PROCEEDINGS B

Research

Author for correspondence:
Scott H. McArt
e-mail: shm33@cornell.edu

THE ROYAL SOCIETY
PUBLISHING

Landscape predictors of pathogen prevalence and range contractions in US bumblebees

Scott H. McArt[1], Christine Urbanowicz[2], Shaun McCoshum[3], Rebecca E. Irwin[4] and Lynn S. Adler[5]

[1]Department of Entomology, Cornell University, Ithaca, NY 14853, USA
[2]Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA
[3]Westchester Land Trust, Bedford Hills, NY 10507, USA
[4]Department of Applied Ecology, North Carolina State University, Raleigh, NC 27695, USA
[5]Department of Biology, University of Massachusetts, Amherst, MA 01003, USA

SHM, 0000-0001-7157-9011

Several species of bumblebees have recently experienced range contractions and possible extinctions. While threats to bees are numerous, few analyses have attempted to understand the relative importance of multiple stressors. Such analyses are critical for prioritizing conservation strategies. Here, we describe a landscape analysis of factors predicted to cause bumblebee declines in the USA. We quantified 24 habitat, land-use and pesticide usage variables across 284 sampling locations, assessing which variables predicted pathogen prevalence and range contractions via machine learning model selection techniques. We found that greater usage of the fungicide chlorothalonil was the best predictor of pathogen (*Nosema bombi*) prevalence in four declining species of bumblebees. *Nosema bombi* has previously been found in greater prevalence in some declining US bumblebee species compared to stable species. Greater usage of total fungicides was the strongest predictor of range contractions in declining species, with bumblebees in the northern USA experiencing greater likelihood of loss from previously occupied areas. These results extend several recent laboratory and semi-field studies that have found surprising links between fungicide exposure and bee health. Specifically, our data suggest landscape-scale connections between fungicide usage, pathogen prevalence and declines of threatened and endangered bumblebees.

1. Introduction

Many bee species are experiencing range contractions and extinction threats, with bumblebees providing some of the most conclusive data [1–5]. Evidence suggests several factors are associated with bee declines, including lack of floral resources, pesticides and pathogens [6]. Furthermore, current studies suggest interactions among stressors are important. For example, pesticide stress can influence susceptibility to pathogens [7], and diet stress can exacerbate the impact of pesticides [8] and pathogens [9,10]. Few studies, however, have assessed the relative importance of multiple factors associated with bee health. This is an important knowledge gap because understanding the relative importance of stressors is crucial for implementing effective conservation programmes.

Landscape-scale analyses can provide broad insight into multiple competing factors that may be impacting bee health. Because bumblebees typically forage several hundred or thousand metres from their nests (see the electronic supplementary material, table S1), landscape characteristics such as habitat composition and fragmentation [11,12], urbanization [13] and agricultural intensification [2] can impact foraging behaviour, pathogen prevalence and hive performance. With this understanding, Williams *et al.* [14] performed to our knowledge the first analysis of multiple potential factors associated with bumblebee declines, considering several

behavioural, physiological, dietary and geographical factors. Their analysis found that declines were greater for species that exhibit greater climatic specialization, in areas closest to the edges of their climatic ranges, and for species that have queens that become active later in the year. Szabo et al. [15] built on this study by explicitly considering landscape factors that could be associated with bumblebee declines, including commercial greenhouse density, pesticide usage summaries and human population density in their models. While this analysis highlighted the potential importance of greenhouse density (and thus possible negative effects of pathogen spillover from managed bumblebees), the link between greenhouse density and pathogens was inferred but not assessed empirically. Furthermore, the small number of variables considered limit a more robust consideration of stressors. For example, could specific pesticides, habitat types and/or land-use practices be important predictors of pathogen prevalence and landscape-wide declines?

Previous data show that several species of US bumblebees are experiencing range contractions, and some declining species have elevated levels of the microsporidian pathogen *Nosema bombi* [5,16,17]. Yet factors that may be causing high pathogen levels have not been elucidated, nor has potential covariation between pathogens, range contractions and landscape characteristics been studied. Because myriad landscape characteristics vary among sites, a common analytical approach for landscape analyses is multiple regression [18]. Yet such analyses are often limited in scope because including large numbers of predictors in traditional multiple regression models poses a statistical challenge. However, such analyses of high dimensional data (i.e. high independent to dependent variable ratio) are possible via modern machine learning techniques [19–21]. Techniques such as stability selection [20] and least absolute shrinkage and selection operator (LASSO) [19], are advantageous for large datasets with many independent variables where some automatic variable selection is desired, or for dealing with highly correlated predictors, where standard multiple regression will usually have regression coefficients that are too large. These techniques are being used extensively in fields such as molecular biology and genetics, for example, to select among large numbers of genes in regulatory network analyses (e.g. [22,23]). However, machine learning multi-variable model selection remains rare in ecological and conservation-oriented studies (e.g. [24,25]), possibly owing to lack of familiarity with the techniques.

Here, we summarize an analysis that uses stability selection [20] and LASSO [19] to assess the potential importance of 24 habitat, land-use and pesticide usage variables for predicting pathogen prevalence and range contractions of bumblebees. We used one of the largest datasets to date of bumblebee declines in North America, a contemporary sampling of 10 745 bumblebees across the USA that was compared to a 73 759-specimen natural history collection database, as described in Cameron et al. [5] and Cordes et al. [16], to ask two questions: (i) what factors predict pathogen (*N. bombi*) prevalence in four focal declining bumblebee species, and (ii) what factors predict range contractions in declining species?

2. Methods

(a) Bumblebee sampling and pathogen screening
For full details of bumblebee sampling and pathogen screening, see Cameron et al. [5] and Cordes et al. [16]. Briefly, between 2007

and 2009, 284 sites in 40 states throughout the USA were sampled from June to August for bumblebees for a period of 1 ± 0.5 s.d. person-hours. This sampling yielded 10 745 individual bees from 36 *Bombus* species. All bees were identified to species and screened for two pathogens, *N. bombi* and *Crithidia bombi*, under phase contrast microscopy ($400\times$ magnification). The sampling of bees was directed to compare pathogen presence and prevalence among eight bumblebee target species hypothesized to be in decline or stable: *Bombus bifarius* (stable), *Bombus occidentalis* (declining) and *Bombus vosnesenskii* (stable) in the western USA; and *Bombus affinis*, *Bombus pensylvanicus* and *Bombus terricola* (all declining), and *Bombus bimaculatus* and *Bombus impatiens* (both stable) in the eastern USA.

(b) Landscape variable characterization and quantification
At each of the 284 sampling sites, we quantified 24 habitat, land-use and pesticide usage variables that could potentially be associated with bee health. We quantified latitude, longitude and elevation at each site using GPS coordinates from the original dataset [5,16]. We used the 2006 National Land Cover Database [26] to compute the areas of five land-use/land-cover metrics: (i) cultivated crops, (ii) agriculture (pasture, hay, cultivated crops), (iii) forest (deciduous forest, evergreen forest, mixed forest), (iv) developed (developed open space, low intensity developed, medium intensity developed, high intensity developed), and (v) natural area (all forest classes, wetland classes, shrub classes, herbaceous grassland). We calculated total areas for each metric within four buffers around each site (500, 1000, 2500 and 5000 m), which are within the range of typical bumblebee foraging distances (electronic supplementary material, table S1). We used 2009 population census data (http://www.census.gov/geo/maps-data/data/tiger-data.html) to estimate human population density in the four buffer areas. Blocks are the smallest territorial unit with reportable population data. Because blocks did not always fully overlap our buffer areas, population density was calculated as [(block area inside buffer/total block area) × block population]. This calculation assumes population is equally distributed across the entire block.

We used the spatial pattern analysis program FRAGSTATS, version 3.4 [27], to quantify the fragmentation of forested and natural areas using a 'clumpiness index' in the four buffer areas around each sampling location. Clumpiness measures habitat contagion and interspersion and varies from −1 (maximally disaggregated) to 1 (maximally aggregated), with zero indicating the expected level of aggregation under a spatially random distribution. This metric was chosen over other contagion metrics because it is not confounded by the total habitat area [28].

Finally, we used the United States Geological Survey (USGS) National Water Quality Assessment database (http://water.usgs.gov/nawqa/pnsp/usage/maps/) to estimate pesticide usage at each sampling location [29]. These data originate from farm surveys and capture use for all methods of application (sprays, treated seed, soil drenches, etc.). We divided the USGS county-level pesticide usage estimates by the area of agricultural land (as classified above) in each county to estimate pesticide use per agricultural cell. For each of our four buffers, we calculated: (i) total predicted pesticide input (including all 483 potential compounds), (ii) total predicted insecticides (137 potential compounds), (iii) total predicted fungicides (124 potential compounds), and (iv) total predicted herbicides (185 potential compounds) using high and low estimates from 2009. We also calculated predicted pesticide inputs for the three most widely used insecticides (chlorpyrifos, aldicarb, clothianidin), fungicides (chlorothalonil, mancozeb, captan) and herbicides (glyphosate, atrazine, 2,4-D) for the four buffers (electronic supplementary material, table

Table 1. Landscape variables predicting *N. bombi* presence in four declining US *Bombus* species (*B. affinis*, *B. occidentalis*, *B. pensylvanicus* and *B. terricola*) as selected by LASSO and stability selection analyses.

variable	LASSO coefficient[a]	stability selection π_{thr}[b]	*p*-value[c]	ΔAIC[d]
log developed area	−0.315	**1.00**	0.002	8.5
latitude	0.033	**1.00**	0.091	0.9
log chlorothalonil	1.825	**0.94**	<0.001	107.0
longitude	−0.013	0.60		
natural area fragmentation	−0.472	0.36		
log human population	0.000	0.22		
log elevation	0.040	0.18		
log 2,4-D	−0.409	0.10		
log captan	−0.240	0.07		
log aldicarb	2.430	0.06		
log agricultural area	−0.019	0.05		
log chlorpyrifos	0.119	0.02		
log natural area	0.003	0.01		
log atrazine	0.215	0.00		
log mancozeb	−1.377	0.00		
log fungicides	0.880	0.00		
log herbicides	−0.086	0.00		

[a]LASSO analysis of landscape variables quantified at a 1000 m buffer; $\lambda + 1$ standard error of the minimum (0.002) produced a 17-variable model.

[b]Per-family error rate (PFER) value of 1.5 used in stability selection analysis with cut-off value of $\pi_{thr} = 0.75$ [20]. Bold values indicate variables above the π_{thr} cut-off.

[c]*p*-values obtained from *glmer* analysis of the three-variable stability-selected model.

[d]ΔAIC = the change in AIC values when each variable is individually removed from the three-variable stability-selected model.

S2). Clothianidin was the fourth most abundant insecticide used in 2009, but it was included to represent one insecticide each from three major chemical families (i.e. organophosphate, carbamate, neonicotinoid).

(c) Landscape analysis of pathogen prevalence in stable and declining bumblebee species

Following Cameron *et al*. [5], we focused our analyses on the eight historically abundant US bumblebee species that were targeted in their study. Previous data suggested these species had experienced demographic trajectories ranging from population declines to possible expansions. These data have since been confirmed, with the four declining species (*B. occidentalis*, *B. affinis*, *B. pensylvanicus* and *B. terricola*) listed as vulnerable or critically endangered and the four stable species (*B. vosnesenskii*, *B. bifarius*, *B. impatiens* and *B. bimaculatus*) listed as stable and of least concern on the current International Union for Conservation of Nature Red List (http://www.iucnredlist.org/). Of the 7831 bees from the 'stable' species that were sampled, 105 bees (1.3%) were positive for *N. bombi* and 246 bees (3.1%) were positive for *C. bombi*. Of the 762 bees from the 'declining' species that were sampled, 140 bees (18.3%) were positive for *N. bombi*, while only five bees (0.7%) were positive for *C. bombi*. Owing to the low incidence of *C. bombi* in the declining species, we did not attempt analyses for this group.

To assess which spatial scale was appropriate to conduct analyses (500, 1000, 2500 or 5000 m buffers), we first conducted a literature review of bumblebee foraging distances. Our review of 15 published studies comprised 13 bumblebee species and three methods of foraging distance estimation (mark–recapture, microsatellites, radio telemetry) found an average maximum worker foraging distance of 2044 m and an average mean foraging distance of 842 m (electronic supplementary material, table S1). Next, to assess how spatial scale and high and low pesticide estimates fit our data, we conducted univariate analyses for each variable as a predictor of *C. bombi* in stable species or *N. bombi* in stable or declining species using the *glm* function in R [30], comparing Akaike information criteria (AIC) values for each model. We log-transformed 18 of the 24 variables to improve normality of the residuals. We found that the 2500 m spatial buffer and low pesticide estimates most often provided the lowest AIC score for each variable (see the electronic supplementary material, tables S3–S6). Thus, based on both the foraging distance literature summary and univariate analyses, we chose to conduct analyses using three spatial buffers (500, 1000, 2500 m) and low pesticide estimates for each variable. We chose to present results from the 1000 m spatial buffer in the main text (table 1 and figure 1) because this distance most closely represents the mean foraging radius of worker bumblebees found in our literature review (electronic supplementary material, table S1); overall conclusions from analyses for the three spatial buffers were very similar (tables 1 and 2; electronic supplementary material, tables S7–10 and figures S1–S12).

We used LASSO for multi-variate model selection [19] via the glmnet package in R. LASSO is a machine learning shrinkage and selection method for multiple regression that penalizes the absolute size of regression coefficients. By penalizing (or equivalently constraining) the sum of absolute values of the estimates, some parameter estimates may be exactly zero. The larger the penalty applied, which is controlled via a tuning parameter λ, the further estimates are shrunk towards zero. We randomly chose two-thirds of each dataset for training and performed cross-validation (function *cvfit* in glmnet) on the training set to obtain the number of variables at λ minimum and within one

4

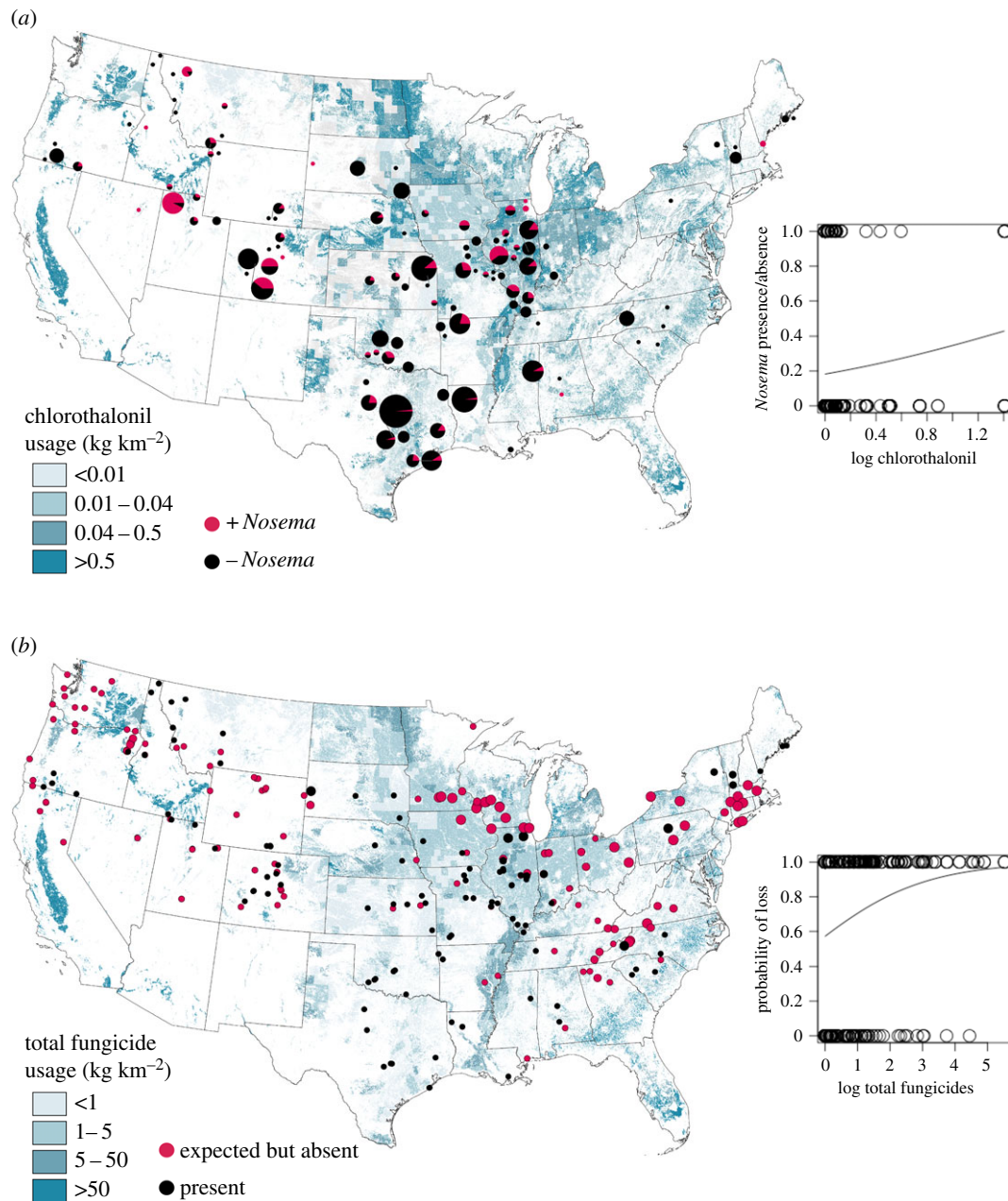rspb.royalsocietypublishing.org   Proc. R. Soc. B 284: 20172181

(a)

(b)

Figure 1. (a) Relationship between *N. bombi* presence/absence in four declining *Bombus* species (*B. affinis*, *B. occidentalis*, *B. pensylvanicus* and *B. terricola*) and chlorothalonil usage in the USA in 2009. Size of circles is proportional to abundance of bees sampled at each site, with red representing proportion of bees infected with *N. bombi*. (b) Relationship between probability of loss of the four declining *Bombus* species and total fungicide usage in the USA in 2009. The size of circles is proportional to the number of declining species expected at each site, with black indicating presence and red indicating expected presence but absent (see Methods for expected presence details).

standard error of the minimum (λ + 1 s.e.). Cross-validation is the standard method for choosing an appropriate λ value [31]. We visualized: (i) coefficient paths by fraction deviance explained as variables were added to each model, and (ii) mis-classification error across the range of λ values using 10-fold cross-validation. We assessed reproducibility of each LASSO model at λ minimum and λ + 1 s.e. by creating pathogen pres-ence/absence predictions for each model (function *predict* in glmnet), then comparing predicted error rate to an error rate expected by chance (i.e. a zero-variable model).

Next, we used the function *stabsel* in the stabs package [32] to test the fraction of times each variable was included in the LASSO-determined multi-variable model. Stability selection is a resampling procedure often employed in combination with techni-ques such as LASSO to evaluate the consistency of individual predictors in multi-variable model selection [20,21]. For example, the technique has been used for selection of environmental vari-ables predictive of species distributions [24]. We used a cut-off

value of $\pi_{\text{thr}} = 0.75$ (i.e. variable selected in 75% of fitted models) and per-family error rate (PFER) value of 1.5 for all stability selec-tion analyses [20]. Following variable selection via LASSO and stability selection, we assessed the importance of individual varia-bles via generalized linear mixed effects models (function *glmer* in the lme4 package) with bee species as a random effect, obtaining *p*-values and ΔAIC values for each variable.

(d) Landscape analysis of range contractions in declining bumblebee species

Owing to recent advances in methods for niche modelling using MAXENT software to reduce sampling bias [33,34], here we update predictions of historical US bumblebee distributions for each of the 'declining' target species summarized in Cameron *et al*. [5]. Our niche model predictions are based on the original dataset, which comprises a 73 759-specimen natural history

Table 2. Landscape variables predicting presence versus expected presence[a] of four declining *Bombus* species (*B. affinis*, *B. occidentalis*, *B. pensylvanicus* and *B. terricola*) across 284 sampling locations as selected by LASSO and stability selection analyses.

variable	LASSO coefficient[b]	stability selection π_{thr}[c]	*p*-value[d]	ΔAIC[e]
latitude	0.161	**1.00**	0.001	8.4
log fungicides	0.342	**0.84**	<0.001	32.5
longitude	0.023	0.64		
log chlorothalonil	2.131	0.59		
log developed area	0.240	0.49		
log 2,4-D	−0.128	0.37		
log aldicarb	−2.636	0.22		
log captan	−0.125	0.06		
log human population	0.013	0.04		
log atrazine	−0.117	0.03		
log agricultural area	−0.009	0.02		
log crop area	0.035	0.01		
log clothianidin	−0.504	0.00		
natural area	0.003	0.00		

[a]Expected presence determined via MAXENT niche modelling of historical specimen database (see Methods). Positive coefficients indicate variables that are positively related to loss from sites.

[b]LASSO analysis of landscape variables quantified at a 1000 m buffer; $\lambda + 1$ standard error of the minimum (0.007) produced a 14-variable model.

[c]PFER value of 1.5 used in stability selection analysis with the cut-off value of $\pi_{thr} = 0.75$ [20]. Bold values indicate variables above the π_{thr} cut-off.

[d]*p*-values obtained from *glmer* analysis of the two-variable stability-selected model.

[e]ΔAIC = the change in AIC values when each variable is individually removed from the two-variable stability-selected model.

collection database (USDA-ARS Bee Biology and Systematics Laboratory: patterns of widespread decline in North American bumblebees. doi:10.15468/kjpwz1, accessed via http://www.gbif.org/dataset/8c04889e-a0ad-4f39-8623-06aa9cf0131d on 20 March 2017). Observations were rarefied by 5 km, which reduced the number of observations per species (*B. occidentalis*: $n = 1,229$; *B. pensylvanicus*: $n = 1,088$; *B. affinis*: $n = 385$; and *B. terricola*: $n = 495$), thus reducing model bias. We created binary presence–absence rasters for each species from the continuous MAXENT models (logistic threshold = 0.20), which produced relatively conservative distribution maps for the four 'declining' target species (i.e. updated maps were slightly narrower than the original maps published in Cameron *et al.* [5]). For each species, contemporary survey sites within the expected historic presence distribution were scored as expected occurrences.

We then compared landscape variables between expected occurrence locations where sampling did not find the species (absent) versus expected occurrence locations where the species was found (present). Similar to the pathogen analysis, we conducted analyses using three spatial buffers (500, 1000, 2500 m) and low pesticide estimates for each variable. We chose to present results from the 1000 m spatial buffer in the main text (table 2 and figure 1) because this most closely represents the mean worker bumblebee foraging radius (electronic supplementary material, table S1); overall conclusions from analyses at the three spatial buffers were very similar (tables 1 and 2; electronic supplementary material, tables S7–10 and figures S1–S12). We used LASSO and stability selection as above, assessing the importance of individual variables via generalized linear mixed effects models with bee species as a random effect.

3. Results

We attempted multi-variable model selection on four datasets: *C. bombi* or *N. bombi* prevalence in the four focal stable species (*B. vosnesenskii*, *B. bifarius*, *B. bimaculatus* and *B. impatiens*), *N. bombi* prevalence in the four focal declining species (*B. occidentalis*, *B. pensylvanicus*, *B. affinis* and *B. terricola*) and presence versus expected presence of the four focal declining species at the 284 sampling locations. LASSO did not provide meaningful models for *C. bombi* or *N. bombi* prevalence in the stable species (electronic supplementary material, figures S13 and S14), probably owing to low prevalence of positive pathogen detections (3.1 and 1.3% positive detections for *C. bombi* and *N. bombi*, respectively). However, LASSO did provide meaningful models for *N. bombi* prevalence in the declining species, and presence versus expected presence in the declining species. At least one of the four declining species was expected to be present at 252 of the 284 locations sampled. We found that 8.6% of expected locations contained *B. affinis*, 25.0% of expected locations contained *B. terricola*, 54.4% of expected locations contained *B. occidentalis* and 77.1% of expected locations contained *B. pensylvanicus*.

When we investigated the 24 variables potentially associated with *N. bombi* prevalence in the four declining species via LASSO using the 1000 m spatial buffer, we found a 17-variable model was optimal at $\lambda + 1$ s.e. (electronic supplementary material, figure S1). The predicted error rate (19.6%) was more accurate than the error rate expected by chance (22.4%). When we performed stability selection on the dataset, we found three variables remained above the cut-off of $\pi_{thr} = 0.75$ (table 1; electronic supplementary material, figure S2). Developed area was negatively related to *N. bombi* prevalence ($z = -3.2$, $p = 0.002$), chlorothalonil usage was positively related to *N. bombi* prevalence ($z = 3.8$, $p < 0.001$) and latitude was marginally positively related to *N. bombi* prevalence ($z = 1.7$, $p = 0.091$) in a generalized

linear mixed model containing all three variables. Developed area and chlorothalonil usage were not correlated ($p = 0.33$), indicating both variables independently predicted *N. bombi* prevalence. ΔAIC values, a measure of the importance of each variable, were greatest for chlorothalonil (107.0) and smaller for developed area (8.5) and latitude (0.9; table 1 and figure 1*a*). The pattern of large ΔAIC values and positive relationships between chlorothalonil usage and *N. bombi* prevalence were similar in parallel analyses conducted at the 500 and 2500 m landscape variable buffers (electronic supplementary material, tables S7 and S8 and figures S5, S6, S9 and S10).

When we investigated the 24 variables potentially associated with loss of the four declining species from sampling locations via LASSO, we found a 14-variable model was optimal at $\lambda + 1$ s.e. (electronic supplementary material, figure S3). The predicted error rate (24.5%) was more accurate than the error rate expected by chance (66.7%). When we performed stability selection, we found two variables remained above the cut-off of $\pi_{thr} = 0.75$ (table 2; electronic supplementary material, figure S4). Both latitude and total fungicides were positively related to losses ($z = 3.2$, $p = 0.001$ and $z = 3.8$, $p < 0.001$, respectively) in a generalized linear mixed model containing both variables. Total fungicides and latitude were not correlated ($p = 0.19$), indicating both variables independently predicted losses. ΔAIC values were larger for total fungicides (32.5) than latitude (8.4; table 2 and figure 1*b*). The pattern of ΔAIC values and positive relationships between losses and both fungicide usage and latitude were similar in parallel analyses conducted at the 500 and 2500 m landscape variable buffers (electronic supplementary material, tables S9 and S10 and figures S7, S8, S11 and S12).

4. Discussion

Effective conservation of pollinators is contingent upon knowing which stressors contribute substantially to declines in bee health. While current research suggests multiple factors and their interactions are important [6], few studies have sought to identify the relative importance of stressors. Here, by analysing numerous habitat, land-use and pesticide usage variables, we find consistent results suggesting fungicides are the strongest predictor of both *N. bombi* prevalence and range contractions in four declining species of US bumblebees. A handful of studies are beginning to find surprising links between fungicides and bee health (e.g. [35,36–39]). Our results extend this literature beyond the laboratory by showing to our knowledge for the first time that landscape-scale connections exist between fungicide usage, pathogen prevalence and declines of four bumblebee species known to be at risk. Analyses such as this can begin to identify the relative importance of variables associated with bee declines, which can direct manipulative follow-up studies and inform more targeted conservation efforts.

While a great amount of attention has focused on the risks posed to bees from certain insecticides (e.g. [40,41]), fungicides are typically the most abundant pesticides found in bumblebee [42] and honeybee [37,43–45] hive material. Chlorothalonil, especially, has been found in great abundance in honeybee hive residues [37,43]. This fungicide has been linked to reduced colony growth in *B. impatiens* [35] and increased likelihood of *Nosema ceranae* infection in honeybees

[36,37]. We found chlorothalonil usage was the strongest predictor of *N. bombi* prevalence in four declining species of bumblebees, while total fungicide usage was the strongest predictor of range contractions. While our analyses do not allow detailed mechanistic explanations for why chlorothalonil specifically, and fungicide exposure in general, may impact bumblebees, there are multiple plausible scenarios. For example, parts per billion levels of fungicides have been shown to kill honeybee midgut cells, resulting in proliferation of *N. ceranae* [38], a close relative of *N. bombi*. Fungicides can also indirectly compromise mitochondrial regeneration and ATP production in honeybees, which may limit bees' ability to extract sufficient energy from food [39]. Furthermore, antimicrobials have recently been shown to alter bee microbiota, resulting in increased susceptibility to pathogens and greater mortality of honeybees [46]. Finally, while most fungicides are relatively non-toxic to bees, many are known to interact synergistically with insecticides, greatly increasing their toxicity (e.g. [47,48]).

In addition to fungicides, our stability-selected models also showed greater *N. bombi* prevalence in the four declining bumblebees in locations with less developed area and a trend for greater *N. bombi* prevalence at higher latitudes (table 1). Furthermore, the four declining species were more likely to be lost from sites at higher latitudes (table 2). A few studies have investigated how urbanization impacts pathogen prevalence in bees, finding that urbanization was positively associated with pathogen prevalence in bumblebees [13,49] and honeybees [50]. While our results contrast with these studies, there are reasons to suspect a negative relationship could occur between urbanization and bee pathogens. For example, greater abundance and diversity of pollen and nectar resources available in urban gardens could improve nutrition and support bee immunity [51,52].

One recent worldwide synthesis has found that bumblebees are generally failing to track warming through time at species' northern range limits, and that range losses are often occurring at southern range limits [53]. Our analyses of four declining species in the USA show that loss from sites has occurred more often towards the species' northern range boundaries and there was a trend for greater *N. bombi* prevalence at higher latitudes. We note that the strongest predictor of *N. bombi* prevalence in our study— chlorothalonil usage—was also greater at higher latitudes ($p < 0.001$, $R^2 = 0.021$). Thus, our data are consistent with that notion that fungicides or some other geographically specific factor may be associated with greater pathogen prevalence in bees, which may in turn be impacting range contractions and limiting the northward movement of bumblebees in response to climate change. Interestingly, while Kerr *et al.* [53] did not find associations between total pesticide or neonicotinoid usage and observed shifts in bumblebee species' historical ranges, the study did not test for associations between fungicide usage and range contractions. Clearly, more work needs to investigate the potential importance of links between geographically specific stressors, pathogen prevalence and latitudinal movement (or lack thereof) of declining bumblebee species.

While suggestive, we caution against over-interpretation of correlational patterns between pathogens and bee declines that may not be indicative of larger patterns or causal mechanisms. For example, while previous data show that several species of declining US bumblebees have elevated levels of

the microsporidian pathogen *N. bombi* [5,16,17], a recent worldwide phylogenetic analysis has found that parasites (including *N. bombi*) are found less often in bumblebee species in decline [54]. This is a striking associational pattern that runs counter to evidence from many individual empirical studies. However, biologically, this result can potentially be explained via non-declining species carrying internal pathogens with low fitness costs, or perhaps the pathogens may be highly lethal to populations of declining species and therefore act as strong ecological filters. This caution in interpreting causation from correlational studies should also be exercised for multi-variable model selection techniques such as LASSO and stability selection [55].

Results from our LASSO and stability selection analyses are a first step towards informing and complementing manipulative follow-up studies. For example, while manipulative studies have found links between chlorothalonil exposure and bee health [35,37], we are unaware of manipulative studies investigating the response of bumblebees to many additional commonly used fungicides in the USA, such as mancozeb, captan and pyraclostrobin. We suggest such studies are

warranted, given our results. With both landscape-scale analyses and controlled manipulative experimental data, conservation biologists and policy makers will have a broader information base upon which to make more informed decisions and improve pollinator health.

References

1. Bartomeus I, Ascher JS, Gibbs J, Danforth BN, Wagner DL, Hedtke SM, Winfree R. 2013 Historical changes in northeastern US bee pollinators related to shared ecological traits. *Proc. Natl Acad. Sci. USA* **110**, 4656–4660. (doi:10.1073/pnas.1218503110)

2. Goulson D, Lye GC, Darvill B. 2008 Decline and conservation of bumble bees. In *Annual review of entomology* (ed. A Fallon), pp. 191–208. Palo Alto, CA: Annual Reviews.

3. Williams PH, Osborne JL. 2009 Bumblebee vulnerability and conservation world-wide. *Apidologie* **40**, 367–387. (doi:10.1051/apido/2009025)

4. Kosior A, Celary W, Olejniczak P, Fijal J, Krol W, Solarz W, Plonka P. 2007 The decline of the bumble bees and cuckoo bees (Hymenoptera: Apidae: Bombini) of Western and Central Europe. *Oryx* **41**, 79–88. (doi:10.1017/S0030605307001597)

5. Cameron SA, Lozier JD, Strange JP, Koch JB, Cordes N, Solter LF, Griswold TL. 2011 Patterns of widespread decline in North American bumble bees. *Proc. Natl Acad. Sci. USA* **108**, 662–667. (doi:10.1073/pnas.1014743108)

6. Goulson D, Nicholls E, Botias C, Rotheray EL. 2015 Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* **347**, 1435–1439. (doi:10.1126/science.1255957)

7. Sánchez-Bayo F, Goulson D, Pennacchio F, Nazzi F, Goka K, Desneux N. 2016 Are bee diseases linked to pesticides?—a brief review. *Environ. Int.* **89–90**, 7–11. (doi:10.1016/j.envint.2016.01.009)

8. Renzi MT, Rodriguez-Gasol N, Medrzycki P, Porrini C, Martini A, Burgio G, Maini S, Sgolastra F. 2016 Combined effect of pollen quality and thiamethoxam on hypopharyngeal gland development and protein content in *Apis mellifera*. *Apidologie* **47**, 779–788. (doi:10.1007/s13592-016-0435-9)

9. Di Pasquale G, Salignon M, Le Conte Y, Belzunces LP, Decourtye A, Kretzschmar A, Suchail S, Brunet JL, Alaux C. 2013 Influence of pollen nutrition on honey bee health: do pollen quality and diversity matter? *PLoS ONE* **8**, e72016. (doi:10.1371/journal.pone.0072016)

10. Brunner FS, Schmid-Hempel P, Barribeau SM. 2014 Protein-poor diet reduces host-specific immune gene expression in *Bombus terrestris*. *Proc. R. Soc. B* **281**, 20140128. (doi:10.1098/rspb.2014.0128)

11. Redhead JW, Dreier S, Bourke AF.G., Heard MS, Jordan WC, Sumner S, Wang JL, Carvell C. 2016 Effects of habitat composition and landscape structure on worker foraging distances of five bumble bee species. *Ecol. Appl.* **26**, 726–739. (doi:10.1890/15-0546)

12. Jha S, Kremen C. 2013 Resource diversity and landscape-level homogeneity drive native bee foraging. *Proc. Natl Acad. Sci. USA* **110**, 555–558. (doi:10.1073/pnas.1208682110)

13. Theodorou P, Radzevičiūtė R, Settele J, Schweiger O, Murray TE, Paxton RJ. 2016 Pollination services enhanced with urbanization despite increasing pollinator parasitism. *Proc. R. Soc. B* **283**, 20160561. (doi:10.1098/rspb.2016.0561)

14. Williams P, Colla S, Xie Z. 2009 Bumblebee vulnerability: common correlates of winners and losers across three continents. *Conserv. Biol.* **23**, 931–940. (doi:10.1111/j.1523-1739.2009.01176.x)

15. Szabo ND, Colla SR, Wagner DL, Gall LF, Kerr JT. 2012 Do pathogen spillover, pesticide use, or habitat loss explain recent North American bumblebee declines? *Conserv. Lett.* **5**, 232–239. (doi:10.1111/j.1755-263X.2012.00234.x)

16. Cordes N, Huang WF, Strange JP, Cameron SA, Griswold TL, Lozier JD, Solter LF. 2012 Interspecific geographic distribution and variation of the

pathogens *Nosema bombi* and *Crithidia* species in United States bumble bee populations. *J. Invertebr. Pathol.* **109**, 209–216. (doi:10.1016/j.jip.2011.11.005)

17. Cameron SA, Lim HC, Lozier JD, Duennes MA, Thorp R. 2016 Test of the invasive pathogen hypothesis of bumble bee decline in North America. *Proc. Natl Acad. Sci. USA* **113**, 4386–4391. (doi:10.1073/pnas.1525266113)

18. Jongman R, Ter Braak C, Van Tongeren O. 1995 *Data analysis in community and landscape ecology*, p. 292. Cambridge, UK: Cambridge University Press.

19. Tibshirani R. 1996 Regression shrinkage and selection via the LASSO. *J. R. Stat. Soc. Ser. B* **58**, 267–288.

20. Meinshausen N, Bühlmann P. 2010 Stability selection. *J. R. Stat. Soc. B* **72**, 417–473. (doi:10.1111/j.1467-9868.2010.00740.x)

21. Hofner B, Boccuto L, Göker M. 2015 Controlling false discoveries in high-dimensional situations: boosting with stability selection. *BMC Bioinformatics* **16**, 144. (doi:10.1186/s12859-015-0575-3)

22. Haury AC, Mordelet F, Vera-Licona P, Vert JP. 2012 TIGRESS: trustful inference of gene regulation using stability selection. *BMC Syst. Biol.* **6**, 145. (doi:10.1186/1752-0509-6-145)

23. Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM. 2012 Wisdom of crowds for robust gene network inference. *Nat. Methods* **9**, 796–804. (doi:10.1038/nmeth.2016)

24. Hothorn T, Müller J, Schröder B, Kneib T, Brandl R. 2011 Decomposing environmental, spatial, and spatiotemporal components of species distributions. *Ecol. Monogr.* **81**, 329–347. (doi:10.1890/10-0602.1)

25. Gerber BD, Kendall WL, Hooten MB, Dubovsky JA, Drewien RC. 2015 Optimal population prediction

of sandhill crane recruitment based on climate-mediated habitat limitations. *J. Anim. Ecol.* **84**, 1299–1310. (doi:10.1111/1365-2656.12370)

26. Fry J, Xian G, Jin S, Dewitz J, Homer C, Yang L, Barnes C, Herold N, Wickham J. 2011 Completion of the 2006 national land cover database for the conterminous United States. *Photogr. Eng. Remote Sens.* **77**, 858–864.

27. McGarigal K, Cushman SA, Neel MC, Ene E. 2002 FRAGSTATS: spatial pattern analysis program for categorical maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. See http://www.umass.edu/landeco/research/fragstats/fragstats.html.

28. Neel MC, McGarigal K, Cushman SA. 2004 Behavior of class-level landscape metrics across gradients of class aggregation and area. *Landscape Ecol.* **19**, 435–455. (doi:10.1023/B:LAND.0000030521.19856.cb)

29. Baker N, Stone W. 2015 Estimated annual agricultural pesticide use for counties of the conterminous United States, 2008–12: U.S. Geological Survey Data Series 907, 9 p. See https://dx.doi.org/10.3133/ds3907.

30. R Core Team. 2015 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. See http://www.r-project.org/.

31. Tibshirani RJ, Tibshirani R. 2009 A bias correction for the minimum error rate in cross-validation. *NASA* **2**, 822–829. (doi:10.1214/08-AOAS224)

32. Hofner B. 2017 Stability selection with error control. See https://github.com/hofnerb/stabs.

33. Boria RA, Olson LE, Goodman SM, Anderson RP. 2014 Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecol. Modell.* **275**, 73–77. (doi:10.1016/j.ecolmodel.2013.12.012)

34. Radosavljevic A, Anderson RP. 2014 Making better Maxent models of species distributions: complexity, overfitting and evaluation. *J. Biogeogr.* **41**, 629–643. (doi:10.1111/jbi.12227)

35. Bernauer MO, Gaines-Day RH, Steffan AS. 2015 Colonies of bumble bees (*Bombus impatiens*) produce fewer workers, less bee biomass, and have smaller mother queens following fungicide

exposure. *Insects* **6**, 478–488. (doi:10.3390/insects6020478)

36. Pettis J, vanEngelsdorp D, Johnson J, Dively G. 2012 Pesticide exposure in honey bees results in increased levels of the gut pathogen *Nosema*. *Naturwissenschaften* **99**, 153–158. (doi:10.1007/s00114-011-0881-1)

37. Pettis JS, Lichtenberg EM, Andree M, Stitzinger J, Rose R, vanEngelsdorp D. 2013 Crop pollination exposes honey bees to pesticides which alters their susceptibility to the gut pathogen *Nosema ceranae*. *PLoS ONE* **8**, e70182. (doi:10.1371/journal.pone.0070182)

38. Huang W-F, Solter LF, Yau PM, Imai BS. 2013 *Nosema ceranae* escapes fumagillin control in honey bees. *PLoS Pathog.* **9**, e1003185. (doi:10.1371/journal.ppat.1003185)

39. Mao WF, Schuler MA, Berenbaum MR. 2017 Disruption of quercetin metabolism by fungicide affects energy production in honey bees (*Apis mellifera*). *Proc. Natl Acad. Sci. USA* **114**, 2538–2543. (doi:10.1073/pnas.1614864114)

40. Rundlof M *et al.* 2015 Seed coating with a neonicotinoid insecticide negatively affects wild bees. *Nature* **521**, 77–80. (doi:10.1038/nature14420)

41. Woodcock BA *et al.* 2017 Country-specific effects of neonicotinoid pesticides on honey bees and wild bees. *Science* **356**, 1393–1395. (doi:10.1126/science.aaa1190)

42. Botías C, David A, Hill EM, Goulson D. 2017 Quantifying exposure of wild bumblebees to mixtures of agrochemicals in agricultural and urban landscapes. *Environ. Pollut.* **222**, 73–82. (doi:10.1016/j.envpol.2017.01.001)

43. Mullin CA, Frazier M, Frazier JL, Ashcraft S, Simonds R, vanEngelsdorp D, Pettis JS. 2010 High levels of miticides and agrochemicals in North American apiaries: implications for honey bee health. *PLoS ONE* **5**, e9754. (doi:10.1371/journal.pone.0009754)

44. Sanchez-Bayo F, Goka K. 2014 Pesticide residues and bees: a risk assessment. *PLoS ONE* **9**, e94482. (doi:10.1371/journal.pone.0094482)

45. McArt SH, Fersch AA, Milano NJ, Truitt LL, Boroczky K. 2017 High pesticide risk to honey bees despite

low focal crop pollen collection during pollination of a mass blooming crop. *Sci. Rep.* **7**, 46554. (doi:10.1038/srep46554)

46. Raymann K, Shaffer Z, Moran NA. 2017 Antibiotic exposure perturbs the gut microbiota and elevates mortality in honeybees. *PLoS Biol.* **15**, e2001861. (doi:10.1371/journal.pbio.2001861)

47. Tsvetkov N, Samson-Robert O, Sood K, Patel HS, Malena DA, Gajiwala PH, Maciukiewicz P, Fournier V, Zayed A. 2017 Chronic exposure to neonicotinoids reduces honey bee health near corn crops. *Science* **356**, 1395–1397. (doi:10.1126/science.aam7470)

48. Pilling ED, Jepson PC. 1993 Synergism between EBI fungicides and a pyrethroid insecticide in the honeybee (*Apis mellifera*). *Pesticide Science* **39**, 293–297. (doi:10.1002/ps.2780390407)

49. Goulson D, Whitehorn P, Fowley M. 2012 Influence of urbanisation on the prevalence of protozoan parasites of bumblebees. *Ecol. Entomol.* **37**, 83–89. (doi:10.1111/j.1365-2311.2011.01334.x)

50. Youngsteadt E, Appler RH, López-Uribe MM, Tarpy DR, Frank SD. 2015 Urbanization increases pathogen pressure on feral and managed honey bees. *PLoS ONE* **10**, e0142031. (doi:10.1371/journal.pone.0142031)

51. Goulson D, Hughes W, Derwent L, Stout J. 2002 Colony growth of the bumblebee, *Bombus terrestris*, in improved and conventional agricultural and suburban habitats. *Oecologia* **130**, 267–273. (doi:10.1007/s004420100803)

52. Alaux C, Ducloz F, Crauser D, Le Conte Y. 2010 Diet effects on honeybee immunocompetence. *Biol. Lett.* **6**, 562–565. (doi:10.1098/rsbl.2009.0986)

53. Kerr JT *et al.* 2015 Climate change impacts on bumblebees converge across continents. *Science* **349**, 177–180. (doi:10.1126/science.aaa7031)

54. Arbetman MP, Gleiser G, Morales CL, Williams P, Aizen MA. 2017 Global decline of bumblebees is phylogenetically structured and inversely related to species range size and pathogen incidence. *Proc. R. Soc. B* **284**, 20170204. (doi:10.1098/rspb.2017.0204)

55. Burnham KP, Anderson DR. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edn. New York, NY: Springer.