

1
2
3
4
5
6 Why are monarch butterflies declining in the West?

7 An assessment of what we can and cannot know from monitoring data

8
9 Elizabeth E. Crone¹, Emma M. Pelton², Leone M. Brown¹, Cameron C. Thomas³ & Cheryl B.
10 Schultz³
11

12 ¹Department of Biology, Tufts University, 200 College Ave, Medford MA 02155 USA

13 ²Xerces Society, 628 NE Broadway Ste 200, Portland OR 97232 USA

14 ³School of Biological Sciences, Washington State University, 14204 NE Salmon Creek Ave.
15 Vancouver, WA 98686 USA

16 *For submission to Ecological Applications*

17 DRAFT – Please do not cite or circulate without permission!
18
19
20

21 After acceptance for publication, data will be archived in the Dryad digital repository
22

23 **Running title:** Threats to monarchs in the West

Abstract

Understanding the factors associated with declines of at-risk species is an important first step in setting management and recovery targets. This step can be challenging when multiple aspects of climate and land use are changing simultaneously, and any or all could be contributing to population declines. We present analyses of population trends of monarch butterflies in western North America in relation to likely environmental drivers. Unlike the larger eastern monarch population, past analyses of western monarchs have emphasized the importance of climate factors as drivers of abundance. We used partial least squares regression (PLSR) to evaluate the potential importance of changes in land use and climate variables. Trends in western monarch abundance were more strongly associated with land use variables than climate variables. Conclusions about importance of climate and land use variables were robust to changes in PLSR model structure. However, individual variables were too collinear to unambiguously separate their effects. We compared these conclusions to the more widely used technique of multiple regression, followed by multi-model inference (MRMI). Naïve interpretation of MRMI results could be misleading, if collinearity were not taken into account. MRMI was also highly sensitive to variation in model construction. Our results suggest a two-pronged approach to monarch conservation, specifically, starting efforts now to restore habitat, while also using experiments to more clearly delineate separate effects of climate and land use factors. They also demonstrate the utility of PLSR, a technique that is growing in use but is still relatively under-appreciated in conservation biology.

Key words: endangered species, threats analysis, PLSR, multimodel inference, multiple regression, land use change, climate change, glyphosate, neonicotinoids, pesticide, habitat loss

Introduction

Conservation biologists often need to understand the factors causing declines of at-risk species, but cannot always isolate or experimentally manipulate hypothesized environmental drivers of population dynamics (Davidson et al. 2012; Earl et al. 2018). This issue has become particularly problematic in recent years because, in many parts of the world, both land use and climate are changing simultaneously in ways that may be unfavorable to many species, and it is difficult to disentangle their effects (Brook et al. 2008; Didham et al. 2005). Even in a carefully planned correlative study, many aspects of land use and climate are likely to be changing together in ways that may affect population viability of focal species (Selwood et al. 2015; Zeigler et al. 2017). An important challenge for conservation biologists is to use such environmental data to evaluate hypothesized drivers of population declines in ways that neither overstate nor understate confidence in the causal relationships.

Understanding contributions of different factors to population declines is essential for mitigating declines and enacting protections. In the United States, this information can help to determine whether a species should be protected under the Endangered Species Act (Smith et al. 2018). Listing in some cases is contingent on whether current practices are deemed sufficient to recover the species without government protection. For example, in 2010, the US Fish and Wildlife Service determined that the greater sage-grouse (*Centrocercus urophasianus*) warranted listing. The Service reversed this decision in 2015, determining that listing was precluded because primary threats had been ameliorated through conservation efforts by public and private landowners (Service 2015). In other words, it was deemed low priority for federal protection because private land management actions seemed adequate to make near-term extinction unlikely. Now, similar criteria may be applied to monarch butterflies (*Danaus plexippus*

plexippus), a formerly widespread species that has been under consideration for listing under the U.S. Endangered Species Act (Service 2018). Prior to the listing decision, public and private organizations have begun extensive programs to protect the breeding and migratory habitat (Pleasants 2017; Thogmartin et al. 2017a), some of which were initiated with the specific goal of pre-emptively preventing listing.

The monarch butterfly has two behaviorally distinct migratory populations which winter in reproductive diapause at overwintering sites, then spread northward over multiple generations for breeding. The western population primarily overwinters in coastal California, and breeds west of the Rocky Mountains. The eastern population overwinters in central Mexico and breeds east of the Rocky Mountains. During the past half-century, the eastern monarch population has declined by ~80% (Semmens et al. 2016), and the western population has declined by ~95% (Schultz et al. 2017). Several recent studies (Thogmartin et al. 2017b) have evaluated the potential contributions of different factors to population dynamics of the eastern monarch population. Here, we evaluate the importance of different aspects of climate and land use change for the western monarch population, with special emphasis on how we make inferences about the importance of these correlated variables, and what actions might mitigate these threats in the short and long term.

We assess land use and climate variables as potential drivers of western monarch abundance on overwintering grounds in California from 1982-2016. During this period, the population dropped from ~1-30 million butterflies in the 1980's to ~200-300 thousand butterflies in the 2000's, with substantial fluctuations in the 1990's (Schultz et al. 2017). We began with a set of seven variables that are the most widely cited threats to monarch populations (Stevens & Frey 2010; Thogmartin et al. 2017b; Zipkin et al. 2012): temperature and precipitation in the

overwintering and breeding grounds (4 variables), land conversion of overwintering sites, and use of two common pesticides: glyphosate, an herbicide that incidentally kills host and nectar plants, and neonicotinoids, a class of insecticides. Hereafter, we refer to this as the *a priori* predictor analysis. Considering a limited number of well-understood causal variables is widely encouraged in wildlife biology to prevent spurious correlations (e.g., Burnham & Anderson 2002). However, although analyses that consider only *a priori* hypotheses are good at determining relative support for different predictor variables, they are inherently subject to the existing biases of expert opinion. Therefore, we also considered a more extended set of 29 predictors, which includes a number of additional weather and pesticide-use variables. Hereafter, we refer to this as the “kitchen sink” analysis, with the recognition that evaluating 29 possible predictor variables for a 36-year time series is likely subject to spurious correlations, as well as new hypotheses.

For both analyses, we used partial least squares regression (PLSR) to investigate the influence of multiple correlated variables on monarch abundance (Carrascal et al. 2009; Thogmartin et al. 2017b). PLSR deals with collinearity by estimating synthetic components, i.e., linear combinations of the predictor variables (here, climate and land use variables) that best explain variation in the response variable (here, monarch butterfly abundance (Carrascal et al. 2009). PLSR is similar to principle components analysis (PCA), a technique that is currently more widely used in population and community ecology. However, PLSR is better suited to identifying environmental drivers because it finds the components that best explain the response variable, whereas PCA finds components that simply capture variation in the predictor variables. PLSR is also well-suited to identifying cases where there is ambiguity in the set of causal variables (see *Results*, below). Finally, for comparison, we include a multiple regression

analysis of the seven *a priori* variables, with multi-model inference across all possible models (hereafter, MRMI). Because MRMI is widely used but performs poorly in this situation, we include a brief exploration of what features of the data cause MRMI to fail in our situation.

Materials and Methods

Monarch butterfly abundance

Western monarchs breed throughout western North America, and primarily overwinter in forested groves in coastal California, though some migrate to Mexico (Jepsen & Black 2015; Morris et al. 2015, Fig 1). In early spring, wintering butterflies leave roosting grounds and breed in California and adjacent Nevada and Arizona (Dingle et al. 2005), Schultz et al. unpubl. data). The population expands to Oregon, Washington, Idaho and British Columbia (as well as other western states) over 2-3 successive generations, the last of which migrates back to overwintering sites in the fall. Because there are no systematic efforts to assess western monarch abundance in the breeding range, we use monarchs in the wintering grounds as our metric of relative abundance through time. Monarchs have been counted in coastal California overwintering groves intermittently from 1981-1996, and systematically from 1997-ongoing as part of the Xerces Society Western Monarch Thanksgiving Count (www.westernmonarchcount.org). We previously analyzed these data with state-space models to obtain an estimate of the total wintering population size in each year (Schultz et al. 2017; their Fig. 2a). Here, we use the estimates from Schultz et al. (2017) as our primary metric of abundance. Because this abundance index comes with uncertainty, we accounted for uncertainty in annual abundance estimates by repeating PLSR analyses (described below) using 5000 bootstrapped time series of

abundance, generated by sampling from the mean and standard error of the estimate of abundance for each year.

Land use change

Land use change is among the primary threats faced by eastern monarchs (Stenoien et al. 2018; Thogmartin et al. 2017b); but see (Inamine et al. 2016); however, it has not (until now) been evaluated in relation to western monarchs. Deforestation of winter roost forests was one of the earliest causes of concern for eastern monarch population viability (Vidal et al. 2013). Agricultural intensification and associated decline in milkweed abundance (*Asclepias* spp., monarch's obligate larval host plants), are also strongly associated with eastern monarch declines (Pleasants & Oberhauser 2013). In one past analysis, increased use of glyphosate explained 77% of annual variation in eastern monarch abundance (Thogmartin et al. 2017b). More generally, use of herbicides, herbicide-resistant crops, and insecticides has increased throughout the United States in recent decades (Benbrook 2016). In addition, neonicotinoid insecticide use is correlated with declines in butterfly diversity in California (Forister et al. 2016). We assessed changes in wintering habitat using coastal development data from the California Farmland Mapping and Monitoring Program (FMMP; <http://www.conservation.ca.gov/dlrp/fmmp>). We assessed changes in breeding habitat by using pesticides as an index of agricultural intensification (See S4 Supplemental Methods).

Climate variables

Relative to land use change, relationships between monarch abundance and climate are more heterogeneous and less understood. In the West, changes in abundance at overwintering

sites from 1998-2007 correlated negatively with summer drought (Stevens & Frey 2010). Breeding season monarchs in central California from 1972-2014 tended to be more abundant in years with warmer and wetter springs (Espeset et al. 2016). These results suggest that the primary threat to western monarchs may be increasing frequency and intensity of drought (Mann & Gleick 2015), with the caveat that land use change in the West had not (prior to our results in this paper) been evaluated as a possible competing driver (cf. Espeset et al. 2016; Malcolm 2018). We assessed associations with climate variables in the overwintering season using data from the PRISM database (<http://www.prism.oregonstate.edu>) and in the breeding season using data from the National Climate Data Center (<https://www.ncdc.noaa.gov>, see S4. Supplemental Methods). For breeding season variables, we first calculated temperature and drought indices separately for four regions defined by their contribution to the western breeding population (Fig. 1, corresponding to hydrogen isotope regions used by Yang et al. (2016) to identify sources of breeding monarchs (Yang et al. 2016). Hereafter, we refer to these four regions as: Y1, Y2, Y3, and Y4 (“Y” for Yang et al. 2016). We also calculated the average of these metrics across the four regions, as a measure of overall conditions (Table 1)

Analyses

Univariate correlations: To aid in interpretation of the multivariate analyses, we calculated all pairwise univariate correlations between the predictor variables used in each analysis, log-transformed monarch abundance in each year, log-transformed population growth rates (change in abundance on a log scale), and year. We visualized and calculated these correlations using base functions in R and the panel.cor function.

Partial Least Squares Regression (PLSR): We used PLSR to predict a single response variable (here, monarch abundance) from a number of predictor variables (here, land use and climate variables). PLSR generates synthetic component predictors that are weighted sums of the original predictor variables, selected to best explain variation in a response variable (Chong & Jun 2005; Mehmood et al. 2012). The contributions of the input variables to each component are quantified by their loading, i.e., the correlation between each predictor variable and each component, and their Variable Importance in the Projection (VIP), a cumulative measure of the relative contributions of each predictor variable to the first n components of the PLSR model. Because VIP is scaled to have an average value of 1, we use VIP as a qualitative measure of relative contribution, not a significance test (Chong & Jun 2005; Mehmood et al. 2012). Unlike some model selection techniques that aim to identify a single best set of predictor variables, PLSR loadings should be interpreted as the potential importance of different factors. If a large number of correlated variables have high PLSR loadings, any or all could be the cause of abundance changes over time.

We implemented PLSR models using the `pls` package (Mevik et al. 2016) in R, and the `VIP.R` function (downloaded from <http://mevik.net/work/software/pls.html>). Models were evaluated using leave-one-out cross validation, and statistical significance of components was evaluated using the randomization test in the `selectNcomp` function. All models included natural log-transformed monarch abundance as the dependent variable, and the log of abundance in the previous year as a predictor variable. Including previous abundance generalizes the model to include effects of environmental variables on growth rate as a nested model; if the slope coefficient for previous abundance were exactly one, and previous abundance was the only variable that determined the first PLSR component, then the second component would be an

analysis of population growth rate, i.e., previous abundance would be equivalent to an offset in linear regression (recall that $\ln(A) - \ln(B) = \ln(A/B)$, so an offset is one way to estimate population growth rates from abundance data). In addition to previous abundance, the *a priori* model included seven environmental predictor variables: the proportion of developed lands near overwintering sites, glyphosate use, neonicotinoid use, average winter minimum temperature and winter precipitation at wintering sites, and average summer maximum temperature and summer drought severity (PDSI) in the breeding range (see Table 1). The “kitchen sink” model included 28 environmental predictor variables, including additional weather and pesticide variables (see Table 3).

Multiple regression (MRMI): We report the results of multiple regression, followed by multimodel inference, for comparison with PLSR. In ecology, MRMI is currently one of the most widely used approaches for evaluating environmental correlates of ecological time series. However, we note from the outset that our *a priori* data set included a large number of correlated variables, a situation that is known to be problematic for MRMI (cf. Burnham & Anderson 2002). In some cases, statisticians suggest removing correlated variables prior to MRMI analysis (Zuur et al. 2010); this solution is not suitable in our situation because the specific question of interest is which of the correlated variables are plausible predictors of monarch abundance.

We conducted MRMI using the *a priori* variable set. We fit a global model with all eight predictors and used the dredge function in the MuMIn package in R (Barton 2018) to search all possible variable combinations. Over this set of all possible models, we calculated importance values for each predictor variable (i.e., Akaike weights (Burnham & Anderson 2002) summed over all models including that predictor) and model-averaged standardized regression coefficients. To aid interpretation of results of variable importance and model-averaged

coefficients, we explored selected subsets of the variable sets included in the global model. For these comparisons, we calculated importance values and model-weighted regression coefficients only over the appropriate subset of models, i.e., over new “global” models with 5, 6 or 7 predictor variables (as appropriate), not the full 8 variables.

Results

Univariate correlations

Nearly all of the variables included in the *a priori* analysis were changing through time, i.e., significantly correlated with year (Figure 2, bottom row and right column). Coastal developed land area, glyphosate use, and neonicotinoid use increased through time (Pearson’s $r = 0.99, 0.98$ and 0.92 , respectively). Breeding season temperature and drought also increased through time, though not as strongly ($r = 0.48$ and -0.42 , noting that PDSI decreases with drought). Coastal winter minimum temperature and precipitation did not change significantly ($r = 0.28$ and -0.05). Monarch abundance also declined ($r = -0.76$), and was negatively associated with coastal development, glyphosate use, neonicotinoid use, breeding season temperature and breeding season drought ($r = -0.77, -0.75, -0.59, -0.33$ and 0.42). Annual population growth rates were higher following years with warmer winters ($r = 0.51$). Since the population was declining overall, monarch abundance in the current and previous years were positively correlated ($r = 0.85$). The set of *a priori* predictor variables was also highly collinear, with very strong correlations ($r > 0.9$) among coastal development, glyphosate use and neonicotinoid use, and weaker but statistically significant correlations among these variables and breeding temperature and drought ($0.70 > |r| > 0.35$). Coastal winter temperature was positively correlated with coastal winter precipitation ($r = 0.47$), but not with other variables. Coastal winter

precipitation was negatively associated with breeding season temperature and drought ($r = -0.34$ and 0.60).

The full set of variables in the “kitchen sink” analysis was also highly collinear, and strongly associated with monarch abundance (Figure S1). Seventeen of the 28 predictor variables were changing through time (Pearson correlation with year, $|r| > 0.34$, $P < 0.05$) and fifteen were correlated with monarch abundance (Pearson correlation with log-abundance, $|r| > 0.34$, $P < 0.05$). About half of the pairwise correlations among the predictor variables were statistically significant (Pearson correlation, $|r| > 0.34$, $P < 0.05$). The only significant correlations with annual growth rate were minimum and maximum coastal winter temperature ($r = 0.51$ and $r = 0.40$).

PLSR analysis

A priori analysis: The first two components of the *a priori* PLSR model were statistically significant ($P < 0.001$ and $P = 0.007$, respectively, Figure S2). The first component explained 64% of variation in monarch abundance, and the second explained 18.5% (Fig. 3). None of the later components were statistically significant, and none explained $> 3\%$ of variation in abundance. The first PLSR component was strongly negatively associated with land use variables (coastal development, glyphosate use, neonicotinoid use) and previous abundance (VIP > 1 , $|loadings| > 0.35$, Fig. 4 A & B), more weakly negatively associated with breeding season temperature and drought (VIP > 0.5 , $|loadings| = 0.3-0.35$) and not associated with coastal winter temperature or precipitation.

Because the first component was strongly associated with previous abundance, the second PLSR component reflects population growth rate, i.e., the annual difference between

current and previous abundance (see *Methods*). Coastal winter temperature and precipitation both had higher VIP values (0.67 and 0.44) when the first two components were considered (Fig 4A & C), capturing the positive association between population growth rate with warmer, and, to a weaker extent, drier winter conditions near overwintering sites (see univariate relationships in Fig 2). Breeding season temperature was also significantly positively associated with the second PLSR component. This combination of a negative relationship between breeding season temperature and the first PLSR component and a positive relationship with the second PLSR component reflects a nonlinear, concave relationship between breeding season temperature and monarch abundance (see univariate relationship in Fig. 2).

Conclusions were also broadly robust to estimation error in monarch abundance (shown as error bars in Fig. 4).

“Kitchen sink” model: The kitchen sink PLSR, with 29 variables (28 environmental drivers + previous abundance) performed slightly worse than the *a priori* model. The first two components were statistically significant ($P < 0.001$ and $P = 0.016$, respectively, Figure S2), and explained slightly less variation in monarch abundance than the *a priori* model with seven variables (57.5% and 19.1% of variation in monarch abundance explained by the 1st and 2nd components). The “kitchen sink” model also had noticeably higher prediction error than the *a priori* model (Fig. S2). As in the *a priori* model, the first component was strongly positively associated with coastal development, glyphosate use (both agricultural and non-agricultural), neonicotinoid use (both agricultural and nonagricultural), and previous abundance. Breeding season temperature and drought were also generally negatively associated with monarch abundance, although the strength of this relationship varied among regions, with the strongest loadings of breeding conditions in Y2 on the first PLSR component. As a measure of

temperature, cooling degree days had slightly higher PLSR loadings than maximum temperature (the metric we chose for our *a priori* analysis). Of the additional pesticide variables, pyrethroids also had strong negative loadings on the first PLSR component. Organochlorines (the pesticide group that includes DDT), and dicamba use had positive loadings on the first component. Use of both has been declining since the 1980's, so they are positively associated with trends in monarch abundance.

The second component in the “kitchen sink” model was nearly identical to the *a priori* model. It revealed positive association of annual population growth rates with warmer and drier conditions at winter roosting sites.

MRMI analysis

MRMI explained a slightly higher proportion of variance (85.9% in the global model and 84.3% in the model with the lowest AICc) than the first two *a priori* PLSR axes (82.5%). However, as described below, the MRMI variable weights and regression coefficients (Table 1) would be misleading if interpreted without accounting for the strong collinearity among predictor variables.

The variable with the highest importance was abundance in the previous year (Akaike weight, $w_i = 1.00$, standardized regression coefficient, averaged across the subset of models including this term, $\beta_i = 1.13$), followed by minimum temperature at overwintering sites ($w_i = 0.92$, $\beta_i = 0.43$). Glyphosate use and coastal development had very similar values ($w_i = 0.58$ and 0.50 and $\beta_i = -0.97$ and -0.72 , respectively). The contrast between the importance of winter temperature vs. glyphosate and coastal development occurs at least in part because trends in statewide glyphosate use and coastal development are so highly correlated ($r = 0.91$) that they are

effectively substitutable. For each, the relative importance across the subset of models including the other variable was small ($w_i = 0.21$ for coastal development across the subset of models constrained to include glyphosate, and $w_i = 0.24$ for glyphosate across the subset of models constrained to include coastal development), whereas the importance across the subset of models without the other variable was high ($w_i = 0.90$ for coastal development for the subset of models without glyphosate, and $w_i = 0.92$ for glyphosate use across the subset of models without coastal development; see Table S3). In contrast, coastal winter temperature is only weakly correlated with any of the other predictor variables (Figure 2), so its influence in statistical models cannot be substituted by other variables (see Table 3).

Neonicotinoid use had a high importance ($w_i = 0.92$), but the model-averaged regression coefficient was positive ($\beta_i = 0.40$), i.e., in the opposite direction of the univariate correlation and PLSR loading in relation to monarch abundance. This positive relationship is a statistical artefact of high collinearity among predictor variables included in the model; when neonicotinoids are considered only across the subset of models without any of the three highly collinear ($r > 0.9$, Fig 1) variables (previous abundance, glyphosate use, and coastal development), i.e., only in the presence of climate variables, its importance was very high, $w_i = 0.99$, and the model-averaged regression coefficient was strongly negative, $\beta_i = -0.94$.

The least important variables in MRMI models were breeding season temperature, drought, and coastal winter precipitation ($w_i = 0.22, 0.35$ and 0.31 and $\beta_i = 0.24, -0.08, -0.20$). However, none of these effects are negligible; $w_i = 0.22$ can be loosely interpreted as a 22% chance that breeding season temperature is included in the best model. All three effects are consistent with conclusions from the univariate correlations and PLSR models, i.e., negative

association of monarch abundance with breeding season temperature, breeding season drought, and coastal winter precipitation.

Discussion

Our results reveal many possible explanations for monarch declines in the West, and indicate that potential environmental drivers are too highly correlated to unambiguously discriminate their effects. Nonetheless, there is stronger support for land use change than climate change as the cause of declines. Past studies have investigated climate variables only as drivers of western monarch abundance (e.g., Dingle et al. 2005; Espeset et al. 2016; Stevens & Frey 2010). These past analyses, as well as naïve interpretation of the MRMI analysis in this paper, could be interpreted as evidence that climate is the primary driver of western monarch populations. In contrast, our PLSR analysis suggests that habitat loss and pesticide use are more likely to be the primary drivers of declines in the West.

In spite of this contrast, our analysis reinforces previously-found associations with environmental variables, including the general importance of central California for monarch abundance (compare Table 2 in Stephens and Frey 2010 to the importance of Y2 in our “kitchen sink” analysis). Similarly, the importance of overwintering habitat loss in the PLSR analysis corroborates speculation by Espeset et al. (2016), who proposed the overwintering stage as an important driver of monarch populations because, at a set of breeding sites in central California, declines in adult monarch abundance were stronger in spring than summer. Espeset et al. (2016) also showed higher abundance of monarchs at breeding sites in years with warmer winter and spring temperatures, and wetter springs. Combined with our results, these results suggest that warm winters and springs generally favor monarchs, but hot summers do not, similar to seasonal

patterns shown by Zipkin et al. (2012) for eastern monarchs. Our results are also partly consistent with Forister et al. (2016), who found strong negative associations between annual neonicotinoid use and butterfly abundance across 67 species in California. Their data did not appear to reveal any association between butterfly abundance and carbamates or pyrethroids, both of which were negatively associated with monarch abundance in our “kitchen sink” analysis. This discrepancy may be due to differences in spatial and temporal scale of the two analyses, and it emphasizes that both results are correlational. Experimental investigation of nontarget effects of pesticides on butterflies is an important area for future research.

We were surprised by the extent to which PLSR outperformed MRMI. Specifically, conclusions about which variables were associated with declines were broadly consistent across the *a priori* and kitchen sink models, with only subtle qualitative differences. In contrast, variable weights and regression coefficients from MRMI differed radically, depending on which variables were included in the models. Simulation studies have shown that PLSR is more robust than MRMI, in the sense that it is better at identifying the correct causal variables, especially when sample sizes are low and/or predictor variables are correlated (Selwood et al. 2015). However, past comparisons of PLSR and MRMI (Carrascal et al. 2009; Hosseini et al. 2017) have emphasized smaller quantitative differences in predictive power, not the stark qualitative differences in variable importance and directions of effects we see here. PLSR is increasingly used in conservation-related fields such as soil science (Shi et al. 2013) and forest and fisheries management (Lecerf & Richardson 2010, Alofs et al. 2014). PLSR may be a useful tool for conservation biologists to adopt more widely, especially in situations where inferences about the potential effects of correlated drivers of species’ distribution and abundance are necessary. Of course, no statistical tool is right for all situations. For example, it is tempting to include long

lists of candidate predictor variables in PLSR models because PLSR is theoretically able to handle very large numbers of predictor variables. In such cases, however, models are subject to overfitting and begin to lose their ability to identify the correct variables, as evidenced by lower performance of our kitchen sink model compared to the *a priori* PLSR model (see also James et al. 2013)

For insect populations in general, the list of at-risk species is growing, and previously widespread species are joining the ranks of at-risk species (Cameron et al. 2011; Van Dyck et al. 2009). In many cases these declines represent some unknown combination of effects of land use and climate change (Casner et al. 2014). The strength of our PLSR analyses is that they deal with multiple correlated variables in a statistically robust way. In this case, because population declines are most strongly associated with land use change, habitat conservation and restoration are a good first management step that may go a long way towards population recovery. At the same time, although PLSR results are informative, they do not replace causal explanations. Like any analysis of monitoring data, PLSR results are best interpreted as suggesting hypotheses for experimental testing (Alofs et al. 2014). The clear next steps are mechanistic experiments to test if habitat protection and restoration opportunities – which hinge on factors identified in analysis such as this – can reverse declines before we lose formerly widespread species like the monarch butterfly.

Acknowledgements

We are grateful to M. Seiderman for assistance with climate data, J. Thorne for introducing us to FMMP data, K. Steinmann for help accessing archival pesticide data and providing corrections to the database, S. Black, A. Code, C. Fallon, R. Hatfield, & S. Jepsen for discussions of possible

threats, and E. Scott for discussion of PLSR. We thank M. Forister, A. Shapiro, W. Thogmartin and L. Yang for valuable comments on a draft of this manuscript. This work was funded by a contract from the US Fish and Wildlife Service to CBS and EEC.

Author contributions

CBS, CCT, EMP and LMB compiled the database. EEC conceived and ran analyses with input from CBS and LMB. EEC drafted the manuscript. All coauthors commented on the manuscript.

References

- Alofs, K. M., D. A. Jackson, and N. P. Lester. 2014. Ontario freshwater fishes demonstrate differing range-boundary shifts in a warming climate. *Diversity and Distributions* **20**:123-136.
- Barton, K. 2018. MuMIn: Multi-Model Inference. R package version 1.42.1. .
- Benbrook, C. M. 2016. Trends in glyphosate herbicide use in the United States and globally. *Environmental Sciences Europe* **28**:1-15.
- Brook, B. W., N. S. Sodhi, and C. J. A. Bradshaw. 2008. Synergies among extinction drivers under global change. *Trends in Ecology & Evolution* **23**:453-460.
- Burnham, K. P., and D. R. Anderson 2002. Model selection and multimodel inference : a practical information-theoretic approach. Springer, New York.
- Cameron, S. A., J. D. Lozier, J. P. Strange, J. B. Koch, N. Cordes, L. F. Solter, and T. L. Griswold. 2011. Patterns of widespread decline in North American bumble bees.

433 Proceedings of the National Academy of Sciences of the United States of America
 434 **108**:662-667.

435 Carrascal, L. M., I. Galvan, and O. Gordo. 2009. Partial least squares regression as an alternative
 436 to current regression methods used in ecology. *Oikos* **118**:681-690.

437 Casner, K. L., M. L. Forister, K. Ram, and A. M. Shapiro. 2014. The utility of repeated presence
 438 data as a surrogate for counts: a case study using butterflies. *Journal of Insect*
 439 *Conservation* **18**:13-27.

440 Chong, I. G., and C. H. Jun. 2005. Performance of some variable selection methods when
 441 multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems* **78**:103-
 442 112.

443 Davidson, A. D., A. G. Boyer, H. Kim, S. Pompa-Mansilla, M. J. Hamilton, D. P. Costa, G.
 444 Ceballos, and J. H. Brown. 2012. Drivers and hotspots of extinction risk in marine
 445 mammals. *Proceedings of the National Academy of Sciences of the United States of*
 446 *America* **109**:3395-3400.

447 Didham, R. K., J. M. Tylianakis, M. A. Hutchison, R. M. Ewers, and N. J. Gemmell. 2005. Are
 448 invasive species the drivers of ecological change? *Trends in Ecology & Evolution*
 449 **20**:470-474.

450 Dingle, H., M. P. Zalucki, W. A. Rochester, and T. Armijo-Prewitt. 2005. Distribution of the
 451 monarch butterfly, *Danaus plexippus* (L.) (Lepidoptera : Nymphalidae), in western North
 452 America. *Biological Journal of the Linnean Society* **85**:491-500.

453 Earl, J. E., S. Nicol, R. Wiederholt, J. E. Diffendorfer, D. Semmens, D. T. T. Flockhart, B. J.
 454 Mattsson, G. McCracken, D. R. Norris, W. E. Thogmartin, and L. Lopez-Hoffman. 2018.

Quantitative tools for implementing the new definition of significant portion of the range in the US Endangered Species Act. *Conservation Biology* **32**:35-49.

Espelet, A. E., J. G. Harrison, A. M. Shapiro, C. C. Nice, J. H. Thorne, D. P. Waetjen, J. A. Fordyce, and M. L. Forister. 2016. Understanding a migratory species in a changing world: climatic effects and demographic declines in the western monarch revealed by four decades of intensive monitoring. *Oecologia* **181**:819-830.

Forister, M. L., B. Cousens, J. G. Harrison, K. Anderson, J. H. Thorne, D. P. Waetjen, C. C. Nice, M. Parsia, M. L. Hladik, R. Meese, H. Vliet, and A. M. Shapiro. 2016. Increasing neonicotinoid use and the declining butterfly fauna of lowland California. *Biology Letters* **12**.

Hosseini, M., S. R. Agereh, Y. Khaledian, H. J. Zoghalchali, E. C. Brevik, and S. A. Naeini. 2017. Comparison of multiple statistical techniques to predict soil phosphorus. *Applied Soil Ecology* **114**:123-131.

Inamine, H., S. P. Ellner, J. P. Springer, and A. A. Agrawal. 2016. Linking the continental migratory cycle of the monarch butterfly to understand its population decline. *Oikos* **125**:1081-1091.

James, G., D. Witten, T. Hastie, and R. Tibshirani 2013. *An Introduction to Statistical Learning with Applications in R*. Springer Science+Business Media, New York.

Jepsen, S., and S. H. Black. 2015. Understanding and conserving the western North American monarch population. Pages 147-156 in K. Oberhauser, K. R. Nail, and S. Altizer, editors. *Monarchs in a changing world*. Cornell University Press, Ithaca.

476 Lecerf, A., and J. S. Richardson. 2010. Litter decomposition can detect effects of high and
 477 moderate levels of forest disturbance on stream condition. *Forest Ecology and*
 478 *Management* **259**:2433-2443.

479 Malcolm, S. B. 2018. Anthropogenic Impacts on Mortality and Population Viability of the
 480 Monarch Butterfly. Pages 277-302 in M. R. Berenbaum, editor. *Annual Review of*
 481 *Entomology*, Vol 63.

482 Mann, M. E., and P. H. Gleick. 2015. Climate change and California drought in the 21st century.
 483 *Proceedings of the National Academy of Sciences* **112**:3858-3859.

484 Mehmood, T., K. H. Liland, L. Snipen, and S. Saebo. 2012. A review of variable selection
 485 methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory*
 486 *Systems* **118**:62-69.

487 Mevik, B.-H., R. Wehrens, and K. H. Liland. 2016. pls: Partial Least Squares and Principal
 488 Component Regression. R package version 2.6-0 [https://CRAN.R-](https://CRAN.R-project.org/package=pls)
 489 [project.org/package=pls](https://CRAN.R-project.org/package=pls).

490 Morris, G. M., C. Kline, and S. M. Morris. 2015. Status of *Danaus plexippus* populations in
 491 Arizona. *Journal of the Lepidopterists Society* **69**:91-107.

492 Pleasants, J. 2017. Milkweed restoration in the Midwest for monarch butterfly recovery:
 493 estimates of milkweeds lost, milkweeds remaining and milkweeds that must be added to
 494 increase the monarch population. *Insect Conservation and Diversity* **10**:42-53.

495 Pleasants, J. M., and K. S. Oberhauser. 2013. Milkweed loss in agricultural fields because of
 496 herbicide use: effect on the monarch butterfly population. *Insect Conservation and*
 497 *Diversity* **6**:135-144.

498 Schultz, C. B., L. M. Brown, E. Pelton, and E. E. Crone. 2017. Citizen science monitoring
 499 demonstrates dramatic declines of monarch butterflies in western North America.
 500 *Biological Conservation* **214**:343-346.

501 Selwood, K. E., M. A. McGeoch, and R. Nally. 2015. The effects of climate change and land-use
 502 change on demographic rates and population viability. *Biological Reviews* **90**:837-853.

503 Semmens, B. X., D. J. Semmens, W. E. Thogmartin, R. Wiederholt, L. Lopez-Hoffman, J. E.
 504 Diffendorfer, J. M. Pleasants, K. S. Oberhauser, and O. R. Taylor. 2016. Quasi-extinction
 505 risk and population targets for the Eastern, migratory population of monarch butterflies
 506 (*Danaus plexippus*). *Scientific Reports* **6**.

507 Service, U. F. a. W. 2018. *Endangered species: U. S. Species*, Falls Church, VA.

508 Service, U. S. F. a. W. 2015. Endangered and Threatened Wildlife and Plants; 12-Month Finding
 509 on a Petition To List Greater Sage-Grouse (*Centrocercus urophasianus*) as an Endangered
 510 or Threatened Species; Proposed Rule. *Federal Register* **80**:59858-59942.

511 Shi, Z. H., L. Ai, X. Li, X. D. Huang, G. L. Wu, and W. Liao. 2013. Partial least-squares
 512 regression for linking land-cover patterns to soil erosion and sediment yield in
 513 watersheds. *Journal of Hydrology* **498**:165-176.

514 Smith, D. R., N. L. Allan, C. P. McGowan, J. A. Szymanski, S. R. Oetker, and H. M. Bell. 2018.
 515 Development of a Species Status Assessment Process for Decisions under the US
 516 Endangered Species Act. *Journal of Fish and Wildlife Management* **9**:302-320.

517 Stenoien, C., K. R. Nail, J. M. Zalucki, H. Parry, K. S. Oberhauser, and M. P. Zalucki. 2018.
 518 Monarchs in decline: a collateral landscape-level effect of modern agriculture. *Insect*
 519 *Science* **25**:528-541.

520 Stevens, S. R., and D. F. Frey. 2010. Host plant pattern and variation in climate predict the
 521 location of natal grounds for migratory monarch butterflies in western North America.
 522 Journal of Insect Conservation **14**:731-744.

523 Thogmartin, W. E., L. Lopez-Hoffman, J. Rohweder, J. Diffendorfer, R. Drum, D. Semmens, S.
 524 Black, I. Caldwell, D. Cotter, P. Drobney, L. L. Jackson, M. Gale, D. Helmers, S.
 525 Hilburger, E. Howard, K. Oberhauser, J. Pleasants, B. Semmens, O. Taylor, P. Ward, J.
 526 F. Weltzin, and R. Wiederholt. 2017a. Restoring monarch butterfly habitat in the
 527 Midwestern US: 'all hands on deck'. Environmental Research Letters **12**.

528 Thogmartin, W. E., R. Wiederholt, K. Oberhauser, R. G. Drum, J. E. Diffendorfer, S. Altizer, O.
 529 R. Taylor, J. Pleasants, D. Semmens, B. Semmens, R. Erickson, K. Libby, and L. Lopez-
 530 Hoffman. 2017b. Monarch butterfly population decline in North America: identifying the
 531 threatening processes. Royal Society Open Science **4**.

532 Van Dyck, H., A. J. Van Strien, D. Maes, and C. A. M. Van Swaay. 2009. Declines in common,
 533 widespread butterflies in a landscape under intense human use. Conservation Biology
 534 **23**:957-965.

535 Vidal, O., J. Lopez-Garcia, and E. Rendon-Salinas. 2013. Trends in Deforestation and Forest
 536 Degradation after a Decade of Monitoring in the Monarch Butterfly Biosphere Reserve in
 537 Mexico. Conservation Biology **28**:177-186.

538 Yang, L. H., D. Ostrovsky, M. C. Rogers, and J. M. Welker. 2016. Intra-population variation in
 539 the natal origins and wing morphology of overwintering western monarch butterflies
 540 *Danaus plexippus*. Ecography **39**:998-1007.

541 Zeigler, S. L., D. H. Catlin, M. B. Brown, J. D. Fraser, L. R. Dinan, K. L. Hunt, J. G. Jorgensen,
542 and S. M. Karpanty. 2017. Effects of climate change and anthropogenic modification on
543 a disturbance-dependent species in a large riverine system. *Ecosphere* **8**.
544 Zipkin, E. F., L. Ries, R. Reeves, J. Regetz, and K. S. Oberhauser. 2012. Tracking climate
545 impacts on the migratory monarch butterfly. *Global Change Biology* **18**:3039-3049.
546 Zuur, A. F., E. N. Ieno, and C. S. Elphick. 2010. A protocol for data exploration to avoid
547 common statistical problems. *Methods in Ecology and Evolution* **1**:3-14.

548

549

Table 1. Summary statistics from secondary analyses, including the “kitchen sink” PLSR and multiple regression, followed by multimodel inference. Related statistics for the *a priori* PLSR are shown in Figure 4.

Variable		Role in <i>a priori</i> PLSR	“kitchen sink” PLSR				Multiple regression	
			VIP 1	VIP 2	Load 1	Load 2	wi	bi
Coastal development		included	1.65	1.45	-0.28	-0.14	0.50	-0.72
Glyphosate use, agricultural		summed	1.53	1.32	-0.28	-0.10	0.58	-0.97
Glyphosate use, non-ag.			1.69	1.56	-0.26	-0.23		
Dicamba use			0.83	0.75	0.13	0.07		
Neonicotinoid use, agricultural		summed	1.16	1.06	-0.25	-0.04	0.63	0.71
Neonicotinoid use, non-ag.			1.44	1.25	-0.26	-0.03		
Carbamates			1.09	0.99	-0.24	0.01		
Organophosphates			0.49	0.68	0.15	-0.05		
Organochlorines (incl. DDT)			1.67	1.57	0.24	0.22		
Pyrethroids			1.51	1.31	-0.27	-0.09		
Coastal winter min. temp		included	0.16	0.44	-0.08	0.07	0.92	0.44
Coastal winter max. temp			0.07	0.62	-0.06	0.24		
Coastal winter avg. temp			0.27	0.64	-0.02	0.33		
Coastal winter precipitation		included	0.05	0.52	0.07	-0.24	0.31	-0.20
Breeding season maximum temperature	Y1	average d	0.65	0.80	-0.18	0.26	0.22	-0.08
	Y2		1.03	0.98	-0.24	0.18		
	Y3		0.60	0.92	-0.20	0.36		
	Y4		0.51	0.93	-0.19	0.36		
Breeding season cooling degree days (CDD)	Y1		0.58	0.82	-0.18	0.27		
	Y2		1.13	1.05	-0.25	0.17		
	Y3		0.74	0.93	-0.21	0.34		
	Y4		0.68	0.83	-0.19	0.32		
Breeding season Palmer Drought Severity Index (PDSI)	Y1	average d	0.33	0.64	0.13	-0.22	0.35	0.24
	Y2		0.73	0.80	0.19	-0.20		
	Y3		0.85	0.76	0.14	-0.12		
	Y4		0.88	0.78	0.14	-0.07		
SOI*			0.25	0.33	-0.07	-0.07		
MEI*			0.30	0.26	0.06	0.05		
Previous abundance		included	1.83	1.88	0.23	0.24	1.00	1.13

*Indices of the El Nino Southern Oscillation; MEI = Multivariate ENSO (El Nino Southern Oscillation) Index and SOI = Southern Oscillation Index. See S4 Supplemental Methods.

DRAFT

Figure 1. Map of western monarch breeding and wintering in the West. The dotted line identifies the eastern limit of coastal winter habitat. Orange and yellow arrows identify hypothesized spring and fall migration routes. Colored bands (blue, green and yellow) correspond to four regions of breeding habitat defined by Yang et al. (2015), used to identify different habitat zones in the kitchen sink PLSR in this paper. Gray areas are outside the known range of western monarch breeding habitat, and white areas do not currently support milkweed.

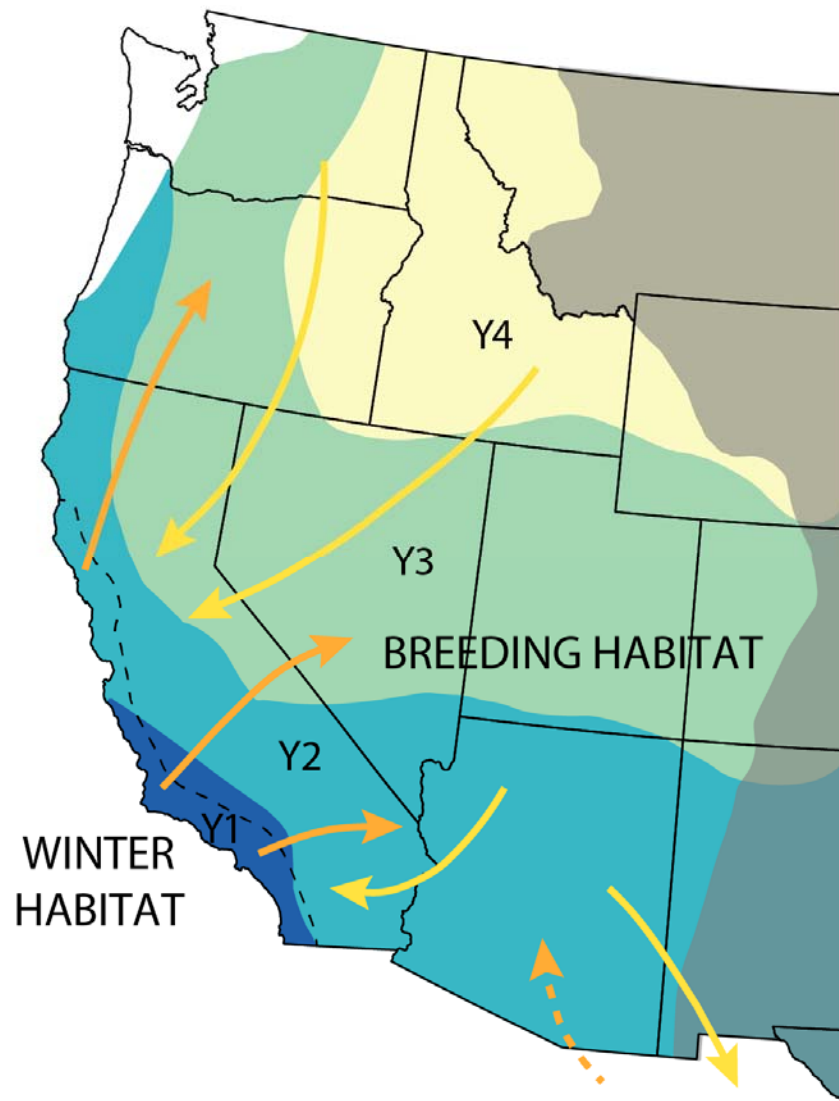


Fig 2: Scatterplot matrix of standardized variables included in the *a priori* PLSR and MRMI analyses. The first two variables are the log-scale abundance index in coastal wintering groves, and log-scale annual growth rate. Other variables follow the order of presentation in Table 1. For graphical presentation, all variables were scaled to have a mean of 0 and standard deviation of 1. Red lines in scatterplots (lower left) identify moving average values. Font size of correlation coefficients (upper right) is proportional to statistical significance of the correlation coefficient.

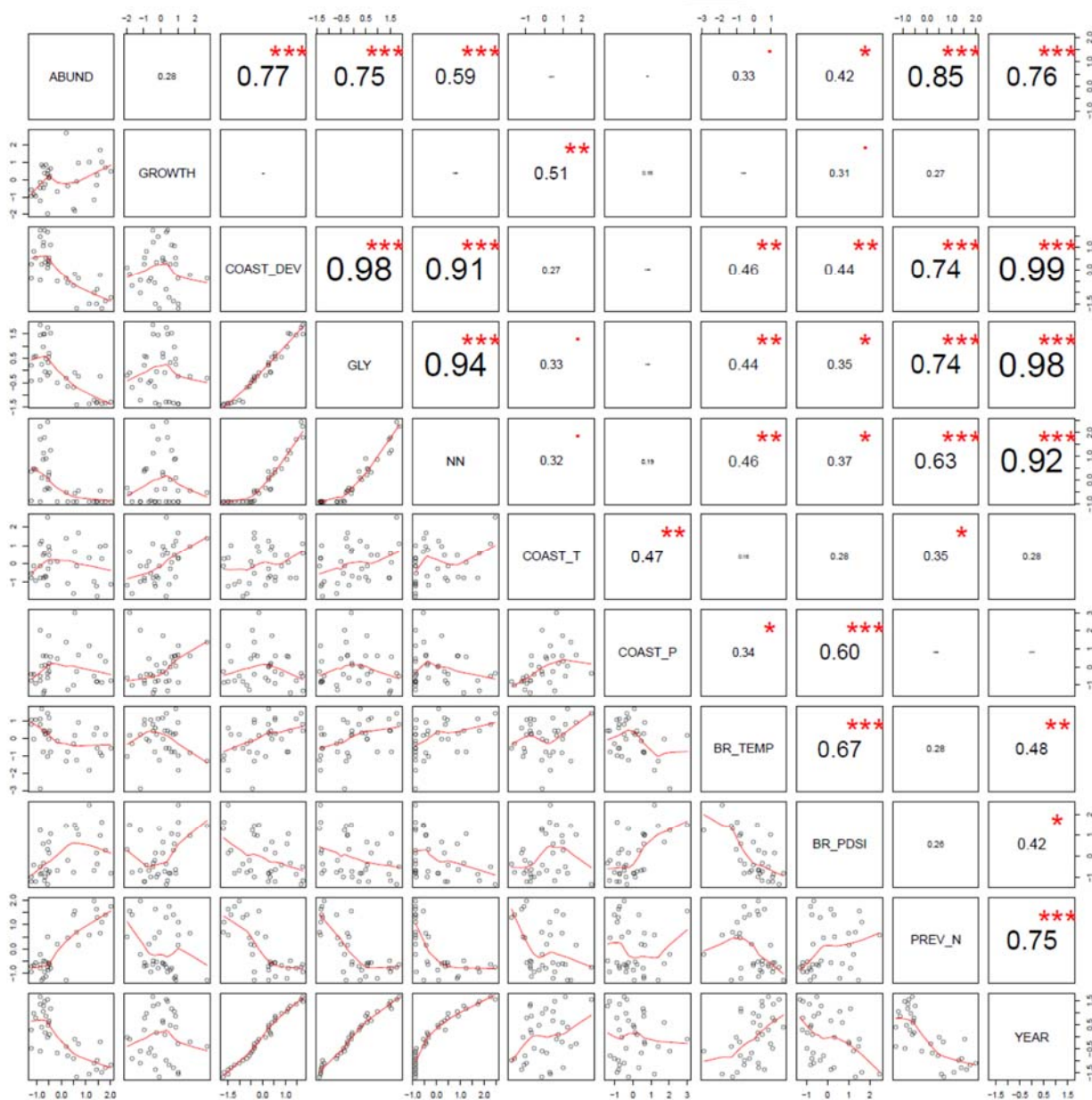


Figure 3. Biplot of *a priori* PLSR. Length of arrows is proportional to loading (i.e., variance explained by each component). Size of points is proportional to log-abundance in each year. Axis length is proportional to the variation in monarch abundance explained by each.

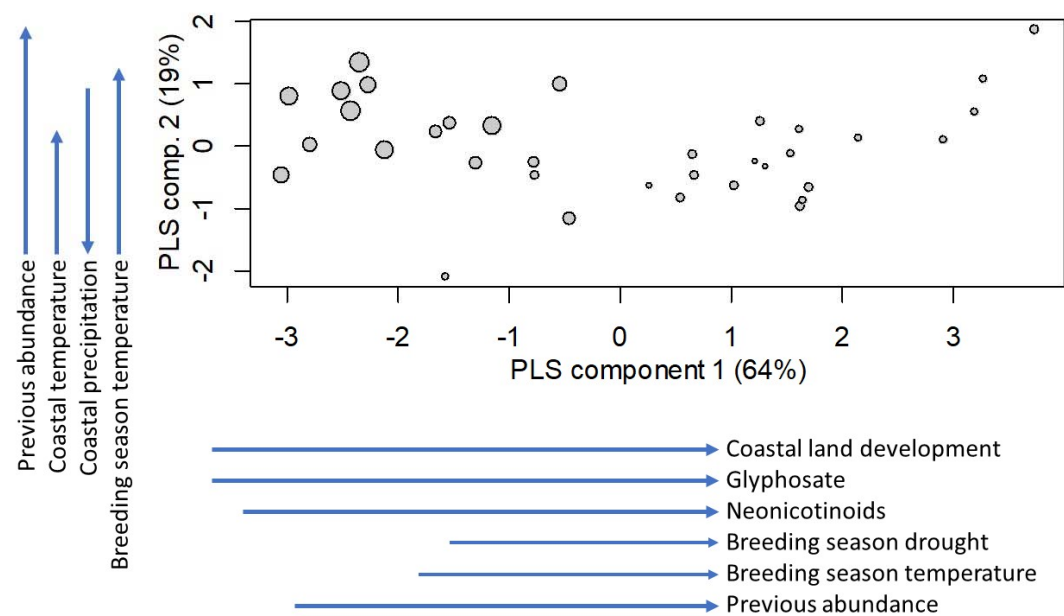


Figure 4. PLSR results for a priori model. (A) Variable importance in the Projection (VIP), with contributions to the first component in gray and contributions summed over the first and second components in white. (B) Loadings of each variable on the first PLSR component, with size of circles proportional to their loading. (C) Loadings of each variable on the second PLSR component, with size of circles proportional to their loading. Error bars on all three panels were calculated by repeating the analysis over 5000 bootstrap time series, sampled from the mean and standard error of the abundance index estimate in each year.

