables (variables with values in Table S2). The GLM employed standard stepwise regression using Akaike's information criterion (AIC); the BRT technique generally does not include non-informative predictors when fitting trees; the MARS adds terms to a model beginning with only the intercept, until there is no longer a reduction in sum-of-squares residual error and then prunes the model until it achieves the best model according to generalized cross-validation.

We developed an ensemble of the results from these four presence/absence techniques following Stohlgren *et al.* (2010) within the Software for Assisted Habitat Modeling (SAHM; http://www.fort.usgs.gov/ram). The SAHM program utilizes modules written in R code to calculate the models, thresholds and assessment statistics. The ensemble was created by developing a model for each of the four techniques, discretizing the continuous predictions produced from each model to binary values representing suitable and unsuitable habitat, and adding the four binary maps together. We used the equal sensitivity and specificity threshold rule (see Liu *et al.*, 2005) to covert the continuous predictions into discrete categories of suitable or unsuitable habitat for each of the four models, where presence locations are just as likely to be erroneously predicted as absence locations. This threshold rule has performed well in a comparison of various threshold selection rules (Liu *et al.*, 2005; Jimenez-Valverde & Lobo, 2007). The ensemble model had values ranging from zero to four, where a zero indicated none of the four models predicted a location as suitable, a value of one indicates a single one of the four models predicted a location as suitable and so on, up to a value of four indicating that all four models predicted a location as suitable. We examined these ensemble values to determine the level of agreement between the four different discretized model predictions, where a value of four would indicate agreed upon suitable habitat across all four models and a value of zero would indicate agreed upon unsuitable habitat across all four models.

We evaluated model performance using the test data withheld from model generation. The evaluation metrics included the receiver operating characteristic area under the curve (AUC) values, R^2 and overall percentage correct. The AUC is a threshold-independent metric with values between 0.5 and 1, where values >0.8 indicate high accuracy (Swets, 1988). The R^2 and overall percentage correct metrics were dependent on the threshold rule we used (the value where sensitivity equalled specificity).

Supplementary absence data

The high ratio of presence to absence locations (641–107) likely reflected a bias in our sampling, which included mainly presence-only data sets as genetic testing is often only conducted when a colony exhibits AHB behavioural traits, rather than a reflection of actual prevalence across the landscape. Thus, we followed the recommendation of McPherson *et al.* (2004) to subsample our location data to include an equal number of presence and absence locations. As doing this greatly reduced our sample size, it was desirable to supplement absence data to alleviate to maintain a higher sample size. Additionally, given that we are dealing with an invasive species and are interested in potential – not current – distribution, any absence locations we have could be viewed as pseudo-absence locations as this invasive species may still be spreading. We generated pseudo-absence data by selecting locations predicted as unsuitable based on data collected at known EHB and AHB hives. We hypothesized that AHB are less likely to overwinter in conditions where a sustained winter dearth interrupts nectar flows and/or regions with only a single spring nectar flow, due to differences in food storage and swarming behaviour between the two groups (Winston, 1992; Schneider *et al.*, 2004). We hypothesize that AHB colonies (which have a higher propensity for swarm production throughout the summer) require at least two nectar flow seasons per year to enable colonies and swarms to survive the winter compared with EHB that generally exhibit only a single swarming period during the spring and early nectar flows. In spring-only nectar flow regions, AHB colonies should exhibit greatly reduced survivorship because the later swarms cannot collect enough nectar to overwinter successfully. Therefore, we postulate that locations with only single-season nectar flows would be good surrogate AHB absence locations. Thus, we examined preliminary model relationships for locations with differing nectar flow phenologies, based on scale hive samples, to test this hypothesis.

We derived nectar flow data from changes in hive weight obtained from the HoneyBeeNet (http://honeybeenet.gsfc.nasa.gov) network of scale hives for eight sites (Fig. 2). These sites were categorized as either having unimodal ('spring or summer only') or multiple nectar flows. For this purpose,

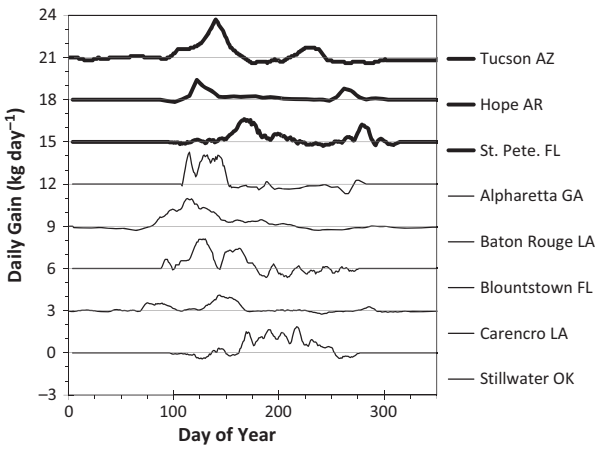

Figure 2 Daily hive gain for sites with bimodal nectar seasons for three locations within known Africanized honeybee (AHB) range (top three, bold lines) and five representative locations with unimodal nectar seasons outside of existing AHB range (bottom five, light lines). Data were smoothed with a seven day running average and an offset value of 3 kg has been applied sequentially to each location for separation.

the seasons were defined by day of year as spring before day 170, summer from day 170 to day 225 and fall after day 225. Daily hive gains above 0.435 kg signified that nectar flow had started (~two standard deviations above environmental noise; Esaias, unpublished data), with unimodal sites defined as those with such a gain during one season while bimodal was defined by such gains during two different seasons (i.e. spring and fall). We generated preliminary models for the three regions (contiguous United States and two subregions), extracted ensemble model values for each scale hive site and averaged the three regional values for each site. Both a two-tailed *t*-test and a Mann–Whitney test showed a significant difference between the values for unimodal and bimodal nectar flow sites (Table 1; $P = 0.0282$ for the US, 0.0003 for the ensemble average; $t = 7.417$ for the ensemble average). Given these relationships, we supplemented the absence data with 37 EHB scale hive locations from single nectar flow regions to improve our sample size (Fig. 1a scale hive absence locations). These locations were selected based on data collected at each hive, rather than information gleaned from remotely sensed imagery. This increased our absence location sample size in the final models we discuss to 180. Thus, after correcting for unequal numbers of presence and absence data, we used a total of 88 locations for the south-west and 228 locations for the south-east. The US extent was composed of the two regional subsets (all 316 locations).

Table 1 The ensemble scores and tests between the nectar flow groups. The Africanized honeybee (AHB) group gives the observed AHB status.

Site	Ensemble US	Ensemble score avg US, SW, SE	Nectar flow season	AHB group
Tucson AZ	4	4	2	Present
St. Petersburg FL	4	3.67	2	Present
Hope AR	4	3.67	2	Present
Alpharetta GA	0	1.33	1	Absent
Baton Rouge LA	0	1.67	1	Absent
Carencro LA	2	2	1	Absent
Blountstown FL	1	1.67	1	Absent
Stillwater OK	0	0.67	1	Absent
Two-tailed P	0.0282	0.0003		

Mann–Whitney test: $t = 7.417$; df = 6; $U = 0.0$; $n = 8$.

RESULTS

In the model for the United States, the climate variables, rather than satellite-derived vegetation variables, were generally selected through the model fitting process (Table S2). Exceptions included inclusion of herbaceous cover in the GLM and vegetation metric variable enhanced vegetation index (EVI) difference from root mean square error (RMSE) in the MARS model. While the RF model does not select variables but rather includes all those provided, the Gini index can be interpreted as a measure of variable importance, and the climate variables all had greater importance than the phenology predictors and the other satellite products. Evaluation of the models with the test data revealed that the overall percentage correct (i.e. observed presence location in area modelled as suitable AHB habitat and observed absence location in area modelled as unsuitable) ranged between 93% and 96% and all AUC values were >0.92 (Table 2). Examination of the ensemble of the four binary maps revealed high agreement among all four models, with 58% of grid cells predicted as suitable by at least one model also being predicted as suitable by all four models (locations with ensemble model value = 4). These high values extended across the south-west and in to Florida, while lower ensemble scores representing disagreement between models covered less area (3 = 17%, 2 = 12% and 1 = 13%; Fig. 3a). These models and reports can be viewed on the AHB website (http://ahb.colostate.edu).

Table 2 Receiver operating characteristic area under the curve (AUC) values (and R^2 values) for the test data for the models for the United States (US), south-west region (SW) and the south-east (SE) for each of the model techniques boosted regression tree (BRT), generalized linear model (GLM), multivariate adaptive regression splines (MARS) and random forest (RF).

Model	US AUC (R^2)	SW AUC (R^2)	SE AUC (R^2)
BRT	0.950 (83.1)	0.997 (85.8)	0.987 (77.9)
GLM	0.930 (72.0)	0.907 (53.1)	0.915 (50.2)
MARS	0.940 (73.4)	1.000 (88.2)	0.935 (63.4)
RF	0.976 (81.8)	1.000 (84.0)	0.950 (68.5)

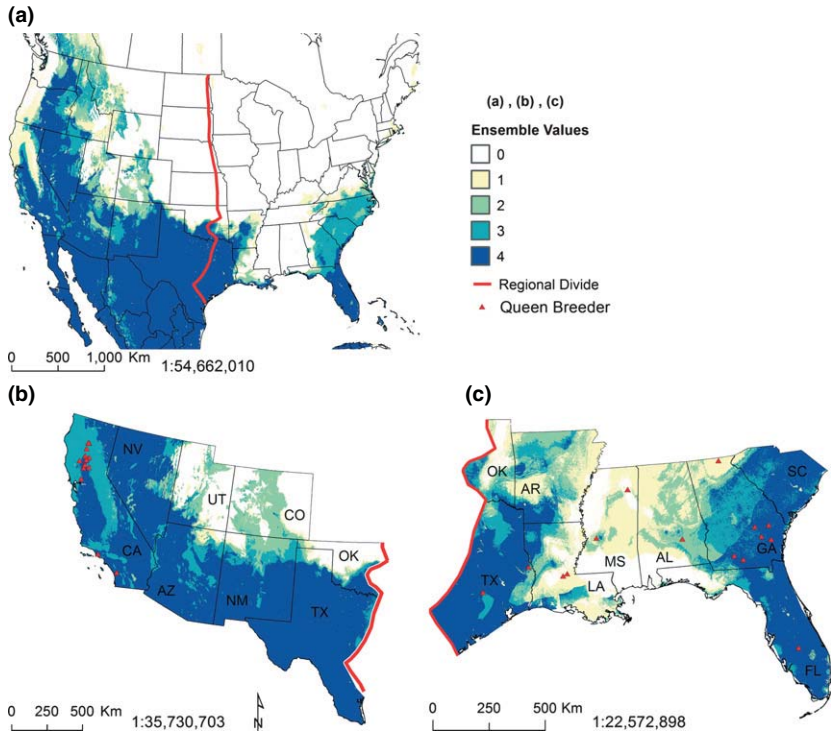

Figure 3 National and regional habitat suitability ensemble models for Africanized honeybee (AHB) (a) the United States, (b) the south-west region and (c) the south-east region with representative queen bee breeder locations overlaid (red triangles) and the red line depicting the regional divide. Queen breeder locations were taken from all advertisements in two issues of the American Bee Journal and two of Bee Culture. Values represent the number (0–4) of models with a prediction of suitable at that location (North America Albers Equal Area Conic projection, Datum NAD83).

The next two sets of predictions were conducted regionally for the south-western and south-eastern United States to capture potential differences in drivers of AHB distribution in smaller, environmentally distinct areas according to bee forage zones. The south-west regional model was superior in predictions according to the assessment metrics, although there was greater variance in the prediction success of the models with overall percentage correct between 88.2% and 100% and AUC values from 0.91 to 1.0 (Table 2). Model agreement was again high, with the greatest two ensemble values accounting for 87% of the grid cells (4 = 64%; 3 = 23%; 2 = 11% and 1 = 3%; Fig. 3b). Total numbers of predictors included in the final GLM, BRT and MARS models were four, two and eight, respectively. Selected predictors by GLM, BRT and MARS were again climate variables with the exception of EVI green up rate with MARS. Within the RF model, phenology predictors again contributed less than climate predictors according to the Gini index and only one phenology predictor had a Gini index >0.5.

Evaluation metrics for the south-east region were not as successful, with percentage correct between 87% and 93%, and AUC values from 0.91 to 0.99. Model agreement was much lower, with similar percentages across the ensemble scores (4 = 42%, 3 = 17%, 2 = 17%, 1 = 24%; Fig. 3c). The south-east had more phenology predictors selected (EVI RMSE and EVI difference from RMSE in GLM; EVI season length in MARS; five phenology variables with mean accuracy >1 in random forest; Table S2). Within RF, the phenology Gini values were again lower than the climate predictors, but were higher than in the other two models (all >0.5 with three >1.0). Because supplemental absence data were greater for the south-east model and may have influenced phenology predictor inclusion, we also examined the variables selected in the preliminary models used to evaluate selection of the supplemental absence data. In these models, more phenology predictors were again selected (peak date in GLM; winter dearth and EVI amplitude in MARS; and winter dearth and EVI base levels in BRT). Similar to the south-east model including the supplemental absence locations, RF phenology Gini index values were higher than the other models, with all >0.5 and two >1.0.

DISCUSSION

Vegetation phenology metric variables were selected as AHB habitat suitability predictors in almost all models. While NDVI has been a commonly used predictor in models, actual phenologic information has not often been used. To our knowledge, few papers exist predicting distributions of pollinators (Hinojosa-Diaz *et al.*, 2009; Kadoya *et al.*, 2009), and to date, no models have used phenology metrics as

predictors. We interpret their selection here as a surrogate for bee forage availability (nectar and pollen) for these generalist pollinators. Our comparison of preliminary models with scale hive locations indicating either single or multiple nectar flows supported our hypothesis that AHB respond to phenology.

Based on these models, it appears that the AHB may continue to extend its range northwards with subtle spatial patterns reflecting influences from both climatic and vegetation conditions (Fig. 3). Within the ensemble approach, each individual model extends AHB expansion farther north than previously expected (Harrison *et al.*, 2006), although this may be somewhat influenced by bias in the available training points and the original bias towards presence locations that we attempted to alleviate by creating additional pseudo-absence locations. This study, which focused on the United States, lacked both presence and absence data points near the core of AHB distribution in the south-western states. Arizona was among the first states invaded by the AHB, and unfortunately, these original data records were lost and data collection ceased. As noted in the methods, we did supplement data in these locations from other sources, but biases may still exist. Despite these data gaps, the models show new areas of concern including EHB queen breeding regions in the south-eastern United States and along the southern Atlantic Coast. The Central Valley in California appears to provide suitable habitat, as does most of the Basin and Range province and Washington State. These new areas of concern are considerably further north of current AHB invaded regions.

Models for the south-west provided the most accurate results of the three according to the evaluation metrics, while the south-east results were least favourable. This result may stem from the divergent variables driving AHB distribution in the two regions; in particular, we hypothesize that phenology variables are more important in the south-east, while climatic variables are more important in the south-west and nationally. Thus, the south-east models may be inferior because of imprecision in the relationship between NDVI phenology metrics and bee forage availability. While all currently available phenology metrics were included in the model, these metrics may not be closely correlated with timing of nectar flows and hence AHB food availability as the metrics are based on total vegetation, and not necessarily blooming of species favourable for honeybee forage which are generally only a fraction of the total vegetation. Scale hive data that provide a measurement of timing and number of nectar flows would be ideal, but such information is unavailable across the United States.

There is a highly significant correlation between suitable AHB habitat and the gross phenology of nectar flows (multiple versus single annual peaks) as determined by daily to weekly weighing of EHB colonies for locations within 200 km of suitable habitat in Fig. 3(a) ($P < 0.001$, $n = 8$, Table 1 and Fig. 2). Nectar flows within the AHB native range in Africa are biannual (Hepburn & Radloff, 1998), and the phenology of the AHB is closely coupled with local plant phenology and phytochoria (Hepburn & Radloff, 1995). The AHB generally live in smaller clusters and have a higher metabolic rate than EHB, and their propensity for reproductive swarming in response to pollen availability appears to require significant nectar availability in the late summer to fall period, with short winter dearth (Winston, 1992; see for example Harrison *et al.*, 2006). These associated AHB traits appear to be highly conserved despite interbreeding (hybridizing) with EHB during its 50 year, 6000 km migration northwards. The AHB requirement for strong fall nectar flows suggests that usurpation of EHB colonies containing large honey stores by the AHB (Schneider *et al.*, 2004) has clear survival value in regions with classic bell-shaped, unimodal nectar phenology. These unfavourable 'spring only' nectar flows were first encountered by the AHB when crossing into Louisiana from eastern Texas. The association between floral type associated with higher rainfall from Louisiana eastward, and the apparent cessation of AHB expansion east from Texas were noted by Villa *et al.* (2002). However, this study additionally explains why South Florida, with multiple annual nectar flows, is suitable AHB habitat despite higher precipitation than in the monsoonal south-western US. Projections of future AHB expansion in response to climate warming could thus also be dependent upon plant succession and/or changes in agriculture that result in bi-modal nectar phenology rather than warming or climate changes per se. Properly defining nectar flow across the broad regions, from the bimodal flows found in the arid south-west with its seasonal monsoons, to non-seasonal southern Florida, and the strong vernal flows found along the east coast will require continued monitoring of the honeybee nectar flow. Different phenology metrics specifically tuned to correlate better with the timing of nectar flows rather than the current greenness-based metrics, if possible in the future, might improve predictions.

There are additional sources of uncertainty in the models. Buisson *et al.* (2009) partitioned various sources of uncertainty and determined that model method introduced the most variability. By creating an ensemble model, we provide information on uncertainty caused by modelling method. Additional sources of uncertainty arise from bias in our location data, both presence and absence, and because AHB may still be spreading, albeit relatively slowly. This study is exploratory and provides a preliminary understanding until additional data are gathered and modelled in an iterative approach (Crall *et al.*, 2013).

These findings have valuable application for predictions related to honeybees and other pollinator species. For example, they could guide where migratory beekeepers might overwinter EHB colonies to minimize potential impact of AHB, and identify where queen breeders may need to pay close attention to hybridization of their European stocks with AHB. The novel use of phenology as predictors here highlights a useful application of remote sensing products. The models underscore the importance of phenology in

understanding the current and potential future distributions of bees and potentially other organisms. The regional model approach also allowed us to distinguish potential geographical differences in factors related to AHB distribution. Continued work investigating nectar flow maps derived using satellite and scale hive data should help improve distribution models and understanding of drivers of distribution for species-dependent on nectar and pollen as a food source – our earth's pollinators.

ACKNOWLEDGEMENTS

This research is funded by the NASA Applied Sciences Program and USGS. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We thank R. Grantham, J. Hayes, A. Szalaski and the USDA for providing AHB location information.

REFERENCES

Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.

Ayers, G.S. & Harman, J.R. (1992) Bee forage of North America and the potential for planting for bees. *The Hive and the Honey Bee* (ed. by J. Graham) pp. 437–493. Dadant and Sons Inc, Hamilton, IL.

Bolliger, J., Kienast, F., Soliva, R. & Rutherford, G. (2007) Spatial sensitivity of species habitat patterns to scenarios of land use change (Switzerland). *Landscape Ecology*, **22**, 773–789.

Bradley, B.A., Olsson, A.D., Wang, O., Dickson, B.G., Pelech, L., Sesnie, S.E. & Zachmann, L.J. (2012) Species detection vs. habitat suitability: are we biasing habitat suitability models with remotely sensed data? *Ecological Modelling*, **244**, 57–64.

Buisson, L., Thuiller, W., Casajus, N., Lek, S. & Grenouillet, G. (2009) Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, **16**, 1145–1157.

Crall, A.W., Jarnevich, C.S., Panke, B., Young, N., Renz, M. & Morisette, J. (2013) Using habitat suitability models to target invasive plant species surveys. *Ecological Applications*, **23**, 60–72.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.

Fei, S.L., Liang, L., Paillet, F.L., Steiner, K.C., Fang, J.Y., Shen, Z.H., Wang, Z.H. & Hebard, F.V. (2012) Modelling chestnut biogeography for American chestnut restoration. *Diversity and Distributions*, **18**, 754–768.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Carroll, M., Dimiceli, C. & Sohlberg, R.A. (2003) Global percent tree cover at a spatial resolution of 500 meters: first results of the MODIS vegetation continuous fields algorithm. *Earth Interactions*, **7**, 1–15.

Harrison, J.F., Fewell, J.H., Anderson, K.E. & Loper, G.M. (2006) Environmental physiology of the invasion of the Americas by Africanized honeybees. *Integrative and Comparative Biology*, **46**, 1110–1122.

Heikkinen, R.K., Luoto, M., Araujo, M.B., Virkkala, R., Thuiller, W. & Sykes, M.T. (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, **30**, 751–777.

Hepburn, H.R. & Radloff, S.E. (1995) First approximation to a phenology of the honeybees (Apis mellifera) and flora of Africa. *Oecologia*, **1010**, 265–273.

Hepburn, H.R. & Radloff, S.E. (1998) *Honeybees of Africa*. Springer, New York.

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

Hinojosa-Diaz, I.A., Feria-Arroyo, T.P. & Engel, M.S. (2009) Potential distribution of orchid bees outside their native range: the cases of Eulaema polychroma (Mocsary) and Euglossa viridissima Friese in the USA (Hymenoptera: Apidae). *Diversity and Distributions*, **15**, 421–428.

Hosmer, D.W. & Lemeshow, S. (2000) *Applied logistic regression*, 2nd edn. Wiley, New York.

Jimenez-Valverde, A. & Lobo, J.M. (2007) Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica-International Journal of Ecology*, **31**, 361–369.

Kadoya, T., Ishii, H.S., Kikuchi, R., Suda, S. & Washitani, I. (2009) Using monitoring data gathered by volunteers to predict the potential distribution of the invasive alien bumblebee Bombus terrestris. *Biological Conservation*, **142**, 1011–1017.

Leathwick, J.R., Elith, J. & Hastie, T. (2006) Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling*, **199**, 188–196.

Liu, C.R., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385–393.

McAlpine, C.A., Rhodes, J.R., Bowen, M.E., Lunney, D., Callaghan, J.G., Mitchell, D.L. & Possingham, H.P. (2008) Can multiscale models of species' distribution be generalized from region to region? A case study of the koala. *Journal of Applied Ecology*, **45**, 558–567.

McPherson, J.M., Jetz, W. & Rogers, D.J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811–823.

Measey, G.J., Rödder, D., Green, S.L., Kobayashi, R., Lillo, F., Lobos, G., Rebelo, R. & Thirion, J.M. (2012) Ongoing

invasions of the African clawed frog, *Xenopus laevis*: a global review. *Biological Invasions*, **14**, 2255–2270.

Prasad, A.M., Iverson, L.R. & Liaw, A. (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, **9**, 181–199.

Schneider, S., DeGrandi-Hoffman, G. & Roan Smith, D. (2004) The African honey bee: factors contributing to a successful biological invasion. *Annual Review of Entomology*, **49**, 351–376.

Southwick, E.E., Roubik, D.W. & Williams, J.M. (1990) Comparative energy balance in groups of Africanized and European honey bees: ecological implications. *Comparative Biochemistry and Physiology Part A: Physiology*, **97**, 1–7.

Stohlgren, T.J., Ma, P., Kumar, S., Rocca, M., Morisette, J.T., Jarnevich, C.S. & Benson, N. (2010) Ensemble habitat mapping of invasive plant species. *Risk Analysis*, **30**, 224–235.

Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.

Tan, B., Morisette, J.T., Wolfe, R., Gao, F., Ederer, G., Nightingale, J. & Pedelty, J.A. (2008) Vegetation phenology metrics derived from temporally smoothed and gap-filled MODIS data. *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, **3**, 593–595.

Taylor, O.R. & Spivak, M. (1984) Climatic limits of tropical African honeybees in the Americas. *Bee World*, **65**, 38–47.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L. & Williams, S.E. (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.

Urbina-Cardona, J.N. & Flores-Villela, O. (2010) Ecological-niche modeling and prioritization of conservation-area networks for Mexican herpetofauna. *Conservation Biology*, **24**, 1031–1041.

Villa, J.D., Rinderer, T.E. & Stelzer, J.A. (2002) Answers to the puzzling distribution of Africanized bees in the United States. *American Bee Journal*, **142**, 480–483.

Winston, M.L. (1992) *Killer bees the Africanized honey bee in the Americas*. Harvard University Press, Cambridge, MA.

Zimmermann, N.E., Edwards, T.C., Moisen, G.G., Frescino, T.S. & Blackard, J.A. (2007) Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology*, **44**, 1057–1067.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1. Data layers used in analysis.

Table S2. Output values for each model for each region.

BIOSKETCH

Catherine S. Jarnevich's research focuses on patterns of invasion. She works with resource managers to develop species distribution models for invasive species to assist in management of invasive species.

Author contributions: All conceived the ideas; P.M., B.T. and J.N. collected the data; C.J., P.M. and W.E. analysed the data; and C.J. and W.E. led the writing.

Editor: Mark Robertson