**Title:** Recent outbreaks of a disease vector detected through analysis of museum specimens

Regarding what journal to submit to, I really don’t have any idea what’s appropriate in this field, so I agree entirely with whatever you decide. (although I will say, it certainly does sound appealing to have something appear in the Proceedings of the Royal Society!)

Another comment that I’ll put here, regarding Appexdix B, the R code. Unless it’s totally against the journal guidelines (or your preferences), I would recommend changing the font and line spacing for the R code in Appexdix B. It would look better using a smaller fixed-width font. I made that change in the modified version that I’m also sending – although without using Word’s track changes, so it doesn’t show up as a change, it just looks different. Take a look and see what you think. At least in statistics, that’s quite common when including computer code.

Another comment about Appendix B (the R code). The NIMBLE library has to be loaded **before** calls like nimbleFunction() and registerDistributions(). So a call to library(nimble) or require(nimble) must be practically the first thing.

**Authors:** Adam R. Zeilinger1,2, Giovanni Rapacciuolo, Daniel Turek, Pete Oboyski, Rodrigo P. P. Almeida, George Roderick

**Affiliations:**

1 Berkeley Initiative for Global Change Biology, University of California Berkeley

2 Department of Environmental Science, Policy, and Management, University of California Berkeley

Daniel’s affiliation:

Department of Mathematics and Statistics, Williams College

**Abstract**

The emergence rate of new plant diseases has been increasing due to a combination of factors including novel introductions, climate change, and changes in vector populations. New diseases pose risks to agricultural sustainability through reduced productivity and increased pesticide use. Since the mid-1990s, potato growers in western United States, Mexico, and Central America have experienced severe yield loss from the novel Zebra Chip disease. In response, growers have greatly increased insecticide use to suppress populations of the insect vector, the potato psyllid or *Bactericera cockerelli*. Despite the sever nature of Zebra Chip outbreaks, the cause(s) of emergence remain unknown. We used a large data set of museum specimen occurrence data and hierarchical Bayesian models to analyze changes in occupancy of *B. cockerelli* across California over the last century, and relate any trends to variation in climate. We found strong evidence that *B. cockerelli* occupancy has increased over the last century. However, these changes appear to be unrelated to climate change. Nonetheless, our analysis provides the first quantitative support for the hypothesis that changes in *B. cockerelli* populations are responsible for Zebra Chip disease emergence. Beyond Zebra Chip disease, our analysis provides a macro-ecological approach to comparative risk assessments of pest and disease vector outbreaks under a changing climate.

**Keywords:** NIMBLE, occupancy model, *Candidatus* Liberibacter solanacearum, list length analysis, etc….

I certainly would like to see “NIMBLE” as a keyword here, although by no means must it be the first keyword listed.

**Introduction**

The emergence of novel infectious diseases of humans, animals, and plants has increased in recent decades (Anderson et al. 2004, Jones et al. 2008). For plant diseases, the majority of recent outbreaks have occurred due to introduction of pathogens into new areas, climate change, or agricultural change (Anderson et al. 2004). For vector-borne plant pathogens, these same processes could also cause disease outbreaks indirectly through changes in vector populations (Anderson et al. 2004, Fereres 2015). Understanding the causes of disease outbreaks will facilitate management of current epidemics and risk assessments of future ones (Chakraborty 2013).

Beginning in the mid-1990s, potato growers in Western US, Mexico, and Central America have experienced severe yield losses from the emergence of Zebra Chip disease, caused by infections of the bacterial pathogen *Candidatus* Liberibacter solanacearum (Lso; = *C.* L. psyllaurous) (Munyaneza et al. 2007, Brown et al. 2010). While Lso causes diseases in other solanaceous agricultural crops—namely tomato and peppers—Zebra Chip disease is the most damaging, causing large losses in potato production in western United States, Mexico, Central America, and New Zealand (Secor and Rivera-Varas 2004, Butler and Trumble 2012, Munyaneza 2012, Horton et al. 2015). Lso is transmitted among plant hosts by the potato psyllid, *Bactericera cockerelli* (Sulc) (Hemiptera: Triozidae) (Buchman et al. 2011, Sengoda et al. 2014). While *B. cockerelli* is likely native to the western United States and Mexico, Zebra Chip disease was first described in 1994 and Lso was first described in 2008 (Munyaneza et al. 2007, Hansen et al. 2008). Since emergence, disease management has primarily involved vector suppression through insecticide use—primarily spirotetramat and imidacloprid (Butler and Trumble 2012, Guenthner et al. 2012).

Despite increasing harms to agricultural production and the environment from ongoing outbreaks of Lso and increasing insecticide applications, the causes of Lso emergence remain largely unknown. Most work has focused on possible changes in the vector population to explain outbreaks. Two specific hypotheses have emerged. First, work by Horton et al. (2015) and Liu and Trumble (2007) have suggested that the emergence of new *B. cockerelli* genotypes are responsible for outbreaks in the northwestern United States (Oregon, Washington, Idaho) and in California, respectively. However, this hypothesis fails to explain the parallel outbreaks of disease in the southwestern and Midwestern states, particularly Texas, Mexico, and Central America, where the “invasive” Northwest and California genotypes are rare or absent.

Second, greater overwintering survival of *B. cockerelli* populations in northern areas of its distribution could be causing Zebra Chip outbreaks. Historically, the phenology and range of *B. cockerelli* was thought to be largely confined by its relatively narrow thermal tolerance and availability of host plants. Based on previous observations, Pletsch (1947) and Wallis (1955) hypothesized that populations were limited in northern latitudes by either cold winter temperatures, a lack of host plants during winter, or both; these regions would be subsequently colonized by populations migrating from the south in springtime. There was also some evidence that southern populations were extirpated by high summer temperatures (Romney 1939). In 2011, populations of *B. cockerelli* were found overwintering in Oregon, Washington, and Idaho (Murphy et al. 2013, Horton et al. 2015), well beyond the species’ (hypothesized) historic northern limits (Wallis 1955), although a lack of historical monitoring leave it unclear whether this is a new development.

Greater overwintering survival of *B. cockerelli* could be caused by two factors: recent invasion of a non-native host plant, *Solanum dulcamara* (Solanaceae) (Horton et al. 2015), or warmer winter temperatures from anthropogenic climate change. Regardless of the specific cause, greater overwintering survival of northern populations would likely cause much greater densities in potato fields earlier in potato plant development, when plants are more susceptible to Lso infection (Gao et al. 2009). Lso infection prevalence in *B. cockerelli* populations appears to be very low, between 0 and 8% across the Midwest (Goolsby et al. 2012). At such low infectivity rates, very large vector populations may be required to cause a disease epidemic (Jeger et al. 2004).

We tested the hypothesis that *B. cockerelli* populations have increased in California over the last century using occupancy modeling of museum specimen occurrence data. Interest in historical ecological data—such as that from natural history collections—has increased in recent years due to increasing concern over the ecological impacts of anthropogenic climate change (Tingley and Beissinger 2009, Pyke and Ehrlich 2010). At the same time, much historical ecological data were opportunistically collected and are rife with biases, making rigorous ecological inference difficult (Newbold 2010). Hierarchical statistical models and Bayesian analytical methods have facilitated the analysis of such opportunistically collected data by explicitly modeling the observation process separate from the ecological process and thereby correcting for known biases in the data (Isaac et al. 2014, Isaac and Pocock 2015).

Here we show that the analysis of museum specimen occurrence data using Bayesian hierarchical models can provide insights into the causes of recent outbreaks of an agricultural pests. Specifically, we provide robust evidence that *B. cockerelli* populations have increased in California over the last century. At the same time, we find no evidence that these increases are related to anthropogenic climate change.

**Materials and Methods**

*Compiling species occurrence records*

We first collected, digitized, and georeferenced all pinned museum specimens of *B. cockerelli* from six major California museums: Bohart Museum of Entomology (University of California Davis), California Academy of Science, California State Collection of Arthropods (CA Dept. of Agriculture), Essig Museum of Entomology (UC Berkeley), Los Angeles County Museum of Natural History, and UC Riverside Entomology Research Museum. For each collection, we identified *B. cockerelli* specimens that had not yet been determined to the species level, housed in the “Undetermined Psyllidae” or Undetermined Triozidae” sections of the collections. We identified undetermined specimens using morphological description for *Paratrioza (= Bactericera) cockerelli* in Tuthill (1945). We digitized all identified *B. cockerelli* specimens, and georeferenced them using the point-radius method of Wieczorek et al. (2004) and the GEOLocate web application (http://www.museum.tulane.edu/geolocate/default.html). We also included slide-mounted nymph specimens from Bohart Museum and records of ethanol-preserved specimens from Essig Museum. All georeferenced specimens were uploaded to the Essig Museum online database (<https://essigdb.berkeley.edu/>).

Our data set of *B. cockerelli* museum specimens represented presence-only data. To generate non-detection (i.e., absence) data, we used the List Length Analysis approach of Phillips et al. (2009). We first built lists of related species collected across the same time and area of our *B. cockerelli* records. According to List Length Analysis, the length of these species lists is then assumed to relate to collecting effort (Roberts et al. 2007), which in turn is related to detection probability—longer lists without *B. cockerelli* are more indicative of a true absence than shorter lists. To generate the lists, we compiled historical occurrence records for all species in the Order Hemiptera, excluding families that were predominantly aquatic or carnivorous. The excluded families were: Anthocoridae, Belostomatidae, Cimicidae, Corixidae, Gelastorocidae, Geocoridae, Gerridae, Hebridae, Hydrometridae, Leptopodidae, Macroveliidae, Mesoveliidae, Nabidae, Naucoridae, Nepidae, Notonectidae, Ochteridae, Phymatidae, Pieidae, Polyctenidae, Reduviidae, Saldidae, and Veliidae. The remaining families of Hemiptera are predominately composed of terrestrial herbivorous species, are likely to exhibit roughly similar life histories and phenologies as *B. cockerelli,* and are likely to be collected by entomologists in similar manners as well, which is a central assumption of List Length Analysis (Szabo et al. 2010).

We compiled digitized and georeferenced California records of our selected Hemipteran families from three sources: the Global Biodiversity Information Facility (GBIF; which includes all records from the Essig Museum online database), the American Museum of Natural History (AMNH), and the Plant Pest Diagnostic Center’s collection database (PPDC; California Department of Agriculture). In the PPDC database, we included only records that were opportunistically collected, excluding those that reported trap monitoring for specific pest species. The data from the PPDC were biased toward agricultural areas; however, this balanced biases against agricultural areas in the GBIF and AMNH data sets. Occurrence records were included in our data set if they were collected within California between 1900 and 2015, included the full species name, and were georeferenced.

To generate species lists, we divided California into 15 km x 15 km grid cells and combined all species records collected within each cell, month, and year. We included only lists that (this sounds informal, maybe saying “had a minimum of three species” ?), following methods of van Strien et al. (2013), and lists in which at least one collector had also collected a *B. cockerelli* specimen in our data set. By filtering lists to a set of common collectors, we reduced the influence of collectors focused on only a single family or other taxon—a common concern in the analysis of opportunistically collected data (Isaac and Pocock 2015).

*Climate data*

Our aim was to model the temporal trends in *B. cockerelli* occupancy and to test if these trends could be explained by local-scale climate change. As such, we compiled estimates of historical climate from the California Basin Characterization Model (CA-BCM), which provides estimates of a range of temperature- and precipitation-derived climatic indices at 270 m x 270 m grid resolution for much of California from 1895 to 2014 (Flint et al. 2013). For each species in a list, we extracted estimates of the actual evapotranspiration (AET) in the month of collection, the water-year annual minimum temperature (Tmin), and the water-year annual maximum temperature (Tmax). For annual minimum and maximum temperatures, we extracted the minimum temperature occurring in the previous winter (December – February) and previous summer (June – September), respectively. Because our species lists were compiled at larger spatial cell sizes than the BCM cells, we extracted the climate values for each specimen within a list and then averaged these estimates to obtain a list-level estimate of climate. The standard deviations of these averages across lists were relatively small for each of the three climate variables (AET mean SD = 2.12; Tmin mean SD = 0.17; Tmax mean SD= 0.26), indicating that the climatic variation within our larger sites—from which our lists were constructed—was relatively small.

AET is an estimate of the water available for plant growth; it is a function of precipitation and temperature, and is more biologically relevant than raw precipitation estimates (Stephenson 1998, Rapacciuolo et al. 2014). Additionally, historical descriptions of the distribution of *B. cockerelli* indicate that the species primarily occurs in arid and semi-arid climates of the southwestern and Rocky Mountain regions of the US (Wallis 1955), which would be captured by estimates of AET. We also considered incorporating estimates of climate water deficit (CWD); however, AET and CWD were highly negatively correlated within our data set (Fig. A1, Appendix A).

We included annual Tmin because the prevailing hypothesis explaining the emergence of Zebra Chip Disease is that *B. cockerelli* populations at the northern limits of the species range have increased due to greater overwintering survival (Wallis 1955, Horton et al. 2015). If true, then *B. cockerelli* occupancy should be positively associated with annual Tmin. Finally, we included annual Tmax because of historical descriptions of *B. cockerelli* phenology suggesting that southern populations are extirpated due to high summer temperatures (Wallis 1955).

*Statistical analyses*

We modeled *B. cockerelli* occupancy using two model structures: a binomial generalized linear mixed-effects model (GLMM) and a hierarchical occupancy model as described in Royle and Kéry (2007). In both models, we included the same covariates: list length, year collected, AET, Tmin, T­max, and a quadratic form for month collected. We also included the interaction between the list length and year covariates to correct for changes in the relationship between list length and occupancy over time, e.g., due to changing collecting and preserving technologies and collector expertise. List length was natural log transformed (Szabo et al. 2010) and all continuous covariates were standardized around the mean and divided by the standard deviation. Spatial cell was included as a random effect in both models.

We used a binomial GLMM model based on previous simulations from Isaac et al. (2014), who showed that such models—when including list length as a linear covariate and site random effects—were robust against many prevalent forms of bias in opportunistic species occurrence data sets. Such models had high Type I error rates only when detection of simulated focal species increased independent of occupancy, and when non-focal species declined over time. Our inclusion of (again, unless this notation is standard in your field, I would instead say “the interaction between list length and year”) corrected for changes in detectability, and given our broad set of background taxa—all herbivorous Hemipteran families—we doubt that all non-focal species experienced declines over the time period of our analysis.

However, our data set was zero-inflated, with 36 detections vs. 864 non-detections (see Results). To model the observation process and account for zero-inflation, we also fit the data using a hierarchical occupancy model—also known as a binomial-binomial mixture model—following the framework of Kéry and Schaub (2012), except that the data did not include repeated visits to a site. Rather, we modeled detection probability as a function of list length and (same comment as before for the following: ) year x list length interaction; occupancy probability was modeled with all other covariates including the site random effects. Because the detection submodel and occupancy submodel had distinct sets of covariates, the latent states were identifiable (Sólymos et al. 2012). According to the simulations of Isaac et al. (2014), occupancy models with list length and site random effects can be overly conservative. Thus we include here the results of both the GLMM and occupancy models.

Both models were fit using the Markov chain Monte Carlo (MCMC) engine provided with the NIMBLE package (NIMBLE Development Team 2015, de Valpine et al. 2016) for R 3.3.0 (R Core Team 2016). MCMC speed and convergence were expedited making use of the flexibility of NIMBLE algorithms. For the occupancy model, latent states representing true (unknown) site occupancy were analytically removed from the model formulation using a custom-specified distribution, as described in Turek et al. (2016). Block sampling (e.g., Roberts and Sahu 1997) was used to jointly sample the coefficients of each linear predictor term in both models, since these will generally exhibit strong posterior correlation. For the occupancy model, the linear predictors for occupancy sub-model and detection sub-model were assigned separate block samplers.In addition, the standard deviation of site random effects was sampled on a logarithmic scale using the generalized Gibbs sampling framework described in Liu and Sabatti (2000). For both models, we used uninformative priors and three MCMC chains each with 150,000 iterations and a burn-in period of 50,000. Convergence was verified by calculating Gelman-Rubin diagnostic (add a reference for the Gelman-Rubin diagnostic?) and effective sample size. R code for fitting models is available in Appendix B; additional code for compiling and filtering data and analyzing model output can be found at <https://github.com/arzeilinger/potato_psyllid_distribution_modeling>.

**Results**

Our final data set of specimen records included 87,035 records of 2,840 species from 73 families; it also included 613 records of *B. cockerelli*. Specimens of all selected Hemiptera were collected throughout California, with the greatest number collected in the 1930s, 1960s, and 1970s (Fig. 1). Specimens of *B. cockerelli* were concentrated in southern California and arid interior regions of state; the majority of specimens were collected from 1950 to 1970 (Fig. 1). From this specimen data set, our resulting data set of species lists included 900 lists, which ranged in length from 3 to 40 species, and with 36 lists containing *B. cockerelli* (Fig. 2).

Convergence of the MCMC chains was successful for both the GLMM and occupancy models. Estimates of the Gelman-Rubin diagnostic for all parameters were less than 1.1 except for the variance of the site random effects in the GLMM (Appendix C). Likewise, effective sample sizes of all parameters were greater than 1,000 except for the site random effect mean and variance of the GLMM (Appendix C).

For the GLMM model, the coefficient estimate for the list length main effect was positive and was marginally significant from 0—indicating that the probability of occupancy, *Poccupancy*, of *B. cockerelli* increased with increasing list length (Table 1, Fig. 3). The coefficient estimate for the year main effect was significantly different from 0 and positively related to *Poccupancy*. For the climate variables, Tmin was significantly different from 0 and negative, whereas AET and Tmax were not significant. Looking at the distribution of *Poccupancy* across years and months, increases in *B. cockerelli* occupancy occurred in the late autumn/early winter in the 1950s and in the late winter/early spring beginning in the 1980s (Fig. 3).

For the occupancy model, the magnitude of the estimated coefficients were larger, but the 95% credible intervals were also larger (Table 1). List length was marginally significant and positively related to *Poccupancy* (Table 1, Fig. 4). Similar to the GLMM model, the year main effect was significant and positive. Unlike the GLMM model results, none of the climate variables were significant. Like the GLMM, the occupancy model predicts an increase in *B. cockerelli* *Poccupancy* in the late autumn/early winter in the 1950s. But the model also predicts increases in *B. cockerelli* throughout all seasons beginning in the 1990s, with particularly high occupancy rates in summers (Fig. 4).

Focusing on the occupancy model results, we generated maps of the estimated *Poccupancy* over three different time periods: 1920-1945, 1950-1975, and 1990-2015 (Fig. 5). Occupancy probability overall increased in southern California in more recent years, with greater variation in occupancy in early time periods. Unfortunately, the paucity of contemporary data from central and northern California precludes any robust predictions about spatial patterns in *B. cockerelli* outbreaks.

**Discussion**

Since emergence in 1994, Zebra Chip disease has devastated potato production in epidemic areas of the United States, Mexico, Central America, and New Zealand (Secor and Rivera-Varas 2004, Rosson et al. 2006, Horton et al. 2015). Whereas Texas potato growers often used no insecticides prior to Zebra Chip emergence, growers currently average 7.7 applications annually (Guenthner et al. 2012). Yet the causes of these outbreaks remain unknown. Using museum specimen occurrence data and Bayesian hierarchical models, we showed that occupancy probability of the Zebra Chip disease vector, *B. cockerelli*, has increased in California over the last century. These results provide the first quantitative support to the dominant hypothesis on the cause of Zebra Chip disease emergence—that *B. cockerelli* populations have increased in density, expanded in range, or both, in epidemic areas.

Our analysis produced strong evidence for increasing occupancy probability over the last century—both the GLMM and occupancy models estimated the year main effect as significant and positive. However, sparse data weaken our inference. Despite our initial large data set (~87,000 Hemipteran records and ~600 *B. cockerelli* records), our final data set consisted of 958 total species lists and only 36 lists containing *B. cockerelli*. The sharp reduction in data when constructing the lists was due to 1) a large number of single species lists, i.e., a single record without any other species collected in the same area at the same time, and 2) a large number of duplicate records; if multiple *B. cockerelli* were collected in the same place at the same time, this reduces to a single detection event. Furthermore, the lack of data in northern California in later decades prevent us from making robust inferences on possible range expansion of *B. cockerelli* (Fig. 5).

Our analysis focuses on occupancy probability of *B. cockerelli*. Patterns of vector occupancy are less relevant to Lso epidemiology than estimating patterns of vector population dynamics directly. Nonetheless, occupancy probability over large spatial scales is positively related to abundance, albeit in complex ways (Freckleton et al. 2005, McCarthy et al. 2013). In other words, increases in occupancy probability that we detected should be associated with increases in vector abundance and potentially enhanced Lso spread. In future analyses, it may be possible to treat replicated species records (i.e., multiple individuals of the same species collected in the same collecting event) as count data and directly model abundance within a list length analysis framework. However, theory relating species list length to count data has not been developed and would likely require stronger statistical assumptions.

While we found robust evidence for increased *B. cockerelli* populations, an equally important problem regards the causes of such population increases. We tested whether occupancy probability was related to climatic variation as measured in monthly actual evapotranspiration (AET), annual minimum temperature (Tmin), and annual maximum temperature (Tmax). The GLMM model indicated significant negative associations between occupancy and Tmin, which runs counter to the overwintering survival hypothesis. Meanwhile, the occupancy model showed no such significant effects. Our occupancy model adjusted for zero-inflation (i.e., inflated non-detections) in our *B. cockerelli* occurrence data, making it a more conservative model (Martin et al. 2005, Isaac et al. 2014). At the same time, the coarse spatial and temporal resolution of our analysis may have made associations between occupancy and climate variation difficult to detect.

In addition to climate change as a cause for *B. cockerelli* outbreaks, previous authors have also suggested that the outbreaks are caused by invasion of a non-native host plant, *Solanum dulcamara*, and accompanied adaptation by *B. cockerelli* populations (Horton et al. 2015). In future analyses, it may be possible to incorporate changes in host plant availability in our analysis, given sufficient historical records of host plant occurrence, particularly of *S. dulcamara*. More broadly, the potential role of biotic interactions in *B. cockerelli* outbreaks remains an important area for future research.

Beyond vector population change, other explanation of Zebra Chip disease emergence are more speculative. Agricultural change is a possible but unlikely explanation for Zebra Chip outbreaks. The outbreaks have occurred over multiple regions of the United States, Mexico, and Central America (Horton et al. 2015). Furthermore, all tested varieties of potato are highly susceptible making it unlikely that adoption of a new susceptible variety may have contributed to outbreaks (Munyaneza et al. 2011).

Likewise, pathogen introduction remains a possible but currently unlikely explanation. Lso has recently been detected in Europe and is associated with the carrot psyllid, *Trioza apicalis* (Munyaneza et al. 2010). Introduction from Europe is possible but Lso populations in Europe appear to be distinct from those in the Americas and appear to infect primarily non-solanaceous host plants, namely carrot and celery (Lin and Civerolo 2014), although the possibility of inter-continental pathogen introduction deserves more attention.

Another possibility is that Lso has co-existed with *B. cockerelli* for a long time and has recently evolved enhanced virulence. In the 1920s, a new disease of potatoes was described, called Psyllid Yellows, which was associated with feeding of *B. cockerelli* (Richards 1928). Through the 1930s and 1940s, there was significant debate whether Psyllid Yellows was caused by an infectious agent or a toxin in the saliva of the psyllid (Ever and Crawford 1933, Pletsch 1947, Carter 1950). While Lso-free *B. cockerelli* cause disease symptoms that are distinct from those of Zebra Chip disease (Sengoda et al. 2009), Psyllid Yellows and Zebra Chip disease symptoms in foliage and tubers are quite similar (Munyaneza et al. 2007, Sengoda et al. 2009). Given that Lso is unculturable (Lin and Gudmestad 2013) and the conflicting early evidence about the cause of Psyllid Yellows, early reports may have been documenting a combination of disease caused by psyllid salivary toxins and Lso infections. Regardless, this hypothesis could at most only partially explain the recent (re-)emergence of Lso. Importantly, none of these possible causes are mutually exclusive to the leading hypothesis of *B. cockerelli* population change. Rather, multiple factors may be contributing to Zebra Chip emergence and may differ in relative importance in different epidemic areas.

While we found no conclusive evidence that climate change in California explains outbreaks of *B. cockerelli* and the emergence of Lso, our test did not falsify the hypothesis either. More work will be needed to explore the possible linkages more fully. More broadly, the potential that climate change could reduce agricultural sustainability through pests and disease outbreaks—and any resulting increases in pesticide use—has become an increasing concern for growers, extension agents, scholars, and other stakeholders. Anthropogenic climate change could induce pest or disease outbreaks through a multitude of pathways (Garrett et al. 2006, Chakraborty 2013). As with ecological risk assessment of other stressors—such as those from pollutants, invasive species, or genetically modified organisms—a key challenge in the analysis of outbreak risks from climate change remains to identify which pest or pathogen species are most likely to become more problematic (USEPA 1998). As a first approximation, this will depend on the species’ life history, the local climatic changes experienced across the species’ range, and any ensuing changes in biotic interactions (Landsberg and Smith 1992). Most work to date has approached the problem from a mechanistic mode, focused on predicting outbreak risks by testing for specific changes in biotic interactions or responses from species with different life history traits under a subset of predicted climatic changes (Garrett et al. 2006).

Our analysis points to an additional, complementary approach. Our data set of species lists contain occurrence data of a wide range of agricultural pest and disease vectors across the Hemipteran families. We applied the same occupancy model to detection and non-detection data for two pest species that were well represented in the data set. *Lygus hesperus* (Miridae) is a common pest of cotton and strawberries in the San Joaquin Valley and Central Coast regions of California, respectively (Allen and Gaede 1963, Rosenheim et al. 2006). *Myzus persicae* (Aphididae) is a vector of potato leaf roll virus and pest of potatoes (Castle and Berger 1993). Our analysis indicates that *L. hesperus* occupancy probability has significantly decreased in California over the last century whereas *M. persicae* occupancy has shown no clear pattern of change (Appendix D).

Although our analysis is spatially and temporally coarse and ignores the roles of biotic interactions and agricultural change, comparisons among species—using a common data set and model—could facilitate more fine-grain investigations into the specific mechanisms that underlie large-scale differences among pest species and their relationships to climatic variation. Linking fine-scale studies of biotic interactions and life history traits to macro-ecological analyses will be necessary to develop rigorous assessment frameworks for the future risks of pest and pathogen outbreaks caused by climate change.

**Acknowledgements**

We thank those who assisted us in sorting through and borrowing specimens from entomological museums: Norm Penny at California Academy of Sciences, Doug Yanega at UC Riverside Entomology Research Museum, Lynn Kimsey at UC Davis Bohart Museum of Entomology, Weiping Xie at LA County Natural History Museum, and Jackie Kishmirian at California State Collection of Arthropods. Stephen Gaimari and Katja Seltmann helped with access to and filtering of the Plant Pest Diagnostic Center data, and Christine Johnson helped with data access from the American Museum of Natural History. Emily Yao, Arsene Gaude, K Kai Din, Gordon Nishida, and Joyce Gross assisted with digitizing and georeferencing of museum specimens. Members of the Almeida Lab and Evo Lab provided helpful comments on early versions of the analysis and paper. This work was supported by the Berkeley Initiative for Global Change Biology and the Gordon and Betty Moore Foundation.

**Literature Cited**

Allen, W. W., and S. E. Gaede. 1963. The Relationship of Lygus Bugs and Thrips to Fruit Deformity in Strawberries. Journal of Economic Entomology 56:823–825.

Anderson, P. K., A. A. Cunningham, N. G. Patel, F. J. Morales, P. R. Epstein, and P. Daszak. 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends in Ecology & Evolution 19:535–544.

Brown, J. K., M. Rehman, D. Rogan, R. R. Martin, and A. M. Idris. 2010. First Report of “Candidatus Liberibacter psyllaurous” (synonym “Ca. L. solanacearum”) Associated with “Tomato Vein-Greening” and “Tomato Psyllid Yellows” Diseases in Commercial Greenhouses in Arizona. Plant Disease 94:376–376.

Buchman, J. L., V. G. Sengoda, and J. E. Munyaneza. 2011. Vector Transmission Efficiency of Liberibacter by Bactericera cockerelli (Hemiptera: Triozidae) in Zebra Chip Potato Disease: Effects of Psyllid Life Stage and Inoculation Access Period. Journal of Economic Entomology 104:1486–1495.

Butler, C. D., and J. T. Trumble. 2012. Identification and Impact of natural enemies of Bactericera cockerelli (Hemiptera: Triozidae) in Southern California. Journal of economic entomology 105:1509–1519.

Carter, R. D. 1950. Toxicity of Paratrioza cockerelli (Sulc) to certain solanaceous plants. Ph.D Dissertation, University of California Berkeley, Berkeley, CA.

Castle, S. J., and P. H. Berger. 1993. Rates of growth and increase of *Myzus persicae* on virus-infected potatoes according to type of virus-vector relationship. Entomologia Experimentalis et Applicata 69:51–60.

Chakraborty, S. 2013. Migrate or evolve: options for plant pathogens under climate change. Global Change Biology 19:1985–2000.

Ever, J. R., and R. F. Crawford. 1933. Observations on the Feeding Habits of the Potato Psyllid ( Paratrioza cockerelli Sulc) and the Pathological History of the “ Psyllid Yellows ” which it produces. Journal of Economic Entomology 26.

Fereres, A. 2015. Insect vectors as drivers of plant virus emergence. Current Opinion in Virology 10:42–46.

Flint, L. E., A. L. Flint, J. H. Thorne, and R. Boynton. 2013. Fine-scale hydrologic modeling for regional landscape applications: the California Basin Characterization Model development and performance. Ecological Processes 2:25.

Freckleton, R. P., J. A. Gill, D. Noble, and A. R. Watkinson. 2005. Large-scale population dynamics, abundance–occupancy relationships and the scaling from local to regional population size. Journal of Animal Ecology 74:353–364.

Gao, F., J. Jifon, X. Yang, and T.-X. Liu. 2009. Zebra chip disease incidence on potato is influenced by timing of potato psyllid infestation, but not by the host plants on which they were reared. Insect Science 16:399–408.

Garrett, K. A., S. P. Dendy, E. E. Frank, M. N. Rouse, and S. E. Travers. 2006. Climate change effects on plant disease: genomes to ecosystems. Annu. Rev. Phytopathol. 44:489–509.

Goolsby, J. A., J. J. Adamczyk Jr, J. M. Crosslin, N. N. Troxclair, J. R. Anciso, G. G. Bester, J. D. Bradshaw, E. D. Bynum, L. A. Carpio, and D. C. Henne. 2012. Seasonal population dynamics of the potato psyllid (Hemiptera: Triozidae) and its associated pathogen “Candidatus Liberibacter solanacearum” in potatoes in the southern Great Plains of North America. Journal of Economic Entomology 105:1268–1276.

Guenthner, J., J. Goolsby, and G. Greenway. 2012. Use and Cost of Insecticides to Control Potato Psyllids and Zebra Chip on Potatoes. Southwestern Entomologist 37:263–270.

Hansen, A. K., J. T. Trumble, R. Stouthamer, and T. D. Paine. 2008. A New Huanglongbing Species, “Candidatus Liberibacter psyllaurous,” Found To Infect Tomato and Potato, Is Vectored by the Psyllid Bactericera cockerelli (Sulc). Applied and Environmental Microbiology 74:5862–5865.

Horton, D. R., W. R. Cooper, J. E. Munyaneza, K. D. Swisher, E. R. Echegaray, A. F. Murphy, S. I. Rondon, C. H. Wohleb, T. D. Waters, and A. S. Jensen. 2015. A New Problem and Old Questions: Potato Psyllid in the Pacific Northwest. American Entomologist 61:234–244.

Isaac, N. J. B., and M. J. O. Pocock. 2015. Bias and information in biological records. Biological Journal of the Linnean Society 115:522–531.

Isaac, N. J. B., A. J. van Strien, T. A. August, M. P. de Zeeuw, and D. B. Roy. 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. Methods in Ecology and Evolution 5:1052–1060.

Jeger, M. J., J. Holt, F. Van Den Bosch, and L. V. Madden. 2004. Epidemiology of insect-transmitted plant viruses: modelling disease dynamics and control interventions. Physiological Entomology 29:291–304.

Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. 2008. Global trends in emerging infectious diseases. Nature 451:990–993.

Kéry, M., and M. Schaub. 2012. Bayesian population analysis using WinBUGS a hierarchical perspective. 1st ed. Academic Press, Waltham, MA.

Landsberg, J., and M. Smith. 1992. A Functional Scheme for Predicting the Outbreak Potential of Herbivorous Insects Under Global Atmospheric Change. Australian Journal of Botany 40:565–577.

Lin, H., and E. L. Civerolo. 2014. Comparative Genomics of the Liberibacteral Plant Pathogens. Pages 203–233*in* D. C. Gross, A. Lichens-Park, and C. Kole, editors.Genomics of Plant-Associated Bacteria. Springer Berlin Heidelberg.

Lin, H., and N. C. Gudmestad. 2013. Aspects of Pathogen Genomics, Diversity, Epidemiology, Vector Dynamics, and Disease Management for a Newly Emerged Disease of Potato: Zebra Chip. Phytopathology 103:524–537.

Liu, D., and J. T. Trumble. 2007. Comparative fitness of invasive and native populations of the potato psyllid (Bactericera cockerelli). ENTOMOLOGIA EXPERIMENTALIS ET APPLICATA 123:35–42.

Liu, J., and C. Sabatti. 2000. Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation. Biometrika 87:353–369.

Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possingham. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecology Letters 8:1235–1246.

McCarthy, M. A., J. L. Moore, W. K. Morris, K. M. Parris, G. E. Garrard, P. A. Vesk, L. Rumpff, K. M. Giljohann, J. S. Camac, S. S. Bau, T. Friend, B. Harrison, and B. Yue. 2013. The influence of abundance on detectability. Oikos 122:717–726.

Munyaneza, J. E. 2012. Zebra chip disease of potato: biology, epidemiology, and management. American Journal of Potato Research:1–22.

Munyaneza, J. E., J. L. Buchman, V. G. Sengoda, T. W. Fisher, and C. C. Pearson. 2011. Susceptibility of selected potato varieties to zebra chip potato disease. American Journal of Potato Research:1–6.

Munyaneza, J. E., J. M. Crosslin, and J. E. Upton. 2007. Association of Bactericera cockerelli (Homoptera : Psyllidae) with “zebra chip,” a new potato disease in southwestern United States and Mexico. JOURNAL OF ECONOMIC ENTOMOLOGY 100:656–663.

Munyaneza, J. E., T. W. Fisher, V. G. Sengoda, S. F. Garczynski, A. Nissinen, and A. Lemmetty. 2010. Association of “Candidatus Liberibacter solanacearum” With the Psyllid, Trioza apicalis (Hemiptera: Triozidae) in Europe. Journal of Economic Entomology 103:1060–1070.

Murphy, A. F., S. I. Rondon, and A. S. Jensen. 2013. First Report of Potato Psyllids, Bactericera cockerelli, Overwintering in the Pacific Northwest. American Journal of Potato Research 90:294–296.

Newbold, T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Progress in Physical Geography 34:3–22.

NIMBLE Development Team. 2015. NIMBLE: An R Package for Programming with BUGS models, Version 0.4.

Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19:181–197.

Pletsch, D. J. 1947. The potato psyllid, Paratrioza cockerelli (Sulc), its biology and control. Bulletin of the Montana Agricultural Experiment Station 446:1–95.

Pyke, G. H., and P. R. Ehrlich. 2010. Biological collections and ecological/environmental research: a review, some observations and a look to the future. BIOLOGICAL REVIEWS 85:247–266.

R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rapacciuolo, G., S. P. Maher, A. C. Schneider, T. T. Hammond, M. D. Jabis, R. E. Walsh, K. J. Iknayan, G. K. Walden, M. F. Oldfather, D. D. Ackerly, and S. R. Beissinger. 2014. Beyond a warming fingerprint: individualistic biogeographic responses to heterogeneous climate change in California. Global Change Biology:n/a-n/a.

Richards, B. L. 1928. A new and destructive disease of the potato in Utah and its relation to the potato psylla. Phytopathology 18:140–141.

Roberts, G. O., and S. K. Sahu. 1997. Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 59:291–317.

Roberts, R. L., P. F. Donald, and R. E. Green. 2007. Using simple species lists to monitor trends in animal populations: new methods and a comparison with independent data. Animal Conservation 10:332–339.

Romney, V. E. 1939. Breeding Areas of the Tomato Psyllid, Paratrioza cockerelli (Šulc). Journal of Economic Entomology 32:150–151.

Rosenheim, J. A., K. Steinmann, G. A. Langellotto, and A. G. Zink. 2006. Estimating the Impact of Lygus hesperus on Cotton: The Insect, Plant, and Human Observer as Sources of Variability. Environmental Entomology 35:1141–1153.

Rosson, P., M. Niemeyer, M. Palma, and L. Ribera. 2006. Economic impacts of zebra chips on the Texas potato industry. Center for North American Studies, Department of Agricultural Economics, Texas A&M University, College Station, TX.

Royle, J. A., and M. Kéry. 2007. A Bayesian state-space formulation of dynamic occupancy models. Ecology 88:1813–1823.

Secor, G. A., and V. V. Rivera-Varas. 2004. Emerging diseases of cultivated potato and their impact on Latin America. Revista Latinoamericana de la Papa (Suplemento) 1:1–8.

Sengoda, V. G., W. R. Cooper, K. D. Swisher, D. C. Henne, and J. E. Munyaneza. 2014. Latent Period and Transmission of “Candidatus Liberibacter solanacearum” by the Potato Psyllid Bactericera cockerelli (Hemiptera: Triozidae). PLoS ONE 9:e93475.

Sengoda, V. G., J. E. Munyaneza, J. M. Crosslin, J. L. Buchman, and H. R. Pappu. 2009. Phenotypic and Etiological Differences Between Psyllid Yellows and Zebra Chip Diseases of Potato. American Journal of Potato Research 87:41–49.

Sólymos, P., S. Lele, and E. Bayne. 2012. Conditional likelihood approach for analyzing single visit abundance survey data in the presence of zero inflation and detection error. Environmetrics 23:197–205.

Stephenson, N. 1998. Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. Journal of Biogeography 25:855–870.

van Strien, A. J., C. A. Swaay, and T. Termaat. 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. Journal of Applied Ecology 50:1450–1458.

Szabo, J. K., P. A. Vesk, P. W. J. Baxter, and H. P. Possingham. 2010. Regional avian species declines estimated from volunteer-collected long-term data using List Length Analysis. Ecological Applications 20:2157–2169.

Tingley, M. W., and S. R. Beissinger. 2009. Detecting range shifts from historical species occurrences: new perspectives on old data. Trends in Ecology & Evolution 24:625–633.

Turek, D., P. de Valpine, and C. J. Paciorek. 2016. Efficient Markov Chain Monte Carlo Sampling for Hierarchical Hidden Markov Models. arXiv preprint arXiv:1601.02698.

Tuthill, L. D. 1945. Contributions to the Knowledge of the Psyllidae of Mexico (Continued). Journal of the Kansas Entomological Society 18:1–29.

[USEPA] U.S. Environmental Protection Agency. 1998. Guidelines for Ecological Risk Assessment. U.S. Environmental Protection Agency, Washington, D.C.

de Valpine, P., D. Turek, C. J. Paciorek, C. Anderson-Bergman, D. T. Lang, and R. Bodik. 2016. Programming with models: writing statistical algorithms for general model structures with NIMBLE. Journal of Computational and Graphical Statistics 0:1–28.

Wallis, R. L. 1955. Ecological studies on the potato psyllid as a pest of potatoes. US Dept. of Agriculture.

Wieczorek, J., Q. Guo, and R. Hijmans. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. International Journal of Geographical Information Science 18:745–767.

**Table 1.** Estimated mean and 95% credible intervals (in brackets) of linear coefficients for GLMM and occupancy model of *B. cockerelli* occurrence.

|  |  |  |
| --- | --- | --- |
| Model term | GLMM | Occupancy model |
| Year | 0.88 [0.41, 1.39]b | 49.56 [14.89, 92.2] |
| Month | -0.13 [-0.53, 0.31] | 4.96 [-35.44, 52.72] |
| Month2 | 0.39 [0.09, 0.69] | 11.07 [-32.1, 56.13] |
| AET | -0.57 [-1.25, 0.01] | -22.79 [-71.61, 29.92] |
| Tmin | -0.48 [-0.89, -0.09] | -15.55 [-66.49, 36.1] |
| Tmax | -0.33 [-0.84, 0.18] | -11.43 [-60.56, 41.12] |
| Detection sub-model intercepta | NA | -2.44 [-2.91, -1.98] |
| List length | 0.34 [-0.05, 0.73] | 0.3 [-0.03, 0.62] |
| List length \* year | 0.31 [-0.1, 0.76] | 0.33 [-0.03, 0.71] |
| αµ | -4.85 [-6.06, -3.93] | -14.06 [-68.65, 45.46] |
| ασ | 1.17 [0.64, 2.06] | 271.68 [49.19, 884.28] |

a Detection sub-model intercept only applicable for Occupancy model

b Mean [2.5% CI, 97.5% CI] estimated from posterior probabilities

αµ = mean of the site random effects; ασ = variance of the site random effects

**Figure Legends**

**Figure 1.** Spatial (A) and temporal (B) distribution of all museum records (grey points and bars) and of *B. cockerelli* museum records (black points and bars).

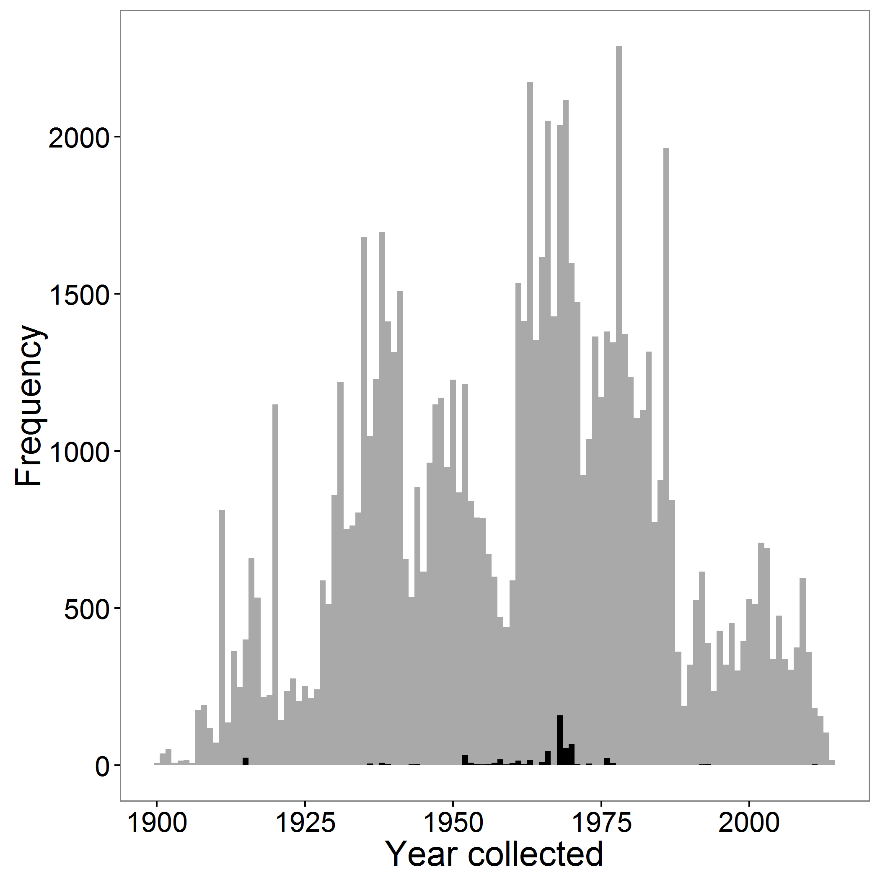
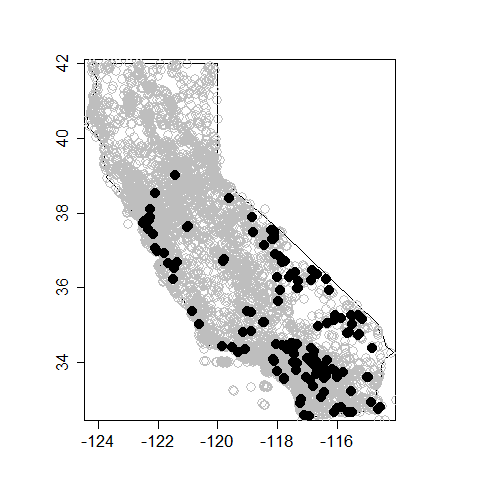
**Figure 2.** Histograms of species list length for all lists (grey bars) and lists that contain *B. cockerelli* (black bars).

**Figure 3.** GLMM model results: Relationships between GLMM-predicted probability of *B. cockerelli* occupancy and list length (A), year collected (B), Tmin (C), as well as the relationships of occupancy, year collected and month collected (D). For the contourplot (E), greyscale values indicate probability of occupancy. Fitted lines on scatterplots are smoothing splines.

**Figure 4.** Occupancy model results: Relationships between occupancy model-predicted probability of *B. cockerelli* occupancy and list length (A) and year collected (B), as well as the relationships of occupancy, year collected and month collected (C). For the contourplot (C), greyscale values indicate probability of occupancy. Fitted lines on scatterplots are smoothing splines.

**Figure 5.** Maps of estimated *B. cockerelli* occupancy across California for three selected time periods. Closed circles indicate species lists containing *B. cockerelli* (i.e., detection events); open circles indicate non-detection events. The size of the circle indicates the estimated probability of occupancy, with larger circles representing greater probabilities. Occupancy probabilities were averaged over years for each site.

*Figure 1. Spatial and temporal distribution of species records*



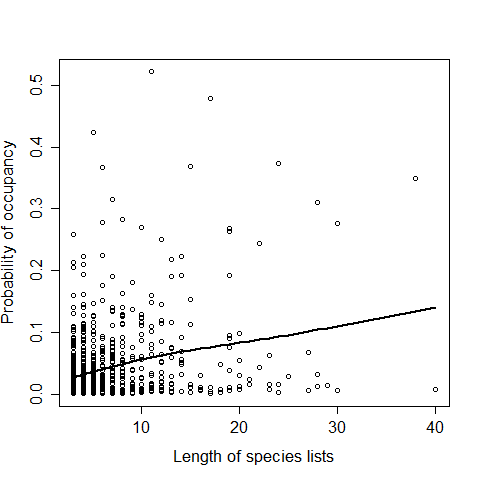
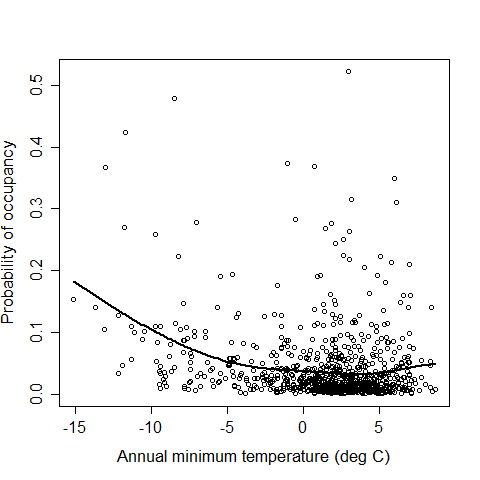
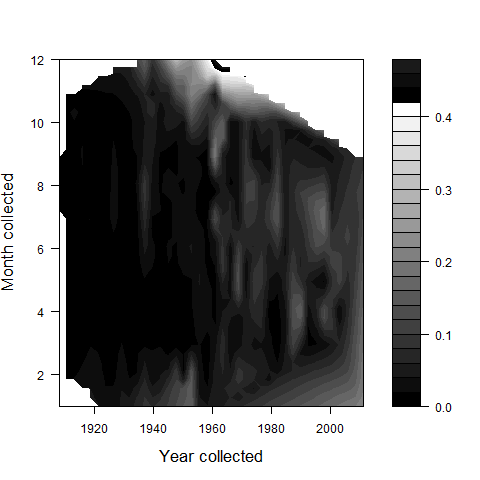
**A**

**B**

*Figure 2. Distribution of species list length*



*Figure 3. GLMM model results*



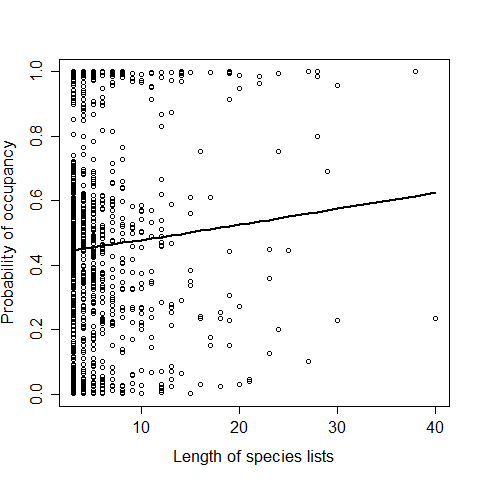
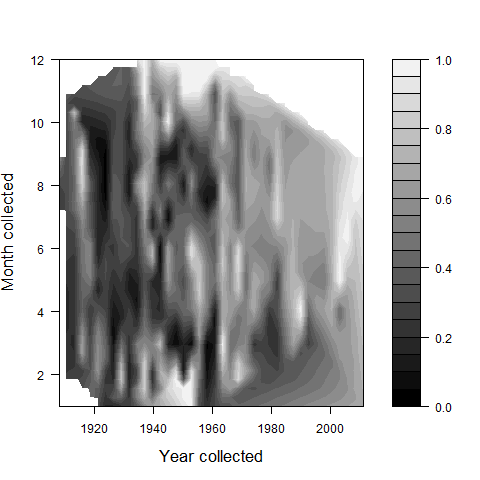
**A**

**B**

**C**

**D**

*Figure 4. Occupancy model results*

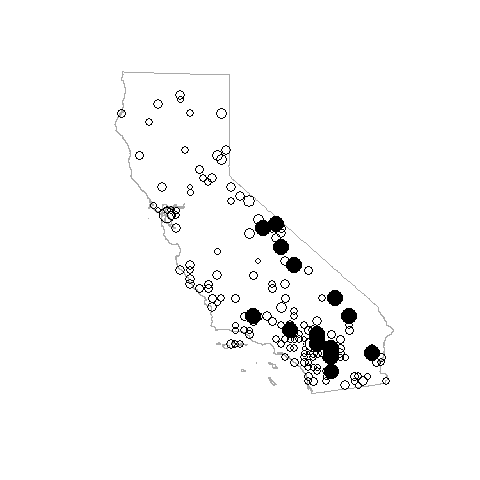


**A**

**B**

**C**

*Figure 5.* *Maps of occupancy model results*



**1920 - 1945**

**1950 - 1975**



**1990 - 2015**

**Appendix A.** Scatter plots of BCM climate data pairs and year collected. Y-axis and x-axis variables for each plot are given along the diagonal.



**Appendix C.** Estimate and 97.5% Credible Interval of the Gelman-Rubin diagnostic () and estimate of effective sample size (ESS) for linear coefficients for the GLMM and Occupancy models

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model term | GLMM | | | Occupancy model | | |
|  | (97.5% CI) | ESS | | (97.5% CI) | ESS | |
| Year | 1 (1.01) | | 5035.65 | 1 (1) | | 1925.35 |
| Month | 1 (1) | | 7803.63 | 1 (1) | | 2669.92 |
| Month2 | 1 (1) | | 3982.4 | 1 (1.01) | | 1747.28 |
| AET | 1 (1) | | 7130.32 | 1 (1) | | 2843.19 |
| Tmin | 1 (1) | | 5113.31 | 1 (1) | | 2136.98 |
| Tmax | 1 (1.01) | | 6539.5 | 1 (1) | | 2223.55 |
| Detection sub-model intercepta | NA | | NA | 1 (1) | | 6630.32 |
| List length | 1 (1) | | 7962.11 | 1 (1) | | 22700.12 |
| List length \* year | 1 (1) | | 8096.92 | 1 (1) | | 22913.79 |
| αµ | 1.03 (1.09) | | 822.26 | 1 (1.01) | | 3966.34 |
| ασ | 1.06 (1.2) | | 123.14 | 1 (1.02) | | 1103.12 |

a Detection sub-model intercept only applicable for Occupancy model

αµ = mean of the site random effects; ασ = variance of the site random effects

**Appendix D.** Results for occupancy of *Lygus hesperus* and *Myzus persicae* using the same species records data set and occupancy model

**Table D1.** Estimated posterior mean and 95% credible intervals (in brackets) of linear coefficients for occupancy model of *Lygus hesperus* occurrence.

|  |  |
| --- | --- |
| Model term | Mean [2.5% CI, 97.5% CI] |
| Year | -52.76 [-81.72, -23.12] |
| Month | 8.91 [-4.72, 29.52] |
| Month2 | -11.15 [-29.24, 8.53] |
| AET | 15.72 [-0.1, 36.78] |
| Tmin | -21.7 [-53.39, 1.39] |
| Tmax | 13.94 [-8.16, 33.92] |
| Detection sub-model intercept | -1.44 [-1.63, -1.24] |
| List length | 0.29 [0.12, 0.46] |
| List length \* year | -0.04 [-0.19, 0.1] |
| αµ | 65.2 [33.15, 99.46] |
| ασ | 63.51 [26.04, 121.38] |

αµ = mean of the site random effects; ασ = variance of the site random effects

**Table D2.** Estimated posterior mean and 95% credible intervals (in brackets) of linear coefficients for occupancy model of *Myzus persicae* occurrence.

|  |  |
| --- | --- |
| Model term | Mean [2.5% CI, 97.5% CI] |
| Year | -7.91 [-66.9, 54.3] |
| Month | -24.13 [-77.05, 32.13] |
| Month2 | -7.34 [-49.79, 38.81] |
| AET | 28.3 [-28.18, 81.32] |
| Tmin | 1.6 [-57.3, 58.53] |
| Tmax | -1.94 [-54.14, 49.21] |
| Detection sub-model intercept | 5.79 [-1.79, 22.18] |
| List length | 24.21 [0.61, 64.66] |
| List length \* year | -19.75 [-55.45, 0.75] |
| αµ | -16.43 [-76.12, 49.2] |
| ασ | 352.24 [79.68, 920.4] |

αµ = mean of the site random effects; ασ = variance of the site random effects