ARTICLE IN PRESS

CellPress
REVIEWS

Opinion

Overcoming the Data Crisis in Biodiversity Conservation

Holly K. Kindsvater,[1,*] Nicholas K. Dulvy,[2] Cat Horswill,[3,4] Maria-José Juan-Jordá,[2,5] Marc Mangel,[6,7] and Jason Matthiopoulos[3]

How can we track population trends when monitoring data are sparse? Population declines can go undetected, despite ongoing threats. For example, only one of every 200 harvested species are monitored. This gap leads to uncertainty about the seriousness of declines and hampers effective conservation. Collecting more data is important, but we can also make better use of existing information. Prior knowledge of physiology, life history, and community ecology can be used to inform population models. Additionally, in multispecies models, information can be shared among taxa based on phylogenetic, spatial, or temporal proximity. By exploiting generalities across species that share evolutionary or ecological characteristics within Bayesian hierarchical models, we can fill crucial gaps in the assessment of species' status with unparalleled quantitative rigor.

The Data Crisis in Biodiversity Conservation

Biodiversity is disappearing at an unprecedented rate [1,2], much faster than the rate with which we can afford to study and monitor threatened populations. Approximately one-quarter of the 87 967 animals, plants and fungi assessed by the International Union for Conservation of Nature (IUCN) are classified as threatened with extinction and another 10% are Data Deficient [3]. Data deficiency is prevalent for fishes, invertebrates, plants, and fungi [1,4,5]. Conserving these data-limited species is a challenge for scientists and policy makers, because they lack the information needed to apply standardized criteria for declines [6]. Precautionary measures of protection could be used if declines are suspected, but precaution is difficult to justify when species exploitation for food, medicine, or other goods sustains many livelihoods.

For many species, we lack evidence of declines that we suspect are happening. The status of terrestrial species hunted for bushmeat or for the luxury health-product market is often unknown [7,8], although demand is undiminished. Large terrestrial mammals are commonly hunted, especially in Southeast Asia, Africa, and the neotropics [9], but population data are scarce and enforcement of management is weak [10]. Likewise, only 10% of the 2000 fish species that are commercially exploited have been assessed for scientific management [11,12]. For fishes that are caught incidentally (as bycatch), the extent of human exploitation is difficult to quantify. For example, of more than 1250 species of sharks, rays, and chimaeras, 17% have been categorized as threatened by the IUCN but another 46% are Data Deficient [13]. We know of only 39 shark or ray populations (spanning 33 species) that are currently managed sustainably [14], a small fraction of the exploited chondrichthyans. The situation is worse in many other terrestrial and marine groups [1,15–17]. Given the scale of this problem, how can environmental scientists reliably advise stewards tasked with sustainable management? The cost of monitoring can be prohibitive: the USA spent US$215 million to assess 121 fish stocks in 2015 alone [18]. Therefore, while more data collection is a key component of the solution, we highlight new statistical approaches that can extend the utility of existing data by making the best use of

Highlights

Diagnosing the conservation status of many species is hampered by insufficient data. Modern computer-intensive fitting methods make it possible to merge mechanistic models and population data on well-studied indicator species, extending the inferences we can make about their data-limited relatives.

Historically, assessments have used data from one population or species to create *ad hoc* proxy values for the life-history traits of relatives, but with modern Bayesian models we can share information in a standardized, coherent way.

Advances in understanding community ecology and life-history evolution can be incorporated into these models as priors, extending statistical power even when data are sparse.

These advances offer new possibilities for the rigorous assessment and protection of populations and species that previously have suffered from policy gaps created by insufficient data.

[1]Department of Ecology, Evolution, and Natural Resources, Rutgers University, 14 College Farm Road, New Brunswick, NJ 08901, USA
[2]Earth to Ocean Research Group, Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
[3]Institute of Biodiversity, Animal Health and Comparative Medicine, Graham Kerr Building, University of Glasgow, Glasgow G12 8QQ, UK
[4]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3DT, UK
[5]AZTI Tecnalia, Herrera Kaia,

Trends in Ecology & Evolution

ecological principles and shared characteristics of species. With these approaches, information can be shared from data-rich to data-limited species, bolstering our assessments of species' status with unprecedented quantitative rigor.

Fundamental Ecological Principles Offer Clues to the Dynamics of Data-Limited Populations

In a classic paper, John Lawton asked 'There may be ten million species of plants and animals on earth; are there really ten million kinds of population dynamics?' [19], making the point that there are common principles that underlie the dynamics of populations of different species. Meta-analyses have confirmed that there is a surprisingly narrow range of biologically plausible population growth and recovery rates [20–23]. Profuse evidence for convergent evolution across ecosystems illustrates that there are a limited number of solutions to ecological problems [24,25]. Therefore, we need tools that make the best use of these fundamental characteristics that are shared among species.

The main strategy is to draw parallels between related species. For example, borrowing information on life-history traits and population connectivity from related species and populations was sufficient to protect specific populations of rockfishes under the US Endangered Species Act in 2010 [26]. Phylogenetic relatedness and geographic proximity have also been used to infer the status of Data Deficient mammals [27]. Recently, nonlinear autoregressive models have also been used to produce more accurate forecasts of single-species population dynamics by pooling simulated time-series data from multiple species in the same ecosystem [28].

Despite this recent progress, we are not yet taking full advantage of the opportunity to combine multiple sources of information and leverage ecological principles when modeling the population dynamics and trajectories of data-limited taxa. Technological advances in modeling population dynamics, using modern computational and statistical methods, provide a powerful convergence between data and mechanistic understanding. These new approaches allow us to separate the main signal in a model from environmental noise and measurement error, inexorably leading to greater predictive power [29–31]. Ecologists seeking to assess data-limited populations now have the opportunity to exploit generalities across species that share evolutionary and ecological characteristics using a seamless combination of theoretical and statistical reasoning.

Integrated Models Use Biological Information to Improve Statistical Model Fits

Here we suggest that modern models of population dynamics, such as **integrated models** (see Glossary) or **state-space models** [32,33], could be used in hierarchical multispecies analyses of population trends. They are a valuable complement to exhaustive data collection and analysis of species in isolation. To show their potential, we focus on the dynamics of marine fish populations, because they have historically led the development of population ecology and life-history theory [34–36]. Globally, exploited fish have the most extensive time-series datasets in existence [12], and models of fish populations routinely incorporate uncertainty [21,37]. Notwithstanding our marine focus, the principles outlined here apply to a wide range of taxa.

The fundamental ecological principle that the dynamics of populations depend on physiology, behavior, life history, and community ecology is central to our premise (Figure 1). Populations and species are phylogenetically related and have spatial structure. Hierarchical statistical models [38,39] that group populations or species by phylogenetic, spatial, and temporal

Gipuzkoa, Spain
[6]Theoretical Ecology Group, Department of Biology, University of Bergen, Bergen 5020, Norway
[7]Institute of Marine Sciences and Department of Applied Mathematics and Statistics, University of California, Santa Cruz, Santa Cruz, CA 95064, USA

*Correspondence:
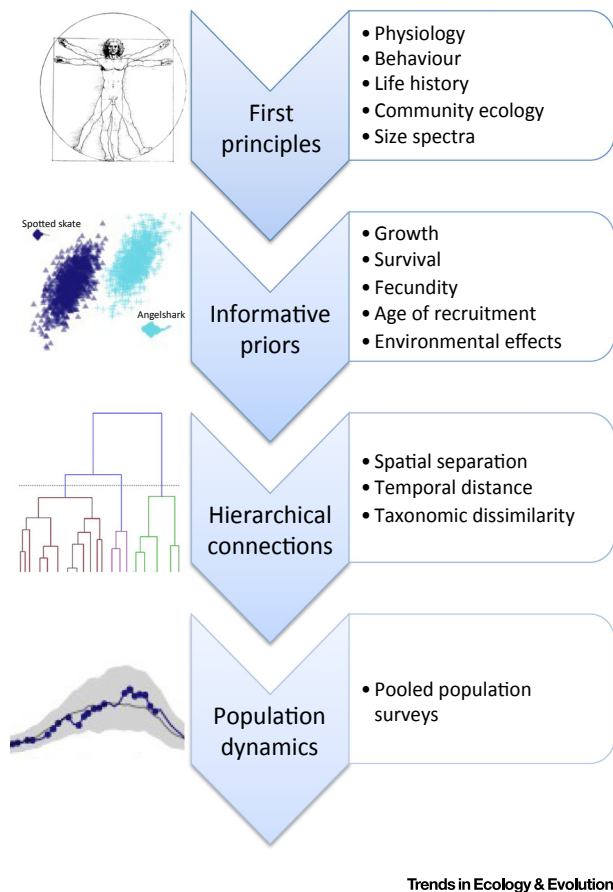holly.kindsvater@rutgers.edu (H.K. Kindsvater).

Trends in Ecology & Evolution

CellPress
REVIEWS

- First principles
  - Physiology
  - Behaviour
  - Life history
  - Community ecology
  - Size spectra

- Informative priors
  - Growth
  - Survival
  - Fecundity
  - Age of recruitment
  - Environmental effects

- Hierarchical connections
  - Spatial separation
  - Temporal distance
  - Taxonomic dissimilarity

- Population dynamics
  - Pooled population surveys

Spotted skate

Angelshark

Trends in Ecology & Evolution

Figure 1. The Ultimate Objective of Assessing the Status of Single Populations Can Be Achieved by Interlocking Different Types of Information. Starting from first principles, we can ask the following questions. How do individual decisions on foraging and reproduction affect survival and births (behavior)? How do tradeoffs and synergies between demographic parameters affect the ability of a population to grow (life-history theory)? What is the multispecies context in which a population is operating (community ecology)? What strategies can be energetically sustained in an ecosystem (size-spectrum theory)? At a second stage, these fundamental results can be combined with data on vital rates and the environment to inform not just the marginal priors but also the correlations between population model parameters. Although these results can bolster the precision and accuracy of individual species assessments, such strengths can be better achieved by pooling information among species. To do this formally it is necessary to acknowledge and represent the hierarchical connections between species and populations, as determined by their degrees of phylogenetic, spatial, and temporal separation. In this way, demographic and environmental parameters can be expressed in terms of baseline values and additional effects whose hierarchy is determined by interpopulation similarity. Creating such parametric connections between populations allows the pooled survey data to disperse their information across all assessments.

Glossary

Bayesian model: a model in which data and prior information concerning unknown parameters or model states are combined using Bayes' theorem to produce a posterior distribution for the parameters and states.
Demographic model: a model of population dynamics that includes birth rates (fertility) and death rates (mortality) and can also include stage or age structure.
Hierarchical model: a model whose structure contains hierarchies of model units (e.g., metapopulation/ population/individual) allowing the variability in the raw data to be apportioned according to the levels of the hierarchy. In a frequentist setting, hierarchical models are often referred to as mixed-effects models.
Initial conditions for model fitting: starting parameter values for a **Markov chain Monte Carlo (MCMC)** fitting algorithm. These candidate values allow the MCMC search to begin from a particular location in parameter space and should become increasingly irrelevant as the MCMC goes through its burn-in phase.
Initial conditions of a state-space model: the initial values of the time series of a state variable (e.g., the size of a given population in the year when observations began). These values could be unknown and would therefore need to be assigned prior distributions.
Integrated model: a model in which multiple streams of data are used either to construct informative prior distributions or to update those into posterior distributions as part of formal model fitting.
Latent (or hidden) variable: a variable that cannot be directly observed but can be inferred through its effect on the behavior of a system. Internal process such as density-dependent survival or reproduction are often latent variables.
Markov Chain Monte Carlo (MCMC): a class of computer-intensive algorithms used for fitting Bayesian models. MCMC attempts to balance the computational efficiency of optimization methods (Markov chain searches of parameter space) with the wealth of information

proximity can integrate these biological principles, resulting in a more comprehensive, realistic model. These statistical tools can revolutionize the way fragmented, error-prone, or incomplete data are used to assess the status of data-limited populations and species.

Data and Information in Models of Population Dynamics

Traditionally, statistical regression methods were reserved for empirical analyses (e.g., linear models) that rely on **model selection** to make inferences, while more **mechanistic models** (e.g., logistic population growth) were predominantly explored using equilibrium and stability analyses. Consequently, it was unclear how mechanistic models should draw information from data, particularly multiple data types (e.g., population time series and demographic rates). Therefore, mechanistic population modeling approaches tended to use data either for parameterization or for validation. Parameterization involved independent analyses of single types of data (e.g., mark–recapture analyses on the probabilities of survival and detection) to estimate **demographic model parameters** and develop corrections for observation errors [33]. Meanwhile, validation compared the predictions of the parameterized models with observed count data.

Trends in Ecology & Evolution

CellPress
REVIEWS

This divide between statistical and mechanistic models has narrowed over the past three decades. Biologists have gained a new statistical lexicon with which they can model the population dynamics of organisms in their environment, including integrated models and state-space models. These models can be constructed in the frequentist or in a **Bayesian modeling** framework [33,40]. Bayesian frameworks offer the opportunity to incorporate different data types to address nontrivial biological hypotheses [39,41].

Modern Models of Population Dynamics Can Incorporate Multiple Data Types

Modern Bayesian modeling [42,43] blends both parameter estimation and validation in a single analysis. Initial parameterization is replaced by Bayesian priors and validation becomes the recursive process of **model fitting**, to confront the model with the data [41]. When multiple types of data are used when constructing priors (e.g., mark–recapture estimates of natural mortality, estimates of age and growth) and fitting the model (e.g., diverse and intermittent data on population size, structure, and environmental covariates), we call it an integrated population model [32] (although we note that integrated models do not have to be Bayesian).

In all integrated population models, **initial conditions**, **state** variables, and parameters are not inherently different in a statistical sense. We no longer discriminate between demographic data for estimating birth and death rates and population models for estimating trends over time. Rather, we must consider the relationship between data and models in the common currency of information. All available types of data can be used simultaneously to inform qualitatively different model parameters in one coherent framework. Hence, having n observations for a particular parameter no longer implies a sample size of n once the parameter becomes part of a population model fitted simultaneously to multiple types of data [44–46].

Modern Statistical Models Can Overcome Problems in Sparse Datasets in Three Ways

The first way to overcome problems of missing data or variables that are unobserved (**latent**) is to differentiate between two sources of variability: process variation and observation error. We call this type of dynamical model a state-space model (Supplement 1 in the supplemental information online). In such formulations, unobserved (latent) quantities can be worth inferring (such as density-dependent survival of juveniles). Others could be **nuisance parameters** or state variables that are modeled by necessity, such as historical population size. State variables are used to model the intrinsic processes controlling the dynamics of the population [31,33,40,47]. These can be more resolved than the data. For instance, in fisheries aggregated data on total biomass or numbers are routinely modeled with an age- or size-structured model that includes parameters describing these processes (e.g., maturation rates of females).

The second way that modern models can overcome the limitations of sparse data is with formalized biological or geographic connections between different species. The existence of hierarchical data structures (e.g., populations nested within species, species nested within genera) prompts the use of **hierarchical model** structures that – by analogy with mixed-effect models in regression [48] – apportion variation in the data to different levels of a hierarchy [38,49]. The hierarchical model structure allows the data-limited parts of the model to borrow information from data-rich parts.

The third way that modern models can overcome the limits of sparse datasets is to use Bayes' theorem [41,43]. In a Bayesian model, the outputs of analysis are **posterior distributions**. In a Bayesian population model, posteriors can be obtained for four distinct parts: (i) the

returned by Bayesian approaches by yielding an approximation of representative random (Monte Carlo) sampling from the posterior distribution. The stochastic components in the implementation of the Markov chain search guarantee that, in the limit, the chain visits different parameter combinations at frequencies proportional to their posterior density.

Mechanistic biological model: a mathematical model containing explicit descriptions of biological subprocesses such as births, predation, or foraging. This is in contrasted to a phenomenological model, where the parameters of the mathematical expressions do not necessarily have a biological interpretation (e.g., a regression line in a scatter plot).

Model fitting: a process by which parameters are estimated (in a frequentist setting) or parameter posterior distributions are summarized (in a Bayesian setting) from data using statistical methods such as least-squares regression or computer intensive computation (e.g., MCMC methods).

Model selection: evaluation and ranking of competing models, via information criteria, likelihood ratio tests, or cross-validation. The process aims to achieve a balance between the quality of model fit and the number of effective degrees of freedom used to achieve it.

Nuisance parameters: parameters that are not scientifically interesting for the problem at hand but nevertheless are required in the analysis. For example, the mean of a sample distribution could be important for inference while the sample variance might be a necessary parameter for some analyses, but less interesting.

Parameters: coefficients or other characteristics specifying the behavior of a mathematical model.

Posterior distribution: the probability distribution that assigns degrees of belief to different candidate values for the parameter obtained by updating the prior distribution with data using Bayes' theorem.

Prior distribution: a distribution of candidate values for a parameter based on expert opinion, biological

ARTICLE IN PRESS

Trends in Ecology & Evolution

CellPress
REVIEWS

parameters; (ii) the unobserved states; (iii) the initial states; and (iv) the observed states. The posterior distribution for a parameter can be interpreted meaningfully. For example, in a linear function the posterior distribution of slope coefficients can be used to answer the question 'is the effect of increasing temperature on fecundity positive, negative, or negligible?' In a Bayesian state-space model, posteriors of unobserved states can be used to reconstruct population trajectories or demographic rates [50]. However, Bayesian models are most powerful when additional knowledge is integrated through informative priors.

Incorporating Ecological and Evolutionary Information into Bayesian Models

The use of informative priors is a natural entry point for multiple types of information in a Bayesian model. Correlations between life-history traits can be used to constrain priors on demographic rates. Life-history theory provides a link between individual-level processes and macroevolutionary patterns. Using information from life-history theory when fitting a Bayesian model of population dynamics can improve the model fit. This is because demographic rates emerge from life-history traits such as maturation and fecundity [35], which evolve predictably according to mortality risk and food availability [51]. For example, high adult mortality selects for earlier maturation [52,53] and high juvenile mortality that depends on size (e.g., competition) can select for larger offspring [54]. Barring constraints, species evolve to allocate their resources to maximize expected reproductive output over a lifetime. There are a limited number of ways to solve this allocation problem, resulting in widespread convergence among life histories. Trait data therefore can provide a basis for **prior distributions** and parameters in an integrated population model.

As one example, we compare fish life-history traits to show the fundamental constraint on viable combinations of juvenile and adult survival (Figure 2). We plot the age of maturation and lifetime fecundity (both transformed) of 204 fish species from 24 clades. These transformed traits serve as proxies for adult and juvenile mortality. The trait space is not entirely filled: there are no species with high juvenile and high adult survival (the upper right corner of Figure 2). We then simulated the population dynamics – with fishing – of four characteristic species at the edges of the extant species' trait space (Supplement 2 in the supplemental information online) to confirm that life-history traits and vulnerability to overfishing are interrelated. Species with late maturation are more vulnerable to overfishing [55,56]. Furthermore, since fishing is size selective in this example (Supplement 2), offspring size and somatic growth rate also contribute to the differences in the relative effects of fishing on population dynamics, such that the angelshark is more vulnerable to fishing mortality than the brown-marbled grouper.

Traditionally, information about demographic parameters is provided to a Bayesian population model as a set of univariate prior distributions. Biologically, this implies no correlations between these parameters. However, as Figure 2 shows, this is not the case; the connection between life-history tradeoffs and demographic parameters motivates the use of multivariate joint priors in population models. Multivariate joint priors, combining information from several demographic parameters, explicitly incorporate covariance among these traits (Box 1).

Fitting a State-Space Model to a Simulated Multispecies Dataset with Informative Priors

To illustrate the power of using multivariate joint priors when fitting population models, we constructed an example. We used a simple stage-structured model to simulate the population dynamics of a group of eight chondrichthyans based on their life-history traits (Supplement 3 in the supplemental information online). We sampled our simulated time series, with observation error, to replicate the nature of the data that come from research trawl surveys. As in an actual

first principles, historical data, or previous analyses.

Size-spectrum theory: a theory relating individual body size to abundance, biomass, and production across trophic levels in a community.

State: a scalar or vector uniquely describing a population at a particular point in time (e.g., the absolute or relative numbers of individuals belonging to different classes of age, size, or sex).

State-space model: a dynamical (i.e., time series) model that comprises two coupled models, one for a biological process (e.g., population dynamics) and one for the method of data collection (e.g., population surveys). A state-space model acknowledges that the biological state variables are partially and imperfectly observed, and the task of model fitting is to reconstruct the underlying state time series, as well as the parameters of the model.

Trends in Ecology & Evolution
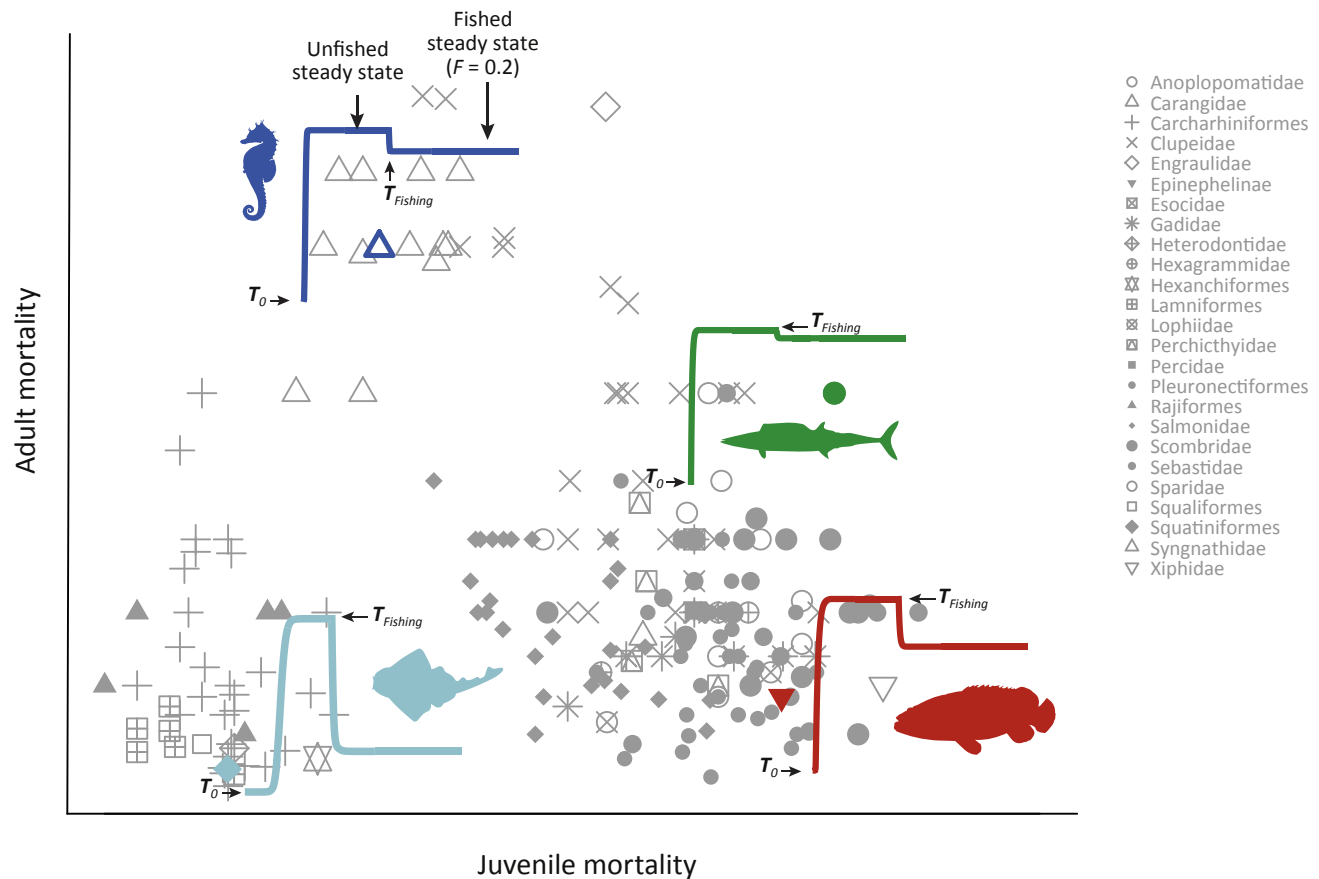
Trends in Ecology & Evolution

Figure 2. Life Histories Underlie Population Dynamics and Resilience. A species' position in trait space can be used to infer its vulnerability to overfishing. To show the range of trait space occupied by fished species, we plot covariance in adult mortality (with age at maturity as a proxy; y-axis) and juvenile mortality (with inverse lifetime fecundity as a proxy; x-axis). Gray points represent fish species in trait space (the figure is based on a similar analysis in [70]). We then address the relative effect of fishing on a characteristic species in each corner of the graph. Colored lines in each quadrant show the connection between a species' position in trait space and its response to equivalent levels of size-selective fishing. Each line represents the deterministic population dynamics of a species with one of four characteristic life histories [55]: precocial (blue) – tiger tail seahorse, *Hippocampus comes*; opportunistic (green) – wahoo, *Acanthocybium solandri*; episodic (red) – brown-marbled grouper, *Epinephelus fuscoguttatus*; and survivor (light blue) – angelshark, *Squatina squatina*. The model is described in Supplement 2 in the supplemental information online and in [55]. The population simulation begins at T_0 at a small population size, then grows until it reaches a steady state. The unfished population size is determined by our assumptions regarding density dependence, which is assumed to be proportional to maximum body length and offspring size, respectively (details in Supplement 2). At T_{Fishing} we add fishing mortality and allow the population to reach a fished steady state. Each population is vulnerable to fishing mortality (F = 0.2) once individuals reach three-quarters of their maximum length.

survey, the abundant species were relatively data rich and the less common species were data limited in our simulated dataset.

We generated joint priors for demographic parameters using multivariate regression [57]. We regressed adult survival, fecundity, and age of maturity against adult body size simultaneously to estimate the size-scaling and empirical covariances among the traits (Box 1). This analysis captured relationships between survival, fecundity, and age at maturity for a given body mass (see Figure I in Box 1), showing that in this group of species, if reproductive investment (fecundity) is high, it is simply not possible to mature early without also trading off survival.

ARTICLE IN PRESS

Trends in Ecology & Evolution

CellPress
REVIEWS

To demonstrate the utility of these multivariate joint priors, we fit population models to simulated survey data for all eight species (Box 2). Following **size-spectrum theory** [58], we assumed that larger species occur at lower densities (and environmental fluctuations act on all the species in the same way). We conducted four fitting experiments, combining naïve versus informed modeling and data-rich versus data-limited species. We fitted naïve models for each species with broad, independent priors. Informed modeling was implemented by fitting hierarchical models to all of the species simultaneously, allowing information to flow from data-rich to data-limited species during model fitting (Box 2; see Supplement 1 for a simple overview of model structure and code). We also provided our informed model with the multivariate joint priors for the demographic traits (Box 1).

For the spotted skate (Figure 3A,C), virtually the same prediction was obtained with either naïve or informed priors. This is because the spotted skate is data rich; that is, given enough data, information from priors will only marginally improve model fits [41]. For the data-limited angelshark (Figure 3B,D) using hierarchical models with informative joint priors improved our estimation of population trends dramatically. Here, the information borrowed from data-rich species, combined with those available for the data-limited species, allows an improved estimate of the true demographic rates (which underlie the observed data).

Box 1. Generating Demographic Priors from Empirical Data

To estimate joint tradeoffs from life-history data (in Table S2 in the supplemental information online), we used pairwise regressions between demographic rates within or between populations, species, or taxa. If we let x, y denote appropriate transformations of the variables [e.g., $x = \log(fecundity)$ or $y = \text{logit}(survival)$], we model survival and fecundity as

$$y \sim N(\mu_y, \sigma)$$
$$\mu_y = a_0 + a_1 x .$$ [I]

The coefficients capture the tradeoff (or synergy) between x and y. We note that Equation 1 does not imply causality so the roles of x and y could have been reversed.

Rather than fitting pairwise models as linear regressions, we envision the relationship between life-history traits as the covariance between them in a bivariate normal distribution (as in Figure I):

$$(y, x) \sim N_2(\mu, \Sigma)$$
$$\mu = (\mu_y, \mu_x),$$
$$\Sigma = \begin{bmatrix} \sigma^2_y & \text{cov}(x, y) \\ \text{cov}(x, y) & \sigma^2_x \end{bmatrix} .$$ [II]

With n fundamental life-history traits, we can envisage $\frac{1}{2}n(n-1)$ possible pairwise relationships. Such a model would be a generalization of Equation 2:

$$(x_1, ..., x_n) \sim N_n(\mu, \Sigma)$$
$$\mu = (\mu_1, ..., \mu_n),$$
$$\Sigma = \begin{bmatrix} \sigma^2_1 & & \text{cov}(x_1, x_n) \\ & \ddots & \\ \text{cov}(x_1, x_n) & & \sigma^2_n \end{bmatrix} .$$ [III]

The life-history relationships are now captured in the covariances of the matrix Σ.

Such models can therefore be used to quantify which life-history profiles are biologically possible and, by incorporating information from different species, allow us to more precisely specify the likely life-history strategy of a data-limited species.
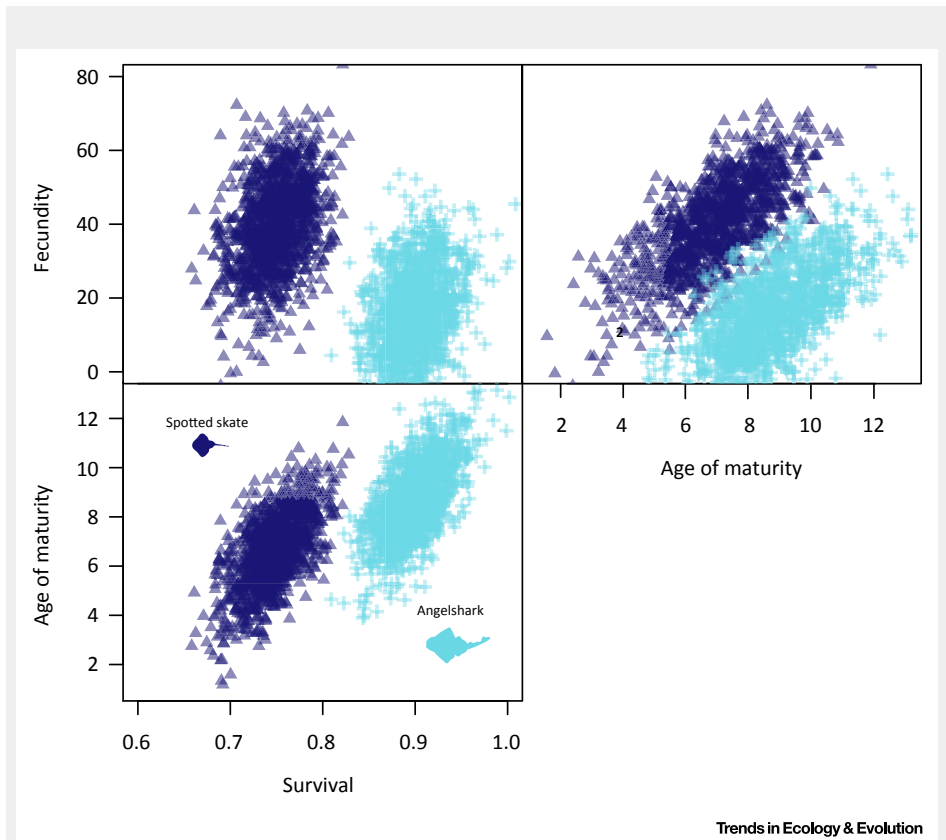
Figure I. Random Life-History Covariances of Angelshark and Spotted Skate. Points are drawn from the multivariate joint probability distributions of fecundity, maturity, and survival for each species. Survival is assumed to be proportional to the inverse of lifespan. The variance and covariance structures connecting the three parameters were estimated via multivariate regression of traits from eight species of chondrichthyan (Table S2.1 in the supplemental information online). Predictions of these models are size specific [this plot depicts plausible life histories for angelshark (*Squatina squatina*) and spotted skate (*Raja montagui*)].

Box 2. Hierarchical Bayesian Models Allow Parameter Estimation for Data-Limited Species

The advantage of Bayesian modeling is that it forces one to be explicit about the model and the data. Incorporating informative priors constrains the model fit to areas that are realistic according to theory. State-space models quantify demographic parameters and hidden states from partial population data [29,50,72]. Hierarchical state-space models allow the estimation of several different parameters from similar taxa, where similarity arises from shared characteristics such as ancestry (e.g. [72]) or geography. Thus, information from data-rich taxa can be shared with data-limited taxa in a hierarchical Bayesian state-space model.

Imagine that the annual rate of survival s_t of a certain age class of a population at time t depends on covariates. Covariates can be: (i) data extrinsic to the model (e.g., temperature data); (ii) autocorrelation (e.g., density dependence can be modeled using previous population densities); or (iii) proxies for unknown covariates (e.g., time can be used as a proxy to capture trends in other variables). These covariates represent scientific hypotheses about what controls population trends.

For example, if T_t denotes temperature in year t, N_{t-1} population size in year $t-1$, and f a flexible function capturing potentially nonlinear survival trends through time, we can write

$$\text{logit}(s_t) = \gamma_0 + \gamma_1 T_t + \gamma_2 N_{t-1} + f(t; \gamma_3). \qquad [\text{I}]$$

ARTICLE IN PRESS

Trends in Ecology & Evolution

The coefficient γ_0 determines baseline survival and is influenced by the priors. The remaining coefficients must be informed indirectly by fitting the model to the population data. If these data contain insufficient information, either the hypotheses (covariates) must be discarded (i.e., returning to an intercept-only model) or additional information is needed.

We can expand this model to represent two species that are likely to respond in similar ways to temperature (e.g., because they are phylogenetically close but in different ecosystems). Assuming evolutionary distance can be quantified (here represented by θ), we write the coefficient of temperature using the following form:

$$\gamma_{i1} = \gamma_{1,0} + \gamma_{1,1}\theta_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma_T), \tag{II}$$

where the coefficients are now fixed effects in the response to temperature and the normally distributed stochastic term quantifies an interspecies random effect relating survival and temperature [73].

Alternatively, we imagine two populations of different species that live in the same ecosystem, experiencing the same environmental drivers at the same time. Spatiotemporal proximity can determine the degree of synchrony in the flexible temporal term of the model $f(t;\gamma_3)$ [74]. Shared ancestry and habitats provide inferential linkages between time series from different populations and allow them to act in a statistically complementary way, so that information flows to data-limited parts of the model from data-rich parts.
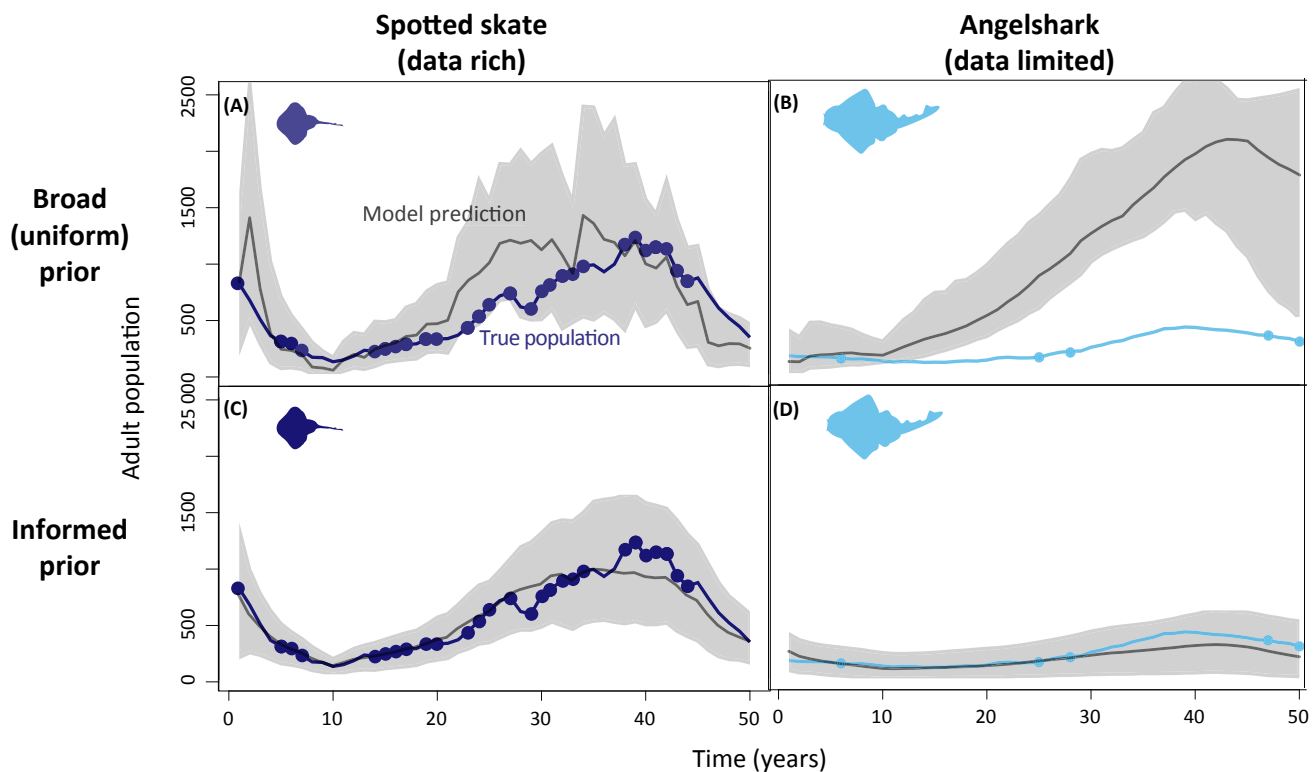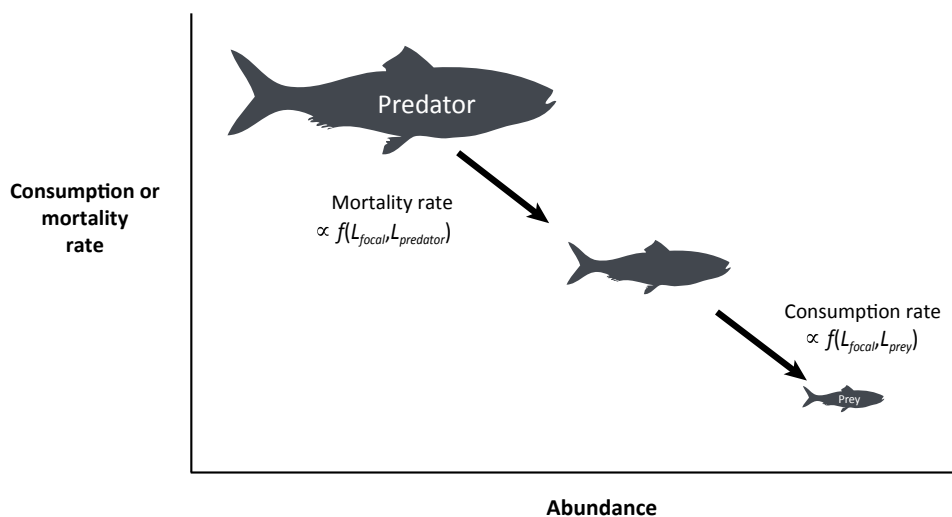

Trends in Ecology & Evolution

Figure 3. Data-Richness and Life-History Covariances Can Inform Predicted Dynamics of Multiple Species at Once. To show the potential of informative priors in model fitting, we fit a Bayesian hierarchical state-space model to multiple species in our simulated dataset. The lines with points are the simulated population trajectory; dots represent time of sample observation (e.g., presence in a trawl survey). The gray line is the model reconstruction; shaded areas are 95% credible intervals. The upper panels (broad, uniform prior) correspond to the situation in which the prior distributions are independent and relatively uninformative; the lower panels (life-history prior) correspond to the situation in which the prior distribution is informed by life history and taxonomic proximity.

Further Sources of Informative Demographic Priors

Construction of informative demographic priors requires considerable intellectual work. The example above used the statistical patterns emerging from observed trait covariation to create informative priors (Box 1). Informative priors can also be constructed by drawing from the burgeoning fields of energetics-based life-history theory [59,60] and community size spectra [61,62]. These fields provide avenues for moving from the qualitative insights in Figure 2 to quantitative demographic priors. We now discuss the exciting possibilities of using first principles from life-history models and community size spectra to generate priors that are sufficiently informative to outperform the empirical multivariate priors in Box 1.

Allocation models that explicitly take into account energetic budgets [63,64] can be used to predict emergent individual life histories including maturation, birth rates, and mortality rates. By incorporating natural variation in available energy, these models predict trait distributions that could be readily transformed into informative priors on demographic traits such as mortality. Individuals with greater energetic demands generally forage more and have larger home ranges; as a result, they frequently have lower survival [65,66]. For example, an energetic model of daily foraging and survival could be used to generate a prior on mortality rate (a process that can be difficult to measure directly) based on the suite of trait combinations observed for a given species [35].

Another source of prior information comes from the predator–prey mass ratio (PPMR). This ratio is a measure of the energetic base available to a predator and can be used to predict reaction norms for traits such as maturation [51]. PPMR is linked to a species' position in its community size spectrum and is thus related to species' abundance and biomass [62,67–69]. Size-spectrum theory can be used to estimate energetic intake (consumption) and natural mortality rate [61]. In Figure 4 we illustrate the relationship

Consumption or mortality rate

Predator

Mortality rate
$\propto f(L_{focal}, L_{predator})$

Consumption rate
$\propto f(L_{focal}, L_{prey})$

Prey

Abundance

Trends in Ecology & Evolution

Figure 4. The Linear Relationship between *Per Capita* Rates of Consumption (of Prey) or Mortality (Due to Predation) and Abundance. From size-spectrum theory, we can assume that there is an inverse relationship between abundance and body mass. Given that, in fish, body mass is a cubic function of length, L, we can assume that $f(L_{focal}, L_{prey})$ or $f(L_{focal}, L_{predator})$ is a ratio or difference function (reviewed in [71]). For simplicity, we have labeled the mortality rate and consumption rate of a focal individual of intermediate size (middle fish), although these rates hold for individuals at the higher and lower trophic levels as well.

Trends in Ecology & Evolution

between predator and prey abundance and the rates of consumption and mortality that follow from size-spectrum theory and PPMR. Notice that body size is implicit in but fundamental to these connections. We suggest that these relationships between body size, physiology, and mortality offer rich information for use in models of population trends.

Concluding Remarks

Scientists and policymakers facing the ongoing challenge of conserving biodiversity with scant data can find utility in modern Bayesian methods, life-history theory, and community ecology (see Outstanding Questions). In addition to continued efforts to survey and assess species, there are underexploited opportunities to apply the methods described here to inform assessments of species' status. Insights from these disparate fields can be incorporated via informative priors when modeling populations (Figure 1 and Boxes 1 and 2). Shared evolutionary and ecological characteristics of populations and species can also be used to formally relate the dynamics of well-studied taxa to data-limited counterparts. While many of these tactics are possible without formal statistical models, the *ad hoc* alternatives would lead to imprecise predictions and lack of standardization between populations and species. By laying out these connections, we hope to fertilize opportunities to study the dynamics of populations or species for which data are limited but threats are ongoing.

Supplemental Information

Supplemental information associated with this article can be found, in the online version, at https://doi.org/10.1016/j.tree.2018.06.004.

References

1. Butchart, S.H.M. *et al.* (2010) Global biodiversity: indicators of recent declines. *Science* 328, 1164–1168
2. Maxwell, S.L. *et al.* (2016) Biodiversity: the ravages of guns, nets and bulldozers. *Nature* 536, 143–145
3. IUCN Red List
4. Collen, B. *et al.* (2008) Broadening the coverage of biodiversity assessments. In *The 2008 Review of the IUCN Red List of Threatened Species* (Vié, J.-C., ed.), IUCN
5. Bland, L.M. *et al.* (2015) Cost-effective assessment of extinction risk with limited information. *J. Appl. Ecol.* 52, 861–870
6. Freese, C.H. (1998) *Wild Species As Commodities: Managing Markets and Ecosystems for Sustainability,* Island Press
7. Nasi, R. *et al.* (2011) Empty forests, empty stomachs? Bushmeat and livelihoods in the Congo and Amazon basins. *Int. For. Rev.* 13, 355–368
8. Ingram, D.J. *et al.* (2018) Assessing Africa-wide pangolin exploitation by scaling local data. *Conserv. Lett.* 11, e12389
9. Schipper, J. *et al.* (2008) The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science* 322, 225–230
10. Bennett, E.L. *et al.* (2002) Hunting the world's wildlife to extinction. *Oryx* 36, 328–329
11. FAO Fisheries and Aquaculture Statistics and Information Service (2013) *SSFIS List of Species for Fishery Statistics Purposes,*

Fisheries and Aquaculture Department, Food and Agriculture Organization of the United Nations. http://www.fao.org/fishery/collection/asfis/en
12. Ricard, D. *et al.* (2012) Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. *Fish Fish.* 13, 380–398
13. Dulvy, N.K. *et al.* (2014) Extinction risk and conservation of the world's sharks and rays. *eLife* 3, e00590
14. Simpfendorfer, C.A. and Dulvy, N.K. (2017) Bright spots of sustainable shark fishing. *Curr. Biol.* 27, R97–R98
15. WWF (2016) *Living Planet Report: Risk and Resilience in a New Era,* WWF
16. McCallum, M.L. (2007) Amphibian decline or extinction? Current declines dwarf background extinction rate. *J. Herpetol.* 41, 483–491
17. North American Bird Conservation Initiative (2016) *The State of North America's Birds,* North American Bird Conservation Initiative
18. Merrick, R. and Methot, R. (2016) *NOAA's Cost of Fish Stock Assessments,* NOAA
19. Lawton, J.H. (1992) There are not 10 million kinds of population dynamics. *Oikos* 63, 337–338
20. Fagan, W.F. *et al.* (2001) Characterizing population vulnerability for 758 species. *Ecol. Lett.* 4, 132–138

Outstanding Questions

In Figure 3 the model fit improves dramatically with a joint prior based on multivariate regression. What is the information content of an evolutionarily informed prior based on theory? In other words, how many survey data points do different priors replace?

In a hierarchical model sharing information from data-rich to data-limited species, we have the opportunity to learn about the hidden intrinsic processes governing their population dynamics. How accurately can we infer the importance of environmental drivers (e.g., resources, mortality) from the life history, metabolism, and position in the size spectrum of a data-limited species?

In Figure 3 we focused on one community of chondrichthyans – mainly skates – with overlapping geographic ranges. Are particular taxa, life-history profiles, or geographic distributions especially conducive to the multispecies, evolutionarily informed fitting approach that we took?

One challenge for policymakers can be gaining intuition about how these model results can be interpreted. Can we create a formal link between the concepts of sensitivity and elasticity in classic population models and the precision of posterior predictions produced by Bayesian state-space models?

21. Ward, E.J. *et al.* (2014) Complexity is costly: a meta-analysis of parametric and non-parametric methods for short-term population forecasting. *Oikos* 123, 652–661

22. Bland, L.M. *et al.* (2015) Predicting the conservation status of data-deficient species. *Conserv. Biol.* 29, 250–259

23. Juan-Jordá, M.J. *et al.* (2015) Population declines of tuna and relatives depend on their speed of life. *Proc. Biol. Sci.* 282, 20150322

24. McGhee, G.R. (2011) *Convergent Evolution: Limited Forms Most Beautiful,* MIT Press

25. Gao, J. *et al.* (2016) Universal resilience patterns in complex networks. *Nature* 530, 307–312

26. Kelly, R.P. *et al.* (2017) Science, policy, and data-driven decisions in a data vacuum. *Ecol. Law Q.* 44, 7

27. Jetz, W. and Freckleton, R.P. (2015) Towards a general framework for predicting threat status of data-deficient species from phylogenetic, spatial and environmental information. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20140016

28. Hsieh, C. *et al.* (2008) Extending nonlinear analysis to short ecological time series. *Am. Nat.* 171, 71–80

29. Horswill, C. *et al.* (2016) Unravelling the relative roles of top-down and bottom-up forces driving population change in an oceanic predator. *Ecology* 97, 1919–1928

30. Andersen, K.H. *et al.* (2017) When in life does density dependence occur in fish populations? *Fish Fish.* 18, 656–667

31. Brooks, M.E. *et al.* (2017) Estimation of individual growth trajectories when repeated measures are missing. *Am. Nat.* 190, 377–388

32. Schaub, M. and Abadi, F. (2011) Integrated population models: a novel analysis framework for deeper insights into population dynamics. *J. Ornithol.* 152, S227–S237

33. Buckland, S.T. *et al.* (2004) State-space models for the dynamics of wild animal populations. *Ecol. Model.* 171, 157–175

34. Beverton, R.J.H. (1987) Longevity in fish: some ecological and evolutionary considerations. In *Evolution of Longevity in Animals* (Woodhead, A.D. and Thompson, K.H., eds), pp. 161–185, Plenum Press

35. Mangel, M. (2017) The inverse life-history problem, size-dependent mortality and two extensions of results of Holt and Beverton. *Fish Fish.* 18, 1192–1200

36. Jensen, O.P. *et al.* (2012) Marine fisheries as ecological experiments. *Theor. Ecol.* 5, 3–22

37. Harwood, J. and Stokes, K. (2003) Coping with uncertainty in ecological advice: lessons from fisheries. *Trends Ecol. Evol.* 18, 617–622

38. Clark, J.S. and Gelfand, A. (2006) *Hierarchical Modelling for the Environmental Sciences: Statistical Methods and Applications,* Oxford University Press

39. Cressie, N.A. *et al.* (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierachical modeling. *Ecol. Appl.* 19, 553–570

40. Newman, K.B. *et al.* (2014) *Modelling Population Dynamics,* Springer New York

41. Hilborn, R. and Mangel, M. (1997) *The Ecological Detective: Confronting Models with Data,* Princeton University Press

42. Cressie, N.A.C. and Wikle, C.K. (2011) *Statistics for Spatio-Temporal Data,* Wiley

43. Gelman, A. *et al.* (2013) *Bayesian Data Analysis.* (3rd edn), Chapman & Hall

44. McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models,* Chapman & Hall

45. Gamerman, D. and Lopes, H.F. (2006) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference,* Taylor & Francis

46. Watanabe, S. (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* 11, 3571–3594

47. Harrison, P.J. *et al.* (2011) Bayesian state-space modeling of metapopulation dynamics in the Glanville fritillary butterfly. *Ecol. Monogr.* 81, 581–598

48. Bolker, B.M. *et al.* (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* 24, 127–135

49. Kéry, M. and Royle, J.A. (2015) *Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS. Volume 1, Prelude and Static Models,* Academic Press

50. Matthiopoulos, J. *et al.* (2014) State-space modelling reveals proximate causes of harbour seal population declines. *Oecologia* 174, 151–162

51. Shuter, B.J. *et al.* (2016) Fish life history dynamics: shifts in prey size structure evoke shifts in predator maturation traits. *Can. J. Fish. Aquat. Sci.* 73, 1–16

52. Abrams, P.A. and Rowe, L. (1996) The effects of predation on the age and size of maturity of prey. *Evolution* 50, 1052–1061

53. Sharpe, D.M.T. and Hendry, A.P. (2009) Life history change in commercially exploited fish stocks: an analysis of trends across studies. *Evol. Appl.* 2, 260–275

54. Kindsvater, H.K. and Otto, S.P. (2014) The evolution of offspring size across life-history stages. *Am. Nat.* 184, 543–555

55. Kindsvater, H.K. *et al.* (2016) Ten principles from evolutionary ecology essential for effective marine conservation. *Ecol. Evol.* 6, 2125–2138

56. Hutchings, J.A. and Kuparinen, A. (2017) Empirical links between natural mortality and recovery in marine fishes. *Proc. Biol. Sci.* 284, 20170693

57. Warton, D.I. *et al.* (2015) So many variables: joint modeling in community ecology. *Trends Ecol. Evol.* 30, 766–779

58. Sprules, W.G. and Barth, L.E. (2016) Surfing the biomass size spectrum: some remarks on history, theory, and application. *Can. J. Fish. Aquat. Sci.* 73, 477–495

59. Nisbet, R.M. *et al.* (2012) Integrating dynamic energy budget (DEB) theory with traditional bioenergetic models. *J. Exp. Biol.* 215, 892–902

60. Stillman, R.A. *et al.* (2015) Making predictions in a changing world: the benefits of individual-based ecology. *Bioscience* 65, 140–150

61. Andersen, K.H. *et al.* (2016) The theoretical foundations for size spectrum models of fish communities. *Can. J. Fish. Aquat. Sci.* 73, 575–588

62. Law, R. *et al.* (2009) Size-spectra dynamics from stochastic predation and growth of individuals. *Ecology* 90, 802–811

63. Houston, A.I. and McNamara, J.M. (1999) *Models of Adaptive Behaviour,* Cambridge University Press

64. Clark, C.W. and Mangel, M. (2000) *Dynamic State Variable Models in Ecology: Methods and Applications,* Oxford University Press

65. Lima, S.L. and Dill, L.M. (1990) Behavioral decisions made under the risk of predation: a review and prospectus. *Can. J. Zool.* 68, 619–640

66. Bull, C.D. *et al.* (1996) Seasonal matching of foraging to anticipated energy requirements in anorexic juvenile salmon. *Proc. Biol. Sci.* 263, 13–18

67. Andersen, K.H. *et al.* (2009) How community ecology links natural mortality, growth, and production of fish populations. *ICES J. Mar. Sci.* 66, 1978–1984

68. Barnes, C. *et al.* (2010) Global patterns in predator–prey size relationships reveal size dependency of trophic transfer efficiency. *Ecology* 91, 222–232

69. Jennings, S. *et al.* (2008) Global-scale predictions of community and ecosystem properties from simple ecological theory. *Proc. Biol. Sci.* 275, 1375–1383

70. Dulvy, N.K. and Kindsvater, H.K. (2017) The future species of Anthropocene seas. In *Conservation for the Anthropocene Ocean* (Levin, P.S. and Poe, M.R., eds), pp. 39–64, Academic Press

Trends in Ecology & Evolution

CellPress
REVIEWS

71. Brodziak, J. *et al.*, eds (2011) *Estimating Natural Mortality in Stock Assessment Applications. NOAA Technical Memo NMFS-F/SPO-119*, US Department of Commerce

72. Dick, E.J. *et al.* (2017) A meta-analysis of fecundity in rockfishes (genus *Sebastes*). *Fish. Res.* 187, 73–85

73. Vincenzi, S. *et al.* (2016) Within- and among-population variation in vital rates and population dynamics in a variable environment. *Ecol. Appl.* 26, 2086–2102

74. Lahoz-Monfort, J.J. *et al.* (2013) Breeding together: modeling synchrony in productivity in a seabird community. *Ecology* 94, 3–10