

## Review



**Cite this article:** Lasky JR, Hooten MB, Adler PB. 2020 What processes must we understand to forecast regional-scale population dynamics? *Proc. R. Soc. B* **287**: 20202219.  
<https://doi.org/10.1098/rspb.2020.2219>

Received: 8 September 2020

Accepted: 12 November 2020

**Subject Category:**

Global change and conservation

**Subject Areas:**

ecology, evolution

**Keywords:**

macroecology, local adaptation, climate change, range dynamics

**Author for correspondence:**

Jesse R. Lasky

e-mail: [jrl35@psu.edu](mailto:jrl35@psu.edu)

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5230324>.

# What processes must we understand to forecast regional-scale population dynamics?

Jesse R. Lasky<sup>1</sup>, Mevin B. Hooten<sup>2,3,4</sup> and Peter B. Adler<sup>5</sup>

<sup>1</sup>Department of Biology, Pennsylvania State University, University Park, PA, USA

<sup>2</sup>U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit,

<sup>3</sup>Department of Fish, Wildlife, and Conservation Biology, and <sup>4</sup>Department of Statistics, Colorado State University, Fort Collins, CO, USA

<sup>5</sup>Department of Wildland Resources and the Ecology Center, Utah State University, Logan, UT, USA

JRL, 0000-0001-7688-5296

An urgent challenge facing biologists is predicting the regional-scale population dynamics of species facing environmental change. Biologists suggest that we must move beyond predictions based on phenomenological models and instead base predictions on underlying processes. For example, population biologists, evolutionary biologists, community ecologists and ecophysiologicalists all argue that the respective processes they study are essential. Must our models include processes from all of these fields? We argue that answering this critical question is ultimately an empirical exercise requiring a substantial amount of data that have not been integrated for any system to date. To motivate and facilitate the necessary data collection and integration, we first review the potential importance of each mechanism for skilful prediction. We then develop a conceptual framework based on reaction norms, and propose a hierarchical Bayesian statistical framework to integrate processes affecting reaction norms at different scales. The ambitious research programme we advocate is rapidly becoming feasible due to novel collaborations, datasets and analytical tools.

## 1. Overview

From systems biology [1] to ecosystem ecology [2], researchers claim that (1) prediction is an acid test of our understanding of biology and (2) the best predictions will come from models rooted in the processes that generate system-wide patterns. This is a growing emphasis among organismal biologists, who often argue that prediction of population- and species-level dynamics should be grounded in genetics, physiology and community ecology [3–7].

A key challenge facing contemporary organismal biologists is predicting or forecasting regional-scale population dynamics under environmental change. We use the terms ‘prediction’ and ‘forecast’ interchangeably, defined as expected states of nature in the future. Reliable predictions could guide biodiversity conservation under future environments [8–10]. For example, a model capable of predicting the dynamics of a threatened or invasive species across a large region could identify future critical habitat or help managers allocate resources for eradication. These predictions have been traditionally based on site-specific population models or large-scale, phenomenological distribution models [11]. However, these approaches are frequently criticized [9,12,13], whereas a number of prominent recent reviews have made the case for models based on lower-level mechanisms [3,5–7]. Phenomenological models are based on empirical correlations between environment and occurrence or abundance without explicit functions representing lower-level biological processes. By contrast, mechanistic approaches have explicit representations of how the environment interacts with genotypes, phenotypes or demography.

But which processes are essential? Many population biologists believe that useful forecasts are impossible without accounting for demographic mechanisms

[14]. Evolutionary biologists argue that models must consider genetic variation and adaptation [3,15]. Community ecologists focus on species interactions [4,5]. Organismal ecologists argue that physiology [6,16–18] or behaviour [19,20] are essential. Must our models include all of these processes?

We suggest that this is an empirical question that can only be answered by making predictions with competing models and validating them with independent observations. However, testing the predictive value of so many processes requires a programme of data collection and integration that, to our knowledge, has never been completed for any system. Our purpose is to motivate and facilitate such an effort. We begin by reviewing the arguments for building multiple mechanisms into forecasts of regional-scale dynamics. We then develop a conceptual framework for integrating diverse mechanisms, and propose a statistical framework to test their impact on predictive skill while accommodating uncertainty.

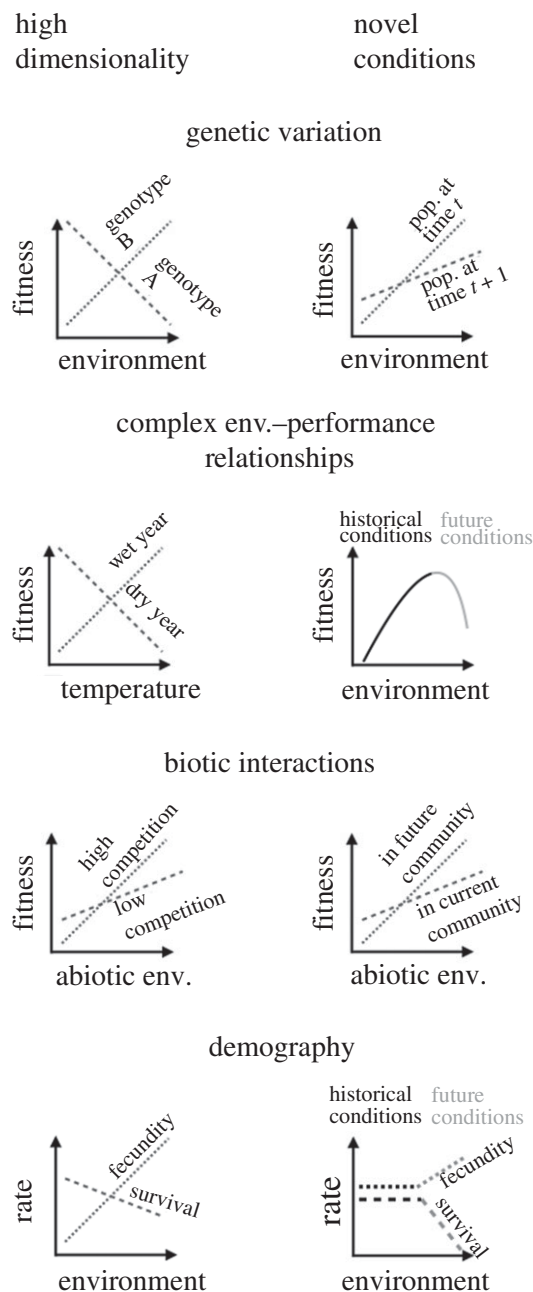
## 2. Toward a process-oriented approach

### (a) Shortcomings of a phenomenological approach

Population predictions based on a purely phenomenological representation of biological systems may lack fidelity under novel conditions or when the underlying drivers are high dimensional. In the first case, the functional form of responses to environment may change as organisms encounter conditions not previously experienced, or as individuals develop trait values and trait combinations not previously observed [21,22]. These novel conditions may arise in space (e.g. for an invasive species) or in future times (e.g. under climate change). Note that forecasts have specific temporal scales: many ecological forecasts apply to the next few years, or to the next few decades. A consideration of novel conditions may be particularly problematic for long-term forecasts as conditions with no contemporary analogue become more common.

In a variety of fields, such as epidemiology of infectious disease, fitting curves to past observations and extrapolating to predict the future has yielded poor results as conditions change [23,24]. In principle, mechanistic models avoid these problems by capturing the processes that lead to novel responses to environment. It is important to note that we focus on approaches that both model underlying processes *and* include lower-level data on these processes. Mechanistic approaches that lack such data may suffer from problems with identifiability, for example, when high-level abundance data can be reproduced by models with very different mechanisms [25,26]. Therefore, a crucial feature of the approach we present below is that mechanistic models are combined with data on underlying mechanisms.

High-dimensionality limits the feasibility of data-driven phenomenological approaches to prediction. When true functional forms are unknown and there are many potentially important and interacting independent variables, it is difficult to collect sufficient high-quality data to describe the true, complex relationships driving regional population dynamics. When multiple population drivers have opposing effects or opposing patterns of variation in nature, observational studies can easily miss these signals. This dimensionality is especially problematic when observations are made at a scale where aggregation obscures the effects of important processes [27–29]. A mechanistic understanding can identify the most



**Figure 1.** Examples of how underlying biological complexity may confound regional population forecasting, due to high dimensionality of the systems or the systems entering novel conditions. For each example, ignoring the complexity shown here could lead to major inaccuracies in forecasts. In the ‘high dimensionality’ examples, relationships between environment and fitness are highly dimensional due to the four different categories of processes. In the ‘novel conditions’ examples, the relationship between environment and fitness observed in contemporary populations change under novel conditions related to the same four categories of processes.

important aspects of organisms, environments and scales of observation. In the following sections, we describe examples of the potential benefits of process-oriented approaches, focusing on four categories of biological complexity that can confound regional population forecasts (figure 1): genetic variation, complex environment–performance relationships, biotic interactions and demography.

### (b) Genetic variation

Genetic variation within species is often overlooked in population, community and macro-ecology. Models and predictions of

species range dynamics or regional population forecasts typically assume a single set of parameters for a species [30].

In reality, species often exhibit intraspecific genetic variation in response to their environment. Evolutionary biology offers numerous examples of genetic variation in trait and demographic responses to environmental variation, such as for abiotic stress response [31–33], host–pathogen interactions [34] and predator–prey interactions [35,36]. Spatially varying environments can favour different traits in different places and, as a result, much of the genetic variation in environmental responses is geographically structured. Thus, individual fitness and population growth rates often respond to the environment in different ways in different locations. Large-scale forecasts typically ignore these differences.

If populations differ genetically in environmental responses, then accurate regional population forecasts may require accounting for genetic variation, as well as genetic change over time (i.e. evolution; figure 1). For example, forecasts of regional population dynamics might assume individuals have as broad an environmental tolerance as the species in aggregate, while in reality locally adapted populations have narrower tolerances [30]. In this case, even though populations suited to future environments might exist within a species' range, specific alleles may not occur in the location where they would be adaptive. As a result, forecasts under environmental change might be overly optimistic and underestimate the threat of extinction [37].

Researchers have begun to incorporate genetic variation into regional population prediction. Some researchers have fit models that identify genotypes associated with specific current environments, assuming local adaptation, and then used predicted future environments to assess mismatch between local genotypes and their future environments [38–41] (reviewed by Capblancq *et al.* [15]). Additionally, researchers have fitted environmental response models using geographically restricted subsets of populations, presumably accounting for geographic genetic variation, to project distributions under future conditions [42–44].

Without rigorous validation, predictions have unknown value. Important hurdles for predictions are success at out-of-sample prediction, and prediction into parameter space where data are sparse. After all, these models are being developed for prediction in a future of novel conditions. Recently, genetic models of environmental response and adaptation have been used in out-of-sample predictions of individual or population performance in response to environmental stressors in both wild and agricultural species [38,41,45–48]. These have generally had modest success, for example, in predicting relative genetic variation in change in performance [46,47] or population change [38] with accuracy (predicted versus observed correlation)  $\approx 0.3$ . Messina *et al.* [48] were able to predict relative genetic variation in performance in novel (non-test) environments, using a combined genomic–physiological model, to  $r \approx 0.6$ .

Incorporating genetic variation in regional population forecasts presents several challenges. First, statistical interactions among genetic loci can be important, but may be unknown *a priori* and too high-dimensional to discover with data-driven approaches. In crop systems, where there have been major advances in using genetics to predict environmental responses, models incorporating genetic variation have largely treated genetics as phenomenological [49]. Opportunities may exist to improve forecasts with largely phenomenological treatment of genetic effects, but focused on capturing some important

aspects of the underlying genetic mechanism. For example, predictions may be based on subsets of genetic loci (e.g. gene regulatory loci) most likely to affect ecologically important traits and environmental responses [50–52].

Second, novel population genetic compositions will evolve through new mutations, gene flow, selection and drift. Some of these processes are particularly challenging to model [30]. For example, gene flow is often poorly observed, and can result in major shifts in genetic composition of populations, potentially alleviating maladaptation to future environments [53]. Predictive modelling of drift may require demographic data (see below). Given the rate of global environmental change and the high standing variation often observed in traits related to abiotic stress response [54], new mutations are likely to be of lesser importance for regional dynamics (although this may not be true at expanding range margins, due to drift) [55]. Selection can be modelled using experimental data relating fitness to genotype [33], although indirect approaches may be required for species not amenable to experimentation [56,57].

### (c) Complex environment–performance relationships

Biologists often study how individuals, populations or species respond to a single environmental gradient, or perhaps a small set of environmental factors, along with their pairwise interactions or linear combinations [58,59]. But interactions among environmental drivers are more complex: precipitation, temperature and soil characteristics determine soil moisture; snow cover and air temperature determine winter soil temperature; solar radiation, albedo, wind and air temperature determine the energy balance of ectothermic plants and animals. Additional complexity arises as organismal trait changes in response to environment moderate the effects of potential stressors. For example, when the timing of developmental transitions (phenology) is sensitive to environmental thresholds, organisms can show sharp trait changes [60]. Animals can use behaviour to avoid thresholds beyond which abiotic conditions can become dangerous [6]. As a result, individual fitness responses to environment are often nonlinear and non-monotonic, as in mortality caused by freezing temperatures, rapid nonlinear declines in performance at high temperatures or hydraulic failure under moisture stress in plants [42,61].

It is difficult to detect thresholds and complex interactions among measured environmental variables, especially given the noisy relationships between those variables and abundance or fitness components. The potential model space to explore is large. While machine learning approaches have been used to predict species or genotype geographical distributions based on complex environmental interactions [62,63], they are not generative models and cannot be easily customized for specific studies, data types and questions. Machine learning approaches can sometimes excel at prediction when they are used in the exact setting they were designed for, but they lack the ability to formally incorporate mechanisms via thoughtfully developed simulation models (e.g. mathematical models, agent-based models, virtual ecologist models [64]). Furthermore, nonlinearities that only emerge under novel conditions may thwart extrapolations of phenomenological relationships. Forecasting models that do not account for thresholds and interactions might not predict large changes in populations moving across thresholds under future environments (figure 1).

Direct observation of individual responses to changing environments, especially under controlled conditions, may



provide more precise information for building this complexity into predictive models [6,19,65]. Complexity may be partly captured in models by relating measured environmental parameters to conditions directly experienced by organisms, such as an organism's body temperature (as opposed to the air temperature in most gridded datasets). Additionally, this complexity may be captured by modelling environmental effects based on developmental, physiological or behavioural first principles, such as energy budget [66] or stem hydraulics models [67]. A mechanistic understanding of these responses might allow prediction of responses to extreme events that are rarely observed even in long-term observational studies.

A major challenge is that these approaches require many detailed observations or background information on the most important aspects of the environment for organisms. For many species and populations this information currently does not exist and will require extensive study of phenology, physiology or behaviour. Even in crops, which are among the most well-studied systems from an ecophysiological and developmental perspective, models struggle to effectively predict across multiple dimensions of novel environmental conditions [49]. Additionally, the fine-scale data required to describe specific aspects of environment that impact fitness might be challenging to collect, model and predict at large scales [61].

#### (d) Biotic interactions

Studies of population response to the abiotic environment and forecasts of large-scale dynamics typically overlook biotic interactions even though they can have large effects on individual fitness and population growth. For example, species with strong interactions (e.g. host–parasite, predator–prey) might have tightly coupled population dynamics, obscuring the effects of fluctuating abiotic conditions. Additionally, a given species might show strong population responses to local gradients in community composition, as in the case of an early or late successional specialist. Biotic constraints such as resource competition may limit organisms' ability to take advantage of favourable conditions, such as greater resource supply (e.g. higher rainfall for plants) [68,69]. Abiotic conditions may determine the nature of species interactions, such as when stress shifts the balance between competition and facilitation [70]. The effects of abiotic change on populations may be constrained and mediated by these biotic interactions.

Incorporating biotic interactions in predictive models is challenging because of the high dimensionality of community ecology (i.e. the number of species and higher-order interactions among them) and the potential for no-analogue future assemblages (figure 1). When species interactions have a strong influence on population growth, shifts in community composition caused by abiotic environmental change can offset or even overwhelm the direct effects of abiotic change [5,71,72]. Unravelling these direct and indirect effects would be straightforward if the effects of abiotic factors and biotic interactions on demographic rates were independent and additive, but we have good reasons to expect complex interactions. For example, abiotic change might reduce fitness of a superior competitor more than an inferior competing species, leading to a population increase of the inferior competitor despite less favourable abiotic conditions [73]. Studies often rely on interannual climate fluctuations to observe abiotic–biotic interactions, but when future abiotic conditions are novel, communities may also enter novel states, thwarting prediction of a focal species's population dynamics based on recent observations (figure 1). Furthermore,

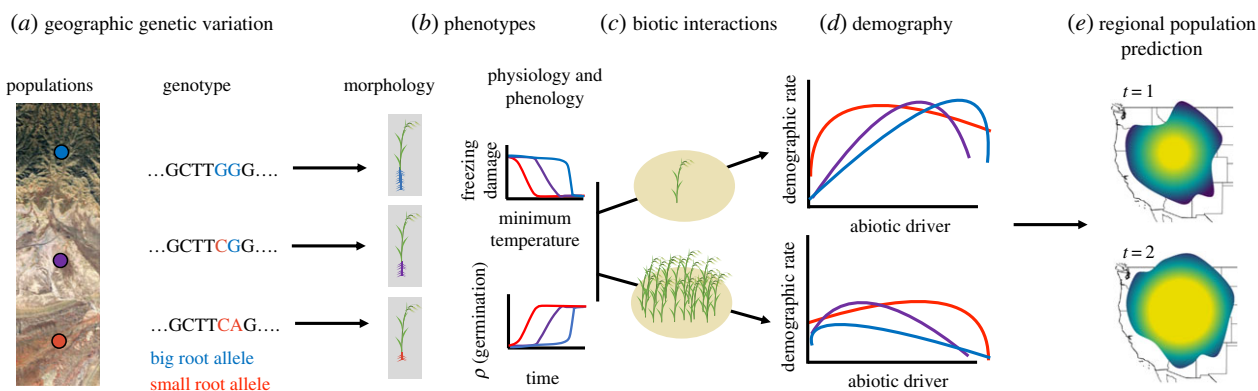
direct human impacts on populations will generate novel communities and biotic interactions [74]. The potential for such interactions illustrates the need for experiments that manipulate both abiotic conditions and biotic interactions.

Some researchers have fit joint models of community members' responses to abiotic conditions while including interactions among species [4,75,76]. However, such an approach requires large amounts of data, and it remains unclear how these efforts will succeed at forecasting under environmental change when models are fitted to noisy observational data [77,78]. A major reason there has been slow empirical progress in this area is the challenge in experimentally manipulating multiple community members (a problem of high dimensionality). Mechanistic models of species interactions are one potential way to include these interactions in predictive models [5,79]. Alternative approaches focus on forecasts of community- or ecosystem-wide responses to environmental change [2]. These aggregate over the many dimensions of community variation, but may offer little information about dynamics of individual species of interest.

#### (e) Demography

Simple, linear and phenomenological models that relate environment to abundance might work well for organisms with very simple life cycles. For organisms with more complex life cycles, ignoring differences in how individuals of different ages, life stages, sizes or sexes respond to the environment may be problematic (figure 1). When these sources of heterogeneity are important, models need to account for how individual vital rates (e.g. survival, growth and fecundity) vary as a function of both demographic state (e.g. age, size, sex) and the environment [80].

To appreciate why accounting for the interactive effects of population structure and the environment is important, consider a phenomenological approach that relates local abundance of a long-lived plant directly to an environmental driver. If the population is well established, it is likely to be dominated by large, mature individuals. Year-to-year variation in abundance will reflect the growth and mortality of these large individuals, and contributions from rare recruitment events will be small. If the mature individuals are stress-tolerant, we might find only a weak correlation between drought and population growth rate. Extrapolating this relationship to project the potential impacts of increasing aridity under climate change would predict minimal impacts on the population. But what if seedlings are less stress tolerant, and recruitment is limited to infrequent cool, wet years [81]? After disturbance of the established stand, the population might have difficulty recovering under future, more arid, conditions. By contrast, a model that accounts for interactions between size or stage structure and the environment could correctly predict both the limited impacts of increasing aridity on a population dominated by mature individuals, as well as a reduced ability of the population to recover following disturbance (when population growth depends most on recruitment). This hypothetical example shows how traditional demographic analyses, such as matrix population models [82] and integral population models [83], can handle both high dimensionality (interactions between the environment and multiple vital rates) and novel conditions (changes in stage structure and environment).



**Figure 2.** Reaction norms (figures with coloured curves) occupy a middle position in a hierarchy of biological complexity among the mechanisms discussed here. Thus, reaction norms can be viewed as a conceptual basis for building integrative models of regional population dynamics. We show a hypothetical plant example. Predicting regional dynamics (e) may require understanding (a) geographical patterns of genetic variation in environmental response, (b) the phenotypes involved and complex environment–performance relationships (such as physiological thresholds), (c) biotic interactions such as intraspecific density dependence or (d) individual demographic rate responses to environment. (Online version in colour.)

While the existing tools for modelling demography and population growth are sophisticated, they have almost always been applied to just one local population at a time. Scaling up to multiple populations along complex environmental gradients represents a much greater challenge. Fortunately, many demographers are now focused on this task, using approaches described as ‘landscape demography’ [84–86] or ‘dynamic range models’ [87–89]. These approaches incorporate differences among populations in environmental context and stage-structure, but have largely ignored genetic or phenotypic variation among populations and species interactions.

### 3. How do we determine which of these processes are most important for prediction?

#### (a) Including all processes in a model may not improve prediction

We have reviewed compelling arguments for why each of the processes described in §2 must be considered in a regional-scale population projection. However, a critical point that this body of the literature has ignored is that including all of these processes in a model may not improve prediction.

Uncertainty in predictive modelling can be divided into two components: bias, caused by incorrect and overly simple models, versus variance, due to low numbers of observations relative to the number of estimated parameters [90]. While it may appeal to biologists to create models that incorporate ever more processes in greater realism, increasing model complexity can increase errors resulting from poorly estimated parameters, lead to overfitting and decrease interpretability (making it harder to troubleshoot models). Optimal models for population prediction must balance bias against variance, and optimal models are often surprisingly simple [91,92]. These problems have already presented themselves in population forecasting, but there has not been a reckoning for the complex models that would be the logical extension of the current movement to incorporate all the processes in §2.

#### (b) Comparing models

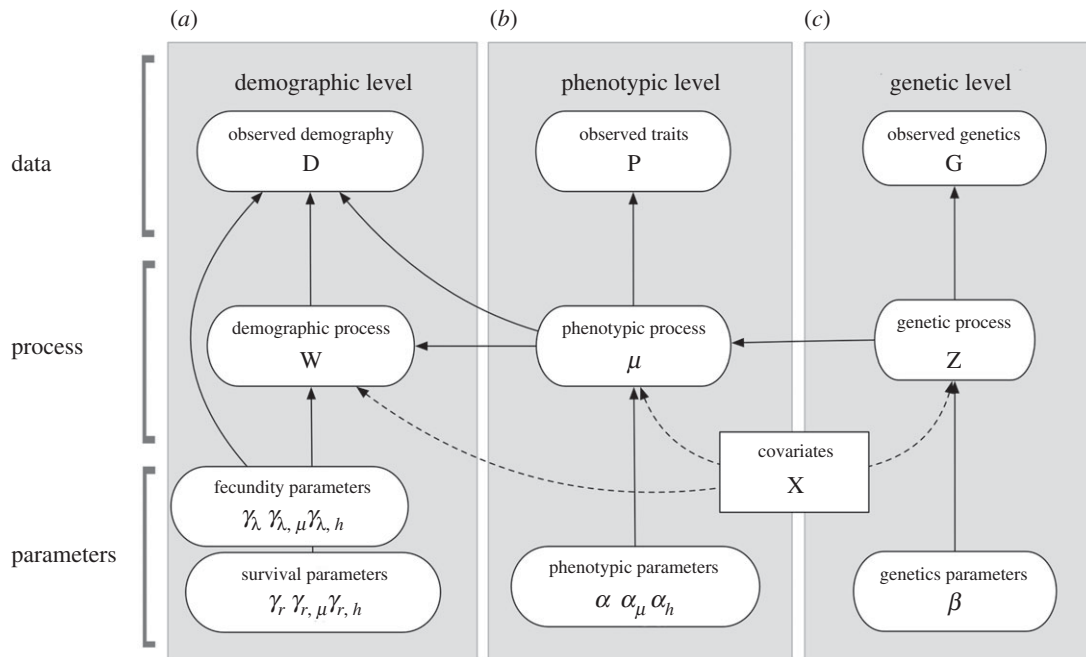
Optimal predictive models are typically selected through empirical comparison or regularization (i.e. reducing complexity of a more general model [93]). While we may have extensive knowledge about some components of the system, we may lack empirical results to constrain certain model formulations. Thus, we must make predictions with a set of models that represent different processes, test those predictions using proper and local scores based on independent data (data not used for model fitting [94]), assess which model works best, and rigorously analyse uncertainties and biases [95].

To date, many efforts to predict species distributional changes or regional population dynamics have been conducted without out-of-sample validation. Without out-of-sample validation, we cannot determine how well predictions will fare under the novel conditions of the future. Experiments that manipulate environments to represent potential future conditions will provide important information. By testing model predictions locally (i.e. for specific locations in geographic or parameter space), we can begin to understand where and why models fail. This kind of validation and model comparison has yet to be performed for a regional-scale population model that incorporates all of the categories of processes described above, meaning that the relative importance of these processes is unknown, as is the utility of an integrative model.

#### (c) Reaction norms as a conceptual basis for integrative models

Traditionally, each of the important candidate processes described above has been studied separately. To test their relative value for predictions, we need new approaches for integrating information about different processes in statistical models. Without an integrative model, we cannot compare models incorporating more than one process, let alone consider interactions among these processes. For example, information about a genetic variant or a physiological threshold that affects organisms of a particular size class might be far more useful in a size-structured demographic model than in a model that simply tracks population density.

Reaction norms, which describe the change in a trait or performance across environments, provide a framework to



**Figure 3.** The directed acyclic graph (e.g. [97]) for our integrated reaction norm model. Solid and dashed arrows indicate stochastic and deterministic connections, respectively. Parameter descriptions in full can also be found in electronic supplementary material, tables S1 and S2.

integrate the processes discussed in the preceding sections. Reaction norms occupy a middle position in a hierarchy of biological complexity among the mechanisms we discussed, and thus they may be viewed as a nexus for integration (figure 2). It is the complexity of these reaction norms that unites the arguments for incorporating mechanism into population forecasts (S2, figure 1). We suggest that reaction norms are an intuitive concept that can be embedded in integrative population models. Genetic variation can be linked to reaction norms of lower-level traits like phenology, behaviour and physiology, which can then be linked to reaction norms for individual demographic rates and fitness (figure 2). However, previous research has described the response of one vital rate or trait to just one abiotic driver measured under controlled conditions. Such idealized reaction norms may be of little use for predicting population dynamics at a single location affected by many interacting environmental drivers, let alone for scaling up population models across a heterogeneous region containing genetically differentiated populations. The complexities we highlight require higher-dimensional reaction norms (figure 1).

One benefit of shifting to models that include reaction norms for key traits is that population models are more easily linked to ecosystem processes. For example, if photosynthetic rates are built into population models of a plant, it is easier to build integrated population–ecosystem models that allow us to assess how environmental changes jointly impact both (e.g. as has been investigated in crop models [49]). The complexity of trait relationships with demography and ecosystem processes requires a substantial integration of underlying mechanisms into population–ecosystem models [96].

#### 4. Integrative hierarchical framework

We propose a hierarchical Bayesian framework to construct an integrated reaction norm model with genetic, phenotypic and demographic components (figure 3). These components correspond to multiple types of experimental and observational data, including DNA sequences, phenotypes, vital rates and

environmental conditions. The components of this model are linked by relationships that have been historically modelled separately. To account for and link the processes described above, an integrative model is required ([98], ch. 25). We propose straightforward extensions and generalizations of past efforts that enhance the power of our inferences and compare the importance of different components, and build on approaches developed for other integrated models that borrow strength from multiple data sources [99,100].

We briefly describe an integrated hierarchical model that comprises three levels (figure 3): genetic, phenotypic and demographic. We offer more detail in the following sections. Our example model connects the three levels using conditional stochastic models with latent processes that depend on each other. In this case, the natural scaling between levels implies the conditional structure with the demographic process depending on the phenotypic process, which in turn depends on the genetic process.

The genetic component allows allele frequencies to change in space and time as a function of underlying environmental selective gradients and gene flow. This part of our approach is similar to genotype–environment association models that identify loci locally adapted to specific environments [63,101], or spatial models of allele frequency turnover [102].

The genetic component is linked to a phenotypic model that determines how genotype and environment determine phenotypes. This part of the model captures processes usually studied by genetic mapping approaches that identify genetic loci causing trait variation [103], or genomic selection approaches that model the whole-genome contribution to quantitative trait variation [104], including models of trait variation across environments [45,48]. This part of the model also captures plastic trait responses to environment (trait reaction norms) that are often a topic in ecophysiology, behavioural ecology and functional ecology.

Finally, the demographic component determines how traits influence vital rates and ultimately fitness. This approach mirrors evolutionary ecology, quantitative genetic analyses of how trait variation is related to vital rates [105], changes in



these relationships across environments (i.e. how traits mediate reaction norms of demography) [106,107], as well as how selection may change across life stages/vital rates [108]. These model components cover the processes discussed in §2 and their integration is described below.

### (a) Modelling genetic processes based on environment and space

We start with an approach to model change in allele frequency along environmental gradients and in space in a way that facilitates prediction of allelic state at genetic markers (e.g. single-nucleotide polymorphisms, SNPs) for a given location. Obtaining the predictive distribution of allele frequencies allows us to make inference on changes in allele frequencies under changing environmental conditions or with gene flow [109]. Parametric models allow us to properly account for, and learn about, uncertainty in a way that can be correctly propagated through to other types of inference. To generalize the concept of genotype–environment associations in a way that accounts for individual-level variability and mechanistic sources of gene flow, we use a hierarchical modelling structure where the individual-based genotype data  $g_{jil}$  for individual  $i=1, \dots, n_j$  in population  $j=1, \dots, J$  and for genetic locus  $l=1, \dots, L$ , arise stochastically as

$$g_{jil} \sim \text{Bern}(z_{jl}). \quad (4.1)$$

The population-level allele frequency is modelled as

$$f(z_{jl}) = \mathbf{x}'_j \beta_l + \eta_{jl}, \quad (4.2)$$

where  $\mathbf{x}_j$  represent a set of covariates (e.g. topography, soil type, precipitation) and  $f$  is an appropriate link function. In this model (4.2), the coefficients  $\beta_l$  connect the environmental variables to allele frequency at SNP  $l$  in a way that accounts for local adaptation.

In cases where we need to account for patterns of gene flow and population structure, the  $\eta_{jl}$  are random effects (e.g. spatially structured) [109–111] that we can model jointly as

$$\eta \sim N(\mathbf{0}, \Sigma), \quad (4.3)$$

where the covariance matrix is parameterized as

$$\Sigma = \sigma^2 \mathbf{R}. \quad (4.4)$$

The  $J \times J$  matrix  $\mathbf{R}$  may be specified as an intrinsic conditional autoregressive (ICAR) [112] correlation matrix that depends on geographic information about the environment. Hanks & Hooten [109] showed that this type of ICAR formulation provides a formal statistical model for gene flow based on circuit theory. By allowing the random effects  $\eta_{jl}$  in (4.2) to be spatially structured, we can accommodate gene flow mechanisms such as isolation by barrier, distance and resistance [111]. Note that historical patterns of gene flow, especially during colonization when drift is strong, may differ from current gene flow and the model can be generalized to accommodate such processes.

In the presence of temporal data, the model could be formulated as

$$g_{jilt} \sim \text{Bern}(z_{jilt}), \quad (4.5)$$

for time  $t=1, \dots, T$ , where the time-specific population-level

allele frequency is modelled as

$$f(z_{jilt}) = \mathbf{x}'_{jt} \beta_{lt} + \eta_{jilt}. \quad (4.6)$$

This extended model formulation explicitly accounts for changes in the environment over time to help us understand how environment aids in the prediction of genotypes. The model is generative and can be used to simulate genetic data to better understand the effects of future environments on genotypes.

This modelling framework can also be generalized to account for multilocus dependence explicitly. To do that, we introduce another random effect to the model in (4.2):

$$f(z_{jl}) = \mathbf{x}'_j \beta_l + \eta_{jl} + \xi_{jl}, \quad (4.7)$$

where, jointly across loci, the random effects are assumed to be structured according to their linkage distance (if on the same chromosome), such that  $\xi_j \sim N(\mathbf{0}, \Sigma_\xi)$ . Not accounting for multilocus associations when they exist can obscure our inference about the importance of physically linked loci. Furthermore, multilocus dependence could improve prediction of genotypes for cases where the genotype is only partially observable (e.g. because of limited resources and genetic markers).

### (b) Incorporating phenotypes

We can integrate genotype–phenotype and environment–phenotype relationships into our model. Consider the measurement of  $p_{jiq}$  for trait  $q$  on individual  $i$  in population  $j$ . To connect the phenotype and genotype in a statistical model, let

$$p_{jiq} \sim [p_{jiq} | \mu_{jiq}, \theta_q]_q, \quad (4.8)$$

such that  $[p_{jiq} | \mu_{jiq}, \theta_q]_q$  is a trait-specific conditional data model with  $\theta_q$  as trait-specific parameters. We let the trait data model vary to accommodate a variety of traits; for example, for binary traits, the distribution in (4.8) should be Bernoulli and  $\theta_q = \emptyset$ .

We now introduce the first reaction norm presented in our model, describing plastic trait change across environments (e.g. figure 2, ‘phenotypes’), as well as a way to accommodate higher dimensionality due to genetic variation in reaction norms. The mean  $\mu_{jiq}$  of the conditional trait distribution is modelled as

$$f(\mu_{jiq}) = \mathbf{x}'_j \alpha + \mathbf{z}'_{ji} \alpha_\mu + h(\mathbf{x}_j, \mathbf{z}_{ji}, \alpha_h) + \varepsilon_{jiq}, \quad (4.9)$$

where  $f$  is a suitable link function and  $\mathbf{x}'_j \alpha$  accounts for phenotypic plasticity (i.e. environmental effects on traits),  $\mathbf{z}'_{ji} \alpha_\mu$  accounts for genetic effects on the trait,  $h(\mathbf{x}_j, \mathbf{z}_{ji}, \alpha_h)$  is a function (e.g. a set of interactions) that accounts for genetic variation in plasticity (i.e. reaction norms), and the error term is correlated by a relevant grouping structure due to genetic similarity among individuals or latent spatial structure, for example. The genetic component of traits changes in space and time due to changes in allele frequency as described in the previous section. Jointly across individuals and groups, the errors are modelled as

$$\varepsilon_q \sim N(\mathbf{0}, \sigma_q^2 \mathbf{I} + \Sigma_y), \quad (4.10)$$

where the variance component  $\sigma_q^2$  may be set to zero as necessary for certain traits  $q$ .

Up to this point, we have described what might be termed an ‘integrated landscape genomic-reaction norm model’, following common usage of terms. We can generalize

this model to account for temporally indexed trait data like the genotype model described in the previous section. Finally, the modular nature of the two models implies that we may be able to apply recursive Bayesian techniques to fit the model in a computationally efficient framework [113,114]. When genetic data are not available or desired, patterns of local adaptation in genetic trait–environment correlations can be accommodated in our framework by treating the genetic process in §4a as latent.

### (c) Incorporating demography

We can learn about population dynamics using demographic models, the specific form of which will vary depending on the study system. For example, at site  $j$  we observe  $n_j$  individual-level survival outcomes ( $w_{ij} = 1$ ) that we model as  $w_{ij} \sim \text{Bern}(r_{ij})$ , where  $r_{ij}$  represents the survival probability of individual  $i$ . By specifying individual-level demography, we can link individual genotype and phenotype to environment and the effects on demography. Our approach is to build reaction norms of demographic vital rates that depend on (possibly latent) phenotype–environment interactions, where the phenotypes themselves are modelled as reaction norms as described in the previous section.

At the end of a reproductive season, we can model the number of offspring of a hermaphroditic individual  $i$ ,  $d_{ij}$ , as arising from the zero-inflated distribution

$$d_{ij} \sim \begin{cases} \text{Pois}(\lambda_{ij}), & w_{ij} = 1 \\ 0, & w_{ij} = 0, \end{cases} \quad (4.11)$$

where  $\lambda_{ij}$  represents the fecundity of individuals surviving to reproduction.

We seek to infer the effect of phenotype and environment on individual-level fecundity and survival at site  $j$ . Consider the latent process model for individual-level survival probability  $r_{ij}$  where

$$f(r_{ij}) = \mathbf{x}'_{ij} \boldsymbol{\gamma}_r + \boldsymbol{\mu}'_{ij} \boldsymbol{\gamma}_{r,\mu} + h(\mathbf{x}_{ij}, \boldsymbol{\mu}_{ij}, \boldsymbol{\gamma}_{r,h}), \quad (4.12)$$

and the covariates and associated regression coefficients are specified in the same manner as in the genetic and phenotypic model components. This model allows us to account for multivariate phenotypic effects  $\boldsymbol{\mu}_{ij}$  on survival. When important traits are unknown or unmeasured (a common occurrence) we may use genotypes in place of or in addition to  $\boldsymbol{\mu}_{ij}$ , to model selection on latent traits. We can specify a similar model for fecundity  $\lambda_{ij}$  as

$$\log(\lambda_{ij}) = \mathbf{x}'_{ij} \boldsymbol{\gamma}_\lambda + \boldsymbol{\mu}'_{ij} \boldsymbol{\gamma}_{\lambda,\mu} + h(\mathbf{x}_{ij}, \boldsymbol{\mu}_{ij}, \boldsymbol{\gamma}_{\lambda,h}). \quad (4.13)$$

Biotic interactions (e.g. conspecific density dependence, parasitism) can be incorporated via  $\mathbf{x}_{ij}$ . This model can be fitted from a hierarchical perspective (perhaps jointly with the previously described genetic and phenotypic components).

To review, we have a genetic model component connected to a phenotypic component, which is then linked to selection and demography. Note that the  $h(\mathbf{x}_{ij}, \boldsymbol{\mu}_{ij}, \boldsymbol{\gamma}_{\lambda,h})$  function in (4.13) (and the similar function in (4.12)) and the  $\mathbf{x}'_j \boldsymbol{\beta}_l$  term in (4.2) both represent changes in selection across environmental gradients  $\mathbf{x}_{ij}$ , and the information on these changes in selection arises from both the distribution of alleles across environments, their effects on phenotypes and the changes in selection on these phenotypes across environments [115]. Together with the genetic and phenotypic model components, we can express the full integrated

hierarchical genetic–phenotypic–demographic–environmental model as a directed acyclic graph (figure 3).

We can generalize the ecological model component by allowing for additional data sources when available. For example, we can fuse presence-only and occupancy (with no individual identification) data, as well as unmarked counts of individuals if replicates are available, by adapting (4.2) to include the point process, occupancy, or N-mixture model frameworks [98]. Furthermore, we can accommodate time series data by indexing the data with a  $t$  subscript as described in §4a.

### (d) Resulting inference

The full hierarchical model in figure 3 is generative and can be used to simulate data to better understand the effects of future environmental regimes on populations, phenotypes and genotypes. For example, by comparing contemporary genotypes in a given location with genotypes under future environments, we can better understand sensitivity to changes in environment [15] and link this inference to spatio-temporal predictor variables and create maps of these sensitivities with uncertainty.

Each model component in figure 3 can be fitted individually to the relevant subset of data while conditioning on known or hypothesized elements from other components. However, the benefit of constructing a jointly specified integrated model is that we can fit the full hierarchical model and simultaneously learn about all the model quantities while allowing for feedbacks to occur among demographic, phenotypic and genetic processes. The full hierarchical model also allows the uncertainties to propagate among model components so that we can make valid conclusions about the relative importance of each process. We can also compare predictions from the full model with predictions from simpler models that ignore one or more components and the processes they represent. These comparisons will identify the optimal level of complexity for prediction. When implementing the model, if additional flexibility is needed, we can use basis function expansions of the environmental, genetic and phenotypic space [116] to account for nonlinearities in the relationships.

To optimize predictive ability, we can hold out data from a subset of sites (or time points) for model validation. We can use the integrated model to obtain the posterior predictions for the latent processes. The direct comparison of observed and predicted population growth is essential for quantifying sources of uncertainty in the model, an important step in guiding future research to improve the model.

### (e) Feasibility

Fitting our integrated model will require an unprecedented, coordinated data collection effort. But unprecedented does not mean impossible. Annual plants may be a good starting point because their simple life cycles facilitate estimation of lifetime fitness, even in experimental settings. For example, existing knowledge and resources make the model annual plant *Arabidopsis thaliana* amenable [40,47]. We are leading a network of researchers tackling the goal of this paper using *Bromus tectorum*, a widely distributed annual grass invasive in western North America. Even some long-lived perennial plants have been studied using common gardens, demographic observations and genomic techniques (especially in species of economic value) [117,118]. While controlled field or lab experiments may even be possible for some small



animals [119], for many species controlled experiments are extremely difficult. Additionally, the complexity of biological systems may resist our efforts to experiment, observe and build mechanistic models. In the face of such challenges, integrative regional forecasting models may be built with fewer of the processes described above, or with phenomenological components. The strength of our suggested approach is its ability to quantify the predictive value of any processes that can be feasibly modelled.

## 5. Conclusion

It is clear that a wide range of biological processes can be incorporated into large-scale forecasts of population dynamics under environmental change. But the route to accurate prediction might not involve building models of ever-increasing complexity. While the integrated model we describe does account for many underlying processes, we emphasize that this full model is intended for development and may not itself be the ideal model for prediction. To advance the field of regional population prediction, we need extensive empirical study and an iterative process of model improvement [120]. Which processes are important for good forecasting may depend on the time scale of the forecast (or forecast horizon) [121].

We presented a statistical framework to integrate these diverse processes in a model for use in prediction. This

framework is novel because it formally unifies genetic, phenotypic and demographic components into an integrated reaction norm model and suggests general computational methods for fitting it. We can compare our methods with other existing, less integrated approaches and quantify the value added when formally borrowing strength across multiple sources of data in a single, cohesive hierarchical modelling framework. The answer to the question we posed, ‘what processes must we understand to forecast regional scale population dynamics?’, will be found only through careful study that generates the extensive multi-dimensional data discussed here combined with an integrative modelling approach.

**Data accessibility.** This article has no additional data.

**Authors' contributions.** All authors contributed to the development of the manuscript. J.R.L. led the writing and M.B.H. led the model development.

**Competing interests.** We declare we have no competing interests.

**Funding.** Funding was provided by NSF DEB grant nos. 1927177, 1927282 and 1927009.

**Acknowledgements.** The authors thank Kristen Ruegg, Erika Zavaleta and the BromeCast Network for early discussions and insights related to this work, and Samuel Scarpino for comments on the manuscript.

**Disclaimer.** Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the US Government.

## References

- Boutillier P, Maasha M, Li X, Medina-Abarca HF, Krivine J, Feret J, Cristescu I, Forbes AG, Fontana W. 2018 The Kappa platform for rule-based modeling. *Bioinformatics* **34**, i583–i592. (doi:10.1093/bioinformatics/bty272)
- Medvigy D, Wofsy SC, Munger JW, Hollinger DY, Moorcroft PR. 2009 Mechanistic scaling of ecosystem function and dynamics in space and time: ecosystem demography model version 2. *J. Geophys. Res. Biogeosci.* **114**, G01002. (doi:10.1029/2008JG000812)
- Bay RA, Rose N, Barrett R, Bernatchez L, Ghalambor CK, Lasky JR, Brem RB, Palumbi SR, Ralph P. 2017 Predicting responses to contemporary environmental change using evolutionary response architectures. *Am. Nat.* **189**, 463–473. (doi:10.1086/691233)
- Clark JS, Gelfand AE, Woodall CW, Zhu K. 2014 More than the sum of the parts: forest climate response from joint species distribution models. *Ecol. Appl.* **24**, 990–999. (doi:10.1890/13-1015.1)
- Buckley LB. 2013 Get real: putting models of climate change and species interactions in practice: thermal, temporal, and spatial niche partitioning. *Ann. N Y Acad. Sci.* **1297**, 126–138. (doi:10.1111/nyas.12175)
- Kearney M, Porter W. 2009 Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecol. Lett.* **12**, 334–350. (doi:10.1111/j.1461-0248.2008.01277.x)
- Urban MC *et al.* 2016 Improving the forecast for biodiversity under climate change. *Science* **353**, aad8466. (doi:10.1126/science.aad8466)
- Dietze MC *et al.* 2018 Iterative near-term ecological forecasting: needs, opportunities, and challenges. *Proc. Natl Acad. Sci. USA* **115**, 1424–1432. (doi:10.1073/pnas.1710231115)
- Dawson TP, Jackson ST, House JI, Prentice IC, Mace GM. 2011 Beyond predictions: biodiversity conservation in a changing climate. *Science* **332**, 53–58. (doi:10.1126/science.1200303)
- Clark J *et al.* 2001 Ecological forecasts: an emerging imperative. *Science* **293**, 657–660. (doi:10.1126/science.293.5530.657)
- Guisan A, Thuiller W. 2005 Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* **8**, 993–1009. (doi:10.1111/j.1461-0248.2005.00792.x)
- Crone EE *et al.* 2013 Ability of matrix models to explain the past and predict the future of plant populations. *Conserv. Biol.* **27**, 968–978. (doi:10.1111/cobi.12049)
- Harris DJ, Taylor SD, White EP. 2018 Forecasting biodiversity in breeding birds using best practices. *PeerJ* **6**, e4278. (doi:10.7717/peerj.4278)
- Merow C, Bois ST, Allen JM, Xie Y, Silander JA. 2017 Climate change both facilitates and inhibits invasive plant ranges in New England. *Proc. Natl Acad. Sci. USA* **114**, E3276–E3284. (doi:10.1073/pnas.1609633114)
- Capblancq T, Fitzpatrick MC, Bay RA, Exposito-Alonso M, Keller SR. 2020 Genomic prediction of (mal)adaptation across current and future climatic landscapes. *Ann. Rev. Ecol. Evol. Syst.* **51**, 245–269. (doi:10.1146/annurev-ecolsys-020720-042553)
- Crozier L, Dwyer G. 2006 Combining population-dynamic and ecophysiological models to predict climate-induced insect range shifts. *Am. Nat.* **167**, 853–866. (doi:10.1086/504848)
- Helmuth B, Kingsolver J, Carrington E. 2005 Biophysics, physiological ecology, and climate change: does mechanism matter? *Annu. Rev. Physiol.* **67**, 177–201. (doi:10.1146/annurev.physiol.67.040403.105027)
- Hooten MB, Scharf HR, Morales JM. 2019 Running on empty: recharge dynamics from animal movement data. *Ecol. Lett.* **22**, 377–389. (doi:10.1111/ele.13198)
- Buckley LB, Urban MC, Angilletta MJ, Crozier LG, Rissler LJ, Sears MW. 2010 Can mechanism inform species' distribution models? *Ecol. Lett.* **13**, 1041–1054. (doi:10.1111/j.1461-0248.2010.01479.x)
- Hefley TJ, Hooten MB, Russell RE, Walsh DP, Powell JA. 2017 When mechanism matters: Bayesian forecasting using models of ecological diffusion. *Ecol. Lett.* **20**, 640–650. (doi:10.1111/ele.12763)
- Moran EV, Alexander JM. 2014 Evolutionary responses to global change: lessons from invasive species. *Ecol. Lett.* **17**, 637–649. (doi:10.1111/ele.12262)

22. Williams JW, Jackson ST. 2007 Novel climates, no-analog communities, and ecological surprises. *Front. Ecol. Environ.* **5**, 475–482. (doi:10.1890/070037)
23. Scarpino SV, Petri G. 2019 On the predictability of infectious disease outbreaks. *Nat. Commun.* **10**, 898–890. (doi:10.1038/s41467-019-08616-0)
24. Pennekamp F *et al.* 2019 The intrinsic predictability of ecological time series and its potential to guide forecasting. *Ecol. Monogr.* **89**, e01359. (doi:10.1002/ecm.1359)
25. Hébert-Dufresne L, Scarpino SV, Young JG. 2020 Macroscopic patterns of interacting contagions are indistinguishable from social reinforcement. *Nat. Phys.* **16**, 426–431. (doi:10.1038/s41567-020-0791-2)
26. Reichert P, Omlin M. 1997 On the usefulness of overparameterized ecological models. *Ecol. Modell.* **95**, 289–299. (doi:10.1016/S0304-3800(96)00043-9)
27. Clark JS, Bell DM, Kwit M, Stine A, Vierra B, Zhu K. 2012 Individual-scale inference to anticipate climate-change vulnerability of biodiversity. *Phil. Trans. R. Soc. B* **367**, 236–246. (doi:10.1098/rstb.2011.0183)
28. Wagner CH. 1982 Simpson's paradox in real life. *Am. Stat.* **36**, 46–48. (doi:10.1080/00031305.1982.10482778)
29. Piantadosi S, Byar DP, Green SB. 1988 The ecological fallacy. *Am. J. Epidemiol.* **127**, 893–904. (doi:10.1093/oxfordjournals.aje.a114892)
30. Peterson ML, Doak DF, Morris WF. 2019 Incorporating local adaptation into forecasts of species' distribution and abundance under climate change. *Glob. Change Biol.* **25**, 775–793. (doi:10.1111/gcb.14562)
31. Josephs EB. 2018 Determining the evolutionary forces shaping  $G \times E$ . *New Phytology* **219**, 31–36. (doi:10.1111/nph.15103)
32. Hoffmann AA, Harshman LG. 1999 Desiccation and starvation resistance in *Drosophila*: patterns of variation at the species, population and intrapopulation levels. *Heredity* **83**, 637–643. (doi:10.1046/j.1365-2540.1999.00649.x)
33. Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011 A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**, 86–89. (doi:10.1126/science.1209271)
34. Carius HJ, Little TJ, Ebert D. 2001 Genetic variation in a host-parasite association: potential for coevolution and frequency-dependent selection. *Evolution* **55**, 1136–1145. (doi:10.1111/j.0014-3820.2001.tb00633.x)
35. Cousyn C, Meester LD, Colbourne JK, Brendonck L, Verschuren D, Volckaert F. 2001 Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proc. Natl Acad. Sci. USA* **98**, 6256–6260. (doi:10.1073/pnas.111606798)
36. Torres-Dowdall J, Handelsman CA, Reznick DN, Ghalambor CK. 2012 Local adaptation and the evolution of phenotypic plasticity in trinidadian guppies (*Poecilia reticulata*). *Evol.: Inter. J. Org. Evol.* **66**, 3432–3443. (doi:10.1111/j.1558-5646.2012.01694.x)
37. Schiffers K, Bourne EC, Lavergne S, Thuiller W, Travis JMJ. 2012 Limited evolutionary rescue of locally adapted populations facing climate change. *Phil. Trans. R. Soc. B* **368**, 20120083. (doi:10.1098/rstb.2012.0083)
38. Bay RA, Harrigan RJ, Underwood VL, Gibbs HL, Smith TB, Ruegg K. 2018 Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **359**, 83–86. (doi:10.1126/science.aan4380)
39. Razgour O *et al.* 2019 Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proc. Natl Acad. Sci. USA* **116**, 10 418–10 423. (doi:10.1073/pnas.1820663116)
40. Exposito-Alonso M, Vasseur F, Ding W, Wang G, Burbano HA, Weigel D. 2018 Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*. *Nat. Ecol. Evol.* **2**, 352–358. (doi:10.1038/s41559-017-0423-0)
41. Rhoné B *et al.* 2020 Pearl millet genomic vulnerability to climate change in West Africa highlights the need for regional collaboration. *Nat. Commun.* **11**, 5274. (doi:10.1038/s41467-020-19066-4)
42. Angert AL, Sheth SN, Paul JR. 2011 Incorporating population-level variation in thermal performance into predictions of geographic range shifts. *Integr. Comp. Biol.* **51**, 733–750. (doi:10.1093/icb/ucr048)
43. Ikeda DH, Max TL, Allan GJ, Lau MK, Shuster SM, Whitham TG. 2017 Genetically informed ecological niche models improve climate change predictions. *Glob. Change Biol.* **23**, 164–176. (doi:10.1111/gcb.13470)
44. Wang T, O'Neill GA, Aitken SN. 2010 Integrating environmental and genetic effects to predict responses of tree populations to climate. *Ecol. Appl.* **20**, 153–163. (doi:10.1890/08-2257.1)
45. Tiezzi F, Parker Gaddis KL, Maltecca C. 2017 Genotype by environment (climate) interaction improves genomic prediction for production traits in US Holstein cattle. *J. Dairy Sci.* **100**, 2042–2056. (doi:10.3168/jds.2016-11543)
46. Lasky JR *et al.* 2015 Genome-environment associations in sorghum landraces predict adaptive traits. *Sci. Adv.* **1**, e1400218. (doi:10.1126/sciadv.1400218)
47. Gienapp P, Fior S, Guillaume F, Lasky JR, Sork VL, Csilléry K. 2017 Genomic quantitative genetics to study evolution in the wild. *Trends Ecol. Evol.* **32**, 897–908. (doi:10.1016/j.tree.2017.09.004)
48. Messina CD, Technow F, Tang T, Totir R, Gho C, Cooper M. 2018 Leveraging biological insight and environmental variation to improve phenotypic prediction: integrating crop growth models (CGM) with whole genome prediction (WGP). *Eur. J. Agron.* **100**, 151–162. (doi:10.1016/j.eja.2018.01.007)
49. Peng B *et al.* 2020 Towards a multiscale crop modelling framework for climate change adaptation assessment. *Nature Plants* **6**, 338–348. (doi:10.1038/s41477-020-0625-3)
50. Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES. 2016 Open chromatin reveals the functional maize genome. *Proc. Natl Acad. Sci. USA* **113**, E3177–E3184. (doi:10.1073/pnas.1525244113)
51. Lasky JR, Des Marais DL, Lowry DB, Povolotskaya I, McKay JK, Richards JH, Keitt TH, Juenger TE. 2014 Natural variation in abiotic stress responsive gene expression and local adaptation to climate in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **31**, 2283–2296. (doi:10.1093/molbev/msu170)
52. Jones FC *et al.* 2012 The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61. (doi:10.1038/nature10944)
53. Polechová J, Barton N, Marion G. 2009 Species' range: adaptation in space and time. *Am. Nat.* **174**, E186–E204. (doi:10.1086/605958)
54. Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. 2008 Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol. Appl.* **1**, 95–111. (doi:10.1111/j.1752-4571.2007.00013.x)
55. Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L. 2013 On the accumulation of deleterious mutations during range expansions. *Mol. Ecol.* **22**, 5972–5982. (doi:10.1111/mec.12524)
56. Gienapp P, Calus MPL, Laine VN, Visser ME. 2019 Genomic selection on breeding time in a wild bird population. *Evol. Lett.* **3**, 142–151. (doi:10.1002/evl3.103)
57. Rudman SM *et al.* 2018 What genomic data can reveal about eco-evolutionary dynamics. *Nat. Ecol. Evol.* **2**, 9–15. (doi:10.1038/s41559-017-0385-2)
58. Whittaker RH, Levin SA, Root RB. 1973 Niche, habitat, and ecotope. *Am. Nat.* **107**, 321–338. (doi:10.1086/282837)
59. John R *et al.* 2007 Soil nutrients influence spatial distributions of tropical tree species. *Proc. Natl Acad. Sci. USA* **104**, 864–869. (doi:10.1073/pnas.0604666104)
60. Donohue K, Burghardt LT, Runcie D, Bradford KJ, Schmitt J. 2015 Applying developmental threshold models to evolutionary ecology. *Trends Ecol. Evol.* **30**, 66–77. (doi:10.1016/j.tree.2014.11.008)
61. Buckley LB, Cannistra AF, John A. 2018 Leveraging organismal biology to forecast the effects of climate change. *Integr. Comp. Biol.* **58**, 38–51. (doi:10.1093/icb/icy018)
62. Phillips SJ, Anderson RP, Schapire RE. 2006 Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* **190**, 231–259. (doi:10.1016/j.ecolmodel.2005.03.026)
63. Fitzpatrick MC, Keller SR. 2015 Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol. Lett.* **18**, 1–16. (doi:10.1111/ele.12376)
64. Hooten M, Wikle C, Schwob M. 2020 Statistical implementations of agent-based demographic models. *Int. Stat. Rev.* **88**, 441–461. (doi:10.1111/insr.12399)
65. Lazarus BE, Castanha C, Germino MJ, Kueppers LM, Moyes AB. 2018 Growth strategies and threshold

- responses to water deficit modulate effects of warming on tree seedlings from forest to alpine. *J. Ecol.* **106**, 571–585. (doi:10.1111/1365-2745.12837)
66. Buckley L. 2008 Linking traits to energetics and population dynamics to predict lizard ranges in changing environments. *Am. Nat.* **171**, E1–E19. (doi:10.1086/523949)
67. Sterck F, Markesteijn L, Schieving F, Poorter L. 2011 Functional traits determine trade-offs and niches in a tropical forest community. *Proc. Natl Acad. Sci. USA* **108**, 20 627–20 632. (doi:10.1073/pnas.1106950108)
68. Lorts CM, Lasky JR. 2020 Competition–drought interactions change phenotypic plasticity and the direction of selection on *Arabidopsis* traits. *New Phytologist* **227**, 1060–1072. (doi:10.1111/nph.16593)
69. Farrior CE, Dybzinski R, Levin SA, Pacala SW. 2013 Competition for water and light in closed-canopy forests: a tractable model of carbon allocation with implications for carbon sinks. *Am. Nat.* **181**, 314–330. (doi:10.1086/669153)
70. Lortie CJ, Callaway RM. 2006 Re-analysis of meta-analysis: support for the stress-gradient hypothesis. *J. Ecol.* **94**, 7–16. (doi:10.1111/j.1365-2745.2005.01066.x)
71. Gilman SE, Urban MC, Tewksbury J, Gilchrist GW, Holt RD. 2010 A framework for community interactions under climate change. *Trends Ecol. Evol.* **25**, 325–331. (doi:10.1016/j.tree.2010.03.002)
72. Suttle K, Thomsen M, Power M. 2007 Species interactions reverse grassland responses to changing climate. *Science* **315**, 640–642. (doi:10.1126/science.1136401)
73. Lasky JR. 2019 Eco-evolutionary community turnover following environmental change. *Evol. Appl.* **12**, 1434–1448. (doi:10.1111/eva.12776)
74. Helmus MR, Mahler DL, Losos JB. 2014 Island biogeography of the Anthropocene. *Nature* **513**, 543–546. (doi:10.1038/nature13739)
75. Ovasikainen O, Abrego N, Halme P, Dunson D. 2016 Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods Ecol. Evol.* **7**, 549–555. (doi:10.1111/2041-210X.12501)
76. Clark JS, Scher CL, Swift M. 2020 The emergent interactions that govern biodiversity change. *Proc. Natl Acad. Sci. USA* **117**, 17 074–17 083. (doi:10.1073/pnas.2003852117)
77. Freckleton RP, Watkinson AR, Green RE, Sutherland WJ. 2006 Census error and the detection of density dependence. *J. Anim. Ecol.* **75**, 837–851. (doi:10.1111/j.1365-2656.2006.01121.x)
78. Detto M, Visser MD, Wright SJ, Pacala SW. 2019 Bias in the detection of negative density dependence in plant communities. *Ecol. Lett.* **22**, 1923–1939. (doi:10.1111/ele.13372)
79. Uriarte M *et al.* 2010 Trait similarity, shared ancestry and the structure of neighbourhood interactions in a subtropical wet forest: implications for community assembly. *Ecol. Lett.* **13**, 1503–1514. (doi:10.1111/j.1461-0248.2010.01541.x)
80. Tredennick AT, Teller B, Adler PB, Hooker G, Ellner SP. 2018 Size-by-environment interactions: a neglected dimension of species' responses to environmental variation. *Ecol. Lett.* **21**, 1757–1770. (doi:10.1111/ele.13154)
81. Niinemets Ü. 2010 Responses of forest trees to single and multiple environmental stresses from seedlings to mature plants: past stress history, stress interactions, tolerance and acclimation. *For. Ecol. Manage.* **260**, 1623–1639. (doi:10.1016/j.foreco.2010.07.054)
82. Caswell H. 2001 *Matrix population models*. Construction, analysis and interpretation, 2nd edn. Sunderland, MA: Sinauer Associates.
83. Ellner SP, Childs DZ, Rees M. 2016 *Data-driven modelling of structured populations*. Lecture Notes on Mathematical Modelling in the Life Sciences. Cham, Switzerland: Springer International Publishing.
84. Ehrlén J, Morris WF. 2015 Predicting changes in the distribution and abundance of species under environmental change. *Ecol. Lett.* **18**, 303–314. (doi:10.1111/ele.12410)
85. Gurevitch J, Fox GA, Fowler NL, Graham CH. 2016 Landscape demography: population change and its drivers across spatial scales. *Q. Rev. Biol.* **91**, 459–485. (doi:10.1086/689560)
86. Gerber BD, Hooten MB, Peck CP, Rice MB, Gammonley JH, Apa AD, Davis AJ. 2019 Extreme site fidelity as an optimal strategy in an unpredictable and homogeneous environment. *Funct. Ecol.* **33**, 1695–1707. (doi:10.1111/1365-2435.13390)
87. Pagel J, Schurr FM. 2012 Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Glob. Ecol. Biogeogr.* **21**, 293–304. (doi:10.1111/j.1466-8238.2011.00663.x)
88. Schurr FM *et al.* 2012 How to understand species niches and range dynamics: a demographic research agenda for biogeography. *J. Biogeogr.* **39**, 2146–2162. (doi:10.1111/j.1365-2699.2012.02737.x)
89. Merow C, Latimer AM, Wilson AM, McMahon SM, Rebelo AG, Silander JA. 2014 On using integral projection models to generate demographically driven predictions of species' distributions: development and validation using sparse data. *Ecography* **37**, 1167–1183. (doi:10.1111/ecog.00839)
90. Geman S, Bienenstock E, Doursat R. 1992 Neural networks and the bias/variance dilemma. *Neural Comput.* **4**, 1–58. (doi:10.1162/neco.1992.4.1.1)
91. Chevalier M, Knappe J. 2020 The cost of complexity in forecasts of population abundances is reduced but not eliminated by borrowing information across space using a hierarchical approach. *Oikos* **129**, 249–260. (doi:10.1111/oik.06401)
92. Ward EJ, Holmes EE, Thorson JT, Collen B. 2014 Complexity is costly: a meta-analysis of parametric and non-parametric methods for short-term population forecasting. *Oikos* **123**, 652–661. (doi:10.1111/j.1600-0706.2014.00916.x)
93. Hooten MB, Cooch EC. 2019 Comparing ecological models. In *Quantitative analysis in wildlife science* (eds LA Brennan, AN Tri, BG Marcot). Baltimore, MD: Johns Hopkins University Press.
94. Hooten MB, Hobbs NT. 2015 A guide to Bayesian model selection for ecologists. *Ecol. Monogr.* **85**, 3–28. (doi:10.1890/14-0661.1)
95. Conn PB, Johnson DS, Williams PJ, Melin SR, Hooten MB. 2018 A guide to Bayesian model checking for ecologists. *Ecol. Monogr.* **88**, 526–542. (doi:10.1002/ecm.1314)
96. Fridley JD. 2017 Plant energetics and the synthesis of population and ecosystem ecology. *J. Ecol.* **105**, 95–110. (doi:10.1111/1365-2745.12693)
97. Hobbs NT, Hooten MB. 2015 *Bayesian models: a statistical primer for ecologists*, student edition. Princeton, NJ: Princeton University Press.
98. Hooten MB, Hefley TJ. 2019 *Bringing bayesian models to life*. New York, NY: CRC Press.
99. Schaub M, Abadi F. 2011 Integrated population models: a novel analysis framework for deeper insights into population dynamics. *J. Ornithol.* **152**, 227–237. (doi:10.1007/s10336-010-0632-7)
100. Hanks EM, Hooten MB, Baker FA. 2011 Reconciling multiple data sources to improve accuracy of large-scale prediction of forest disease incidence. *Ecol. Appl.* **21**, 1173–1188. (doi:10.1890/09-1549.1)
101. Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010 Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**, 1411–1423. (doi:10.1534/genetics.110.114819)
102. Yang WY, Novembre J, Eskin E, Halperin E. 2012 A model-based approach for analysis of spatial structure in genetic data. *Nat. Genet.* **44**, 725–731. (doi:10.1038/ng.2285)
103. Hirschhorn JN, Daly MJ. 2005 Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6**, 95–108. (doi:10.1038/nrg1521)
104. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS. 2009 Invited Review: reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* **92**, 16–24. (doi:10.3168/jds.2008-1514)
105. Lande R, Arnold SJ. 1983 The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226. (doi:10.1111/j.1558-5646.1983.tb00236.x)
106. Wade MJ, Kalisz S. 1990 The causes of natural selection. *Evolution* **44**, 1947–1955. (doi:10.1111/j.1558-5646.1990.tb04301.x)
107. Hall MC, Willis JH. 2006 Divergent selection on flowering time contributes to local adaptation in *Mimulus guttatus* populations. *Evolution* **60**, 2466–2477. (doi:10.1111/j.0014-3820.2006.tb01882.x)
108. Schluter D, Price TD, Rowe L, Grant PR. 1991 Conflicting selection pressures and life history trade-offs. *Proc. R. Soc. Lond. B* **246**, 11–17. (doi:10.1098/rspb.1991.0118)



109. Hanks EM, Hooten MB. 2013 Circuit theory and model-based inference for landscape connectivity. *J. Am. Stat. Assoc.* **108**, 22–33. (doi:10.1080/01621459.2012.724647)
110. Hanks EM, Hooten MB, Knick ST, Oyler-McCance SJ, Fike JA, Cross TB, Schwartz MK. 2016 Latent spatial models and sampling design for landscape genetics. *Ann. Appl. Stat.* **10**, 1041–1062. (doi:10.1214/16-AOAS929)
111. Peterson EE, Hanks EM, Hooten MB, Ver Hoef JM, Fortin MJ. 2019 Spatially structured statistical network models for landscape genetics. *Ecol. Monogr.* **89**, e01355. (doi:10.1002/ecm.1355)
112. Ver Hoef JM, Peterson EE, Hooten MB, Hanks EM, Fortin MJ. 2018 Spatial autoregressive models for statistical inference from ecological data. *Ecol. Monogr.* **88**, 36–59. (doi:10.1002/ecm.1283)
113. Hooten MB, Buderman FE, Brost BM, Hanks EM, Ivan JS. 2016 Hierarchical animal movement models for population-level inference. *Environmetrics* **27**, 322–333. (doi:10.1002/env.2402)
114. Hooten MB, Johnson DS, Brost BM. In press. Making recursive Bayesian inference accessible. *Am. Stat.* (doi:10.1080/00031305.2019.1665584)
115. Lasky JR, Forester BR, Reimherr M. 2018 Coherent synthesis of genomic associations with phenotypes and home environments. *Mol. Ecol. Resour.* **18**, 91–106. (doi:10.1111/1755-0998.12714)
116. Hefley TJ, Broms KM, Brost BM, Buderman FE, Kay SL, Scharf HR, Tipton JR, Williams PJ, Hooten MB. 2017 The basis function approach for modeling autocorrelation in ecological data. *Ecology* **98**, 632–646. (doi:10.1002/ecy.1674)
117. Yeaman S *et al.* 2016 Convergent local adaptation to climate in distantly related conifers. *Science* **353**, 1431–1433. (doi:10.1126/science.aaf7812)
118. Lowry DB *et al.* 2019 QTL  $\times$  environment interactions underlie adaptive divergence in switchgrass across a large latitudinal gradient. *Proc. Natl Acad. Sci. USA* **116**, 12 933–12 941. (doi:10.1073/pnas.1821543116)
119. Rudman SM *et al.* 2019 Microbiome composition shapes rapid genomic adaptation of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **116**, 20 025–20 032. (doi:10.1073/pnas.1907787116)
120. Ver Hoef JM, Boveng PL. 2015 Iterating on a single model is a viable alternative to multimodel inference. *J. Wildlife Manage.* **79**, 719–729. (doi:10.1002/jwmg.891)
121. Adler PB, White EP, Cortez MH. 2020 Matching the forecast horizon with the relevant spatial and temporal processes and data sources. *Ecography* **43**, 1729–1739. (doi:10.1111/ecog.05271)