

Abiotic factors

Early biogeographers (e.g. Von Humbolt or Darwin) observed that a same species could occur in sites with different environmental conditions (each combination is a distinct habitat), and thereby occupy a range of different habitats.

Yet, most species have limited geographic ranges. Thus, while a species can colonize a range of conditions along environmental gradients, in most cases its range only occupies a proportion of all the possible habitat conditions available, resulting in the species occupying only a limited geographic and environmental range where these specific conditions are met.

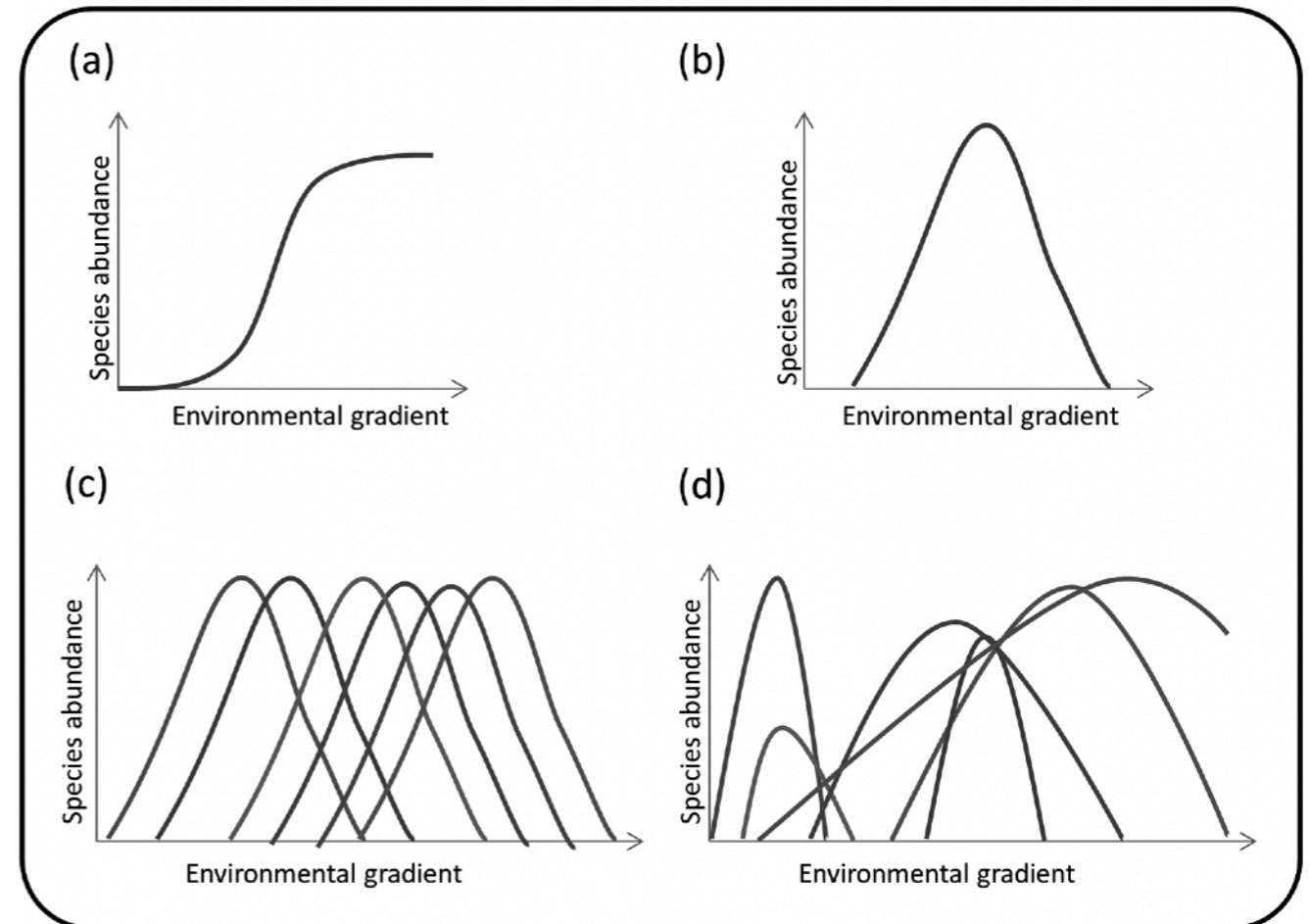
This happens since the specialized physiological adaptations that most species have undergone through evolution in order to survive and be competitive in specific habitats. These adaptations come at the cost of being maladapted (and thus unable to survive) or less competitive (and thus excluded) in other habitats, representing adaptation or functional tradeoffs.

Abiotic factors

Physiological specialization usually results in different shapes of responses along environmental gradients

Generally, there is an optimum, where species performs the best, and a gradual decrease in performance the further it moves away from this optimum, in either direction.

Such physiological response curves can therefore be represented as sigmoidal or unimodal shapes. The width of the curve is the physiological tolerance of the species along the gradient. Different species have different optima and tolerance. A species that has a very broad tolerance along a specific gradient is a generalist species.



Abiotic factors

Transition from optimal to poor performance can be smooth or abrupt, depending on the types of physiological mechanisms involved.

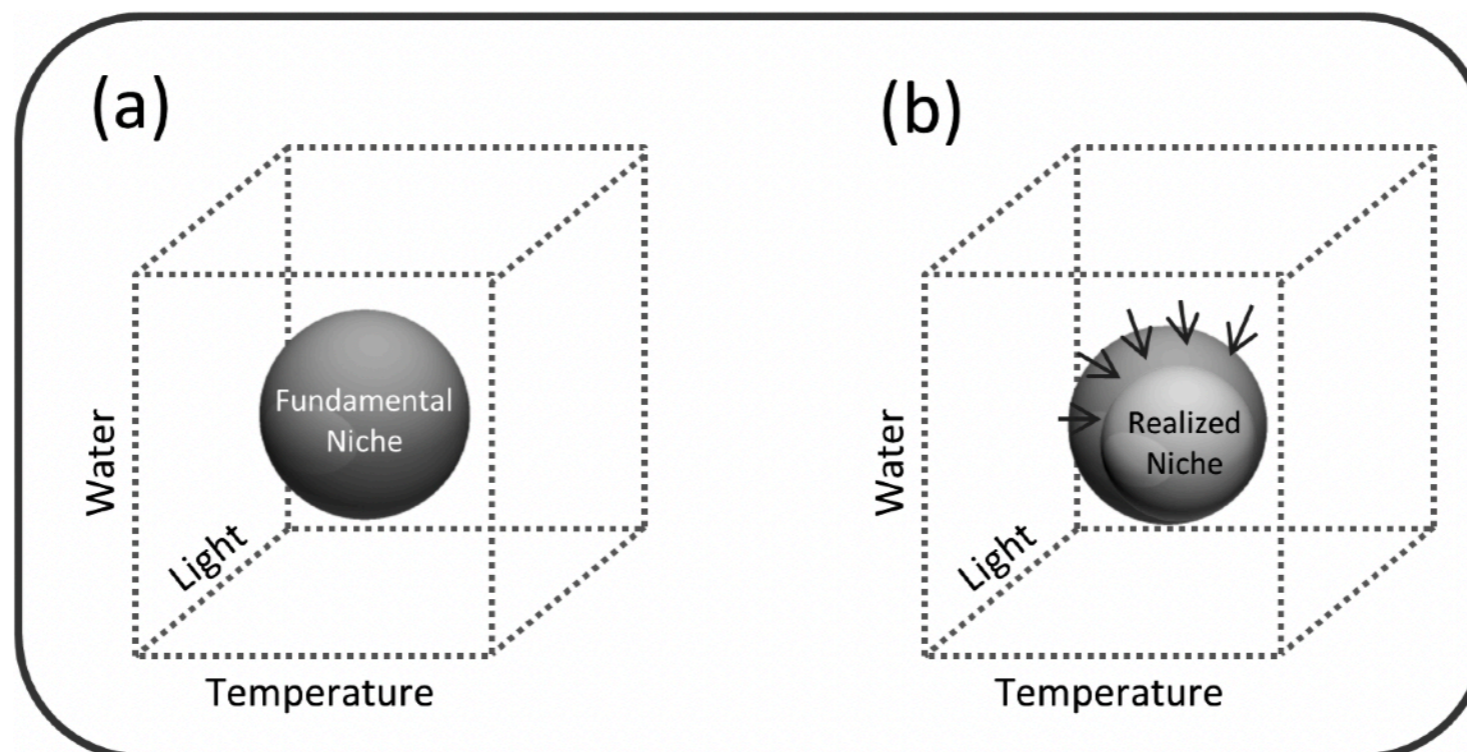
An abrupt transition can result from limiting factors with threshold effects below or above which some metabolic pathways abruptly change. As an example, in the conifer *Pinus cembra* in the Alps, the cambium activity allowing root growth tends to stop rather abruptly below some threshold value of soil temperature (around 7°C at -10 cm).

Smoother transitions occur when the gradient has a more or less linear effect on some metabolic rates (e.g. carbon sequestration, water use efficiency), progressively lowering individuals' fitness.

However, the effect of environmental variable rarely affects a species independently. On the contrary, synergic effects are often observed. Interactions may dampen or amplify the effect of each other's variable on species physiology. Therefore, all important variables should be considered jointly in an analysis, in order to define what is known as the environmental niche of species, a term initially coined by Joseph Grinnell at the beginning of the 20th century.

Abiotic factors

When considered jointly, the physiological responses of a given species to several environmental variables define a multidimensional volume called a species' fundamental environmental niche, which concept was quantitatively formalized by Hutchinson in 1957, as an n-dimensional hypervolume in a space defined using environmental variables which to have an influence on a species' physiology, and within which the population growth rate is positive. The realized niche, on the contrary, is a portion of the fundamental niche, when biotic interaction are taken into account as well.



Abiotic factors

Knowing which variables have a direct physiological effect requires prior experimental laboratory measurements, as an example measuring metabolic rates and individual fitness while varying environmental variables, and this could not be feasible for all species.

Furthermore, it is even more difficult if such responses must be measured over a large number of populations, to account for genetic differences among populations.

It is therefore often easier to use measurements of surrogate environmental variables that are hypothesized to best correlate with the physiologically meaningful variables (e.g. altitude for temperature, or minimum of mean monthly temperature for absolute minimum temperature), but with the consequence especially of reducing the level of generalization of the model (e.g. altitude cannot be compared between population in cold environments in the Alps, and the Arctic).

The different variables can also be classified, depending on their effect or use by the target species, as limiting factors (causing linear or step responses), regulators (modulating the organism physiology, gradual response) or resources (consumed by the organism).

Biotic factors

Biotic environment enlists all possible interactions among species, either within the same, or between different trophic levels. It can constrain a species' distribution in space, in time, or along environmental gradients.

Competition has been the first biotic factor discussed by the proponents of the environmental niche concept (e.g. Hutchinson). However, interactions between trophic levels (e.g. predator–prey, plant–pollinator, or plant–herbivore) or functional groups (e.g. host–parasite, symbiosis) can be as important.

The fundamental niche a species can occupy is defined by all locations where the abiotic, environmental conditions allow positive fitness. However, species do not usually have uniform responses along environmental gradients, thus fitness, as well as their competitive potential, also varies. The “competitive exclusion” is a direct consequence of this fact.

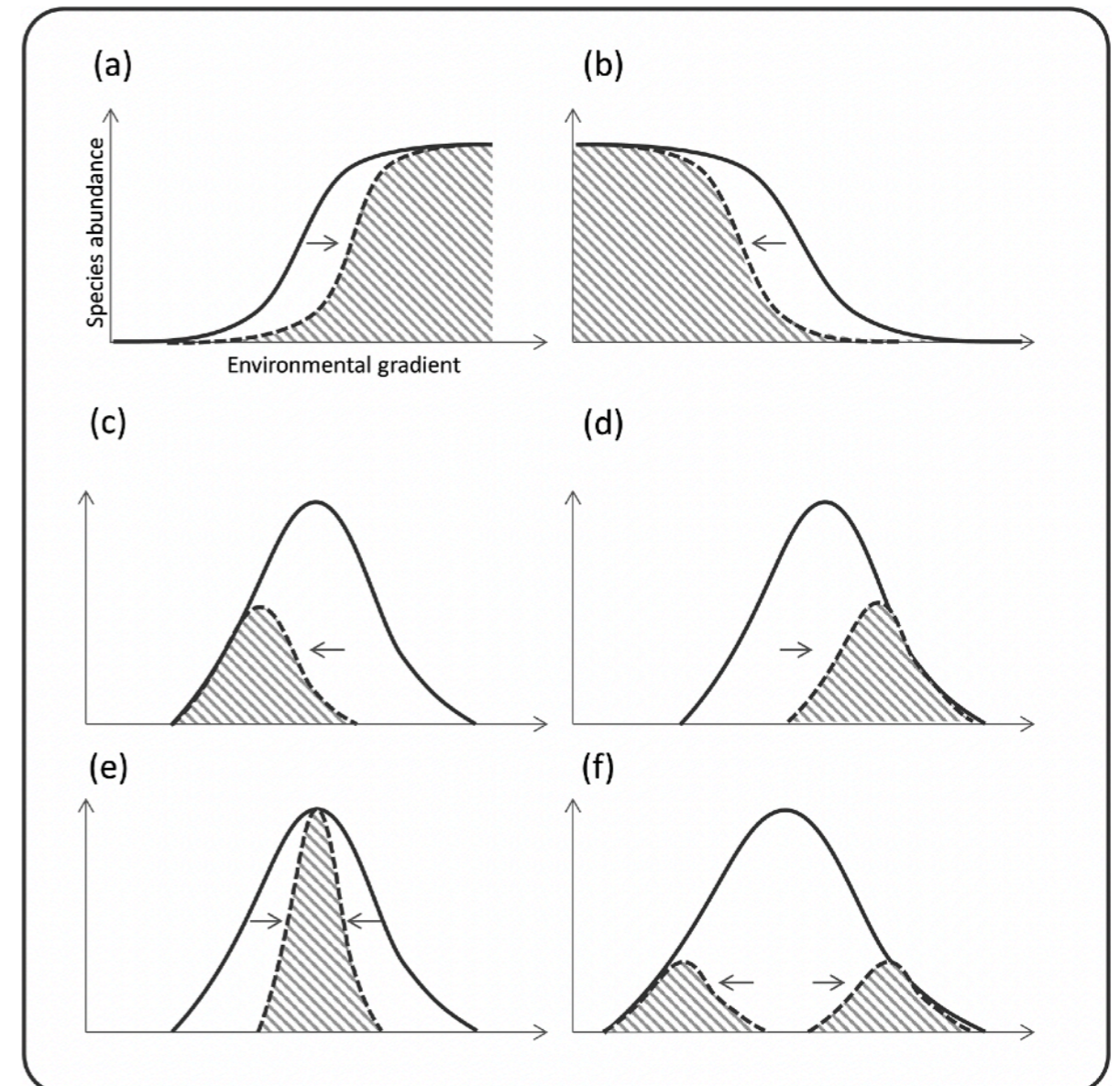
One evidence can be the fact that a number of high-elevation alpine plants can be grown at low-elevation botanical gardens in the absence of competition, while at low elevations in natural systems they are outcompeted.

Biotic factors

When a species is systematically excluded from parts of an environmental gradient, the response along these gradients revealed from field observations only gives a partial view of the full physiological (i.e. fundamental) response. The species response from field observations along gradients thus depends on a particular biotic configuration.

This type of field observation-based response along a single gradient has been called the “realized response” or “ecological response”, in contrast to the “physiological response” or “fundamental response”.

Depending on which part of a fundamental response is excluded by competition, different shapes are obtained for the realized response.



Biotic factors

Thus, the occurrences of species collected in the field do not tell us all the truth about that species physiological potential, i.e. on its fundamental niche, but can inform us on its realized niche, where biotic, and abiotic factors do interact. This can be relevant for predicting the distribution of a taxon in its normal range, where we could assume it could find similar interactions with other species, at the same, or at different trophic levels. However, it could badly bias our predictions when alien invasive species are investigated. In the latter case, it is possible that a species will not find in its new environment similar interactions as in its original range. Thus, assumptions could be made only by mean of a mechanistic approach, when investigating its fundamental niche.

Biotic interactions however may either be negative, by excluding a species from sites that are *a priori* environmentally suitable (i.e. within its fundamental niche) or facilitate a species at sites that appear environmentally unsuitable based on measured average site conditions.

Thus, we cannot exclude that an habitat, predicted suitable by mean of a mechanistic approach, could be unsuitable because of the lack of some positive, and fundamental, biotic interaction.

Biotic factors

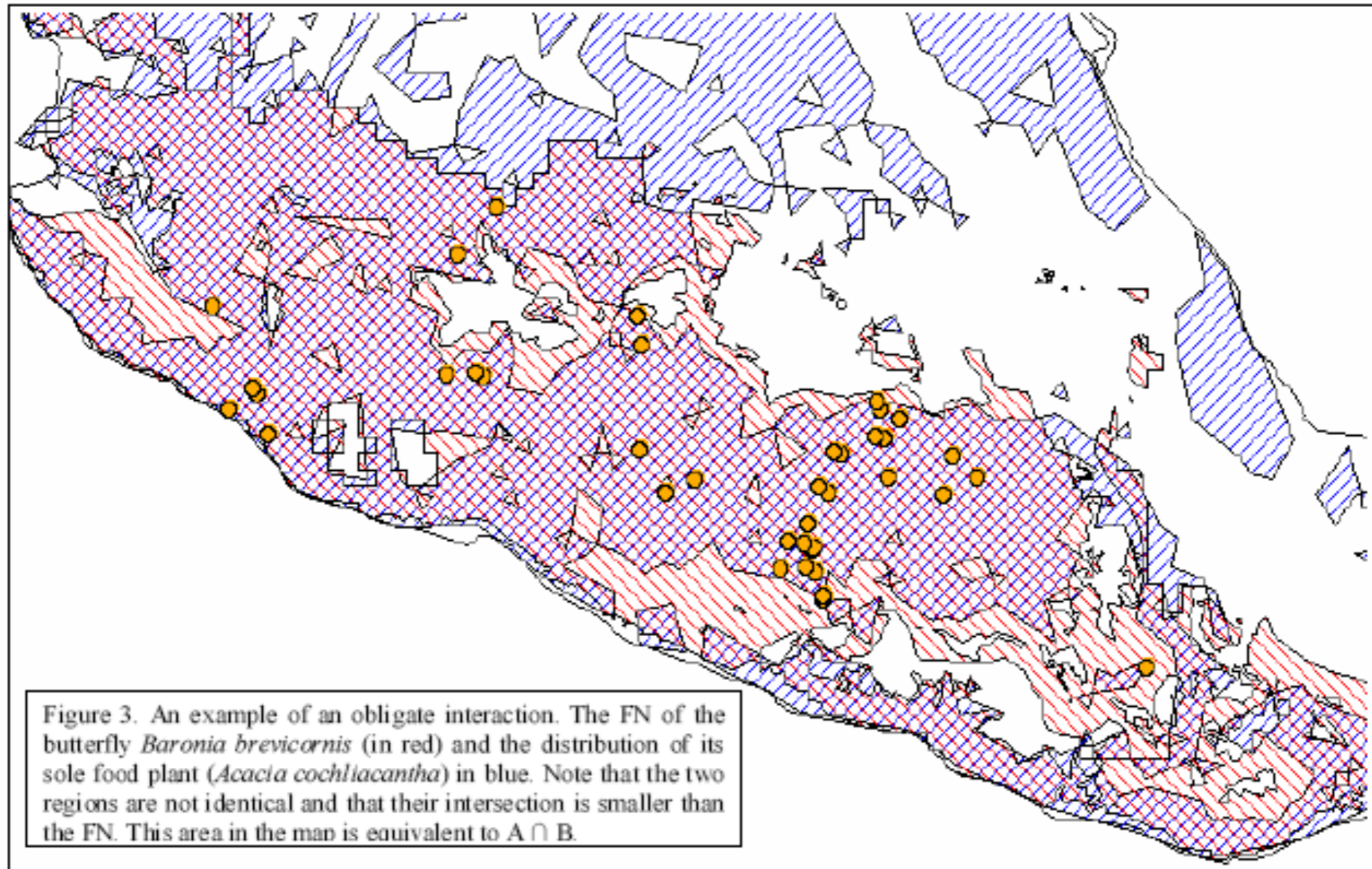
Positive interactions include commensalism, mutualism, biotic engineering (for instance forest understory species that cannot grow in plain light benefiting from the shade from the surrounding tree canopy).

Negative interactions are more common and include competitive exclusion, predation (incl. herbivory), or parasitism when the host is sufficiently affected by its parasite to be removed from entire parts of its range.

Interactions often involved pairs of species (or functional groups of species), but they also naturally take place within more complex biotic frameworks such as food webs and large interaction networks.

From a geographic perspective, the interactions derived from these interactive systems can be used as predictors of individual species distributions. The way biotic interactions influence the presence (positive interactions), absence (negative interactions) or abundance (both) of a given species therefore also influences the final composition of a community. Quantifying these interactions into assembly rules is therefore required in order to understand how communities assemble, and – if the quantified rules allow – to ultimately predict assemblages.

Biotic factors



Niche concept: some history

The ecological niche

The concept of ecological niche is detectable in **Aristotele** (*Historia Animalium*), who depicted it for a multiplicity of life forms with accurate descriptions of their habits, distribution, seasonal, climatic and other abiotic factors. Later, the concept is present in the work of **Theophrastus** (*Historia Plantarum*), who described relationships between organisms and environment. During the 18th century, **Linnaeus** merged the Christian tradition with coeval naturalists' work, defining the theory of the "economy of nature" (1749-), which depicts nature as balanced, rational, and ordered by divine laws, and organisms as members of species, with well defined roles in the maintenance of order in the world.

A fundamental difference between Linnaeus' conception of an ecological science and ours is that in his, biotic interrelationships were designed by God to work harmoniously and permanently and for the benefit of humanity, whereas in ours, interrelationships evolve and can lead to extinction of species.

Niche concept: some history

His conception was part of a general outlook in science: for example, in astronomy, celestial bodies were unchanging in substance and orbits; and in geology, ongoing changes in the landscape were considered minor compared to the changes caused by God in the Flood of Noah. Scientists' study of a "static" universe gradually revealed that it is not static. This even happened to Linnaeus. In his *Systema Naturae* he confidently claimed that all species had been created by God at the beginning and no new ones had since appeared.

However, the discovery of *Peloria* in 1741—so similar to *Linaria*, yet an apparently different species—shook his belief in the constancy of species. He eventually suspected that God had created only a few species, which later hybridized to form the great variety now seen. On 18 August 1764, he explained this idea in a letter to Johannes Burmann:

Let us suppose God made a Ranunculus [and that] this species is crossed with a Helleborus, and Aquilegia, or a Nigella in hybrid generations. Through Divine Law the descendants of these hybrids will have, as in animals, the mother's medulla and father's cortex. As a result, there are so many of Ranuncula with either aquilegous leaves or nigellous ones that you could not separate them into arbitrary genera...

Niche concept: some history

During the 19th century, the idea of interdependent relationships between elements in the environment arose in a variety of embryonal forms, such as the observation of plant-climate dependence, and the consequent description of plants distribution areas using latitude and longitude by **Alexander von Humboldt**, the father of biogeography.

Other novel ideas were the definition of biotic relation types (van Beneden, 1876), the concept of biocoenosis (Möbius, 1877), the study of food web dynamics (Forbes, 1880), the study of vegetal (then ecological) successions (Thoreau, 1860), the notion of limiting factor (Liebig, 1840), and, influenced by the investigation on population dynamics (Malthus, 1798), the concept of adaptation to environment (Darwin, 1859).

Darwin added to the vision of Linnaeus the idea that living organisms have a place in the economy of nature *to which they are adapted by natural selection*, a concept he defined “line of life”, in parallel to the concept of “line of work”, which refers to the profession of a person.

Niche concept: some history

The term “niche”, intended as the place of a species in the environment, was used for the first time by the naturalist **Roswell Hill Johnson** (1910), but it was **Joseph Grinnell** that included, for the first time, this concept in his research framework (1917).

With the term “niche” Grinnell encompassed all the factors which influenced the presence of a species at a given location, both abiotic (such as temperature, rainfall, humidity, etc.), and biotic (such as the presence of food, competitors, predators, etc.). In his concept, there was a close link to the idea of competitive exclusion, which was formalized by **Gause** in 1934. The niche was a complex of environmental factors, a place, according to which species would evolve and exclude each other.

In order to explain the features of a species, Grinnell elaborated the ecological hierarchy, in which he placed biotic and abiotic factors at different levels (from realms to niches). In this view, a niche is considered as the smallest unit of a system, which is defined by the characteristics and relationships of a species with regard to surrounding environment and organisms. In his hierarchy, higher levels were associated with abiotic factors, while lower levels were rather associated with biotic factors, thus had no explicit geographical connotation.

Niche concept: some history

Grinnell also focused on the comparison of communities living in different environments, elaborating the concept of ecological equivalents, i.e. species or associations which share the same ecological traits and strategies, in different geographic areas and environments.

His idea was that a niche could also be occupied in a region, and vacant in another, because of dispersal limitations.

Charles Elton also focused on ecological equivalents (1927). His research focused on the invariance of community structures.

Elton investigated the common traits of associations which occur in different environments or areas. His approach led him to study mainly trophic relationships, and to view the niche of a species essentially as its trophic position, i.e. “[the] place in the biotic environment, its relations to food and enemies, and to some extent to other factors also”.

In Elton’s view, the niche was defined mostly by the position of the organism in a trophic chain.

Niche concept: some history

However, the concepts of both Grinnell, and Elton, were quite similar. For both, the ecological equivalents were the rationale for the concept, since they demonstrated that similar niches exist in different places. **The niche was perceived as a place that existed independently of its occupants.** Food was a major component of the niche, but the niche is not restricted to food, since it includes the micro-habitat factors, and the relationships to predators.

A major and revolutionary change was that introduced by **George Hutchinson**, in 1957, **who shifted the concept of niche as an attribute of the species, and not of the environment.**

The niche was defined as a n-dimensional space (*hypervolume*) of environmental variables, biotic and abiotic, some of which representing the limits of species viability.

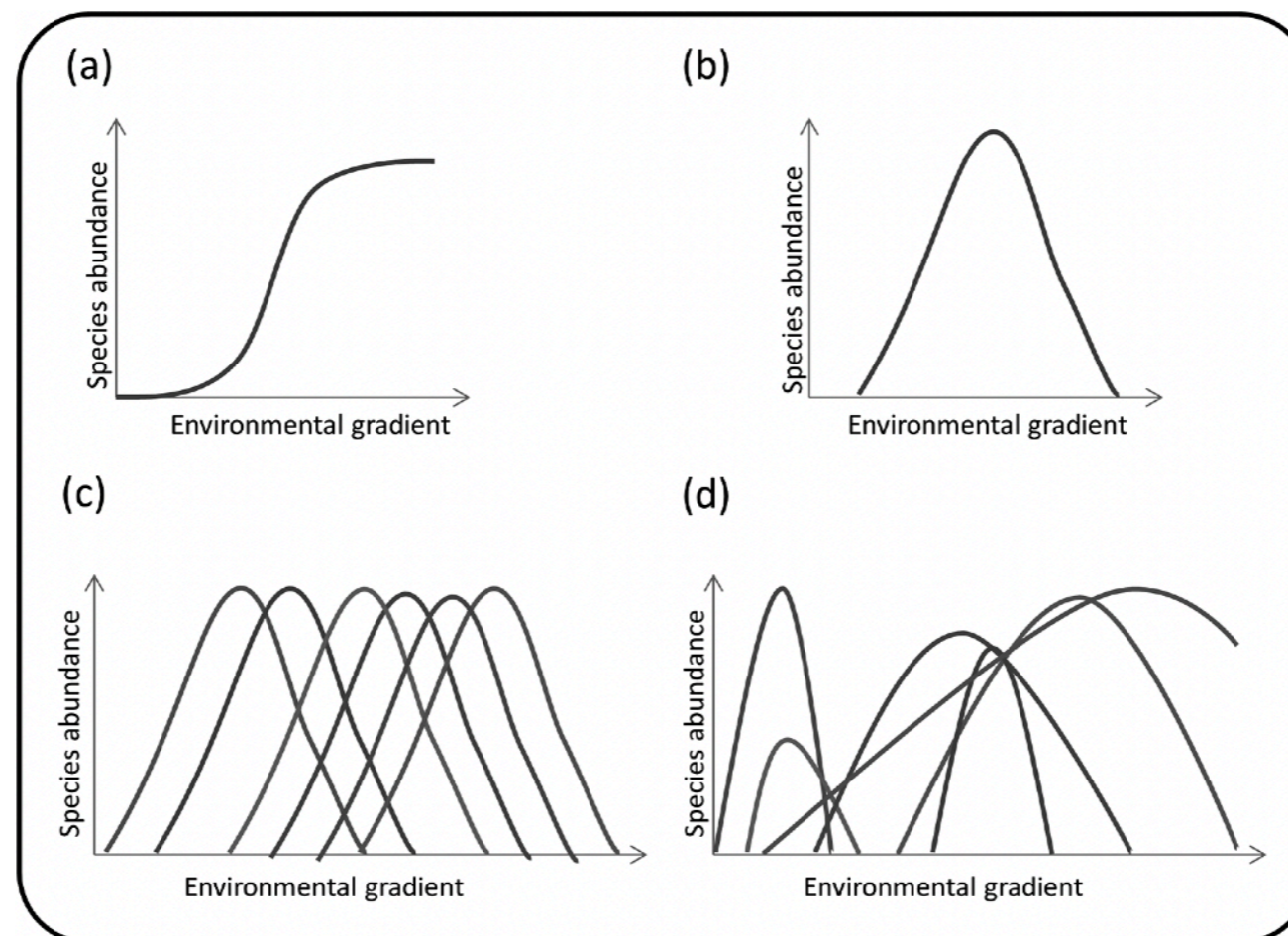
The area included in these limits was “a state of the environment which would permit the species to exist indefinitely”. This definition was named the *fundamental niche*.

The *realized niche*, on the contrary, is the niche actually occupied by a species, i.e. the regions of the fundamental niche from which the species is not excluded by its competitors.

Niche concept: some history

Since the niche is a characteristic of a species, and not of the environment, in Hutchinson's vision the competition (for resources) could change the niche of a species, in the sense of a reduction of similarity from another species' niche.

It is however evident that a flaw exist in this vision of a n-dimensional space. This is, the viability of a species does not behave in a binary mode. Species normally respond to each factor with a uni- or bimodal curve.



Niche concept: some history

Thus, it is a simplification to state that inside the fundamental niche a species could survive indefinitely, while outside it cannot.

Plus, going from optima to other portions of the distribution for one or more variable may impose switches in the utilization of resources, or in the intensity of predation/competition. All these switches are often gradual, and do not work in binary mode.

In consequence, the concept of Hutchinson was replaced in the sixties by the concept of resource utilization distribution. The niche, defined for a specific population, was equivalent to the frequency of utilization of a resource, and could be simply be represented by an histogram. The niche as a utilization distribution was eminently an operational concept, since it can be easily measured.

Overlaps of utilization niches were used to calculate coefficients of competition among populations. This led to a major focus on competition alone, focus which was weakened in the late eighties alone.

A renovation of the concept, even if strictly connected to that of Hutchinson, was made by Chase and Leibold (2003).

Niche concept: some history

They showed that, within the ecology of an organism, we must distinguish the impact of a given ecological factor on the organism itself, i.e. the response of the organism to the factor, and the impact of the organism on the ecological factor.

Thus, the niche can be defined as the union of the responses of the organism, and its impacts.

To cite them: “[the niche is] *the joint description of the environmental conditions that allow a species to satisfy its minimum requirements so that the birth rate of a local population is equal to or greater than its death rate along with the set of per capita effects of that species on these environmental conditions*”.

The concept of ecological niche is however far from being completely frozen, and it is still undergoing discussion and revision, even in recent times. In any case, its multiple interpretations all revolve around the Darwinian view of ecosystems structured by the struggle for survival.

Whichever the concept, the niche is a model of the relationship between the organism and its environment.

Modelling approaches

When facing the problem of modeling the geographic distribution of a species, there are two commonly adopted approaches:

Mechanistic approach, which implies the measurement of the responses of a species to several factors, such temperature, pH, etc. Subsequently, these data are used in combination with climatic data in order to depict the fundamental niche of a species.

Correlative approach, which correlate occurrence data with climatic variables, in order to depict the realized niche of a species

Both approaches lead eventually to the projection in the geographic space of the niche of a species.

A third, more complex approach, is that focusing on population dynamics in space occupancy, called **process-oriented modeling**.

The mechanistic approach

Measuring the tolerance limits of a species with respect to a set of environmental dimensions via physiological experiments lead the measurement of the multidimensional projection of a theoretical, fundamental niche onto a simpler space consisting of the set of variables that is actually being measured.

Such efforts are invariably carried out in relatively few environmental dimensions (e.g., temperature, water stress), as the complexity involved either in carrying out all of the possible measurements cannot be easily addressed.

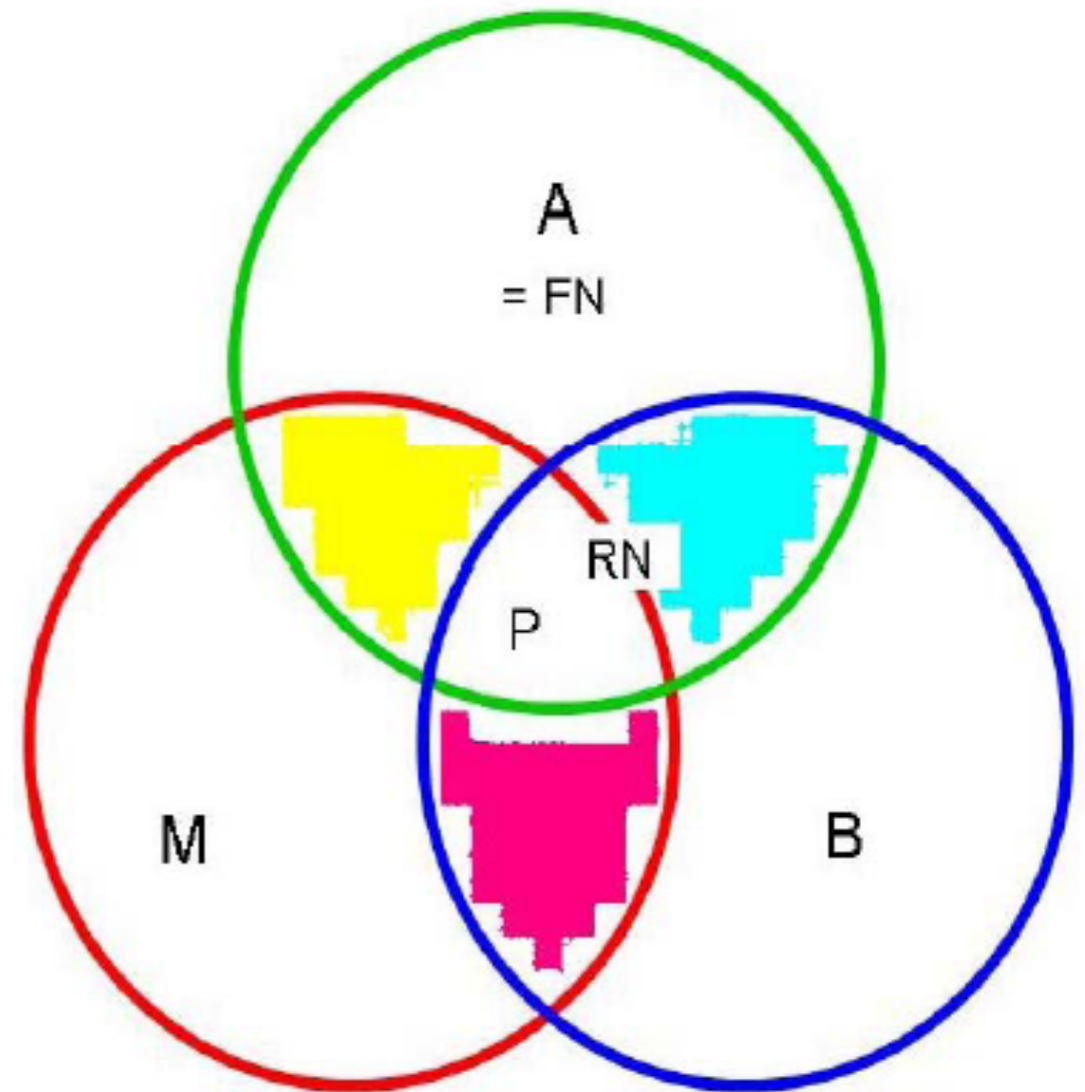
This approach has the significant advantage of having direct ties to the physiology of the species under consideration. As such, it has the potential for direct measurement of dimensions of the fundamental niche without the confounding effects of accessible areas, and interacting species.

In this sense, this approach allows the identification of the spatial footprint of A, without the complications related to M, or B which affect other approaches.

This could be an advantage when dealing with climate change, or with alien species.

Modelling the future distribution of a species in a climate change scenario cannot take into account current biotic interactions, or dispersion limitations, since they could abruptly change in the future.

On the other hand, when a species is moved from its original habitat to a new one, it is impossible to know whether biotic interactions will limit its invasion, or whether it will face dispersal limitations.



Mechanistic approaches: issues

Can the estimate of the niche of a single, or a few individuals, suffice to characterize the ecological niche constraints on the entire geographic range of a species?

The reliability of physiological parameters depends on their source, whether gathered in natural settings or through laboratory experiments. The latter are potentially influenced by limitations specific to experimental situations such as acclimatization, sample sizes, and length of life cycle.

A further major constraint is in the scaling challenges: mechanistic models and measurements represent individual characteristics that are relevant on extremely fine spatial scales, but are then applied to estimate geographic phenomena at coarse spatial and temporal scales.

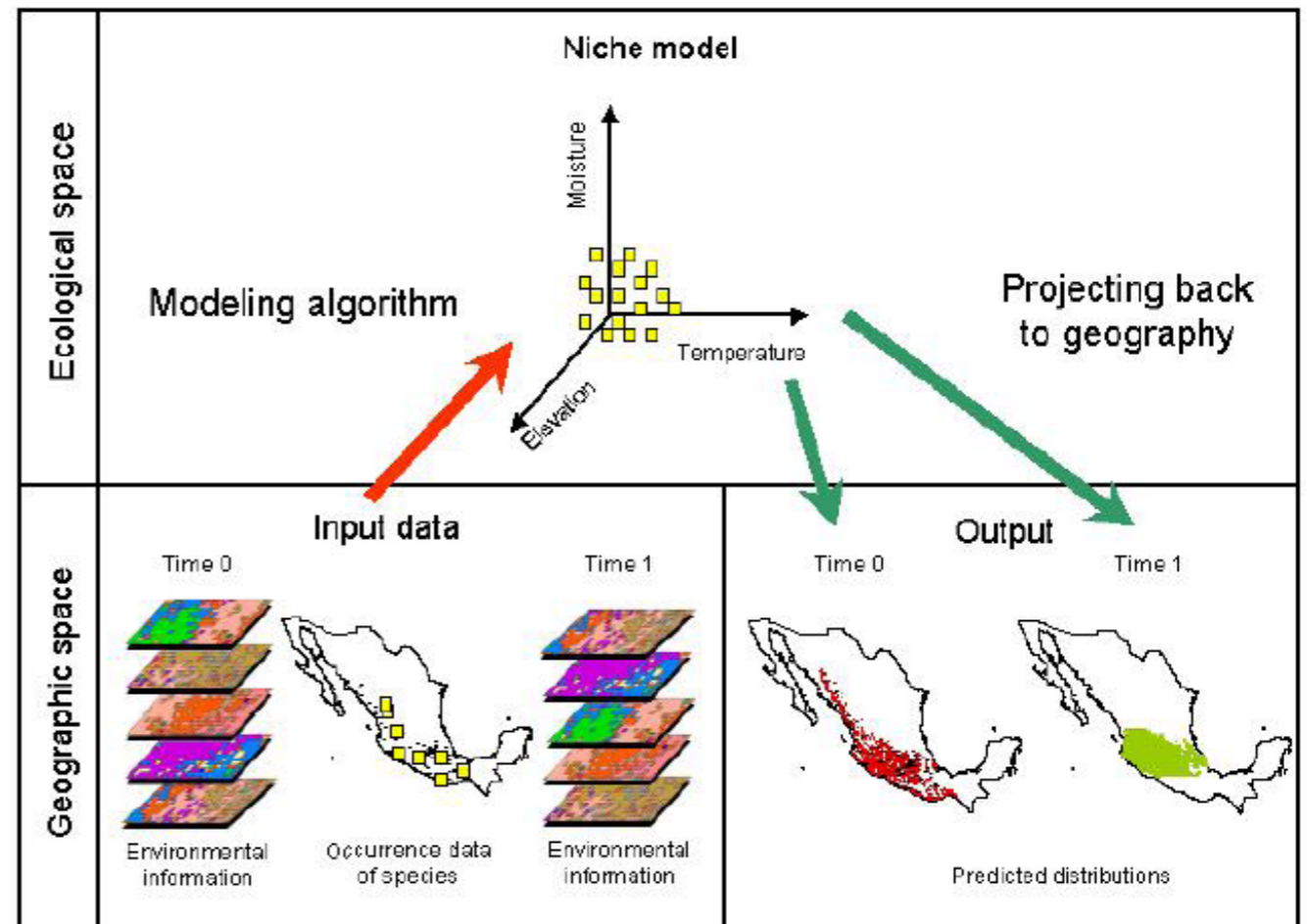
Finally, if biotic interactions or dispersal limitation play significant roles in distributional ecology, then large portions of the geographic area identified by mechanistic models may not be, in reality, suitable, or accessible by a species.

The correlative approach

Data characterizing sites where the species is known to occur (occurrence data) are related to the environmental characteristics of those localities.

Although several diverse correlative approaches exist, most compare environments associated with sites of known presence to those associated with sets of sites that do not (or are not known to) hold populations of the species. Thus, it is possible to identify sets of conditions under which the species could be able to maintain viable populations.

It estimates a 'niche' that is almost certainly intermediate between the realized and the fundamental niches.



The advantages of correlative approaches centre on their low input data needs, since they require few predictor variables for calibration.

Furthermore, since correlational models are keyed at the outset to geographic occurrences, and coarse-resolution environmental dimensions, these approaches avoid the challenges of scaling up from individual phenomena to landscape-level phenomena.

Their reliance on sampling of occurrences across the geographic distribution of the species provides a set of partial tests of niche tolerance limits: in this sense, correlational approaches avoid the single-niche assumptions of mechanistic models and can incorporate effects of numerous causal factors in a single 'model.'

At the same time, it is simultaneously an advantage and a disadvantage that correlative approaches measure something less than the full fundamental niche. Correlative model outputs express a complex combination of factors, including the effects of the frequently unknown biotic interactions, and dispersal, whose combined restrictive effects (plus the effects of the sampling that underlies the occurrence data) shape species' distributions and constrain their estimation.

Correlative approaches have limited use in pinpointing fine-grain, individual-level tolerance limits.

Thus, when local adaptation in niche-relevant dimensions has occurred, correlative models cast at the level of the species will overestimate the niche of any single population.

Finally, the vagaries of the sampling of biodiversity (e.g., spatial bias to accessible areas, incomplete sampling across an area, incomplete detectability of individuals) will have considerable potential to translate into new biases and problems in model outcomes.

In sum, returning to the question of what is being estimated, 'niches' estimated by correlative approaches are complicated in terms of their interpretation. They will generally identify some suite of environmental conditions that fall in between fundamental and realized niches.

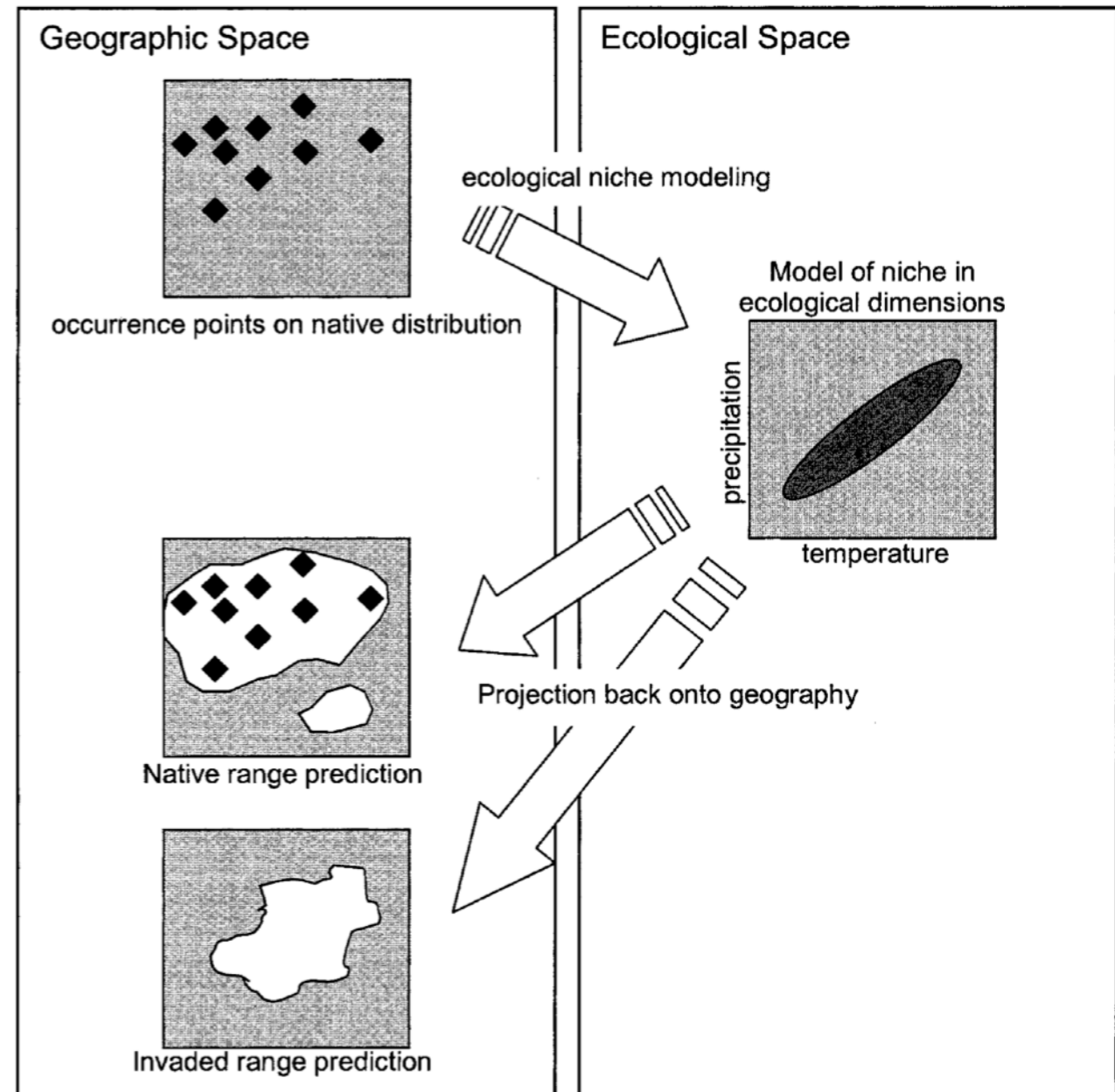
Full interpretation of such models is complex and still requires additional assumptions and hypotheses to be able to interpret model outputs clearly as distributional predictions.

Modelling approaches: focus on invasive species

Correlative approaches can be used for predicting the invasive area of a species, when it is transferred in an area which can be potentially invaded.

Occurrence points from native range are correlated to climate variables.

Then, the model is projected in the geographic space, highlighting areas where the climatic fitness is positive.



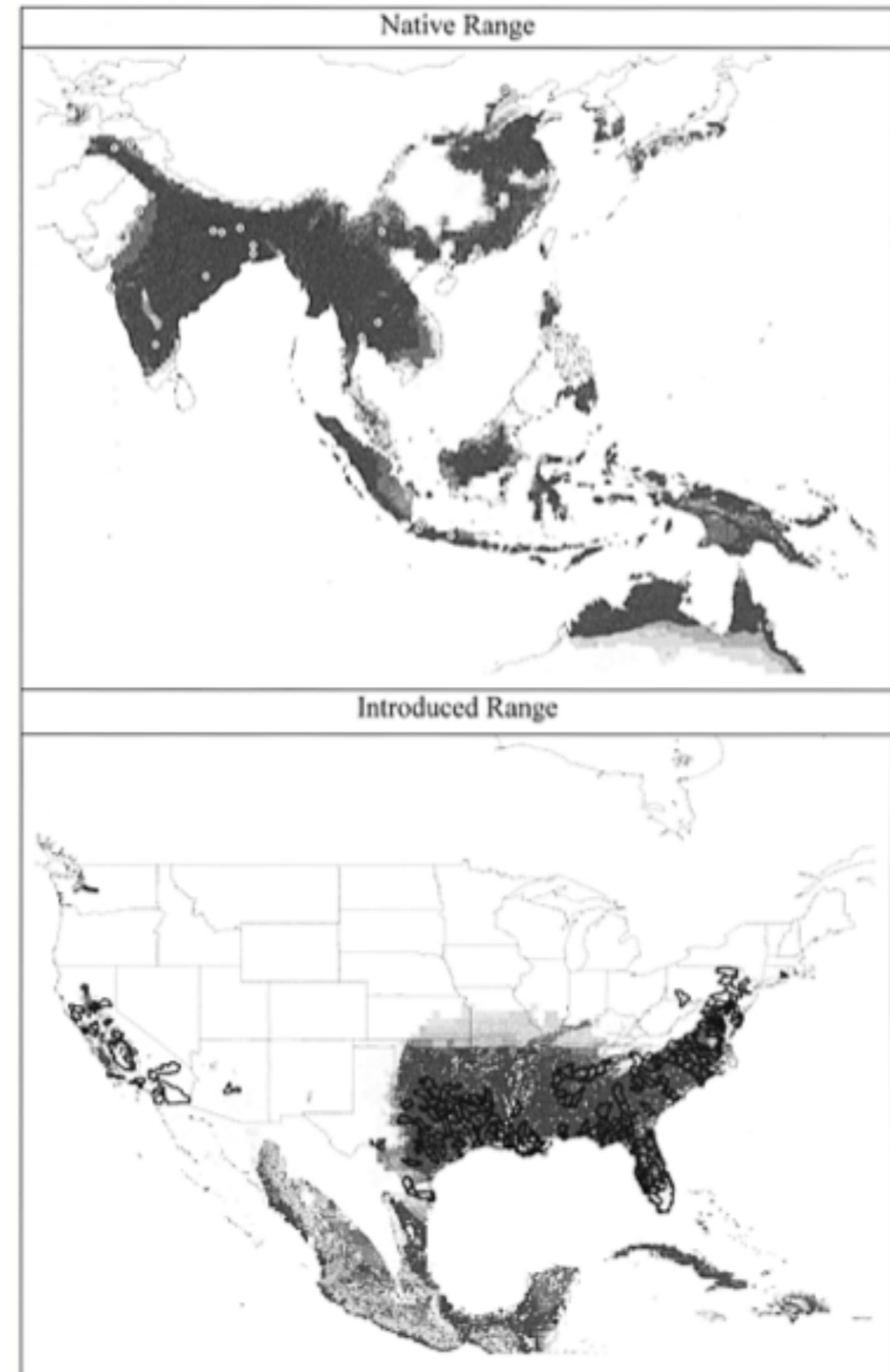
Modelling approaches: focus on invasive species

The case of *Hydrilla verticillata*



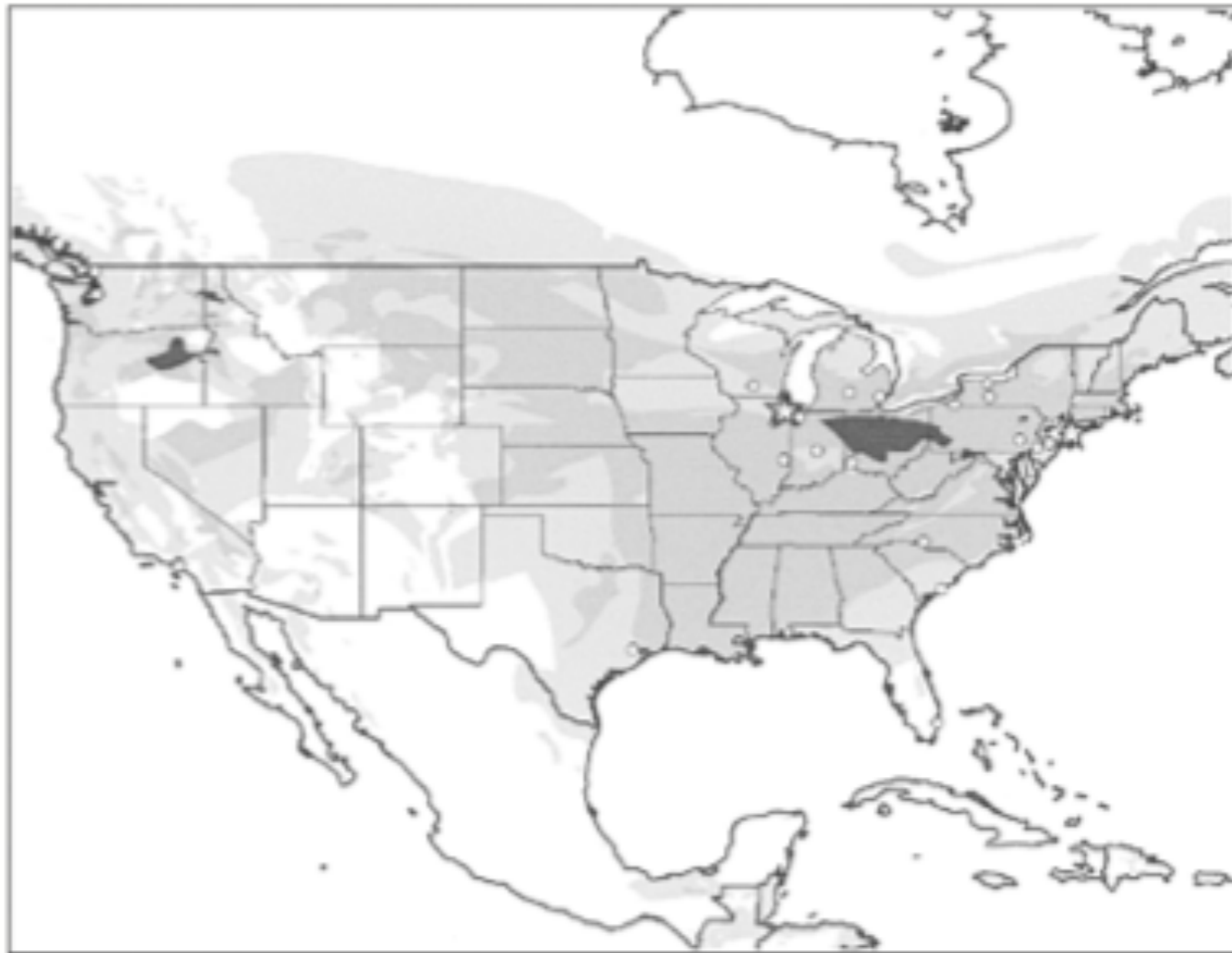
shutterstock.com • 1555036142

NB: Availability of point occurrence data can be a serious limitation



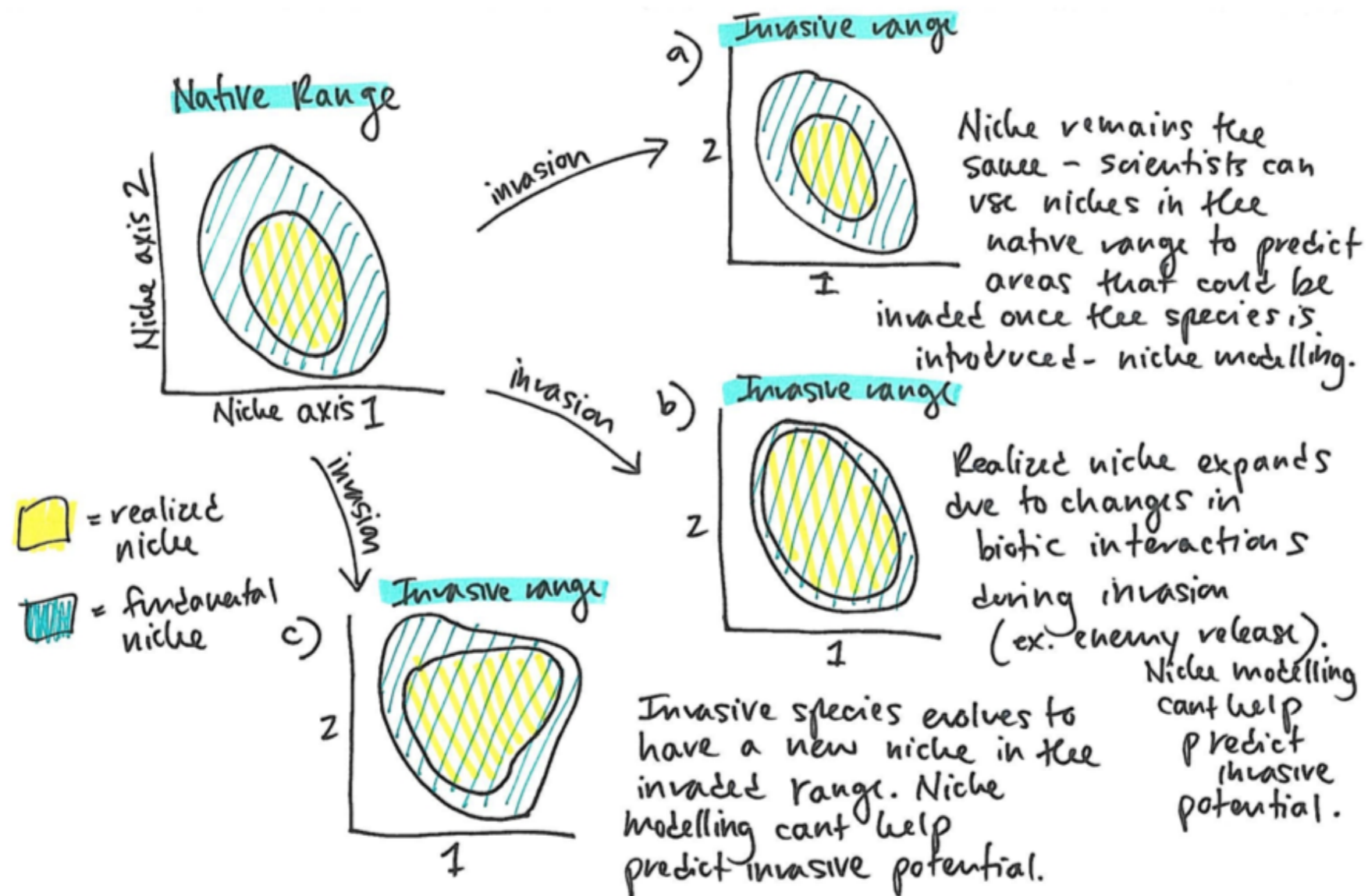
Modelling approaches: focus on invasive species

The case of the Asian longhorn beetle *Anoplophora glabripennis*



Modelling approaches: focus on invasive species

However, correlative models characterize bioclimatic envelopes based on the *realized* niche, since observed species' distributions are, in reality, constrained by non-climatic factors, including biotic interactions. What if the latter disappear?



Modelling approaches: focus on invasive species

Species' invasions is influenced by four factors:

- (1) colonization opportunity to permit species to reach new areas;
- (2) avoidance of demographic problems of small population size early in the establishment of invasive populations;
- (3) ecological appropriateness of the new landscape;
- (4) contagion and extension across that landscape.

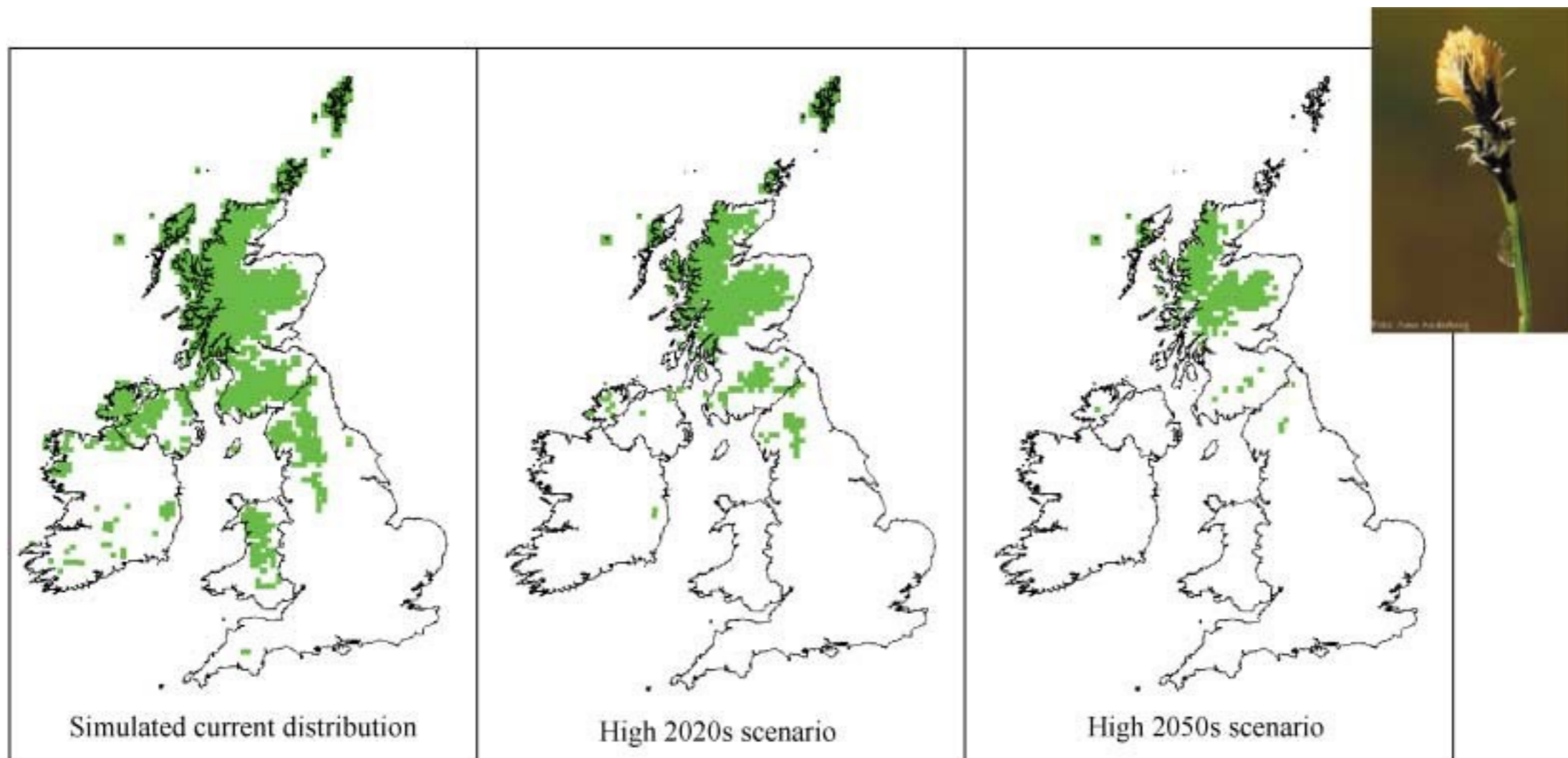
The ecological niche modeling approach focuses solely on the third issue.

Colonization opportunities depend on human economic activity and transportation patterns, wind and ocean currents, and other potential vectors of long-distance movement.

Demographic challenges depend on a series of well-known stochastic factors associated with small population size that can extinguish a small population in spite of the presence of appropriate conditions.

Contagion and extension of species' distributions across the new landscape in turn depend on dispersal ability, movements of environmental media or other species that may help to disperse the invader, and landscape continuity or fragmentation.

Carex bigelowii under future climate scenarios in Great Britain and Ireland



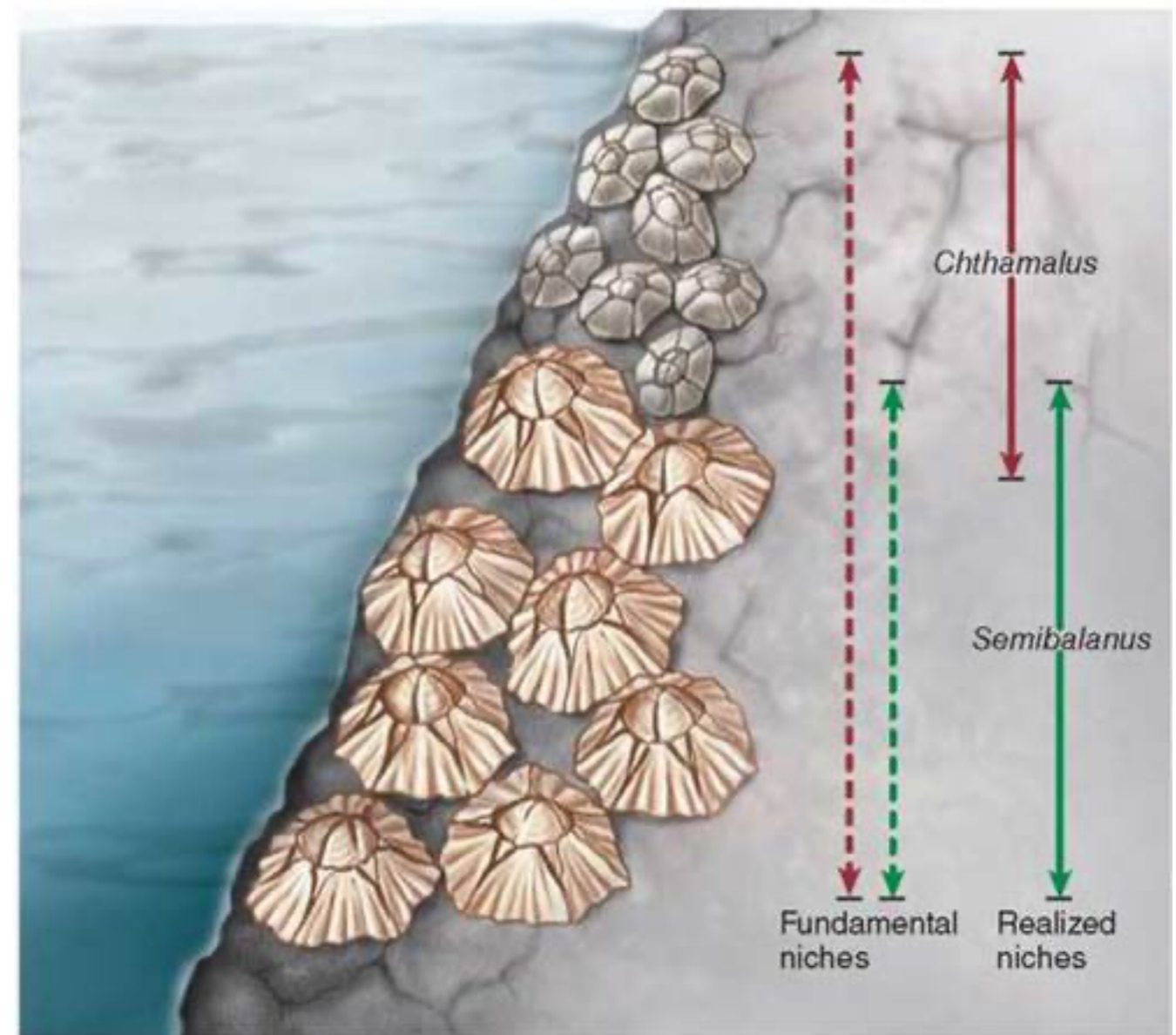
Modelling approaches: climate change

There are fundamental limitations to the predictive capacity of bioclimatic models, especially when dealing with global change scenarios, regardless of the methodology (correlative or mechanistic approaches).

Biotic interactions, evolutionary change, and species dispersal can decrease the predictive power of a model.

Biotic interactions, such as competition, predation, mutualism, parasitism, symbiosis can greatly affect the realized niche of a species, which could range from a full exploitation of the fundamental niche (or more?), down to a very limited distribution in the geographic space.

The example of the competition between *Chthamalus stellatus* and *Balanus balanoides* in the intertidal zone perfectly depicts the concept.



Modelling approaches: focus on invasive species

Plus, especially in certain groups, adaptive changes to species in response to climate change presents a huge challenge to modellers.

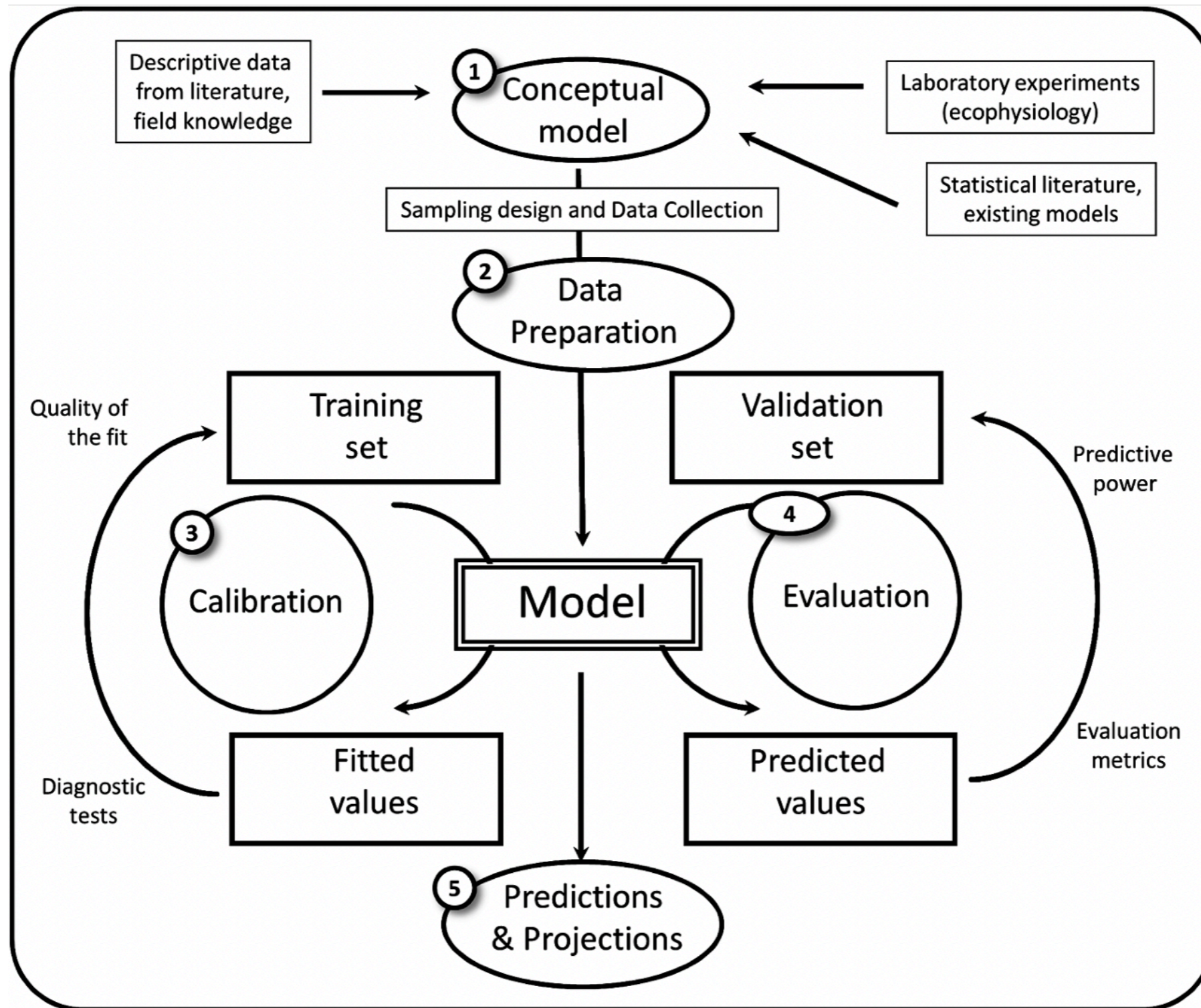
Applications of bioclimate models for predicting distribution changes over the next century are most appropriate for species not expected to be able to undergo rapid evolutionary change over this timescale. This is most likely to be the case for long-lived species and poor dispersers.

On the other hand, the ability of a species to migrate at a sufficient rate to keep up with the changing climate is dependent on the dispersal characteristics of individual species.

Bioclimate niche models do not account for species dispersal, but instead aim at predicting the *potential* range of organisms under changed climate. Though there is great potential to couple bioclimate niche models and dispersal simulations, it is apparent that current predictions of potential distributions may differ greatly from actual future distributions due to migration limitations.

Modelling: how?

Modelling: how?



Spatial distribution modeling ideally follows five steps:

1. Conceptualization
2. Data preparation
3. Model calibration (fitting)
4. Model evaluation
5. Spatial predictions

As in other sciences, SDMs must rely both on robust methodological principles, as well as sound biogeographical, ecological, and evolutionary theory to explain the patterns and causes of species distributions, and from these, of community assembly and distribution.

The success of SDMs is largely down to the increasing availability of spatially explicit biological and environmental data at different spatial and temporal scales.

The **conceptual phase** should serve to identify all the aspects requiring methodological decision to be taken, at the earliest possible stage of the process. These aspects can be divided into two main categories:

(1) *Theory and data*: One of the first requirements is to define: (i) clear scientific question(s) and objectives for the study; (ii) a good conceptual view of the model system used to answer the question(s); (iii) the main underlying assumptions made when building the model, and identifying the necessary proximal environmental predictors for the focal species, including which of these are available or missing. Furthermore, it requires identifying, if necessary, an appropriate sampling strategy for collecting species observations, and choosing the appropriate spatio-temporal resolution and geographic extent for the study.

(2) *Modeling methods*: The second requirement is to identify: (i) the most appropriate method(s) for modeling the response variable; (ii) the optimal evaluation framework; (iii) the statistics needed to assess the predictive accuracy of the model; and (iv) the methods to be used to derive spatial and temporal predictions.

Numerous other conceptual features – methodological, statistical, or theoretical – relating to the different steps of the process need to be assessed as early as possible, ideally during the conceptual phase.

However, it is not always possible to make all the necessary decisions at the very beginning of a study.

This might be due to a lack of knowledge of the target organisms, or of the study area and related data. For instance, the choice of the appropriate spatial resolution might depend on the size of a species' home range, and the way this species uses resources in the landscape. The choice of the geographic extent might depend on prior knowledge of environmental gradients in the study area, to ensure that complete gradients are sampled. For animal species, males vs. females, or summer vs. winter habitats might require separate models. Answers to these questions usually require collecting preliminary field observations, running exploratory analyses on existing data, or conducting experiments.

Theoretical and methodological assumptions

Theoretical and methodological assumptions

Since they aim at representing simplified models of the real world, SDMs are based on strong assumptions, both theoretical and methodological, that are implicitly considered, and should be reviewed before the models and their predictions can be used to answer basic and applied questions.

Theoretical assumptions

The three most important assumptions for applications of SDMs in the present time are:

- (i) the species–environment relation needs to be considered to be at equilibrium (or pseudo-equilibrium)
- (ii) all important environmental predictors required to capture the desired niche of the modeled species are assumed to be available at the resolution relevant for the organism being modeled
- (iii) species observations (simple occurrences, frequencies, abundance, etc.) need to be suited to the later use of the model to answer the initial aims of the study

Species–environment equilibrium assumption

Species data are usually sampled over a limited period of time. Therefore, they are an instance of the species–environment relationship. A practical working postulate is to assume that the modeled species is in pseudo-equilibrium with its environment. Under this assumption, the models capture the fundamental, or something closer to the realized environmental niche of species, and then project it elsewhere, or into a different time period. It is, therefore, expected that the species–environment relationship will not change in space or time.

In other words, the species is expected to have colonized most of its suitable habitats in the studied area. However, there are obvious circumstances in which this assumption does not hold, for instance during biological invasions, or when species are still recolonizing a territory after major environmental changes. Many invasive species are not in equilibrium with their environment in the invaded range, and should thus preferably be modeled using data from their native range, or from both the native and the invaded ranges.

Species–environment equilibrium assumption

In the case of post-glacial recolonization, limited range filling (or overfilling) – calculated as the realized/potential range size ratio – for many European tree species has been evidenced. Many of these species still appear to be strongly controlled by dispersal constraints since post-glacial expansion, and thus might not be in full equilibrium with their environment throughout their whole range. Hence, using models that fit the observations too closely might lead to underestimating the true potential range of the species. Nevertheless, a limited range filling does not imply necessarily that a species' niche cannot be captured from its current distribution. For instance, if all the possible environmental combinations that make the niche of a species are represented by species occurrences, then the niche can be fitted successfully.

However, when limited range filling is also a limited niche filling, the realized niche of the species is smaller than the one that can be expected once the whole territory is (re)colonized. Only fitting part of the realized niche in the area will then logically result in underestimated range sizes.

Theoretical and methodological assumptions

Availability of all important predictors for the niche being captured

An absence of important predictors when modeling species leaves us with unexplained variance.

Therefore, discussions of model predictions and their use to test theories and hypotheses should always clearly refer to the model being used and which predictors it includes.

Important predictors that are unavailable should be identified prior to model fitting and implications anticipated to ensure successful predictions and avoid drawing spurious conclusions.

Using a partial set of predictors might be acceptable if it is clearly stated that the study intends to consider only a subset of the environmental niche, for instance the climatic niche.

Theoretical and methodological assumptions

Appropriateness of species observations

Whether species observations are appropriate can only be determined if we know what the model will ultimately be used for. For instance, identifying potential locations of population persistence requires estimations of population fitness at observation sites. However, most species observation data do not account for population fitness (e.g. if simple presence–absence is available). If this is the case, one cannot exclude sink populations (i.e. outside the species' fundamental niche, where mortality is greater than fecundity and populations cannot maintain viable populations without constant immigration). If data that include sink populations are used, this may seriously mislead some further applications, for instance if the predictions are used to guide conservation decision-making.

Ideally, habitat suitability should be based on measurements of population fitness at each geographic location, but this would prevent the use of most of the data available in natural history collections, or in the GBIF, as these usually do not contain such information.

However, depending on the type of organisms, sink populations may be difficult to detect when modeling over large areas or at coarse resolution (e.g. 10 km resolution). Nevertheless, this remains a potentially important issue, especially when species are modeled at fine spatial resolution.

Theoretical and methodological assumptions

Fundamental methodological assumptions are:

(iv) the statistical modelling methods need to be appropriate for the data being modeled

(v) predictors need to be measured without error

(vi) species data need to be unbiased

(vii) species observations need to be independent

Theoretical and methodological assumptions

Appropriateness of the statistical methods

Different types of response variables require different types of statistical models. For instance, semi-quantitative data require very specific modeling techniques, such as ordinal regressions. Quantitative data are easier to model and there are numerous techniques available for doing so but, different types of quantitative responses will still require different types of statistical models.

Counts of species or of individuals usually require specifying Poisson, negative binomial, or other probability distribution functions for discrete positive values, a requirement that can be met using several modeling techniques, such as generalized regressions or boosted regression trees.

Binary response variables, such as species' presence–absence data, require binomial probability distribution functions, and logistic transformation, and are by far the data for which the largest range of modeling techniques is available in ready-to-use packages.

Once a statistical method has been chosen, it is assumed to be the right one for the data in hand. Failure to identify the correct method can lead to errors and uncertainty in the predictions.

Theoretical and methodological assumptions

Predictors measured without error

This issue is rarely assessed in studies of SDMs, although errors are an inherent factor in each GIS predictor layer. Although it is essentially impossible to guarantee zero errors in mapped environmental predictors, estimates of spatial distribution of errors could be associated with each layer, and used to calculate spatial uncertainty in the model predictions. There is still a need for a proper method for combining errors from the different environmental variables in the model.

Unbiased species data

SDMs attempt to quantify the environmental niche of species through models. Therefore, the data need to include all possible environments that represent suitable habitats for the species modeled, at least within its colonizable range. Bias is likely to arise when the chosen sampling design lacks a random component, or when the data are gathered without employing a designed sampling strategy (i.e. subjective sampling). The latter typically results in data being clustered in more accessible areas, for instance along communication axes (this is often true especially for citizen science data), or in some habitats being preferentially sampled, or outside others being left out based on prior knowledge, or the observer's judgment. Any bias can potentially lead to partial niche quantification and thus to models that fail to identify all suitable habitats of a species, with obvious consequences for the spatial predictions.

Independence of species observations

When species observations are not independent (e.g. they are spatially autocorrelated), the actual number of degrees of freedom used in many statistics associated with the models no longer corresponds to the apparent number of observations.

There is no simple way of checking whether observations are independent, especially in biological systems where species interact with each other, and disperse into neighboring sites. It is possible to assess spatial autocorrelation, but this only informs on spatial patterns, not on the processes behind these patterns. This means no one can know for certain if the data are dependent due to biological processes, or if the observed patterns simply result from hidden spatially clustered important environmental variables that affect species distributions.

If spatial autocorrelation is not corrected, this might cause additional problems especially if the SDMs are projected into the future.