Species distribution models of a predator-prey system under climate change

Xuezhen Ge\textsuperscript{1}, Cort Griswold\textsuperscript{1}, and Jonathan Newman\textsuperscript{2}

\textsuperscript{1}University of Guelph  
\textsuperscript{2}Wilfrid Laurier University

July 28, 2023

Abstract

Mechanistic and correlative models are two types of species distribution models (SDMs). They each have distinct foci, conceptual foundations, and levels of dependency on data availability, leading to potentially different estimates of species' ecological niches and distributions. Mechanistic SDMs integrate detailed biological processes, making it possible to account for species' biotic interactions. Despite their assumed importance, interactions in species distribution modeling remain uncommon. In this study, we applied an ensemble model of multiple correlative SDMs, a mechanistic SDM of the focal species (prey) alone, and a mechanistic SDM of the predator-prey interactions, to compare the predictions of correlative and mechanistic approaches and assess their relative strengths and limitations. We predict there are considerable and subtle differences in various predictions generated by the correlative and mechanistic approaches for each aphid species, which call for prior knowledge concerning species' presence data or life histories. Our mechanistic SDMs allowed for the assessment of the relative significance of abiotic and biotic factors, along with their interactions, in determining species' habitat suitability. Additionally, we predict aphid habitat suitability decreases across continents due to the effect of predation. However, this decrease may be offset or enhanced by the interaction effect between predation and climate change in different regions. This suggests the necessity of accounting for biotic interactions and the interplay between abiotic and biotic factors in mechanistic approaches. Our research highlights the impact of model philosophies in SDM studies and addresses the importance of selecting an appropriate modeling approach in line with the study's objectives. Furthermore, our study suggests that mechanistic SDMs could serve as a valuable addition for assessing the robustness of correlative SDM predictions.
Species distribution models of a predator-prey system under climate change

Author One

Abstract

Mechanistic and correlative models are two types of species distribution models (SDMs). They each have distinct foci, conceptual foundations, and levels of dependency on data availability, leading to potentially different estimates of species’ ecological niches and distributions. Mechanistic SDMs integrate detailed biological processes, making it possible to account for species’ biotic interactions. Despite their assumed importance, interactions in species distribution modeling remain uncommon. In this study, we applied an ensemble model of multiple correlative SDMs, a mechanistic SDM of the focal species (prey) alone, and a mechanistic SDM of the predator-prey interactions, to compare the predictions of correlative and mechanistic approaches and assess their relative strengths and limitations. We predict there are considerable and subtle differences in various predictions generated by the correlative and mechanistic approaches for each aphid species, which call for prior knowledge concerning species’ presence data or life histories. Our mechanistic SDMs allowed for the assessment of the relative significance of abiotic and biotic factors, along with their interactions, in determining species’ habitat suitability. Additionally, we predict aphid habitat suitability decreases across continents due to the effect of predation. However, this decrease may be offset or enhanced by the interaction effect between predation and climate change in different regions. This suggests the necessity of accounting for biotic interactions and the interplay between abiotic and biotic factors in mechanistic approaches. Our research highlights the impact of model philosophies in SDM studies and addresses the importance of selecting an appropriate modeling approach in line with the study’s objectives. Furthermore, our study suggests that mechanistic SDMs could serve as a valuable addition for assessing the robustness of correlative SDM predictions.
1 Introduction

Global climate change is a significant concern for species worldwide and may result in an increase in extinction rates, changes to species abundances, range shifts, and range contractions in many taxa, such as plants, birds, terrestrial and marine species (Sekercioglu et al., 2008; Chen et al., 2011; Urban, 2015; Soroye et al., 2020; Pinsky et al., 2020; Hastings et al., 2020; Zu et al., 2021). The responses of species to climate change may cause irreversible effects on natural ecosystems and threaten biodiversity, food production, and biosecurity (Malhi et al., 2020; Román-Palacios and Wiens, 2020; Weiskopf et al., 2020). Estimating how species’ abundance and distribution respond to climate change lays the foundation of predicting future global changes.

1.1 Types of species distribution models

Species distribution models (SDMs), theoretically grounded in Hutchinson’s (1957) ‘Ecological Niche Theory’, are commonly used to predict contemporary and future changes in species’ distributions across terrestrial, freshwater, and marine environments (Robinson et al., 2017; Zurell et al., 2020). These models delineate a species’ fundamental niche and predict the species’ potential geographical distribution through space and time, based on abiotic factors.

SDMs are typically either correlative models (i.e., statistical models) or mechanistic models (i.e., process-based models). The comparison between the correlative and mechanistic approaches has been of interest in recent years (Kearney and Porter, 2009; Dormann et al., 2012; Pacifici et al., 2015; Johnston et al., 2019; Tourinho and Vale, 2023). Dormann et al. (2012) comprehensively summarized thirteen features that are either common to, or different between, the two approaches, including Assumptions, Information required, Determination of model structure, etc. More recently, Tourinho and Vale (2023) conducted a systematic review focused on 18 modeling studies and drew distinctions amongst the conceptual, methodological, and forecasting aspects of the two approaches, as well as the hybrid approach which integrates correlative and mechanistic SDMs.

Correlative SDMs (‘classic SDMs’) generally use different statistical or machine learning methods to correlate species’ occurrence records and environmental conditions. They have been widely applied to predict species’ range shifts under climate change, thanks to readily available biogeographical (e.g., presence records) and environmental data (Desjonquères et al., 2022). Despite their popularity, there are several significant drawbacks associated with correlative SDMs. Correlative SDMs do not directly model eco-physiological processes that underpin species’ distributions, making it challenging to discern the factors governing a species’
fundamental niche and those driving range shifts (Dormann et al., 2012). Furthermore, the transferability of
correlative SDMs is limited spatially and temporally due to the assumption of niche conservatism (Fernández
and Hamilton, 2015; Martínez-Freiría et al., 2016; Yates et al., 2018; Liu et al., 2020; Newman et al., 2022).
Heterogeneity and incomplete sampling will likely cause statistical artifacts, which bias model predictions
(Guillaumot et al., 2022).

Compared with correlative SDMs, mechanistic SDMs are alternative approaches that have been recog-
nized to overcome the constraints stated above (Desjonquères et al., 2022). They explicitly incorporate bio-
logical processes through empirically derived relationships between species’ eco-physiological information
and environmental conditions (Tourinho and Vale, 2023), such that they may exhibit a greater potential for
transferability than correlative SDMs across time and space (Yates et al., 2018). Furthermore, since mech-
anicistic SDMs rely on intrinsic mechanisms instead of occurrences, they do not suffer from the sampling
bias limitations of correlative SDMs (Chiu-Valderrama et al., 2022). Despite these advantages, mechanistic
SDMs are still rarely used for SDMs due to the need for detailed knowledge of species-specific physiolog-
ical data and other biological processes, modeling expertise, and a more significant investment of time and
effort (Kearney and Porter, 2009; Dormann et al., 2012; Urban et al., 2016; Chiu-Valderrama et al., 2022).
Nevertheless, the advantages of mechanistic models promote their growing application, especially for ecto-
otherms (Newman et al., 2022). Most mechanistic SDMs can be classified among five types: occupancy
dynamics models, abundance dynamics models, coupled SDM-population models, demographic distribution
models (DDM), eco-physiological models, and individual-based models (Briscoe et al., 2019). Among these,
eco-physiological models are commonly used for ectotherms by accounting for species’ thermal performance
curves and critical thermal temperatures to make predictions (Tourinho and Vale, 2023).

Both correlative and mechanistic models have their advantages and disadvantages. Hybrid models attempt
to combine the two approaches in various ways to take advantage of these differences (Dormann et al., 2012;
Peterson et al., 2015; Tourinho and Vale, 2023). Tourinho and Vale (2023) summarized the goals of hybrid
approaches as: (1) using the prediction from mechanistic model as input in correlative model; (2) using the
prediction from correlative model as input in mechanistic model; (3) considering only suitable regions pre-
dicted by correlative model for mechanistic model; (4) using the unsuitable regions predicted by mechanistic
model as absence data for correlative model, and (5) considering only the overlapping regions predicted by
two types of models. Hybrid models can compensate for some of the limitations of both approaches (Chiu-
Valderrama et al., 2022), but there is an ongoing debate over whether it is appropriate to integrate the two.
Some claim that due to differences in conceptual foundations and scientific goals, combining them may lead
to misinterpretation (Coelho et al., 2019; Tourinho and Vale, 2023). On the other hand, others believe that they are not inherently incompatible and can be integrated to make predictions. In fact, hybrid models tend to exhibit higher or similar accuracy than either approach alone (Briscoe et al., 2019; Tourinho and Vale, 2023). While many studies advocate for integrating both approaches rather than comparing them, the systematic review conducted by Tourinho and Vale (2023) suggests that the best approach should be selected based on the study’s objective.

1.2 Incorporating biotic interactions

Biotic interaction is a notable omission in SDM studies that focus on abiotic variables (Wisz et al., 2013; Dormann et al., 2018). A large body of work reveals that abiotic environmental changes can influence the nature of species’ biotic interactions. Biotic interactions can likewise affect species responses to abiotic environmental changes (Jaeschke et al., 2012; Blois et al., 2013). Some empirical evidence suggests that the performance of SDMs could be improved by explicitly incorporating biotic variables (Palacio and Girini, 2018; Atauchi et al., 2018). An increasing number of studies have attempted to incorporate biotic interactions into SDMs. In previous studies, biotic interactions were rarely modeled explicitly; interacting species were modeled separately, and the role of interactions was indicated by superimposing the species’ distributions (Anderson et al., 2002; Jaeschke et al., 2012; Giannini et al., 2013; Berzitis et al., 2014). Dormann et al. (2018) presented five main approaches for incorporating biotic interactions in (i) correlative methods, (ii) interactor-as-predictor approaches, (iii) ordination approaches, (iv) spatial factor analysis, and (v) joint hierarchical multispecies models. Other than these statistical modeling methods, various dynamic models integrate biotic interactors as well as abiotic factors, such as Dynamic vegetation models (DVMs) and BioMove (Midgley et al., 2010; Wisz et al., 2013; Snell et al., 2014). Despite these advances, the inclusion of biotic interactions is still uncommon in species distribution modeling, due to a variety of challenges, such as multicollinearity for correlative models (Graham, 2003; Wisz et al., 2013; Godsoe et al., 2017), parameterization of mechanistic models (Dormann et al., 2012; Wisz et al., 2013), and species’ evolutionary responses to climate change. Up to now, studies incorporating biotic interactions under climate change that go on to predict the global spatial distributions of species are rare (Cosentino et al., 2023). Few studies have quantitatively identified the relative importance of abiotic and biotic factors in affecting species abundance and distribution at macro scales.

In this study, we are primarily interested in the differences in predictions between correlative and mechanistic SDMs applied to the same focal species, as well as the importance of predator-prey interactions in SDM
studies. We constructed three models: (a) an ensemble model of multiple correlative SDMs; (b) a mechanistic model of the focal species (prey) alone; and (c) a mechanistic model of the predator-prey interaction. The conceptual diagram of the correlative and mechanistic approaches that we used in this study is illustrated in Figure 1. We sought to address the following questions by comparing the predictions from the three models:

• How similar are the predictions of an ensemble correlative SDM to mechanistic SDMs of the same focal species?

• How similar are predictions of models that include biotic interactions to those that exclude such interactions?

• Given that mechanistic models often take a considerably large investment of time and energy to construct and evaluate as compared to correlative SDMs, is that investment worthwhile in terms of differences in predicted outcomes?

2 Method

2.1 Species occurrence data and life history data

Cereal aphids (Aphis spp.) can pose severe threats to many crop species. It is of great economic importance to study their potential distributions. We selected five cereal aphids which are major agricultural pests, including the soybean aphid (Aphis glycines Matsumura), the cotton aphid (Aphis gossypii Glover), the green peach aphid (Myzus persicae Sulz.), the bird cherry-oat aphid (Rhopalosiphum padi L.), and the English grain Aphid (Sitobion avenae Fabricius) as the focal species for this study, and one ladybird species, the Asian lady beetle (Harmonia axyridis Pallas), as their predator species to explore the importance of predator-prey interactions in SDMs. We collected occurrence data and life history data for all the species to use as model inputs. Species occurrence data were used to identify the environmental niche of the species through correlative SDMs, whereas species life history data were employed to parameterize the mechanistic models.

For the correlative approach, we obtained the occurrence records of all the species from the Global Biodiversity Information Facility (GBIF, http://www.gbif.org) using rbgif package. Then, we removed the records with spatial and temporal errors and the duplicate records using the CoordinateCleaner package (Zizka et al., 2019). We spatially thinned the raw occurrence data by randomly selecting a presence point within a single grid cell (i.e., 10× 10 km) using the BiodiversityR package to reduce sampling
Figure 1: Conceptual diagram of correlative and mechanistic approaches of studying potential distributions for the five aphid species. The abbreviations for correlative model algorithms are defined in section 2.3, while the abbreviations for climate data are defined in section 2.3. TPC represents the temperature performance curve.
bias. Depending upon the correlative model algorithm and the number of occurrence records (Barbet-Massin et al., 2012), we used different strategies (Figure S1) to generate the pseudo-absence data for each species by using the biomod2 package (Thuiller et al., 2023). Then, we split the occurrence and pseudo-absence datasets into training sets for model fitting and test sets for model evaluation using the block cross-validation technique, implemented in the blockCV package (Valavi et al., 2019). The decision-making processes of generating pseudo-absence data, splitting training, and test sets are included in (Figure S2).

To parameterize the mechanistic SDMs which are stage-structured and temperature-dependent, we estimated the temperature-dependent vital rates (i.e., fecundity rates, developmental rates, mortality rates, and predation rates) for each species across various developmental stages from experimental studies (AMA, 2006; Embarak et al., 2020; Bai et al., 2022; Saleesha et al., 2022; Islam et al., 2022). Then, we fit the thermal performance curves (TPCs) (Supplement 1) for these vital rates using rTPC and nls.multstart packages. The rTPC package contains 24 mathematical TPC models, but does not include the TPC models we used in the mechanistic model, so we customized the functions in rTPC package and used them for estimating the TPC parameters in our model. (Padfield et al., 2021) to estimate the TPC parameters. All the estimated parameter values for the six species are listed in Table S3.

2.2 Climate data

The correlative SDMs often use bioclimate variables, which are derived from monthly temperature and precipitation data to generate more biologically meaningful variables. These are used to estimate the statistical relationship between environmental conditions and species distribution patterns. We downloaded 19 raster-based bioclimate variables to represent the historical (1970–2000) and future (2081–2100) climates from WorldClim version 2 (Fick and Hijmans, 2017). The historical climate data included in WorldClim2 was generated based on a gridded time-series dataset, which contains temperature and precipitation records from worldwide climate stations from 1901–2009 (CRU TS3.10) (Harris et al., 2014; Fick and Hijmans, 2017). WorldClim2 also includes future bioclimate variables which were derived for 23 global climate models (GCMs) and four Shared Socio-economic Pathways (SSPs): SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5, as well as for four time periods (2021–2040, 2041–2060, 2061–2080, 2081–2100). We downloaded bioclimate variables for 16 GCMs (which have available daily temperature predictions among the 23 GCMs, Table S2), and one most extreme warming scenario (SSP5-8.5) from 2081–2100 and calculated the multi-model ensemble mean using Climate Data Operator (CDO; Schulzweida, 2019) to represent the future climate data. The multi-model ensemble can reduce the uncertainty stemming from the differences in GCMs (Flato et al., 2014;
Kamworapan and Surussavadee, 2019). The spatial resolutions for both historical and future climate data are 5-minutes of a degree of longitude and latitude.

Unlike correlative SDMs which require monthly climate data, our mechanistic SDM used daily temperature data for model simulations. For the historical climate data, we downloaded the hourly 2-meter temperature for 1980–2000 from ERA5 (Hersbach et al., 2023), generated the daily mean temperature data, and interpolated this data set to a 1° × 1° resolution using CDO. For the future predictions, to keep consistent with the climate data we used for correlative SDMs, we downloaded the hindcast (1980-2000) and future (2080-2100) daily temperature predictions under SSP5-8.5 for 16 GCMs (Table S2), interpolated all the temperature data to a 1° × 1° resolution to match the resolution of the historical temperature dataset, then generated the multi-model ensembles of the 16 GCMs. By comparing hindcast and future projections, we calculated the relative change in daily temperature over 100 years and added it to the historical data to generate the calibrated future climate data. The historical and future climate datasets both include 15,478 grid cells that cover the entire global land area.

2.3 Ensemble correlative SDM

Among numerous correlative SDMs, we selected eight commonly used model algorithms provided by Biomod2 package: generalized linear models (GLM, McCullagh and Nelder, 1989), generalized additive models (GAM, Hastie and Tibshirani, 1986), generalized boosted models (GBM, Ridgeway, 1999), multivariate adaptive regression splines (MARS, Friedman, 1991), classification tree analysis (CTA, Breiman, 1984), flexible discriminant analysis (FDA, Hastie et al., 1994) and maximum entropy (MaxEnt, Phillips et al., 2006).

For each species, we applied model-specific training and test sets for presence and pseudo-absence data for different model algorithms. For each presence training set and its corresponding historical climate data, we used the variance inflation factor (VIF) to estimate the multicollinearity of the 19 bioclimate variables and excluded variables with VIF value higher than 5 (Dormann et al., 2013; Sillero et al., 2021). The remaining variables for historical climate data (1970–2000) were used for model fitting, and we used the same variables during the future period (2081–2100) for model predictions.

As stated in section 2.1, we generated multiple pseudo-absence datasets and training-test sets for each model algorithm as replicates. Each replicate was treated as an individual model simulation for these model algorithms. We evaluated each simulation by using the 10% omission rate (OR), which is the threshold value that would exclude 10 percent of training presences having the lowest predicted values, and used to
separate ‘suitable’ and ‘unsuitable’ habitats (Radomski et al., 2022). If more than 10% of presence data in a test set are in an ‘unsuitable’ habitat, we excluded this simulation to avoid overfitting the model. Then, we calculated the ensemble model mean values of the suitability scores from the remaining simulations for all the localities and used them to represent the potential distribution of each species under historical climate condition. For the future time period, we ran the remaining simulations with future climate data (2081-2100, SSP5-8.5) and predicted the future potential distribution by calculating the ensemble model mean suitability values (Figure S3).

2.4 Mechanistic SDMs

We developed a temperature-dependent mechanistic model for the aphid-ladybird predator-prey system (Ge et al., 2023). This model is stage-structured and implemented in continuous time by a coupled set of ordinary differential equations. From this, we can track the demography of aphids and ladybirds over time. We made some modifications to this model to align it with our specific requirements (Appendix A3.1). Then, we treated this model as a mechanistic SDM which includes predator-prey interactions to examine the importance of incorporating species interactions into SDMs. In this model, all the vital rates for these species are temperature-dependent. The species interaction between aphids and ladybirds is determined by the ladybirds’ attack rate and handling time, which are typically measured under laboratory conditions and would overestimate the predation rate compare to field conditions. Due to the lack of accurate estimates of the predation rate, we adjusted the magnitude of the predation rate in the model by introducing a coefficient (ξ) to reduce the magnitude of attack rate, and explored the spatial pattern of predation effects on aphid population abundance as the predation rate increases. By doing the sensitivity test of ξ (Figure S5), we choose ξ = 0.000001 to allow for the persistence for both species.

We selected five aphid species as the prey, and one ladybird species as the common predator, which forms five pairs of predator-prey species. For each species pair, we used the historical (1980–2000) and future (2080–2100) temperature profiles across the global land area (i.e., 15 478 grid cells) as model input, and simulated yearly population dynamics for aphids and ladybirds across a twenty-year time period. To ensure the simulation in each grid cell starts from the winter season, we treated July 1 (the current year) as the simulation start date and June 30 (the next year) as the ending date of a year for all the grid cells in the Southern Hemisphere, while the grid cells in the Northern Hemisphere followed the usual Gregorian calendar. The overwintering stages of aphids and ladybirds are ignored in the present model, assuming that neither species will enter diapause and both species will only face intrinsic mortality rates when the
temperature drops below their lower temperature thresholds. The simulations for each year have the same
initial condition. Each simulation began with $1 \times 10^7$ aperous aphids and $5 \times 10^4$ adult ladybird females (per
$1 \times 10^4{\text{m}}^{-2}$, split equally among all stages) introduced separately on the dates for which temperatures were
‘warm enough’ to support positive aphid and ladybird population growth rates (for five consecutive days)
respectively. This model can be also treated as one mechanistic SDM of the focus species alone by setting
the initial population abundance of ladybirds as zero.

For each grid cell, we summed the aphid’s daily population abundance to calculate the annual aphid pres-
sures ($\overline{AAP}$) under historical and future climate conditions when ladybirds are absent (historical: $\overline{AAP}_{np,h}$;
future: $\overline{AAP}_{np,f}$) or present (historical: $\overline{AAP}_{p,h}$; future: $\overline{AAP}_{p,f}$) in the model. To facilitate analysis, we
used the mean normalized aphid pressure ($\overline{AAP}$), normalized by dividing by the maximum value over all
the model runs \(^1\) and averaged these normalized values over the 20 years in each time period. This metric
represents the population growth potential of the aphid, and will be referred as aphid population abundance
throughout this Chapter. This value can be viewed as the habitat suitability value for a species, which is
comparable to the metric used in the ensemble correlative SDM.

We created heat maps of $\overline{AAP}$ under different climate scenarios and species interaction scenarios at a
global scale, demonstrating the potential distributions of the aphid and the spatial patterns of their potential
pressure on the crops (or alternative host plants) under different scenarios. By comparing the $\overline{AAP}$ under
different scenarios, we examined the effect of climate change, predation, and their interaction on the aphid
population. We defined $\overline{AAP}_{np,h}$ as the reference value of $\overline{AAP}$ under the historical and no predation sce-
nario, $\Delta \overline{AAP}_C$ to be the change in $\overline{AAP}$ due to climate change, $\Delta \overline{AAP}_P$ to be the change in $\overline{AAP}$ due to
predation, $\Delta \overline{AAP}_I$ to be the change in $\overline{AAP}$ due to an interaction between predation and climate change and
$\Delta \overline{AAP}_{CPI}$ to be the combined effect, which is the combination of the effects of climate change, predation,
and their interaction, such that,

$$\Delta \overline{AAP}_C = \overline{AAP}_{np,f} - \overline{AAP}_{np,h},$$
$$\Delta \overline{AAP}_P = \overline{AAP}_{p,h} - \overline{AAP}_{np,h},$$
$$\Delta \overline{AAP}_I = (\overline{AAP}_{p,f} - \overline{AAP}_{np,f}) - (\overline{AAP}_{p,h} - \overline{AAP}_{np,f}),$$
$$\Delta \overline{AAP}_{CPI} = \Delta \overline{AAP}_C + \Delta \overline{AAP}_P + \Delta \overline{AAP}_I,$$
$$= \overline{AAP}_{p,f} - \overline{AAP}_{np,h},$$

\(^1\)Including all the grid cells, both historical and future climate conditions, species interaction scenarios (no predation and with
predation).
where the subscripts denote: future (f) or historical (h) and no predation (np) or predation (p).

3 Results

3.1 Prediction of an ensemble correlative SDM

Based on the global occurrence records of the five aphid species (Figure 2a1-a5) and nine correlative model algorithms, we predicted the global habitat suitability values for these species and plot their potential distribution under historical (Figure 2b1-b5, HSV\textsubscript{h}) and future (Figure 2c1-c5, HSV\textsubscript{f}) climate conditions. The patterns of HSV are quite different among these species under both climate conditions. The soybean aphid has the narrowest occurrence range (Figure 2a1). The cotton aphid has the widest occurrence range, with its southern boundary reaching further south compared to other species (Figure 2a2), while the northern boundary for the English grain aphid reaches further north (Figure 2a5). As a result, our predictions show that the habitat suitability values for the cotton aphid are greater in the Southern Hemisphere compared to other species (Figure 2b2 and c2). Meanwhile, the values of HSV for the English grain aphid are significantly higher in the Northern Hemisphere (Figure 2b5 and c5). In contrast, the soybean aphid should have significantly lower habitat suitability values in most regions, particularly in the Southern Hemisphere (Figure 2b1 and c1). For all species, regions with higher HSVs align closely to regions with concentrated occurrence records.

Additionally, we plot the differences between HSV\textsubscript{f} and HSV\textsubscript{h} to demonstrate the spatial pattern of climate change impacts on species habitat suitability (Figure 2d1 and d5). With climate warming, the habitat suitabilities for all five aphid species will increase in high-latitude regions, such as Canada and northern Asia, and decrease in low and mid-latitude regions. In particular, the extent and strength of climate change impacts vary among species. Climate change should have more widespread impacts on the cotton aphid’ suitability (Figure 2d2) compared to the other species, while its effects are relatively more severe on the green peach aphid and the bird cherry-oat aphid (Figure 2d3 and d4).

3.2 Predictions of mechanistic SDMs

Figure 3 shows the global patterns of aphid population abundance (AAP) under different conditions. Similar to HSV, the color depths of AAP indicate the habitat suitability for aphid species. Under historical climate and with the absence of ladybirds in the model, temperature is the sole factor influencing the magnitude
Bird cherry-oat aphid (Rhopalosiphum padi)

Occurrence records on GBIF website

Figure 2: Global occurrence records and predicted patterns of habitat suitability values (HSV) for five aphid species under historical (1971-2000, HSV_h) and future (2081-2000, HSV_f) climate conditions. Each row displays plots for each aphid species. (a1)-(a5) represent the global occurrence records for the five aphid species obtained from the GBIF website. (b1)-(b5) and (c1)-(c5) demonstrate the global potential distributions for aphid species under historical and future climate conditions, with color depths indicating habitat suitability value magnitudes. (d1)-(d5) demonstrate the climate change effects on global habitat suitability for aphid species, with color depths indicating the magnitudes of ΔHSV (HSV_f − HSV_h).
of aphid population abundance ($\overline{\text{AAP}}_{\text{np,h}}$) worldwide. Among the five aphid species, we predict distinct differences in their suitability at low latitudes (Figure 3a1-a5). Both of the soybean aphid and the cotton aphid are predicted with the highest suitability in these regions (Figure 3a1 and a2), whereas the English grain aphid should have a much lower suitability, particularly in areas such as North Africa, Brazil, and Northern Australia (Figure 3a5).

Under climate change, we predict aphid abundance increases at the ‘cold edge’ of the species range (higher latitudes and elevations) and population decrease in abundance at the ‘warm edge’ (low latitudes and elevations), which can be seen clearly in North Africa (Figure 3c1-c5). Generally, climate change tends to decrease species’ suitability (i.e., a negative effect) in low latitudes while enhancing suitability (i.e., a positive effect) in mid and high latitudes (Figure 4a1-a5). However, the strength and spatial pattern of climate change effects varies among species. For instance, the English grain aphid experiences the most severe negative impact in Central Africa (Figure 4a5), while the soybean aphid faces a weaker negative effect, primarily occurring in North Africa (Figure 3a1). Notably, the magnitude of negative climate effects is typically larger than that of positive climate effects.
Figure 4: Global patterns of climate effects, predation effect, and their interacting effects, and combined effects on aphid population abundances. (a1)-(a5) demonstrate the effect of climate change, $\Delta \Delta \text{AAP}_C = \Delta \text{AAP}_{np,f} - \Delta \text{AAP}_{np,h}$ (Equation 1). (b1)-(b5) demonstrate the effect of predation (caused by ladybirds), $\Delta \Delta \text{AAP}_C = \Delta \text{AAP}_{p,h} - \Delta \text{AAP}_{np,h}$ (Equation 2). (c1)-(c5) demonstrate the interaction between climate and predation, $\Delta \Delta \text{AAP}_I = (\Delta \text{AAP}_{p,f} - \Delta \text{AAP}_{np,f}) - (\Delta \text{AAP}_{p,h} - \Delta \text{AAP}_{np,h})$ (Equation 3). (d1)-(d5) show the combined effect, $\Delta \Delta \text{AAP}_{CPI} = \Delta \Delta \text{AAP}_C + \Delta \Delta \text{AAP}_P + \Delta \Delta \text{AAP}_I$ (Equation 4). Each row displays plots for each aphid species. The color depths indicate the magnitudes of these effects.
Not surprisingly, predation should have a negative effect on AAP in most of the habitable regions for aphids (Figure 3b1-b5 and Figure 4b1-b5). The effect of predation varies spatially, with the strongest effects occurring in low latitude areas. Furthermore, the strength of predation varies among the five species studied. For instance, predation has a stronger effect on the soybean aphid and the green peach aphid (Figure 3b1 and b3) but a weaker effect on the English grain aphid (Figure 3b5).

The interaction effects of climate change and predation (ΔAAP) differ geographically, exhibiting stronger positive effects in lower latitudes, and weaker and narrower negative effects in middle and higher latitudes (Figure 4c1-c5). High latitude areas rarely exhibit such interactions due to the weak predation effects in these regions. The strength of the interaction also varies among the five species, with the green peach aphid exhibiting stronger interactions in Brazil (Figure 4c3), and the English grain aphid demonstrating the most significant interactions in Central Africa (Figure 4c5).

Overall, the effects of climate change, predation, and their interaction demonstrate distinct spatial patterns and vary across the five species. Their combined effects (ΔAAPCP) are primarily negative in low latitudes but positive in the mid and high latitudes of the Northern Hemisphere (Figure 4d1-d5). The positive climate effects in the middle and high latitudes are somewhat mitigated by negative interacting effects, while the negative effects of climate change and predation in low latitudes are partially offset by positive interacting effects. Of the five species, the green peach aphid and English grain aphid (Figure 4d3 and d5) exhibit more pronounced responses to climate change and predation compared to the other species, which are largely attributable to their greater sensitivity to climate change (Figure 4a3 and a5).

3.3 Predictions comparison between ensemble correlative SDM and mechanistic SDMs

To evaluate the predictions of the correlative and mechanistic SDMs, we calculated the differences in habitat suitability predictions between the two approaches under both historical and future climatic conditions when predators are absent (Figure 5) or present (Figure 6) in the model. The difference plots demonstrate noticeable differences in the predictions of the two approaches under both time periods for all five aphid species. Notably, for the soybean aphid, which has the fewest occurrence records (Figure 2a1), the differences between the AAP and HSV that are predicted by the two approaches are the most pronounced. Conversely, the cotton aphid, which has the most occurrence records (Figure 2a2), should have the smallest differences in AAP and HSV under historical climate condition. The variation among the five species suggests that sample bias in the occurrence records could potentially impact the accuracy of the predictions.

Across all five species, the AAPnp,h and AAPp,h values are generally greater than HSVh values in most
areas, notably in South America, Africa, and Australia. However, there are subtle differences between the two approaches in the regions with abundant occurrence records (e.g., the south of the United States, Europe). HSV values are greater than $AAP_{np,h}$ and $AAP_{p,h}$ values. With climate change, the differences between the two approaches’ predictions become smaller. Compare the climate effects on aphid habitat suitability projected by the correlative and mechanistic approaches when predators are absent (Figure 5), we see that the positive value of $AAP_{p,f} - AAP_{p,h}$ in the Northern Hemisphere (Figure 4a1-a5) is larger than $HSV$, where the negative value of $AAP_{p,f} - AAP_{p,h}$ in the Southern Hemisphere is larger than $HSV$. This indicates that the mechanistic SDM generally predicts stronger positive and negative climate effects on aphid habitat suitability compared to the ensemble correlative SDM. By using the correlative model predations as a reference to compare the effect of predation on species’ responses to climate change (Figure 5c1-c5 and Figure 6c1-c5), we predict that predator-prey interaction solely affects the relative abundance of aphid species and does not induce their presence or absence.

4 Discussion

In the preceding section, we described and compared the predictions generated by the ensemble correlative SDM and the two mechanistic SDMs. In this section, we reflect on the three questions listed in the introduction section in light of our results.

4.1 How similar are the results from the two approaches?

Our study indicates there are differences in the geographic patterns for all five aphid species predicted by the two different approaches. The ensemble correlative SDM tends to predict higher habitat suitability values in regions with a larger number of occurrence records. Conversely, the mechanistic SDM typically predicts higher values across most other regions. The variation in the five species in terms of the model prediction difference indicates the potential impacts of sample size and sample bias arising from the correlative SDM approach.

Differences in habitat suitability predictions of the five aphid species generated by the ensemble correlative SDM were driven by differences in presence data and their associated realized climate niches (Figure S4). The soybean aphid displays the narrowest niche, while the cotton aphid has the broadest niche. Consequently, the potential global distribution for the soybean aphid is the most restricted, while the English grain aphid should have the widest potential distribution (Figure 2). The green peach aphid and the bird cherry-oat aphid
Figure 5: Prediction comparisons between an ensemble correlative SDM and a mechanistic SDM without predator-prey interactions. (a1)-(a5) show the difference between aphid population abundance predicted by the mechanistic SDM (AAP_{np,h}) and habitat suitability value predicted by the ensemble correlative SDM (HSV_h) under historical climate conditions. (b1)-(b5) show the difference between AAP_{np,f} and HSV_f (future climate). (c1)-(c5) show the climate effects on aphid habitat suitability predicted by correlative and mechanistic SDMs when predators are absent. Each row presents plots for each aphid species, with color depths indicating the magnitudes of the difference values.
Figure 6: Prediction comparisons between an ensemble correlative SDM and a mechanistic SDM with predator-prey interactions. (a1)-(a5) show the difference between aphid population abundance predicted by the mechanistic SDM (AAP<sub>p,h</sub>) and habitat suitability value predicted by the ensemble correlative SDM (HSV<sub>h</sub>) under historical climate condition. (b1)-(b5) show the difference between AAP<sub>p,f</sub> and HSV<sub>f</sub> (future climate). (c1)-(c5) show the climate effects on aphid habitat suitability predicted by correlative and mechanistic SDMs when predators are present. Each row presents plots for each aphid species, with color depths indicating the magnitudes of the difference values.
share the most similar niches among the five species, resulting in the closest projected distributions for them. These variations in niche breadths have a direct impact on the projected suitable habitats for each species. Our predictions indicate the important role of quantity and quality of species presence data, which is a widely recognized characteristic of correlative SDMs (Aubry et al., 2017; Gaul et al., 2020; Radomski et al., 2022).

In terms of the mechanistic SDM, the predictions are determined by species thermal performances. The five species have different thermal performance curves, resulting in differences in their fundamental niches, which ultimately influence the geographical patterns of population abundance for each species. The English grain aphid exhibits the lowest tolerance of high temperatures and has a relatively lower suitable temperature range (AMA, 2006), so it is predicted with relatively lower habitat suitability values in low latitudes (Figure 2a5).

Overall, the predictions generated by the two model approaches are comparable, but they employed distinct data requirements and adopted different modeling techniques with distinct modeling philosophies. In correlative SDMs, we obtained monthly environmental variables over a specific time period along with species occurrence records as inputs to the model. Subsequently, we employed various model algorithms to estimate the statistical relationships between the species’ realized niche and environmental conditions. These models are focused on observing patterns, rather than providing explanations for underlying mechanisms (Tourinho and Vale, 2023). Conversely, the mechanistic SDMs didn’t use species occurrence records. Instead, they incorporated continuous climate data and climate-dependent life history traits, enabling the species to respond to varying climate conditions over time. These models explicitly simulate the biological processes of each life stage of the focus species, allowing us to explore the mechanisms that affect the species’ habitat suitability.

4.2 What does adding biotic interactions to the model get us?

Our mechanistic SDM of predator-prey interactions accounts for detailed biological processes, enabling us to examine the role of abiotic and biotic factors, as well as their interactions, in the context of climate change. Our predictions demonstrate predation and interaction between climate change and predation can strongly affect the population abundance (i.e., habitat suitability) of aphids (Figure 4b1-b5 and c1-c5). These effects are sensitive to the parameters that govern ladybirds’ predation rate. For instance, the predation effects become stronger as we increase ladybirds’ attack rate (Figure S6). Additionally, the predation effect on each species is different even though all aphid species are assumed to have the same predators, with the same functional and numerical responses.
Our results therefore contradict the Elton’s Noise Hypothesis (ENH), which suggests that biotic interactions have a limited influence on determining species distributions across extensive geographic areas (Soberón and Nakamura, 2009; Palacio and Girini, 2018). The importance of biotic interactions and their interactions with abiotic factors in determining species’ habitat suitability emphasizes the necessity of incorporating biotic interactions to enhance model accuracy.

4.3 Is the added effort worth it?

While correlative SDMs provide estimates of the realized niche, taking into account species interactions and accessibility implicitly (Tourinho and Vale, 2023), it remains difficult to separate and quantify the individual impacts of these factors as we can in mechanistic SDMs. In addition, the model accuracy of correlative SDMs is limited by the sample size and sample bias of the presence data. For poorly known or rare species, a sufficient quantity of occurrence records (i.e., sample size) still remains one of the biggest challenges for modeling their distributions (Guisan et al., 2017; Gaul et al., 2020). Small sample size affects model accuracy, particularly when there are fewer than 30 records (Wisz et al., 2008). Sample bias is another important challenge for correlative models, affecting species with both limited and abundant occurrences. Incomplete sampling and over-sampling can both contribute to unquantifiable spatial bias within presence data, which in turn leads to an inaccurate representation of the species’ realized niche and adversely affects the model accuracy (Inman et al., 2021).

The mechanistic SDMs used in this study do not exhibit the above issues. However, it is crucial to recognize that this approach is more data-demanding and time-consuming than the correlative SDM approach, depending on the complexity of the model (Kearney and Porter, 2009; Dormann et al., 2012), as well as subject to other limitations, such as not accounting for geographic variation in the TPCs or evolution of the TPCs. Maintaining the interpretation of a mechanistic model while including multiple aspects can be challenging. The key is to include features that are informative for a model’s goals. For example, a goal of this paper was to assess how predation affects aphid distributions using correlative SDMs as a point of reference. Figures 5 and 6 indicate the effects of predation decrease the magnitude of the habitat suitability, but didn’t alter the geographic pattern. This is informative because it indicates that while predation induces a numerical shift in the abundance of aphids, it did not induce a shift in aphids’ presence/absence.

We suggest that asking whether the added effort of a mechanistic model is worth the investment is misguided because a correlative model in the absence of the independent assessment of a mechanistic model lacks robustness. The mechanistic SDMs we implemented in this study serve as a valuable addition to correlative
SDMs. The decision to invest efforts to construct a mechanistic SDM should be made on a case-by-case basis, recognizing that one will have greater confidence if both model results are similar and have less confidence if they do not.

4.4 Future directions: hybrid model

Our goal wasn’t to assess the model accuracy, so we focused only on comparing the results from the models, without integrating the two approaches. Recently, an increasing number of studies have recommended the integration of the two approaches, rather than just simply comparing them (Meineri et al., 2015; Peterson et al., 2015).

The combined approach (i.e., hybrid model) is most valuable if the study objective is to make predictions. By incorporating prior information on the relationship between environments and species’ performance provided by a mechanistic model, the predictive power of the correlative SDM can be enhanced. This approach is especially valuable in non-static states when environmental conditions undergo rapid changes, or when the ecological niche evolves in response to these environmental shifts (Guillaumot et al., 2022). Nevertheless, this approach may face some challenges when species do not inhabit their entire fundamental niche. Predictions from mechanistic models are occasionally considered lacking in information value since they estimate the fundamental niche, which often includes many regions that are unoccupied by the species (Newman et al., 2022). However, combining the two approaches encompass the complexities of both of them, making it difficult to understand the underlying mechanisms, which increases the risk of being misinterpreted (Tourinho and Vale, 2023). Moving forward, observations of species abundance at multiple time points may facilitate the integration of correlative and mechanistic approaches. All together, the optimal approach depends on the objective of the study, specifically whether its primary focus is on model predictions or underlying mechanisms.

5 Conclusions

In our study, we applied a stage-structured predator-prey population dynamics model as a mechanistic SDM to project global patterns in the population abundance of five aphid species in response to climate change. This modeling approach allowed us to examine the impacts of climate change, predation, and their interactions on the distribution and population abundance of aphids across various geographical regions, thereby emphasizing the importance of considering biotic interactions and evaluating the interplay between abiotic and biotic
factors when making projections. Furthermore, we utilized multiple correlative model algorithms to generate predictions based solely on abiotic factors. Our comparisons of the predictions from the mechanistic model and the correlative models reveal large differences in model outcomes, underlining the crucial influence of model philosophies in SDM studies.

References


Godsoe, W., Franklin, J., and Blanchet, F. G. (2017). Effects of biotic interactions on modeled species’

84(11):2809–2815.

(2022). Classic or hybrid? the performance of next generation ecological models to study the response of
southern ocean species to changing environmental conditions. *Diversity and Distributions*, 28(11):2286–
2302.

applications in R*. Cambridge University Press.


https://doi.org/10.1214/ss/1177013604.

the American statistical association*, 89(428):1255–1270.

Climate change drives poleward increases and equatorward declines in marine species. *Current Biology*,

Hersbach, H., Bell, B., Berrisford, P., Biavati, G. and Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey,
hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data
Store (CDS).


