Title

Towards process-based range modeling of many species

Authors

Margaret Evans[1, 2], Cory Merow[3], Sydne Record[4], Sean McMahon[5], Brian Enquist[2, 6]

1. Laboratory of Tree-Ring Research and Department of Ecology & Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, U. S. A.

2. Department of Ecology & Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, U. S. A.

3. Department of Ecology & Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269, U. S. A.

4. Department of Biology, Bryn Mawr College, Bryn Mawr, Pennsylvania, 19010, U. S. A.

5. Smithsonian Environmental Research Center, Smithsonian Institution Global Earth Observatory, Edgewater, Maryland, 21307, U. S. A.

6. The Santa Fe Institute, Santa Fe, New Mexico, 87501, U. S. A.

Corresponding author: Evans, M. E. K. (mekevans@email.arizona.edu)

Keywords

data fusion, ecological forecasting, hierarchical model, inverse modeling, species distribution models

Abstract

Understanding and forecasting species' geographic distributions in the face of global change is a central priority in biodiversity science. The existing view is that one must choose between correlative models for many species vs. process-based models for few species. We suggest that opportunities exist to produce process-based range models for many species, by using hierarchical and inverse modeling to borrow strength across species, fill data gaps, fuse diverse data sets, and model across biological and spatial scales. We review the statistical ecology and population and range modeling literature, illustrating these modeling strategies in action. A variety of large, coordinated ecological datasets that can feed into these modeling solutions already exist, and we highlight organisms that seem ripe for the challenge.

From trade-off to fusion

Multiple global change factors, particularly climate change, drive the need for better understanding and prediction of species' geographic distributions. Thomas et al. [1] estimated that 15-37% of species are "committed to extinction" by climate change. Their study also brought one of the tools they used, species distribution models (SDMs), to the forefront of biodiversity science. The appeal of SDMs is their ability to describe spatial occurrence patterns (presence-only or presence-absence data) in terms of environmental predictors (temperature, precipitation, etc.) and then forecast a species' future geographic distribution based on expected environmental change. Because occurrence data are widely available, forecasts can be made for hundreds or thousands of species [2-3]. However, occurrence patterns are the net result of many underlying processes and factors, operating across spatial and biological scales, including physiology, demography, dispersal, biotic interactions, disturbance, adaptive evolution, *etc*., in addition to and interacting with climate. Thus ecologists have increasingly recognized that SDMs are incomplete, questioning the reliability of their forecasts [4-8]. Contemporary correlations between distribution and climate are expected to unravel, given change in the complex, underlying drivers of distributions.

The alternative to correlative SDMs is to use mechanistic or **process-based range models** to project dynamics into the future. Pioneering examples include range models based on physiology [9,10], demography [11], and phenology [12,13]; more recent examples use metapopulation [14,15], metacommunity [16], dynamic vegetation [17,18], and food web models [19,20]. Process-based approaches are expected to yield better prediction of range dynamics because they explicitly model the biology underpinning distributions (*e.g.*, biotic interactions, disturbance, climate, and interactions between them), thus they have the potential to disentangle different drivers of range patterns and allow them to recombine in ways that may better reflect responses to global change. Further, process-based range models offer greater insight into the forces governing geographic ranges, which remain poorly understood [21], and they produce output that is more useful than correlative SDM output. For example, demographic models can yield a detailed understanding of population dynamics and its drivers [22,23], including sensitivities of various demographic processes to climate, and thus predictions of spatial variation in population growth rate, time-to-extinction, or other metrics of threat associated with anticipated climate change. However, this kind of process-based approach requires detailed observations, across years, sites, and throughout the life cycle. Such data are available for but a fraction of all species. Further, process-based range modeling requires more effort.

This trade-off, between modeling few species in detail vs. many species superficially (Figure 1a), reflects overarching trade-offs in modeling, between precision, generality, and realism, identified a half century ago by Levins [24,25]. Addressing this trade-off, and making process-based range models for many species, is a grand challenge in biodiversity science. We suggest that traction can be gained on this challenge, regardless of the study group and process-based range modeling framework, by using modeling strategies that borrow strength across species, fill data gaps, fuse different data sources, and cross scales (Figure 1b). We review the recent literature in statistical ecology and population and range modeling with an eye towards the goal of process-based range modeling *en masse*, highlighting promising approaches already in use. We also highlight the data that are available to feed into those modeling strategies, and groups that seem best suited for the challenge of *en masse* process-based range modeling.

Models that borrow strength, fill gaps, fuse data, and cross scales

Among the hurdles that obstruct *en masse* process-based range modeling, we focus here on two: (1) the data-hungry nature of process-based models, *i.e.*, the need for estimates of **niche parameters** for many species, and (2) the need to account for processes occurring across a range of spatial and biological scales. The first barrier is partially overcome by the increasing number of large ecological datasets available (Box 1), some of which offer estimates of niche parameters [26-30], others of which can be used to estimate these or fit process-based models [31,32]. In the remainder of this section, we describe how **hierarchical models** and **inverse modeling** offer ways to make best use of these data sources, by borrowing strength across species, filling data gaps, and combining different kinds of data, allowing one to gain inference across biological and spatial scales.

Borrowing strength, filling data gaps. Creating process-based range models for many species is challenging because there are many parameters to be estimated. One solution is to *borrow strength across species*: estimate critical niche parameters for all species in a single statistical model, treating species as random effects (Box 2). Rüger et al. [33] nicely illustrate this approach with respect to individual growth: they used a hierarchical model to estimate how light environment and individual size influence the growth of 171 tree species on Barro Colorado Island. The down-side of this approach is that parameter estimates for rare species can be overly influenced by the data for the most abundant species [34,35]. But this can be counteracted by incorporating species-level data on functional traits (Box 2), which have been shown to explain variation in vital rates, life history strategy, and range boundaries [36-38]. In Rüger et al.'s [33] model, species-level parameters, *i.e.*, a species' intrinsic growth rate, and the size- and environment-dependence of its growth, were predicted by functional trait and phylogenetic data. This improved inference of growth parameters, particularly for rare species. Vital rates (survival, growth, and fertility) inferred in this way for many species could then be used to form population models to project range dynamics into the future (Box 2). Uriarte et al. [39] used such a hierarchical, trait-mediated approach to project how the composition of a second-growth tropical forest (171 tree species) might be influenced by climate change.

Further, functional trait, life history, and phylogenetic data, incorporated into a multilevel modeling framework, can be used to impute (*i.e.*, interpolate) the niche parameters so critical for process-based forecasting for those species missing data. Many data types offer information that can be used to impute niche parameters, such as body size (*i.e.*, via allometric or metabolic scaling laws), phylogenetic information, and trait-trait correlations, much in the way that functional trait data are being imputed in increasingly sophisticated ways, and for large numbers of species [40-42]. Such gap-filling methods are receiving more attention now, but the underlying principles are not new: Caswell [43] simulated vital rates of a cetacean for which data were lacking based on data from other monovular mammals, rescaled by age at first reproduction. The emerging availability of comparative data on functional traits [44], life histories [26-29], and trophic niches [30], combined with phylogeny [45] (see Box 1), brings these gap-filling strategies increasingly within reach.

Joint Species Distribution Models (jSDM) offer another example of borrowing strength across species. Information can be gained by modeling species as a joint distribution because their responses are correlated via interactions (*i.e.*, competition, facilitation) and convergent dependencies on the environment [46-48]. Latent variable models (also hierarchical models) have been used to model abundance data for many species at a time, accounting for correlations among species using relatively few parameters [46]. A time-evolving version of abundance-based jSDMs, using simple population or metapopulation models, could be a powerful tool for modeling range change.

A second solution to the problem of estimating many parameters is the use of *inverse modeling* [49,50] (see Glossary). For example, occurrence data can be used to infer the values of key niche parameters (Figure 2b). Higgins et al [51] adopted a physiological model of plant growth, defined which physiological rates were sensitive to the environment and the functional form of those sensitivities, then fit parameters to presence-absence data for 22 tree species. Thus detailed physiological data were not needed to build physiology-based range models (Figure 2b). Hartig et al. [52] used inverse modeling to fit an individual-based model of tropical forest gap dynamics. Inverse modeling could be used to fit process-based range models explicitly accounting for biotic interactions, using time series data on the abundances of interacting species. As with any process-based approach, it is critical that the process model used is appropriate; that is, it requires *a priori* knowledge of which processes might be most important in governing range dynamics.

Fusing Data. In the absence of costly, time-consuming collection of additional data, the challenge of *en masse* process-based range forecasting can only be met by careful fusion of diverse data: occurrence (presence-only and presence-absence) and abundance; demographic, phenological, functional trait, and phylogenetic; experimental and observational; remotely sensed data, genetic data, natural archives and fossils. These data contain complementary information, sampled at different spatial, temporal, or taxonomic scales. Combining data sources also can address weak parameter identifiability, *i.e.*, caused by collinearity among climate predictors [53].

One promising way to fuse data, already used in population and range modeling, is **hidden process models** [54-56] (another hierarchical model). In hidden process models, the observation process is parsed from the latent (unknown) dynamics of the system, so that multiple data sets can be treated as observations arising from a common process of interest. One example of data fusion using a hidden process model is the inference of individual tree growth (an important niche parameter) from both tree-ring and forest inventory data [57], which have annual resolution and long time series vs. strong spatial representation, respectively – complementary temporal and spatial characteristics. In another example, called "dynamic range modeling", a hidden process model was used to combine presence-absence data and abundance time-series data, and from those data, infer abundance on a grid through time, under an explicit model of demography (the Ricker model) and dispersal (a dispersal kernel) [58]. **Integrated population models** [59] also use hidden process models to fuse two types of data – mark-recapture data and abundance time series – to infer the vital rates underlying population dynamics (Figure 2c). Multi-site and spatially-explicit integrated population models [60-62] suggest that process-based range modeling using integrated population models is on the horizon.

A key fusion to be made is between distribution data and data on individual performance (physiological, phenological, or demographic data). Unfortunately, presence-only data are of limited use for this purpose: they do not allow inference of absolute probability of presence or abundance [63,64]. Presence/absence data, however, can be used to infer abundance [65]. The solution then is to use presence-absence or abundance data for range modeling. In a "metamodel" approach [66], a distribution-scale model relying on presence-absence data is conditional upon a submodel. The submodel might include experimental data on the limits of the niche, or a process-based range model (*e.g.*, based on phenology). The next step would be to scale up the metamodel approach to handle many species at once, making a presence-absence based jSDM conditional upon submodel data for the same species. Key to the metamodel approach is the delineation of a biologically-justified way of linking the two data types, *i.e.*, a scaling function that relates 1) the probability that the population growth rate is positive to 2) the probability of presence (from presence-absence data). Also critical is the criterion for weighting the likelihoods associated with the different data types.

Where abundance data are available, the focus of range modeling can shift from occurrence towards abundance data [7,67]. Distribution models of abundance can then be fused with demographic, phenological, or physiological data. Integrated population models [59-62] applied to birds exemplify this: abundance data on birds are widely collected by amateurs and relatively simple demographic models can be fitted, in combination with observations on vital rates (*i.e.*, mark-recapture data). Given the amount of data collected on birds (especially in North America and Europe) the development of multi-species integrated population models is a promising area for future research.

A final approach for combining data is the use of inverse modeling. Hartig et al. [68] envision a wide range of data types – vegetation inventories, trait data, distribution data, remote sensing, eddy flux measurements, and paleorecords – combined together via Bayesian inverse modeling, with "bottom-up" data on parameters (*e.g.*, metabolic rates) captured in informative priors, and "top-down" data (*e.g.*, distribution data) captured in the likelihood function (similar to Figure 2c). Bayes' theorem combines the two (priors and likelihood) to generate a posterior distribution of model parameters reflecting all data sources. The use of informative priors addresses the problem of equifiniality (or "non-identifiability") associated with inverse parameter estimation in complex, process-based models, *i.e.*, that multiple combinations of parameter values can lead to the same observed data. Gonzalez and Martorell [69] exemplify the spirit of Bayesian inverse modeling: they used the fit of observed size structure and density data across chronosequences of environmental change to infer vital rates and the impact of directional environmental change on vital rates. Alternative values of parameters were eliminated by using *a priori* biological knowledge – analogous to strong priors. More recently, Gonzalez et al. [70] showed how inverse modeling can be used to infer vital rates by combining time series data on population size, population structure, and limited individual-level data.

Crossing Scales. The processes influencing a species' abundance and distribution operate at a variety of scales – from microsite, to site, to landscape, to the macroecological scale of atmospheric circulation patterns and ocean currents [71-73]. Indeed, recent macrosystems research has highlighted the potential for cross-scale interactions and cross-scale emergence [72,73]. Even the inference of population-level parameters from individual-level data can be difficult [74,75]. For all these reasons, process-based range modeling relying on data at just one scale (*e.g.*, individual-level performance) is unlikely to be robust across scales.

A central challenge then is to build process-based range models that operate across spatial and biological scales. Key to crossing scales will be the modeling strategies described above that combine different kinds of data – hierarchical modeling and inverse parameter estimation. Hierarchical models are, beyond their ability to combine data at different scales, inherently capable of capturing the hierarchical heterogeneity of ecological systems, including both spatial and evolutionary nestedness, thus they can provide better inference of fixed (climate) effects, better propagation of uncertainty, *etc* [34,35]. Inverse parameter estimation, combined with forward parameter estimation, allows data at multiple scales to inform and better constrain critical niche parameters (Fig 2c).

We envision two approaches to crossing scales aimed at multi-species modeling. First is a "scaling down" approach – starting with models at the landscape scale, and adding more detailed data about individual species or lower-scale processes. Broad spatial scale, process-based models that can account for multiple species include dynamic vegetation models [76] (DVMs), macroecological models [16], metapopulation models [14,15], and food web models [19,20]. DVMs, for example, range in scale from a forest stand to a landscape or region to the global coupled vegetation-atmosphere system [76-78]. For example, Vanderwel et al. [17] parameterized an individual-based forest gap model with repeat measurements of $>10^6$ trees, and then modeled growth, mortality, and recruitment at the scale of half a continent. Strategies that permitted individual-based modeling (considered computationally-intensive) at this scale included grouping species into plant functional types and individuals into cohorts. This model was then used to forecast the distributions of major forest types under projected warming [79]. To derive species-specific forecasts of geographic distributions, which would allow land managers to anticipate and adapt to climate change, the challenge then is to parameterize such a model at the species level.

Second is a "scaling up" approach – merging individual- or population-level models (physiology, phenology, or demography) into landscape-scale models or SDMs. For example, Merow et al. [11] used sparse demographic data for a shrub, sampled across environmental gradients, to build integral projection models, and then map predicted asymptotic population growth rate. Their predictions corresponded remarkably well with presence-absence data. The challenge then is to build such a model for many species, and incorporate distribution data in model fitting. One possibility for doing so is a hidden process state space modeling framework that explicitly models abundance and the processes governing its dynamics [58]. Implemented in a hierarchical modeling framework, it has the potential to integrate lower-level data that relate to abundance – demographic data on individual survival and fecundity, and in turn data on phenology or physiology that relate to vital rates. A second possibility is to combine forward and inverse parameter estimation to fuse individual-level and population-level (abundance) data, similar to integrated population modeling (Fig 2c).

Concluding Remarks

Robust process-based forecasting *en masse* will require flexible modeling tools that enable the fusion of multiple sources of data, relevant to ecological processes operating across a range of scales, and the estimation (or imputation) of process parameters for many species. Hierarchical and inverse modeling can help meet these needs. Testing these modeling approaches with data – both real and simulated – is of the utmost priority. What are the limits of "borrowing strength" across species and estimating niche parameters with the help of coarser forms of information (*e.g.*, phylogenetic, life history, and functional trait)? As an example, abundant data are collected on butterflies in the United Kingdom, compared to less in the United States. Could process-based range models based on presence-absence and abundance data of U.K. butterflies (building on [58] and [80]) be combined with coarser information from butterflies in both places to develop range models for U.S. butterflies? Further exploration of the variety of ways that different data sources can be combined is critical as well – via informative priors, by weighting likelihoods, through the combination of forward and inverse parameter estimation, *etc.* Model parsimony is an open question: how complex do models need to be to adequately capture range dynamics? Is an individual-based model necessary, or will a metapopulation model suffice? Addressing this question will require range modelers to compare the performance of process-based range models of differing complexity for the same organisms, which has not yet been done. Related to the issue of model complexity is the problem of computing power. Recent advances in flexible implementation of hierarchical models [81,82], strategies for writing efficient code [83], and recommendations for scaling up [76] offer hope, but will require further progress to fully realize the vision of process-based range modeling *en masse*. Box 3 highlights these and additional unresolved questions needing attention, including the choice of predictors, the identification of range-limiting factors that vary over space, and accounting for biotic interactions and evolutionary processes.

While we view the challenge of creating process-based models of species' ranges *en masse* positively in the sense that an alternative to correlative, occurrence-based SDMs is clearly needed, this approach is not without pitfalls. Correlative SDMs are able to make reasonable-looking predictions simply by fitting occurrence data well. The bar is higher for process-based range models: the key processes influencing species' distributions must be correctly specified, or the models fail. In other words, process-based models have the potential to make better forecasts, but they also have the potential to make worse forecasts. Among the few direct comparisons between process-based and correlative range models, results thus far are mixed (Online Supplementary Material Table S1). These comparisons typically use fit to the current distribution as the metric of success (*i.e.*, model performance). A better comparison might be the ability of distribution models (correlative vs. process-based) to project range change, or underlying performance, under changing environmental conditions. Evaluating the performance of forecasts, *i.e.*, projections about the future, is problematic. For this reason, opportunities to validate distribution models or their subcomponents using hindcasting exercises should be given high priority (see [84]). Ultimately, a process-based approach offers the chance for improved understanding, which is critical for forecasting in a world requiring robust extrapolation, *i.e.*, climatic conditions outside the historical range of variation and no-analogue climates [85]. At the same time, distribution data will continue to play a key role, even in process-based range modeling, because they reflect all the processes that shape distributions – across spatial and temporal scales. Those processes or factors not captured well by plot-level or experimental data (*e.g.*, dispersal when it is a slow process, spatial variation at a scale greater than available data) are reflected in distribution data. And while presence-only data cannot inform absolute abundance, they are useful for diagnosing the forces that shape distributions, offering clues as to what processes are necessary to explain species' distributions.

Along with general ecosystem models [86,87], process-based range modeling of many species is part of a trend to both scale up and add detail to models of complex ecological systems. Opportunities exist to produce process-based range models for many species – hierarchical and inverse models can fuse multiple sources of data, cross scales, borrow strength across species, and fill data gaps. Building on this foundation of strategies, ecologists can expand upon the existing diversity of range models, and progress towards process-based range modeling *en masse*. Though process-based range models will continue to be more difficult to implement for the foreseeable future, even with the strategies that we suggest, the need to advance ecological forecasting tools justifies the effort.

Acknowledgements

The authors benefitted from discussions surrounding an Organized Oral Session titled “Novel Approaches for Process-Based Species Distribution Models” at the 2014 annual meeting of the Ecological Society of America, and a workshop “Ecological insights from integrating diverse data with hierarchical models” sponsored by the Berkeley Initiative on Global Change Biology in October, 2014. In particular, the authors gratefully acknowledge discussions with Tom Miller, Rob Salguero-Gomez, Mark Vanderwel, Matt Talluto, Jorn Pagel, Frank Schurr, Katy Prudic, Andrew Rominger, Dominic LaRoche, Regis Ferriere, Kent McFarland, Dave Moore, Jeff Oliver, and Noah Charney. John Shaw (Interior West-Forest Inventory and Analysis, USDA Forest Service) and Andrew Gray (Pacific Northwest-Forest Inventory and Analysis, USDA Forest Service) provided data on the abundance of U. S. tree species. We thank Ben Olimpio for his assistance producing panel *ii* in Box 2. CM acknowledges funding from USDA-NRI grant 2008-35615-19014 and DEB-1137366.

Glossary

Hidden process model: examples most familiar to ecologists are mark-recapture modeling [56] and occupancy modeling [54]. Hidden process models track two time series, one of the true, underlying dynamics of a system, and a second consisting of observations of the true state of the system [55,56]. For example, a species' true abundance through time might be the variable of interest; the observations might be counts of that species. Because the true abundance is unknown, it is referred to as *latent variable*. *Hidden Markov models* and *state space models* are both examples of hidden process models, with the former governed by a Markov process and the latter governed by a process that may or may not be Markovian [56].

Hierarchical model: in a hierarchical model, the joint probability distribution of data and parameters is decomposed into multiple conditional probabilities. This may take the form, for example, of a data model, process model, and parameter model, such that the model itself has a hierarchical structure [35]. Alternatively, the data may have a nested or hierarchical structure, so that parameters (*e.g.*, regression coefficients) vary among groups or subgroups; that variation is modeled probabilistically [33]. Taking the example in Box 2, the growth model presumes that individual tree growth increments (G) vary according to a normal distribution, with growth dependent on the tree's size (B) at the previous time step and July maximum temperature (T). The trait-based parameter model presumes that the intercept and coefficients in the growth model (a, b, c) vary among species or subspecies according to functional traits. Hierarchical models can take a great variety of forms, including hidden process models, latent variable models, measurement error models, and others [100].

Integrated Population Model: simultaneous analysis of individual-level demographic data (mark-recapture, fecundity) and population-level abundance data to infer vital rates and population dynamics [59]. Inference is based on the joint likelihood, which is created by the multiplication of the likelihoods from each dataset. Key to the fusion of these different datasets is that one or more parameters are shared across the different likelihoods, typically via a demographic model such as a Leslie or Lefkovitch matrix model. Integrated population models use a state space model, or hidden process model, as described above, to parse error associated with observation from process variability [59].

Inverse modeling: an approach to parameter estimation common across many scientific disciplines. Inverse modeling finds values of parameters that are consistent with the observed data, by varying the values of unknown parameters over a broad range of parameter space and producing simulated observations [49]. From these simulations, sets of parameter values are selected that produce outcomes that are similar to the observed data (*e.g.*, occurrence data). In practice, efficient sampling of high-dimensional parameter space is accomplished through stochastic simulation (*e.g.*, a Markov chain Monte Carlo sampling algorithm) [50].

Niche parameters: Parameters that influence the long-term persistence of a population or metapopulation under varying environmental conditions. This includes, for example, physiological, phenological, and demographic parameters and their environmental sensitivities, parameters governing metapopulation persistence (extinction-colonization dynamics), as well as parameters influencing a species' place in successional dynamics, coexistence with other species in a community, and role within a food web.

Process-based range model: A model of a species' geographic range or distribution that specifies underlying processes or mechanisms. This might include physiological, phenological, or demographic processes, for example, or metapopulation, metacommunity, or food web processes.

References

1 Thomas, C. D. et al. (2004) Extinction risk due to climate change. *Nature* 427, 145-148

2 Jetz, W. et al. (2007) Projected impacts of climate and land-use change on the global diversity of birds. *PLoS Biol.* 5, 1211-1219

3 Thuiller, W. et al. (2014) Consequences of climate change on the tree of life in Europe. *Nature* 470, 531-534

4 Guisan, A. et al. (2006) Making better biogeographical predictions of species' distributions. *J App. Ecol.* 43, 386-392

5 McMahon, S. M. et al. (2011) Improving assessments of climate-change impacts on global biodiversity. *Trends Ecol. Evol.* 26, 249-259

6 Dormann, C. F. et al. (2012) Correlation and process in species distribution models: bridging a dichotomy. *J. Biogeog.* 39, 2119-2131

7 Ehrlen, J. and W. F. Morris. (2015) Predicting changes in the distribution and abundance of species under environmental change. *Ecol. Lett.* 18, 303-314

8 Normand, S. et al. (2014) Demography as the basis for understanding and predicting range dynamics. *Ecography* 37, 1149-1154

9 Buckley, L. B. (2008) Linking traits to energetics and population dynamics to predict lizard ranges in changing environments. *Am. Nat.* 171, E1-E19

10 Kearney, M. R. et al. (2009) Mechanistic niche modeling: combining physiological and spatial data to predict species' ranges. *Ecol. Lett.* 12, 334-350

11 Merow, C. et al. (2014) On using integral projection models to generate demographically driven predictions of species' distributions: development and validation using sparse data. *Ecography* 37, 001-017

12 Chuine, I. and E.G. Beaubien. (2001) Phenology is a major determinant of tree species range. *Ecol. Lett.* 4, 500-510

13 Chapman, D. S. et al. (2014) Phenology predicts the native and invasive range limits of common ragweed. *Global Change Biol.* 20, 192-202

14 Garcia-Valdez, R. et al. (2013) Chasing a moving target: projecting climate change-induced shifts in non-equilibrial tree species distributions. *J. Ecol.* 101, 441-453

15 Yackulic, C. B. et al. (2014) To predict the niche, model colonization and extinction. *Ecology* 96, 16-23

16 Mokany, K. et al. (2012) Dynamic macroecology and the future for biodiversity. *Global Change Biol.* 18, 3149-3159

17 Vanderwel, M. C. et al. (2013) Climate-related variation in mortality and recruitment determine regional forest-type distributions. *Global Ecol. Biogeog.* 22, 1192-1203

18 Gutierrez, A. G., R. S. Snell, and H. Bugmann. (2016) Using a dynamic forest model to predict tree species distributions. *Glob. Ecol. Biogeogr.* 25, 347-358.

19 Pellissier, L. et al. (2013) Combining food web and species distribution models for improved community projections. *Ecol. Evol.* 3, 4572-4583

20 Albouy, C. et al. (2014) From projected species distribution to food-web structure under climate change. *Global Change Biol.* 20, 730-741

21 Sexton, J. P. et al. (2009) Evolution and ecology of species range limits. *Ann. Rev. Ecol. Evol. Syst.* 40, 415-436

22 Doak, D. F. and W. F. Morris. (2010) Demographic compensation and tipping points in climate-induced range shifts. *Nature* 467, 959-962.

23 Ozgul, A. et al. (2010) Coupled dynamics of body mass and population growth in response to environmental change. *Nature* 466, 482-485

24 Levins, R. (1966) The strategy of model building in population ecology. *Amer. Sci.* 54, 421-431

25 Evans. M. R. (2012) Modelling ecological systems in a changing world. *Philos. Trans. R. Soc. B: Biol. Sci.* 367, 181-190

26 Jones, K. E. et al. (2009) PanTHERIA: a species-level database of life-history, ecology, and geography of extant and recently extinct mammals. *Ecology* 90, 2648

27 Myhrvold, N. P. et al. (2015) An amniotic life-history database to perform comparative analyses with birds, mammals, and reptiles. *Ecology* 96, 3109

28 Salguero-Gomez, R. et al. (2015) The COMPADRE plant matrix database: an open online repository for plant demography. *J. Ecol.* 103, 202-218

29 Salguero-Gomez, R. et al. (2015) COMADRE: a global database of animal demography. bioRxiv doi: http://dx.doi.org/10.1101/027821

30 Wilman, H. et al. (2014) EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology* 95, 2027.

31 Bechtold, W. A. and P. L. Paterson, eds. (2005) The enhanced Forest Inventory and Analysis program: national sampling design and estimation procedure. General Technical Report SRS-80. USDA Forest Service, Southern Research Station, Asheville, North Carolina, USA.

32 Sullivan, B. L. et al. (2014) The eBird enterprise: an integrated approach to development and application of citizen science. *Biol. Conserv.* 169, 31-40

33 Rüger, N. et al. (2012) Functional traits explain light and size response of growth rates in tropical tree species. *Ecology* 93, 2626-2636

34 Gelman, A. & J. Hill. (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, NY, USA.

35 Cressie, N. et al. (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecol. Appl.* 19, 553-570

36 Adler, P. B. et al. (2014). Functional traits explain variation in plant life history strategies. *Proc. Natl. Acad. Sci. U.S.A.* 111, 740-745

37 Stahl, U. et al. (2014) Predicting species' range limits from functional traits for the tree flora of North America. *Proc. Natl. Acad. Sci. U.S.A.* 111, 13739-13744

38 Visser, M. D. et al. (2016) Functional traits as predictors of vital rates across the life-cycle of tropical trees. *Funct. Ecol.* 30: 168-180

39 Uriarte, M. et al. (2016) A trait-mediated, neighborhood approach to quantify climate impacts on successional dynamics of tropical rainforests. *Funct. Ecol.* 30: 157-167

40 Kissling, W. D. et al. (2014) Establishing macroecological trait datasets: digitalization, extrapolation, and validation of diet preferences in terrestrial mammals worldwide. *Ecol. Evol.* 4, 2913-2930

41 Penone, C. et al. (2014) Imputation of missing data in life-history trait datasets: which approach performs the best? *Methods Ecol. Evol.* 5, 961-970

42 Schrodt, F. et al. (2015) BHPMF–a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. *Global Ecol. Biogeog.* 24, 1510-1521

43 Caswell, H. et al. (1998) Harbor porpoise and fisheries: an uncertainty analysis of incidental mortality. *Ecol. Appl.* 8, 1226-1238

44 Kattge, J. et al. (2011) TRY – a global database of plant traits. *Global Change Biol.* 17, 2905-2935

45 Hinchliff, C. E. et al. (2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc. Natl. Acad. Sci. U.S.A.* 112, 12764-12769

46 Warton, D. I. et al. (2015) So many variables: joint modeling in community ecology. *Trends Ecol. Evol.* 30, 766-778

47 Clark, J. S. et al. (2014) More than the sum of the parts: forest climate response from joint species distribution models. *Ecol. Appl.* 24, 990-999

48 Fithian, W. et al. (2014) Bias correction in species distribution models: pooling survey and collection data for multiple species. *Methods Ecol. Evol.* 6, 424-438

49 Grimm, V. and S.F. Railsbeck. (2005) Individual-based modeling in ecology. Princeton University Press, Princeton, NJ, USA.

50 Hartig, F. et al. (2011) Statistical inference for stochastic simulation models – theory and application. *Ecol. Lett.* 14, 816-827

51 Higgins, S. I. et al. (2012) A physiological analogy of the niche for projecting the potential distribution of plants. *J. Biogeog.* 39, 2132-2145

52 Hartig, F. et al. (2014) Approximate Bayesian parameterization of a process-based tropical forest model. *Biogeosciences* 11, 1261-1272

53 Dormann, C. F. et al. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36: 27-46

54 Royle, J. A. et al. (2007) Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecol. Monog.* 77, 465-481

55 Marion, G. et al. (2012) Parameter and uncertainty estimation for process-oriented population and distribution models: data, statistics, and the niche. *J Biogeog.* 39, 2225-2239

56 Gimenez, O. et al. (2013) Estimating demographic parameters using hidden process dynamic models. *Theor. Pop. Biol.* 82, 307-316

57 Clark, J. S. et al. (2007) Tree growth inference and prediction from diameter censuses and ring widths. *Ecol. Appl.* 17: 1942-1953.

58 Pagel, J. and F. M. Schurr. (2011) Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecol. Biogeog.* 21, 293-304

59 Schaub, M. and F. Abadi. (2011) Integrated population models: a novel analysis framework for deeper insights into population dynamics. *J. Ornith.* 153, S227-S237

60 McCrea, R. S. et al. (2010) Multi-site integrated population modeling. *J. Agr. Biol. Env. Stat.* 15, 539-561

61 Chandler, R. B. and J. D. Clark. (2014) Spatially explicit integrated population models. *Methods Ecol. Evol.* 5, 1351-1360

62 Robinson, R. A. et al. (2014) Integrating demographic data: towards a framework for monitoring wildlife populations at large spatial scales. *Methods Ecol. Evol.* 5, 1361-1372

63 Hastie, T. and W. Fithian. (2013) Inference from presence-only data - the ongoing controversy. *Ecography* 36, 864-867

64 Renner, I. W. et al. (2015) Point process models for presence-only analysis. *Methods Ecol. Evol.* 6, 366-379

65 Hwang, W.-H. and F. He. (2011) Estimating abundance from presence/absence maps. *Methods Ecol. Evol.* 2, 550-559

66 Talluto, M. et al. (2015) Cross-scale integration of knowledge for predicting species ranges: a metamodeling framework. *Global Ecol. Biogeog.* 25, 238-249.

67 Howard, C. et al. (2014) Improving species distribution models: the value of data on abundance. *Methods Ecol. Evol.* 5, 506-513

68 Hartig, F. et al. (2012) Connecting dynamic vegetation models to data – an inverse perspective. *J. Biogeog.* 39, 2240-2252

69 Gonzalez, E. J. and C. Martorell. (2013) Reconstructing shifts in vital rates driven by long-term environmental change: a new demographic method based on readily available data. *Ecol. Evol.* 3, 2273-2284

70 Gonzalez, E. J. et al. (2016) Inverse estimation of integral projection model parameters using time series of population-level data. *Methods Ecol. Evol.* 7, 147-156.

71 McGill, B. J. 2010. Matters of scale. *Science* 328, 575-576

72 Sorrano, P. A. et al. (2014) Cross-scale interactions: quantifying multi-scaled cause-effect relationships in macrosystems. *Front. Ecol. Environ.* 12, 65-73

73 Heffernan, J. B. et al. (2014) Macrosystems ecology: understanding ecological patterns and processes at continental scales. *Front. Ecol. Environ.* 12, 5-14

74 Ellner, S. P. (2012) Comments on: Inference for size demography from point pattern data using integral projection models. *J. Agr. Biol. Env. Stat.* 17, 682-689

75 Clark, J.S. et al. (2012) Individual-scale inference to anticipate climate-change vulnerability of biodiversity. *Philos. Trans. R. Soc. B: Biol. Sci.* 367, 236-246.

76 Snell, R. S. et al. (2014) Using dynamic vegetation models to simulate plant range shifts. *Ecography* 37, 1184-1197

77 Medvigy, D. et al. (2009) Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography mode version 2. *J. Geophys. Res.* 114, G01002, doi:10.1029/2008JG000812.

78 Keane, R. E. et al. (2011) The FireBGCv2 landscape fire and succession model: a research simulation platform for exploring fire and vegetation dynamics. Gen. Tech. Rep. RMRS-GTR-255. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 137 p.

79 Vanderwel, M.C. and Purves, D.W. (2014) How do disturbances and environmental heterogeneity affect the pace of forest distribution shifts under climate change? *Ecography* 37, 10-20

80 Pagel, J. et al. (2014) Quantifying range-wide variation in population trends from local abundance surveys and widespread opportunistic occurrence records. *Methods Ecol. Evol.* 5, 751-760.

81 Gelman, A. et al. (2015) Stan: a probabilistic programming language for Bayesian inference and optimization. *J. Ed. Behav. Stat.* 40: 530-543.

82 De Valpine, P. et al. (2016) Programming with models: writing statistical algorithms for general model structures with NIMBLE. Journal of Computational and Graphical Statistics.

83 Visser, M. D. et al. (2015) Speeding up ecological and evolutionary computations in R: essentials of high performance computing for biologists. *PLoS Comput. Biol.* 11: e1004140

84 Record, S. et al. (2013) Should species distribution models account for spatial autocorrelation? A test across eight millennia of climate change. *Global Ecol. Biogeog.* 22, 760-771

85 Veloz, S. D. et al. (2012). No-analog climates and shifting realized niches during the late quaternary: implications for 21st-century predictions by species distribution models. *Global Change Biol.* 18, 1698-1713

86 Purves, D. et al. (2013) Time to model all life on Earth. *Nature* 493, 295-297

87 Harfoot, M. B. J. et al. (2014) Emergent global patterns of ecosystem structure and function from a Mechanistic General Ecosystem Model. *PLoS Biol.* 12, e1001841

88 Merow, C. et al. (2014) Advancing population ecology with integral projection models: a practical guide. *Methods Ecol. Evol.* 5, 99-110

89 Merow, C. et al. (2011) Developing dynamic mechanistic species distribution models: predicting bird-mediated spread of invasive plants across northeastern North America. *Am. Nat.* 178, 30-43

90 Araujo, M.B. and Rozenfeld, A. (2014) The geographic scaling of biotic interactions. *Ecography* 37, 406–415

91 Belmaker, J. et al. (2015) Empirical evidence for the scale dependence of biotic interactions. *Global Ecol. Biogeog* 24, 750-761

92 Wisz, M. S. et al. (2013) The role of biotic interactions in shaping distributions and realized assemblages of species: implications for distribution modelling. *Biol. Rev.* 88, 15-30

93 Singer, A. et al. (2016) Community dynamics under environmental change: How can next generation mechanistic models improve projections of species distributions? *Ecol Modelling* 326: 63-74.

94 Alexander, J. M. et al. (2015) Novel competitors shape species' responses to climate change. *Nature* 525, 515–518

95 Bolnick, D.I. et al. (2011) Why intraspecific trait variation matters in ecology. *Trends Ecol. Evol.* 26, 183-192

96 Fordham, D. A. et al. (2014) Better forecasts of range dynamics with genetic data. *Trends Ecol. Evol.* 29, 436-443

97 Sork, V.L. et al. (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Trees Genetics Genomes* 9, 901-911

98 Fitzpatrick, M.C. and Keller, S. R. (2015) Ecological genomics meets community-level modeling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol. Lett.* 18, 1-16

99 Bocedi, G. et al. (2014) RangeShifter: a platform for modelling spatial eco-evolutionary dynamics and species' responses to environmental changes. *Methods Ecol. Evol.* 5, 388-396

100 Clark, J.S. (2005) Why environmental scientists are becoming Bayesians. Ecology Letters, 8: 2-14

Figure Legends

Figure 1. From trade off to fusion. The existing view (panel a) is that models projecting species' future geographic distributions fall along a trade-off between phenomenological or correlative models for many species vs. mechanistic or process-based models for few species (arrow), relying upon a single data type, ranging from presence-only to physiological data (green box), respectively. An alternative view is that models combining these different types of data will enable process-based range modeling of many species (panel b). Examples of models that combine different data types are listed below the green data box in panel b, with points indicating the data used. These include dynamic range models [58], certain joint SDMs [47,48], integrated population models [59], a metamodel approach [66], and Bayesian inverse modeling [69].

Figure 2. Forward, inverse, and combined inference of parameters. a) Forward estimation of parameters for process-based range modeling might use data (box outlined in black) on physiological rates and their environmental sensitivities to parameterize a physiological model of plant growth (the Thornley transport resistance model, center panel, where B is biomass, C is carbon, and N is nitrogen), to then project a species' current and future geographic distribution. b) Inverse estimation of parameters uses distribution data to find the best-fit values of those same physiological rates and sensitivities [51]. c) Combining both "bottom-up" and "top-down" sources of data to better constrain parameters is exemplified by integrated population modeling [59]. Mark-recapture data and counts of one-year olds (N_1) vs. adults (N_a) both inform estimates

of fecundity (b), survival (φ), and immigration (ω), the parameters of a population projection matrix.

Box 1: What Data? Which Organisms?

Process-based range modeling depends on the availability of data. Large repositories of presence-only data (*e.g.*, the Global Biodiversity Information Facility [GBIF] and Biodiversity Information Serving Our Nation [BISON]) will continue to play an important role in range modeling. Some presence-only data sources are generating new data through citizen science initiatives (iNaturalist, iPANE). Presence-absence and abundance data make a more informative basis for range modeling than presence-only data; they can be extracted from certain atlas projects (*e.g.*, the *Protea* Atlas) and citizen science checklist initiatives (*e.g.*, the U. S. Breeding Bird Survey, eBird [32], and eButterfly). Where individuals are marked and censused, forest and other plot networks (*e.g.*, U. S. Forest Inventory and Analysis Program [31], the Amazon Forest Inventory Network RAINFOR, and VEGBANK) can be used to estimate vital rates. Estimates of vital rates are available in an open-access, standardized form from COMPADRE (637 plant species [28]) and COMADRE (345 animal species [29]) and other databases (*e. g.*, PanTHERIA [26]). Life history and Eltonian niche databases have been published for 21,322 species of amniotes [27] and 15,393 species of birds and mammals [30], respectively. Phenology and functional trait databases include the U. S. National Phenology Network (NPN) and the TRY plant trait database (~100,000 plant species [44]). Phylogenetic data are available from the Open Tree of Life project, with a draft phylogeny of 2.3 million tips [45]. Enormous amounts of data already exist, thus the challenge is to use them intelligently – to bridge the gap from what is known about individual species in a given ecological system or functional group to what is not known about other species.

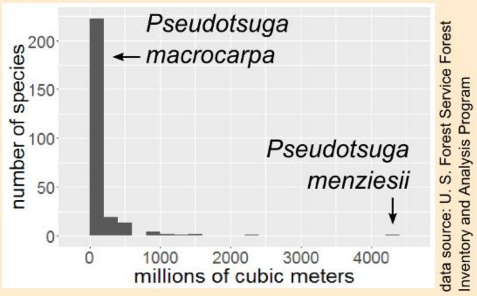
Certain functional groups or clades lend themselves to the first attempts at *en masse* process-based range modeling. These include organisms that are economically important, like trees and fishes, and others that are charismatic, like birds, butterflies, marine mammals, carnivores, and primates. Insects are ripe for exploration especially in light of global change: their physiology is strongly temperature-driven; there is a substantial experimental literature testing their physiological limits, in part because some are economically important as pests, and some of those pests are monitored at a landscape scale; other insects have been important study organisms for the development of structured population models (*i.e.*, *Tribolium castaneum*) or metapopulation models (*i.e.*, butterflies); butterflies have been the subject of citizen science initiatives in several countries.

Box 2

Box 2. Borrowing strength, filling data gaps

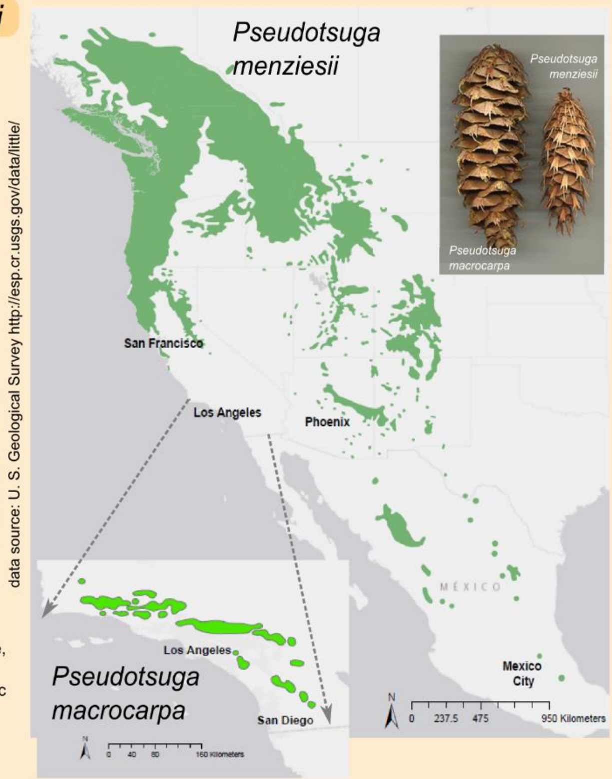
The detailed data needed for mechanistic range modeling are available for a relatively small number of species, leaving many (often rare) species for which we have little data. An important step towards process-based forecasting *en masse* will be to leverage functional trait, phylogenetic, and other kinds of data to estimate or impute niche parameters. To illustrate this idea, we use the example of trees in the United States. Most U. S. tree species are rare, and few are abundant (panel *i*). The single most abundant tree species in the U. S. is *Pseudotsuga menziesii.* Its congener, *P. macrocarpa*, restricted to mountain ranges in southern California (panel *ii*), ranks 217th in abundance. Data to estimate key forecasting parameters, such as vital rates (growth, survival, and fecundity) and how they respond to climate, might be limited or even absent for rare species like *P. macrocarpa*. Using functional trait and phylogenetic data, a two-level hierarchical model can improve estimates of (or impute) vital rates for rare species, borrowing strength from abundant species (panel *iii*). The phrase "borrowing strength" refers to the fact that parameter estimates for any one species in a hierarchical model are the weighted average of "no pooling" vs. "complete pooling" estimates [53], where "no pooling" estimates rely only on the data for a given species, whereas "complete pooling" estimates ignore species identity and use the data for all species together. The weighted average of the two (weighting the "no pooling" estimate by the amount of data available for each species) improves estimates of parameters for rare species, reducing extreme values caused by low samples size. Countering this, the use of functional trait data (see parameter model) would then prevent parameter values for rare species from being overly influenced by those of the most abundant species. For example, data on wood density of *P. macrocarpa* might suggest that its baseline growth rate (parameter a) is substantially different from that of *P. menziesii*, on a functional basis, not simply because data are limited for *P. macrocarpa*. Vital rates estimated for many species together in hierarchical regression models could then be combined into species-specific, spatially-projected demographic models, to then simulate abundance and range dynamics as a function of climate change (panel *iv*) [88].

i Species Abundance Distribution



data source: U. S. Forest Service Forest Inventory and Analysis Program

ii



data source: U. S. Geological Survey http://esp.cr.usgs.gov/data/little/

iii Hierarchical vital rate regression model

growth model $\log(G_{ij}) \sim N(a_j + b_j \log(B_{ij}) + c_j T_{ij}\ , \sigma_g)$

trait-based parameter model

$$a_j \sim N(\alpha_0 + \alpha_1 WD_j + \alpha_2 H_j + \alpha_3 SLA_j\ , \sigma_a)$$
$$b_j \sim N(\beta_0 + \beta_1 WD_j + \beta_2 H_j + \beta_3 SLA_j\ , \sigma_b)$$
$$c_j \sim N(\gamma_0 + \gamma_1 WD_j + \gamma_2 H_j + \gamma_3 SLA_j\ , \sigma_c)$$

G is growth increment of individual *i* of species *j*, *T* is July maximum temperature during the growth interval, *B* is size (biomass) at the start of the interval over which growth is measured, *a* is the species-level intrinsic individual growth rate, *b* describes how growth increment changes with size, *c* describes how growth increment changes with increased July maximum temperature, *WD* is wood density, *H* is maximum height, and *SLA* is specific leaf area, following Rüger et al. [52]. Variances (σ) are specified for each normal distribution.

iv Demographic model

$$n(y,t+1) = \int_L^U [P(y,x) + F(y,x)]n(x,t)dx \qquad \text{where} \quad P(y,x) = surv(x)growth(y,x)$$

An integral projection model projects the number of individuals (*n*) of size *y* at time *t+1* from the number of individuals (*n*) of size *x* at time *t*, based on a survival-growth subkernel *P* and a fertility subkernel *F*. Surv(x) is the probability of survival of x-sized individuals from time t to t+1, and growth is the probability density of sizes reached at time t+1 by individuals starting at size x at time t, conditional upon their survival. Survival-growth and fertility subkernels can be built from hierarchical regression models as in panel *iii* [88].

Box 3. Outstanding Questions

How detailed do models of species' distributions need to be?

An important challenge posed by the desire to create process-based range models for many species is the need to create "efficient" process-based models. How little detail can these models include, both in terms of mechanism and data, and still reliably forecast species' distributions? One approach is to specify those processes understood to be important, while phenomenologically characterizing remaining pattern, *e.g.*, in the form of spatial random effects [84]. If phenomenological characterization does not suffice for short-term forecasting, its pattern may provide clues to missing factors. For example, a range model that does not explicitly account for the process of dispersal, but includes spatial random effects, may suffice for forecasting of a species with limited dispersal capacity over the course of the 21st century. In contrast, a species with greater dispersal capacity, or a forecasting window that is longer relative to the organism's dispersal capacity, suggests the need for explicit modeling of spatial dynamics (as is the case for invasive species [89]).

What kinds of process-based range models work, where, and why?

The literature suggests that there is more than one path forward to capture important processes governing range dynamics – *e.g.*, physiology-, demography-, and phenology-based range models. Are all of these approaches equally effective for a given clade or functional group or ecosystem, and if not, why not? Can we identify the situations in which one approach is more likely to be successful over another?

To the degree that every species' niche is unique, how can niches (and distributions) be modeled *en masse*?

When built for many species at a time, correlative (phenomenological) models of species' ranges often use broad-stroke environmental predictors such as mean annual temperature and precipitation. Can these variables be used in the context of more mechanistic or process-based range models? Or, is it necessary to identify more specific limiting factors for each species (using machine-learning, data-mining, or empirical approaches)?

How do the variety of factors limiting a given species combine to yield the observed variation in its abundance and its geographic limits?

The factors limiting a species' distribution and abundance, including interactions between factors, are not the same throughout its range. For example, it's been suggested that climate is a limiting factor at the high latitude edges of species' ranges, whereas competition is a limiting factor at the low-latitude edges [21]. Mechanistic models of species' ranges cannot be built to capture the limiting factors operating at just one location or one edge of a species' distribution, they must be able to capture all of the limiting factors that define range boundaries, including interaction effects.

Is it necessary to model biotic interactions explicitly in process-based range models?

Evidence has emerged for the importance of biotic interactions in shaping the distributions of various taxa at coarse grains (*e.g.*, birds, butterflies and their host plants, and frugivores and trees [90,91]). Though important progress has been made in suggesting how biotic interactions might be incorporated into process-based range models [92,93], great challenges remain. There is the issue of accumulating sufficient data on the strength and direction of interactions across the geographic ranges of interacting species. There is also the problem of determining which kinds of interactions are most important for structuring communities, within and across trophic levels. The emerging phenomenon of novel species interactions presents the challenge of dealing with no-analog situations where observational data on contemporary species interactions may be less informative than manipulative data from experiments where species that do not currently interact (*e.g.*, due to range limits) are allowed to co-occur [94].

Do process-based range models need to account for plasticity, local adaptation, genomics, or eco-evolutionary dynamics?

Range models typically ignore within-species variation, even though plasticity and genetic variation due to local adaptation are well-established phenomena [95,96]. Recently, there has been interest in using cost-effective, high-throughput sequencing to understand local adaptation at the molecular level [97]. The use of ecological genomic data at the landscape scale (landscape genomics) to make spatially-explicit projections given different climate change scenarios remains largely uncharted territory [98]. The case (and a modeling platform) has been made for eco-evolutionary models of ranges dynamics, incorporating population dynamics, dispersal, plasticity, and local adaptation, all in a spatially-explicit context [99]. Each of these added complexities increases the parameter burden for process-based range modeling, thus they deserve critical evaluation in terms of cost/benefit ratio.

Figure 1

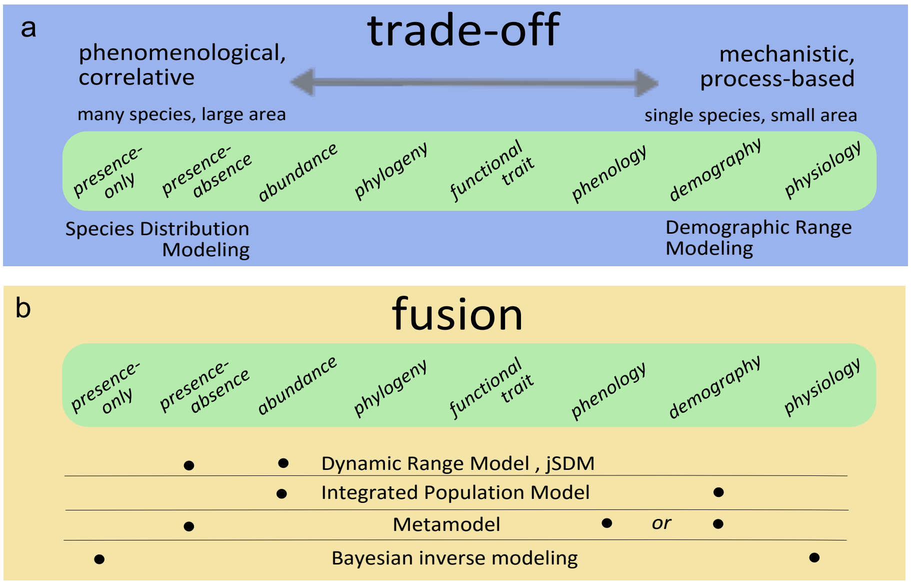
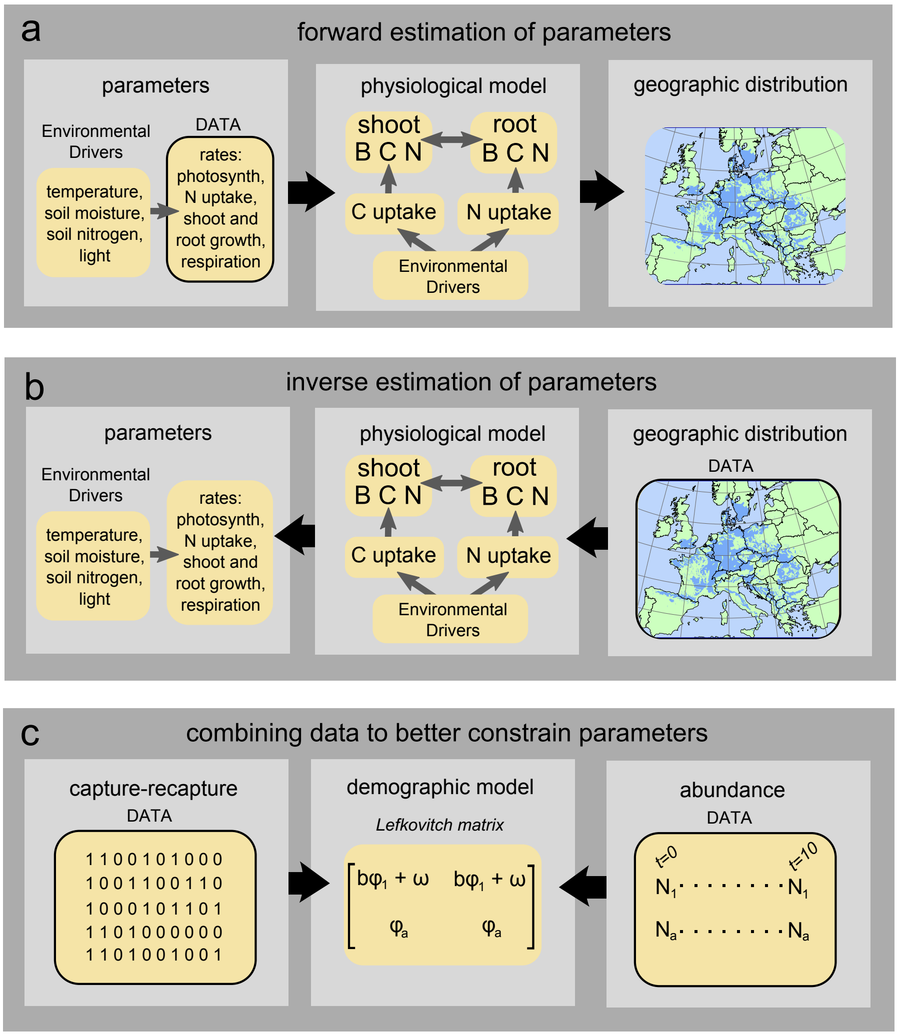

a
trade-off
phenomenological, correlative
mechanistic, process-based
many species, large area
single species, small area
presence-only
presence-absence
abundance
phylogeny
functional trait
phenology
demography
physiology
Species Distribution Modeling
Demographic Range Modeling
b
fusion
presence-only
presence-absence
abundance
phylogeny
functional trait
phenology
demography
physiology
Dynamic Range Model , jSDM
Integrated Population Model
Metamodel
or
Bayesian inverse modeling

Figure 2


a
forward estimation of parameters
parameters
Environmental Drivers
DATA
temperature, soil moisture, soil nitrogen, light
rates: photosynth, N uptake, shoot and root growth, respiration
physiological model
shoot B C N
root B C N
C uptake
N uptake
Environmental Drivers
geographic distribution
b
inverse estimation of parameters
parameters
Environmental Drivers
temperature, soil moisture, soil nitrogen, light
rates: photosynth, N uptake, shoot and root growth, respiration
physiological model
shoot B C N
root B C N
C uptake
N uptake
Environmental Drivers
geographic distribution
DATA
c
combining data to better constrain parameters
capture-recapture
DATA
1 1 0 0 1 0 1 0 0 0
1 0 0 1 1 0 0 1 1 0
1 0 0 0 1 0 1 1 0 1
1 1 0 1 0 0 0 0 0 0
1 1 0 1 0 0 1 0 0 1
demographic model
Lefkovitch matrix
$b\varphi_1 + \omega$
$b\varphi_1 + \omega$
φ_a
φ_a
abundance
DATA
$t=0$
$t=10$
$N_1 \cdots N_1$
$N_a \cdots N_a$