Guest Editorial - *Journal of Biogeography* (2011) – in press

Running header: Predicting spatio-temporal patterns of species assemblages

SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages

Antoine Guisan[1*] & Carsten Rahbek[2*]

[1] Department of Ecology & Evolution, University of Lausanne, Switzerland,
[2] Center for Macroecology, Evolution and Climate, University of Copenhagen, Denmark

* Correspondence (joint first authorship):
Antoine Guisan, Department of Ecology & Evolution, University of Lausanne, 1015 Lausanne, Switzerland. Email: antoine.guisan@unil.ch
Carsten Rahbek, Center for Macroecology, Evolution and Climate, Department of Biology, University of Copenhagen, Universitetsparken 15, DK-2100, Denmark. Email: crahbek@bio.ku.dk

ABSTRACT

Two different approaches currently prevail for predicting spatial patterns of species assemblages. The first approach (macroecological modelling, MEM) focuses directly on realized properties of species assemblages, whereas the second approach (stacked species distribution modelling, S-SDM) starts with constituent species to approximate assemblage properties. Here, we propose to unify the two approaches in a single 'spatially-explicit species assemblage modelling' (SESAM) framework. This framework uses relevant designations of initial species source pools for modelling, macroecological variables, and ecological assembly rules to constrain predictions of the richness and composition of species assemblages obtained by stacking predictions of individual species distributions. We believe that such a framework could prove useful in many theoretical and applied disciplines of ecology and evolution, both for improving our basic understanding of species assembly across spatio-temporal scales and for anticipating expected consequences of local, regional or global environmental changes. In this paper, we propose such a framework and call for further developments and testing across a broad range of community types in a variety of environments.

Keywords: Biodiversity, community properties, ecological assembly rules, ecological niche modelling, macroecological constraints, source species pool, species richness, species sorting, stacked species predictions.

SETTING THE SCENE

Explaining the spatial and temporal distribution of biological diversity on Earth has been a research focus since the days of Alexander von Humboldt, Augustin Pyramus de Candolle, Alfred Russel Wallace and Charles Darwin, and it remains one of the major focuses in biogeography and macroecology. Understanding the processes governing the distribution and assembly of biological communities has become a prerequisite for successfully predicting how the world will look in the wake of global environmental changes. Currently, two distinct predictive spatial modelling approaches prevail (Ferrier & Guisan, 2006), which rely on two theoretical paradigms. The first paradigm focuses directly on realized properties of species assemblages (e.g. Brown, 1995), such as richness, and the methods used include macroecological modelling (MEM, see Gotelli *et al.*, 2009). The second paradigm focuses on aggregate properties of individual constituent species, used to reveal assemblage properties (e.g. Lortie *et al.*, 2004; Ackerly & Cornwell, 2007) and applies species distribution modelling (SDM, see Guisan & Thuiller, 2005; Elith & Leathwick, 2009) to a spatial stack of species. Properties of species assemblages include the number of co-occurring species (richness), inter-specifc abundance patterns, and compositional (e.g. community types), functional and structural characteristics. Hereafter, all of our examples use species richness, the simplest measure of biodiversity and the most commonly considered property of species assemblages (Whittaker *et al.*, 2001).

In MEM, species richness is predicted directly, either based on theoretical expectations or from various factors thought to control the number of species able to coexist in a geographical unit (Fig. 1, top). The main controlling factors are typically hypothesized to be available energy, environmental heterogeneity, disturbance or history, with scale effects and some level of stochasticity (Whittaker *et al.*, 2001; Currie *et al.*, 2004; Mittelbach *et al.*, 2007; Field *et al.*, 2009; Gotelli *et al.*, 2009). The same approach can be used to model any other property of communities, although different hypotheses and explanatory variables are likely to apply to each property. MEM is typically applied either using a curve-fitting correlative approach, predictive simulation or other modelling techniques (Hawkins *et al.*, 2003; Currie *et al.*, 2004; Rahbek *et al.*, 2007; Gotelli *et al.*, 2009). This approach has been widely used in the last two decades (e.g. Pausas & Austin, 2001; Currie *et al.*, 2004; Thuiller *et al.*, 2006; Rahbek *et al.*, 2007; Currie & Kerr, 2008; Algar *et al.*, 2009).

By contrast, the stacked SDM approach (S-SDM) first predicts the distribution of a suite of individual species in a given study area and then assembles them to reconstruct community properties in each geographical unit (Ferrier & Guisan, 2006). Here, species richness is simply predicted by stacking presence–absence predictions of all species. This approach thus relies on our ability to model the distributions of individual species, a field that has greatly matured over the last two decades (see Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005; Elith & Leathwick, 2009; Franklin, 2010). Thus, in SDMs, the main factors are those controlling individual species distributions and often purely abiotic variables are used. Hereafter, the terms 'habitat' and 'environment' always refer to the abiotic environment (although we are aware that habitat may be used elsewhere to refer to some biotic properties). Estimating species richness using S-SDMs (e.g. Guisan & Theurillat, 2000; Lehmann *et al.*, 2002; Wisz *et al.*, 2007; Algar *et al.*, 2009) is more recent than the MEM approach. It originated from the many attempts to map current patterns of species richness for reserve design (e.g. Young *et al.*, 2009) and to forecast the likely impact of climate or land-use change on biodiversity (e.g. Guisan & Theurillat, 2000; see Ferrier & Guisan, 2006). This approach of stacking single-species predictions arose from the view that species respond individualistically to environmental changes and that the structure of future communities may

differ from contemporary ones (Huntley, 1991; Williams & Jackson, 2007), in turn yielding different spatial patterns of diversity (Stralberg *et al.*, 2009).

Currently, the two approaches can be used separately to make the same prediction (Figs 2 and 3). Although the two approaches build on different philosophies, both can potentially use similar environmental factors (e.g. climatic) but with different hypotheses. For instance, in S-SDM, climate constrains the boundaries of ranges, whereas in MEM, climate constrains the number of species that can coexist in an area. In contrast, environmental heterogeneity has a recognized effect on species richness (Gaston, 2000) but is a less meaningful predictor of single species distributions.

MEM and S-SDM both have strengths and limits for predicting patterns of species assemblages. A strength of the MEM approach is that it forces us to seek general rules driving ecosystem functioning (e.g. energetics) and constraining the assemblage of species. A major drawback is that it cannot derive any information about the composition of species assemblages and the biotic interactions that shape them from the predicted patterns of species richness. For example, a geographical unit may experience a complete turnover (i.e. a change of all species) but maintain the same value of species richness, as revealed by analyses of assemblage composition change (see Ferrier & Guisan, 2006). In contrast, the S-SDM approach can predict which species co-occur in a given unit or area. However, the SDM approach is limited because: (1) without adding a dispersal filter, it may incorrectly predict species in areas that appear environmentally suitable but that are outside their colonizable or historical range (Wisz *et al.*, 2007); (2) it does not consider any *constraints* based on the carrying capacity of the local environment that determine the maximum number of species that may co-occur (e.g. species–energy or metabolic theory; Brown *et al.*, 2004; Currie *et al.*, 2004); and (3) it does not explicitly consider any *rules* based on biotic interactions that control species co-occurrences and can exclude species from a community (e.g. through competitive exclusion; Anderson *et al.*, 2002). Due to these three issues, too many species can easily be predicted to occur in an geographical unit by S-SDMs (e.g. Graham & Hijmans, 2006; Pineda & Lobo, 2009; see also Figs 2 & 3).

The best way to understand differences between MEMs and S-SDMs is to examine possible mismatches between their predictions using real datasets. Surprisingly, however, predictions by MEM and S-SDM have so far been mainly used and discussed separately rather than compared or integrated (but see Guisan & Theurillat, 2000; Ferrier *et al.*, 2002; Algar *et al.*, 2009; see table 2 in Ferrier & Guisan, 2006).

CONTRASTING MEM AND S-SDM: ASSUMPTIONS AND EXAMPLES

What differences should we expect between MEM and S-SDM predictions? Given that constraints exist on the maximum number of species that can coexist in a given geographical unit but no macroecological constraints have been set on the realized assemblage, the pool of species predicted by the S-SDM approach should generally tend to exceed the actual number of species found in each unit of analysis. We expect this overprediction of species richness by S-SDMs to occur at any spatial resolution, from fine to coarse scales, whether a geographical unit encompasses a single community or several; if overprediction occurs for one community it should logically be observed for several. In contrast to S-SDM, the MEM approach should predict values around the actual number of species in the unit, by modelling species richness directly.

Therefore, the following *assumption* can be made: in an integrated modelling framework, MEM predictions should allow the maximum number of species predicted by the

S-SDM approach in each spatial unit to be constrained. For the integrated modelling framework to be valid, the S-SDM-based predictions should, on average, overpredict species richness compared to observations (as found in Algar *et al.*, 2009; Pineda & Lobo, 2009) and significantly exceed the MEM-based predictions of species richness.

We examine this assumption with two representative examples. The first uses a large dataset at high-resolution and local extent to assess assemblages of *c*. 300 plant species in the western Swiss Alps (see Dubuis *et al.*, in press), whereas the second uses a dataset at coarse resolution and continental extent to assess assemblages of *c*. 4000 bird species throughout the New World (Diniz-Filho *et al.,* 2009). Here, we only provide information useful for discussing our assumption. Details of methods and results are provided in the respective papers.

In the Swiss Alps plant study, MEM and S-SDM predictions were strongly correlated with observed species richness, but the S-SDM approach consistently overpredicted species richness (Fig. 2e), whereas the MEM approach better approximated SR values around the observed mean (Fig. 2d). Thus, as expected, species richness predictions from S-SDM largely exceeded those from MEM (Fig. 2f). Similarly, in the New World birds example (Fig. 3), both MEM and S-SDM predictions had a very tight relationship with observed species richness (Fig. 3d and 3e), but while the MEM predicted species richness values around the mean (Fig. 3d), the S-SDM dramatically overpredicted species richness in most units (see Fig. 3e). In fact, S-SDM overpredicted species richness in 98% of all grid cells, with an overprediction of $\geq$ 100% in 25% of cells. In comparison, MEM over- or underpredicted species richness by 100% in a total of only 3% of cells. These two examples thus support the assumption outlined above, and pave the way towards a new framework combining the strengths of S-SDM and MEM to predict spatial patterns of species assemblages.

TOWARDS A NEW FRAMEWORK FOR PREDICTING SPATIAL PATTERNS OF SPECIES ASSEMBLAGES

Identifying the components of assemblage predictions

We propose a new modelling framework – spatially-explicit species assemblage modelling (SESAM) – integrating MEM and S-SDM by using MEM to set a limit to the number of species predicted by S-SDM. Adding this macroecological constraint is expected to increase the prediction success of realized species richness. However, because historic and biotic factors also play a role in shaping the distribution of species (MacArthur, 1972; Pulliam, 2000; Soberón, 2007) and communities (Ricklefs & Schluter, 1993; Leibold *et al.*, 2004; Lortie *et al.*, 2004), applying environmental filtering only (i.e. using simple SDMs) is not enough; two other filters – dispersal and biotic interactions – are also needed. To our knowledge, these four components have never been fully integrated in the same modelling framework.

We have previously seen that environmental filtering may be achieved for multiple species using S-SDMs and that the macroecological constraint can be addressed using MEM. Dispersal filtering (i.e. removing species outside their historical range) can most adequately be addressed by defining the appropriate source species pool (SSP) and biotic filtering by defining ecological assembly rules (EAR) (see below). The four steps of our SESAM framework can now be defined (Fig. 4). Note that early ideas on this framework are outlined in Ferrier & Guisan (2006, in their perspectives and Fig. S1).

4

Formalizing the SESAM framework

To illustrate the framework, let us first assume that species data are a set of geographical units (e.g. quadrats, plots, grid cells) that each have an exhaustive species list for the group(s) considered and that all species are included in a larger regional species pool (i.e. aggregated across all geographic units). The domain used for our predictions is a grid of units (i.e. raster maps) or possibly a set of irregular units (vector map, such as a set of islands) with environmental information covering the area of interest.

The SESAM framework unifies the four independent approaches to biodiversity prediction, i.e. SSP, S-SDM, MEM and EAR, by a process of four main steps (Fig. 4) that sequentially filter, for each analytical unit, the species assemblage from the initial regional species pool to the local source pool (step 1), to the environmental pool (step 2) and finally to the realized species assemblage (steps 3 and 4). The difference between the successive species pools can be substantial at different stages of the filtering process. The steps and filters between the successive species pools are:

Step 1 *Dispersal filtering – source species pools (SSP).* This step consists of defining the local source species pool from the regional source pool, as a way to account for historical and contemporary dispersal limitations. Using an empirical approach, this aims to determine which species in the regional source pool could have dispersed to a given unit, ignoring species interactions and species-specific abiotic habitat conditions that are considered in the next steps. This step can be ignored if the same regional source pool can reasonably be applied to all modelled units (i.e. in the absence of evidence of dispersal limitation within the study area).

Step 2 *Abiotic habitat filtering – stacked species distribution modelling (S-SDM).* This step consists of using SDMs to model abiotic habitat suitability (hereafter simply 'habitat') for all species in the previous local source pool (step 1) across the study area and stacking predictions to define the environmentally filtered species pool in each unit.

Step 3 *Biotic filtering – macroecological modelling constraints (MEM).* This step consists of predicting the distribution of assemblage properties, such as species richness, from macroecological modelling and using these predictions as constraints on the community to be assembled from the species habitat pool. To predict the species richness of each geographical unit (e.g. grid cell), this step determines what controls the number of species expected to co-occur within each unit as a way of imposing a constraint based on the resources or environmental heterogeneity characterizing the unit (i.e. environmental carrying capacity) and without which no biotic interaction can take place. This step thus defines *how many* species will be allowed in each unit but does not define which ones.

Step 4 *Biotic filtering – ecological assembly rules (EAR).* This step consists of using ecological assembly rules (EAR) to select *which* species from the environmental-filtered pool will be able to coexist in each unit, conditional to the value (e.g. maximum value of species richness) defined in step 3. It requires that EARs have been previously developed. This step yields the final species assemblage predictions for each unit (Fig. 4).

Summarizing the SESAM framework, steps 1 and 2 filter the global species pool down (in each unit) to a pool of potential species for a given region and abiotic habitat. Step 3 then tells us how many species must be filtered out to end up with a fairly realistic number of species (or any other community property) that makes ecological sense within each mapping unit (i.e.

5

for which there is support from a well-formulated macroecological model). Finally, step 4 simply filters which of the species from the larger habitat species pool yielded in step 2 can co-occur in each unit based on pre-defined rules to end up with the right number of species (or the right community property) defined in step 3. Applying this sequence of steps should allow a more accurate prediction of the realized species assemblage in each geographical analytical unit (Fig. 4).

An essential additional step to validating the framework is of course the comparison of the final assemblages predicted after step 4 with actual assemblages from an independent dataset that was not used for fitting any stage of the framework (see later section on testing the framework).

THE COMPONENTS OF THE FRAMEWORK: CURRENT KNOWLEDGE AND FUTURE NEEDS

The four components of the proposed SESAM framework are all anchored in robust and tested bodies of ecological theory, but the practical formulation of predictive models based on these theories varies among them. Here we discuss current understanding and identify future needs with regard to each step of the suggested framework, with a view to the application of SESAM.

Defining the source species pool (step 1)

The size of the source species pool (SSP) ultimately constrains the number of species expected in a given unit (Ricklefs & Schluter, 1993; Valone & Hoffman, 2002; Graves & Rahbek, 2005). Thus, it is important to determine a meaningful source pool for each unit to generate realistic expected values of, for example, the number of species given certain environmental conditions. The best design of a source pool is case-specific and often varies with the spatial extent and grain size of the analyses. For analyses covering a small spatial extent, the source pool for the units of analysis is most often assumed to be the total number of species occurring within the study area and is thus the same for all modelled units. However, for analyses covering larger spatial extents, units of analysis within the modelled area may vary significantly in their potential SSP, as a result of divergent histories of speciation, extinction and colonization (Graves & Gotelli, 1993; Lyons, 2003). Hence, at a large geographical extent, the composition of the source pool may have to be defined separately for each analytical unit using, for instance, smoothed distribution data and/or the dispersion field approach (e.g. Graves & Rahbek, 2005). This per-unit source pool issue has a clear biological foundation (e.g. in historical biogeography) and should not be viewed only as a technical problem.

Although already useful, the species pool concept requires further development for optimally defining the set of species that can occur in a given geographical unit and for linking species pools at different scales. Furthermore, local source pools may be appropriately set by dispersion fields (Graves & Rahbek, 2005, see also Arita *et al.*, 2008), but delimiting the geographical extent of the latter remains unresolved.

Defining environmental suitability (step 2)

Species distribution models essentially perform abiotic environmental filtering (SDM; Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005; Elith & Leathwick, 2009; Franklin, 2010).

SDMs statistically relate mapped environmental characteristics of the abiotic habitat with observed occurrences, presence–absence, abundance or fitness of the species, thus giving an approximation of its ecological niche (i.e. the realized envelope of all suitable abiotic habitats). Using the quantified species–environment relationship combined with environmental maps then allows prediction of the potential distribution of the species across the whole area or in a different area for which the same environmental information is available (e.g. to predict biological invasion or to test model transferability). Changing the input maps using different scenarios (e.g. for land use or climate change) allows future projections of species' range changes to be derived (e.g. Engler *et al.*, 2009) and thus theoretically allows the framework to be used for global change applications. Factors identified as possibly influencing the accuracy of SDMs include, in order of importance, quality of predictors, intrinsic properties of species (e.g. successional status), modelling techniques, and data properties such as sample size, location error, or scale (Austin, 2007; Guisan *et al.*, 2007; Elith & Leathwick, 2009; Franklin, 2010).

The following improvements of SDM are particularly challenging: (1) adding more ecological realism and mechanistic explanation (Guisan & Thuiller, 2005; Austin, 2007) by, for example, fitting abundance models; (2) using multi-technique 'ensemble modelling' approaches and associated probabilistic predictions (Araújo & New, 2007; Diniz-Filho *et al.*, 2009) to better handle model uncertainty (related to data, resolution, techniques etc.) and improve predictions of rare species (Lomba *et al.*, 2010); the latter can prove important components of diversity (Sizling *et al.*, 2009) but are currently difficult to model with SDMs due to their low prevalence in datasets (Stockwell & Peterson, 2002); and (3) assessing what probability threshold should be used for converting predicted species probabilities into presence–absence information (Pineda & Lobo, 2009), and its implications for defining the species pool or developing approaches to use probabilistic predictions to reconstruct community properties directly.

Defining macroecological constraints (step 3)

Defining a source species pool for each modelled unit could be a way to account for dispersal limitation. However, besides dispersal and environmental filtering, we know little about how individual species assemble from a species pool to form a community. Macroecological models may be used to impose a limit on the number of species that can theoretically co-occur in a given unit based on the space and/or resources available (often interpreted as a rough expression of the environmental carrying capacity of a unit, determining the level of community saturation; Loreau, 2000; see also Currie *et al.*, 2004). The existence and conditions of such a theoretical limit still need to be assessed. Additional insights may come from studies of biological invasions because community saturation may be an important factor preventing invasions from occurring. However, contrasting evidence exists for and against saturation, depending on the ecosystem type and organism being studied (Sax *et al.*, 2007). Assuming that one can successfully use, for example, MEM to determine the limit to the number of species that can co-occur in an area, the next step is to select which species are most likely to coexist in each geographical unit and corresponding species pool. This essentially means incorporating the potential effect of biotic interactions in predicting the composition of species assemblages (Lortie *et al.*, 2004; Gotelli *et al.*, 2010). This may be achieved by quantifying EARs (next section).

Traditionally, MEMs have been based on correlative models that did not explicitly incorporate the mechanisms responsible for the distributions (see discussion in Colwell *et al.*, 2004). Recently introduced spatially explicit Monte Carlo models of the placement of

geographical ranges in an environmentally heterogeneous landscape (Rahbek *et al.*, 2007; Gotelli *et al.*, 2009) can potentially overcome some of the major obstacles of the correlative approach. However, the major challenge remains to define more precisely whether MEM (or alternative models) predictions can be used to constrain the assembly of species into realized communities and, if so, how.

Defining ecological assembly rules (step 4)

EARs are defined as ecological restrictions on the observed patterns of species assemblages that are based on one or more other species or groups of species (Diamond 1975; Wilson & Gitay, 1995). Existence of EARs can be tested through species-based approaches using co-occurrence data, such as total mutual exclusion of species pairs (checkerboard patterns) or, conversely, quantification of species pairwise combinations. In Gotelli & McCabe's (2002) meta-analysis of EARs and their null model testing, the majority of cases strongly supported the existence of EARs. In cases where the approach proves too difficult (e.g. too many species, as in some tropical ecosystems), recourse can be taken to functional approaches based on functional groups, guilds or quantitative traits of species (Graves & Gotelli, 1993; Wilson & Roxburgh, 1994; Shipley *et al.*, 2006), such as testing for patterns of species within functional groups (guild proportionality) or patterns of non-similar resource acquisition traits within communities where high levels of competition are expected (limiting similarity). Nevertheless, this is a re-emerging field, and very few examples of studies exist where communities are recomposed from a abiotic habitat-filtered species pool using assembly rules (e.g. Mouillot *et al.*, 2007; Cornwell & Ackerly, 2009; Ingram & Shurin, 2009). Likewise, documentation of the existence of patterns of species aggregation and segregation at spatial scales larger than local assemblages (e.g. continental scales at which many macroecological and SDMs operate) remains relatively rare (Gotelli *et al.*, 1997; Gotelli *et al.*, 2010). In the context of SDM, the importance of biotic interactions in predicting species distributions has been demonstrated, for example, by incorporating a dominant species as a predictor in a model for another species (Leathwick & Austin, 2001; Pellissier *et al.*, 2010).

EARs are certainly the component of the framework that requires most development, including (1) identifying key species traits that play important roles in community assembly and/or sorting species into (meaningful) functional groups or guilds; (2) quantifying anddefining rules between species or functional groups and testing them (e.g. McGill *et al.*, 2006; Ackerly & Cornwell, 2007); and (3) defining how to use EARs to predict the correct number of species in a community given the macroecological constraint that has been previously defined. EARs must not only be considered within a single taxonomic group but also between distinct groups (e.g. plants and animals; Bascompte *et al.*, 2003; Sargent & Ackerly, 2008). As a first approximation, EARs could be developed further from existing empirical data on species co-existence after environmental filtering (Peres-Neto *et al.*, 2001). However, these data need to be complemented with novel data from well-designed field surveys and from experiments (e.g. to measure species' relative competitive ability; Keddy *et al.*, 2002) with eco-physiological measurements (e.g. to better assess the fundamental response of species along environmental gradients in the absence of biotic interactions; Araújo & Guisan, 2006) and phylogenetic data (e.g. Donoghue, 2008; Graham *et al.*, 2009). We also need to develop methods to incorporate identified EARs into the predictive process that allow, for instance, predictions of the most probable assemblage composition (Weiher & Keddy, 1999) or a list of equiprobable assemblages, possibly resulting in different assembly endpoints (Law & Morton, 1993). Leaving the heated controversy over EARs that took place two decades ago aside (see Gotelli *et al.*, 1997), developing EARs in a broader context is one

of the most promising but also one of the most demanding directions for spatial ecology research.

CRITICAL ASSESSMENT OF THE SESAM FRAMEWORK

By suggesting the integration of four main components and the combination of many analytical approaches, the SESAM framework is by nature complicated. However, it can be broken down into each of its constitutive steps (none of which is overly complex, per se), which can be investigated independently. Below we discuss the practicalities and limitations for each of these.

Practicalities of the SESAM framework

In step 1, the assignment of individual species to the regional species pool is taken as all species in the area considered under stable environmental conditions. Historical barriers to migration or dispersal limitations within the study area can be elucidated and quantified by spatial autocorrelation analyses (see Diniz-Filho *et al.*, 2003; Dormann *et al.*, 2007). When these barriers are considered along with the regional species pool, the source pool for each unit of analysis can then be estimated by one of several available methods, including the dispersion fields approach suggested by Graves & Rahbek (2005) or by interpolating species occurrences (Wisz *et al.*, 2007). However, under changing environmental conditions, such a changing climate, dispersal constraints also need to be considered as transient when running SDM using an appropriate dispersal model (e.g. Engler & Guisan, 2009; Midgley *et al.*, 2010). However, this approach does not consider the arrival of new species to the study area (e.g. as a consequence of climate change) that yield new communities (e.g. Guisan & Theurillat, 2000; Williams & Jackson, 2007; Stralberg *et al.*, 2009) and possibly novel ecosystems (e.g. Seastedt *et al.*, 2008). One solution to this problem may be to fit SDMs at a larger and coarser scale to redefine the species source pools after climate change, given some dispersal constraints.

In step 2, habitat suitability must be considered individually for each species. Thus, community modelling approaches (Elith *et al.*, 2006; Baselga & Araújo, 2009) that use information from one species to model other species should not be used to avoid accounting twice for biotic interactions in the framework. Even though species are modelled individually, by doing so one already accounts for part of the biotic interactions that constrain/favour their distribution by fitting their realized niche. In this case, EARs in step 4 can only attempt to capture the remaining biotic interactions signal. A more thoughtful approach would be to map species distributions based on their fundamental niche (Kearney & Porter, 2004; Kearney & Porter, 2009), use experimental measurements of their ecophysiological requirements, and then use EARs to constrain/expand the predicted distribution. However, appropriate ecophysiological knowledge of species currently exists mostly for cultivated plants as well as for some vertebrate species and pathogenic organisms (e.g. insect species that are disease vectors). Although this approach is currently not feasible for most species hopefully this will change with increasing recognition of the importance of modelling fundamental niches. Nevertheless, even if empirical data used to fit SDMs are already partially shaped by dispersal and biotic interactions, predictions can still be additionally filtered by these two factors; habitat-suitability models are still able to predict a species in environmentally suitable areas outside its colonizing range, or where biotic agents exclude, or do not facilitate, its presence (Pulliam, 2000; Soberón, 2007).

In step 3, the expected value of species richness (or another assemblage property) per analytical unit may be determined theoretically by applying a function based on macroecological theory (Rahbek *et al.*, 2007; Field *et al.*, 2009; Gotelli *et al.*, 2009) or empirically by using any MEM approach. Advantages and limitations of the various MEM approaches are thoroughly discussed in the macroecological literature (e.g. Gotelli *et al.*, 2009 and references therein).

Step 4 consists of identifying members of a given assemblage using EARs derived from community analyses at the species (e.g. from theory, co-occurrence data or experiments) or functional trait (Shipley *et al.*, 2006) levels. When modelling patterns of species or community properties (e.g. species richness) using MEM and SDM, biotic interaction processes have traditionally been overlooked. This is despite the fact that they can potentially influence geographical co-existence patterns of species (Diamond, 1975; Leathwick & Austin, 2001; Gotelli & McCabe, 2002; Pellissier *et al.*, 2010), thus shaping communities and influencing the composition of species assemblages, even for continental assemblages at larger geographical scales (Gotelli *et al.*, 2010). Step 4 will not prove useful if assemblages only result from a purely Gleasonian perspective of communities, i.e. being only the coincidence of species sharing similar environmental requirements, thus not requiring any constraint or assembly rules (i.e. avoiding steps 3 and 4).

Testing and improving the SESAM framework

Constraining stacked predictions of individual SDMs by MEM to predict community properties (like species richness) remains a largely unexplored field. Therefore, testing the SESAM framework on independent data is necessary to ensure its proper evaluation. Good agreement between predicted and observed assemblages will provide strong support for the proposed framework. Knowing when to consider predictions as satisfactorily matching observations remains an open question in predictive modelling, and such decisions are dependent on the study objectives and the expected level of agreement between observed and expected outcomes.

Furthermore, if an incorrect prediction is obtained at the end, which component(s) of the framework most contributed to the failure? Each of the four steps should thus also be evaluated independently; each of them bears its own uncertainty and contributes to the final prediction error, but they may do so unequally. For instance, EARs may need to be developed on one dataset and then tested on a second dataset, independent of the former. Finally, if successful, they may need to be applied to a third dataset to test the whole SESAM framework. The same process can be applied to the three other steps.

Although these important evaluation questions deserve further investigation, answering them is not our primary purpose here. We acknowledge that specific questions related to various parts of the SESAM framework must be solved with regard to how to evaluate each modelling step and the final predictions of the whole framework. Questions that must be addressed include: what if different EAR sets yield the same final predictions, and what if different approaches to the definition of species pool also yield the same final predictions? In fact, such findings, if observed, will pose new questions and pave the way for in-depth analyses of the SESAM framework and more generally of community assembly in space and time. For instance, can different assembly end points be obtained from the same initial pool of habitat-filtered species (Law & Morton, 1993)? Metacommunity theory suggests that this may occur in high-productivity sites where both alternative stable states and cyclical changes in composition are more likely (Leibold *et al.*, 2004). However, larger and more accurate datasets on species' distributions and ecology (including ecophysiology, in an attempt to

capture the fundamental environmental niche) for entire regional species pools are most needed to allow fitting and testing each step of the framework as independently as possible.

CONCLUDING REMARKS

Ecological science is under pressure to predict the potential consequences of global changes on biodiversity and ecosystem services. Species distribution modelling and macroecological modelling are two major approaches currently used to predict biodiversity patterns. Both have progressed significantly in recent years, but thus far, they have been used independently of each other. In this guest editorial essay, we have explained why we think that a major step forward should now be to integrate them into a single unifying framework by using the latter to constrain predictions of the former. Such a framework must consider additional components, including (1) source species pools to account for dispersal limitations and the speciation-extinction history of areas, and (2) ecological assembly rules to account for biotic interactions. We invite others to contribute ideas and solutions to specifically improve the suggested framework, both theoretically and practically. The framework can undoubtedly be fine-tuned, and we encourage any such attempt using novel data.

REFERENCES

Ackerly, D.D. & Cornwell, W.K. (2007) A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. *Ecology Letters*, **10**, 135-145

Algar, A.C., Kharouba, H.M., Young, E.R. & Kerr, J.T. (2009) Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography*, **32**, 22-33

Anderson, R.P., Peterson, A.T. & Gómez-Laverde, M. (2002) Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, **98**, 3-16

Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677-1688

Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42-47

Arita, H.T., Christen, J.A., Rodriguez, P. & Soberón, J. (2008) Species diversity and distribution in presence-absence matrices: mathematical relationships and biological implications. *The American Naturalist*, **172**, 519-532

Austin, M. (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1-19

Bascompte, J., Jordano, P., Melian, C.J. & Olesen, J.M. (2003) The nested assembly of plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences USA*, **100**, 9383-9387

Baselga, A. & Araújo, M.B. (2009) Individualistic vs community modelling of species distributions under climate change. *Ecography*, **32**, 55-65

Brown, J.H. (1995) *Macroecology*. The University of Chicago Press, Chicago.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M. & West, G.B. (2004) Toward a metabolic theory of ecology. *Ecology*, **85**, 1771-1789

Colwell, R.K., Rahbek, C. & Gotelli, N.J. (2004) The mid-domain effect and species richness patterns: what have we learned so far? *The American Naturalist*, **163**, E1-E23

Cornwell, W.K. & Ackerly, D.D. (2009) Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecological Monographs*, **79**, 109-126

Currie, D.J. & Kerr, J.T. (2008) Tests of the mid-domain hypothesis: a review of the evidence. *Ecological Monographs*, **78**, 3-18

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guégan, J.F., Hawkins, B.A., Kaufman, D.M., Kerr, J.T., Oberdorff, T., O'Brien, E. & Turner, J.R.G. (2004) Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, **7**, 1121-1134

Diamond, J.M. (1975) Assembly of species communities. *Ecology and evolution of communities* (ed. by M.L. Cody and J.M. Diamond), pp. 342-444. Harvard University Press, Cambridge, MA.

Diniz-Filho, J.A.F., Bini, L.M. & Hawkins, B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, **12**, 53-64

Diniz-Filho, J.A.F., Bini, L.M., Rangel, T.F., Loyola, R.D., Hof, C., Nogués-Bravo, D. & Araújo, M.B. (2009) Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography*, **32**, 897-906

Donoghue, M.J. (2008) A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences USA*, **105**, 11549-11555

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609-628

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P. & Guisan, A. (in press) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking approaches. *Diversity and Distributions*

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology Evolution and Systematics*, **40**, 677–697

Engler, R. & Guisan, A. (2009) MigClim: predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, **15**, 591-601

Engler, R., Randin, C.F., Vittoz, P., Czáka, T., Beniston, M., Zimmermann, N.E. & Guisan, A. (2009) Predicting future distribution of mountain plants under climate change: does dispersal capacity matter? *Ecography*, **32**, 34-45

Ferrier, S., Drielsma, M., Manion, G. & Watson, G. (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity and Conservation*, **11**, 2309-2338

Ferrier, S. & Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393-404

Field, R., Hawkins, B.A., Cornell, H.V., Currie, D.J., Diniz-Filho, J.A.F., Guégan, J.F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M. & Turner, J.R.G. (2009) Spatial species-richness gradients across scales: a meta-analysis. *Journal of Biogeography*, **36**, 132-147

Franklin, J. (2010) *Mapping species distribution: spatial inference and prediction*. Cambridge University Press, Cambridge.

Gaston, K.J. (2000) Global patterns in biodiversity. *Nature*, **405**, 220-227

Gotelli, N.J., Anderson, M.J., Arita, H.T. *et al.* (2009) Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters*, **12**, 873–886

Gotelli, N.J., Buckley, N.J. & Wiens, J.A. (1997) Co-occurrence of Australian land birds: Diamond's assembly rules revisited. *Oikos*, **80**, 311-324

Gotelli, N.J., Graves, G.R. & Rahbek, C. (2010) Macroecological signals of species interactions in the Danish avifauna. *Proceedings of the National Academy of Sciences USA*, **107**, 5030-5035

Gotelli, N.J. & McCabe, D.J. (2002) Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. *Ecology*, **83**, 2091-2096

Graham, C.H. & Hijmans, R.J. (2006) A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography*, **15**, 578-587

Graham, C.H., Parra, J.L., Rahbek, C. & McGuire, J.A. (2009) Phylogenetic structure in tropical hummingbird communities. *Proceedings of the National Academy of Sciences USA*, **106**, 19673-19678

Graves, G.R. & Gotelli, N.J. (1993) Assembly of avian mixed-species flocks in Amazonia. *Proceedings of the National Academy of Sciences USA*, **90**, 1388-1391

Graves, G.R. & Rahbek, C. (2005) Source pool geometry and the assembly of continental avifaunas. *Proceedings of the National Academy of Sciences USA*, **102**, 7871-7876

Guisan, A. & Theurillat, J.P. (2000) Equilibrium modeling of alpine plant distribution: how far can we go? *Phytocoenologia*, **30**, 353-384

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S. & Peterson, A.T. (2007) What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecological Monographs*, **77**, 615-630

Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guégan, J.F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M., Porter, E.E. & Turner, J.R.G. (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105-3117

Huntley, B. (1991) How plants respond to climate change: migration rates, individualism and the consequences for plant communities. *Annals of Botany*, **67**, 15-22

Ingram, T. & Shurin, J.B. (2009) Trait-based assembly and phylogenetic structure in northeast Pacific rockfish assemblages. *Ecology*, **90**, 2444-2453

Kearney, M. & Porter, W. (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334-350

Kearney, M. & Porter, W.P. (2004) Mapping the fundamental niche: physiology, climate, and the distribution of a nocturnal lizard. *Ecology*, **85**, 3119-3131

Keddy, P., Nielsen, K., Weiher, E. & Lawson, R. (2002) Relative competitive performance of 63 species of terrestrial herbaceous plants. *Journal of Vegetation Science*, **13**, 5-16

Law, R. & Morton, R.D. (1993) Alternative permanent states of ecological communities. *Ecology*, **74**, 1347-1361

Leathwick, J.R. & Austin, M.P. (2001) Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology*, **82**, 2560-2573

Lehmann, A., Leathwick, J.R. & Overton, J.M. (2002) Assessing New Zealand fern diversity from spatial predictions of species assemblages. *Biodiversity and Conservation*, **11**, 2217-2238

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., Holt, R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M. & Gonzalez, A. (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, **7**, 601-613

Lomba, A., Pellissier, L., Randin, C.R., Vicente, J., Moreira, F., Honrado, J. & Guisan, A. (2010) Overcoming the rare species modelling paradox: test of a novel hierarchical framework with an Iberian endemic plant. *Biological Conservation*, **143**, 2647-2657

Loreau, M. (2000) Are communities saturated? On the relationship between alpha, beta and gamma diversity. *Ecology Letters*, **3**, 73-76

Lortie, C.J., Brooker, R.W., Choler, P., Kikvidze, Z., Michalet, R., Pugnaire, F.I. & Callaway, R.M. (2004) Rethinking plant community theory. *Oikos*, **107**, 433-438

Lyons, S.K. (2003) A quantitative assessment of the range shifts of Pleistocene mammals. *Journal of Mammalogy*, **82**, 385-402

MacArthur, R.H. (1972) *Geographical Ecology*. Harper & Row, New-York.

McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006) Rebuilding community ecology from functional traits. *Trends in Ecology and Evolution*, **21**, 178-185

Midgley, G.F., Davies, I.D., Albert, C.H., Altwegg, R., Hannah, L., Hughes, G.O., O'Halloran, L.R., Seo, C., Thorne, J.H. & Thuiller, W. (2010) BioMove − an integrated platform simulating the dynamic response of species to environmental change. *Ecography*, **33**, 612-616

Mittelbach, G.G., Schemske, D.W., Cornell, H.V. *et al.* (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315-331

Mouillot, D., Mason, N.W.H. & Wilson, J.B. (2007) Is the abundance of species determined by their functional traits? A new method with a test using plant communities. *Oecologia*, **152**, 729-737

Pausas, J.G. & Austin, M.P. (2001) Patterns of plant species richness in relation to different environments: an appraisal. *Journal of Vegetation Science*, **12**, 153-166

Pellissier, L., Brathen, K.-A., Pottier, J., Randin, C., Vittoz, P., Dubuis, A., Yoccoz, N.G., Torbjørn, A., Zimmermann, N.E. & Guisan, A. (2010) Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography*, **33**, 1004-1014

Peres-Neto, P.R., Olden, J.D. & Jackson, D.A. (2001) Environmentally constrained null models: site suitability as occupancy criterion. *Oikos*, **93**, 110-120

Pineda, E. & Lobo, J.M. (2009) Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology*, **78**, 182-190

Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349-361

Rahbek, C., Gotelli, N.J., Colwell, R.K., Entsminger, G.L., Rangel, T. & Graves, G.R. (2007) Predicting continental-scale patterns of bird species richness with spatially explicit models. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 165-174

Ricklefs, R.E. & Schluter, D. (1993) *Species diversity in ecological communities: historical and geographical perspectives*. The University of Chicago Press, Chicago.

Sargent, R.D. & Ackerly, D.D. (2008) Plant–pollinator interactions and the assembly of plant communities. *Trends in Ecology and Evolution*, **23**, 123-130

Sax, D.F., Stachowicz, J.J., Brown, J.H., Bruno, J.F., Dawson, M.N., Gaines, S.D., Grosberg, R.K., HastingS, A., Holt, R.D., Mayfield, M.M., O'Connor, M.I. & Rice, W.R. (2007) Ecological and evolutionary insights from species invasions. *Trends in Ecology and Evolution*, **22**, 465-471

Seastedt, T.R., Hobbs, R.J. & Suding, K.N. (2008) Management of novel ecosystems: are novel approaches required? *Frontiers in Ecology and the Environment*, **6**, 547-553

Shipley, B., Vile, D. & Garnier, E. (2006) From plant traits to plant communities: a statistical mechanistic approach to biodiversity. *Science*, **314**, 812

Sizling, A.L., Sizlingova, E., Storch, D., Reif, J. & Gaston, K.J. (2009) Rarity, commonness, and the contribution of individual species to species richness patterns. *The American Naturalist*, **174**, 82-93

Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115-1123

Stockwell, D.R.B. & Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148**, 1-13

Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., Alexander, J.D., Wiens, J.A. & Root, T.L. (2009) Re-shuffling of species with climate disruption: a no-analog future for California birds? *PLoS ONE*, **4**, e6825

Thuiller, W., Midgley, G.F., Rouget, M. & Cowling, R.M. (2006) Predicting patterns of plant species richness in megadiverse South Africa. *Ecography*, **29**, 733-744

Valone, T.J. & Hoffman, C.D. (2002) Effects of regional pool size on local diversity in small-scale annual plant communities. *Ecology Letters*, **5**, 477-480

Weiher, E. & Keddy, P. (1999) The scope and goals of research on assembly rules. *Ecological assembly rules: perspectives, advances, retreats*. Cambridge University Press, Cambridge.

Whittaker, R.J., Willis, K.J. & Field, R. (2001) Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, **28**, 453-470

Williams, J.W. & Jackson, S.T. (2007) Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, **5**, 475-482

Wilson, J.B. & Gitay, H. (1995) Community structure and assembly rules in a dune slack: variance in richness, guild proportionality, biomass constancy and dominance/diversity relations. *Vegetatio*, **116**, 93-106

Wilson, J.B. & Roxburgh, S.H. (1994) A demonstration of guild-based assembly rules for a plant community, and determination of intrinsic guilds. *Oikos*, **69**, 267-276

Wisz, M.S., Walther, B.A. & Rahbek, C. (2007) Using potential distributions to explore determinants of Western Palaearctic migratory songbird species richness in sub-Saharan Africa. *Journal of Biogeography*, **34**, 828-841

Young, B.F., Franke, I., Hernandez, P.A., Herzog, S.K., Paniagua, L., Tovar, C. & Valqui, T. (2009) Using spatial models to predict areas of endemism and gaps in the protection of Andean slope birds. *The Auk*, **126**, 554-565

BIOSKETCHES

Antoine Guisan is an associate professor at the University of Lausanne, Switzerland, where he leads the Spatial Ecology group (http://www.unil.ch/ecospat). His main research interests are in biogeography and what drives species distributions, with a main focus on ecological niche and habitat suitability modelling and a secondary focus on mountain environments, climate change and invasive species. This paper emerges from a new, developing interest in predicting species assemblages.

Carsten Rahbek is a professor at the University of Copenhagen and Director of the Center for Macroecology, Evolution and Climate. His main research interests include patterns of species distributions, species range sizes, species assemblages, species richness and what determines such patterns. Recent focus has been on climate change and the role of scale, formulation and practical design of null and predictive models for hypothesis testing as well as on how evolutionary and ecographical principles can be used to identify robust priorities for conservation of biodiversity.

FIGURE LEGENDS

Figure 1 Illustration of the two approaches – macroecological controls and individual species assembly – used to predict species richness (first-level predictions) and how they may be combined (second-level predictions) using the former to constrain the latter.

Figure 2 Predicted and observed plant species richness at 912 plots in open habitats in the western Swiss Alps (260 species) at a resolution of 25×25 m. (a) Estimated richness by macroecological modelling (MEM) using five variables – growing degree-days above 0° C, moisture index, global solar radiation, slope and topographic position – and the average of four modelling techniques [generalized linear modelling (GLM), generalized additive modelling (GAM), gradient boosting model (GBM) and random forest (RF)]; see Dubuis *et al*., in press); (b) estimated richness by stacked species distribution modelling (S-SDM) using the same five variables and generating species distributions for each species using the average of the same four techniques used for MEMs; (c) observed richness; (d) scatterplot of MEM versus observed richness; (e) scatterplot of S-SDM versus observed richness; (f) scatterplot of MEM versus S-SDM richness. In scatterplots (d) to (f), the dotted red line represents the 1:1 relationship, and the plain blue line represents the regression line of a standard ordinary linear squares (OLS) regression across the cloud of points. Maps were generated and provided by Anne Dubuis (Dubuis *et al*., in press).

Figure 3 Empirical and estimated levels of species richness of New World birds (3837 species) at a resolution of 1×1 degree latitude/longitude across the whole New World (all cells). (a) Estimated richness by macroecological modelling (MEM) using four variables: mean annual maximum and minimum temperature, mean annual precipitation and range in precipitation; (b) estimated richness by species distribution modelling (SDM) using the same four variables generating species distributions for each species combined into 15 models (different variable combinations) and extracted from five different Global Circulation Models

(GCMs; current climate) using seven SDM techniques; (c) observed species richness; (d) scatterplot of MEM versus observed richness; (e) scatterplot of stacked species distribution modelling (S-SDM) versus observed richness; (f) scatterplot of MEM versus S-SDM richness. In scatterplots (d) to (f), the dotted red line represents the 1:1 relationship, and the plain blue line represents the regression line across the cloud of points. Maps were generated and provided by J.A.F. Diniz-Filho, Universidade Federal de Goiás, Instituto de Ciências Biológicas, Departamento de Biologia Geral. See Diniz-Filho *et al.*, 2009) for details on data and modelling.

Figure 4 The proposed unified framework for spatial modelling of realized species assemblages (extant communities). The initial pool of species is successively filtered to account for dispersal limitation, habitat suitability and species interactions up to a value determined for the geographical analytical unit by a macroecological model (MEM). Dispersal filtering can be done through dispersion fields or interpolations to define the species source pool for each analytical unit. Habitat filtering is performed by spatial stacking of predictions of individual species distribution models (S-SDM). Biotic filtering requires ecological assembly rules (EAR) to be developed and applied. All steps need to be critically assessed and the entire framework needs to be tested with independent data. See main text for detailed explanations.

Fig. 1

DATA ----------► MODELS ----------► INTERMEDIATE PREDICTIONS ----------► FINAL PREDICTION

SPECIES RICHNESS ----► MEM ----------► **PREDICTED SPECIES RICHNESS**

• Energy / Productivity
• Climatic stability
• Habitat heterogeneity
• History / Disturbances

Macroecological Theory

MACROECOLOGICAL CONSTRAINTS
on total source pool

SPECIES COUNT

Statistical models +
use of GIS predictors
(e.g. climate)

PATTERNS OF SPECIES RICHNESS

**ASSEMBLAGE of individual
species predictions**

Filtered by dispersal and
ecological assembly rules
(EARs)

species

map cells

**presence /
absence** ---► S-SDM ----------► **PREDICTED SPECIES DISTRIBUTIONS**

Fig. 2

Fig. 3

(a) MEM predicted species richness

0 [gradient] 680 species

(b) S-SDM predicted species richness

35 [gradient] 1488 species

(c) Observed species richness

31 [gradient] 999 species

(d) MEM species richness vs Observed species richness

$y = 0.75\,z + 88$

(e) S-SDM species richness vs Observed species richness

$y = 2.01\,z - 79.26$

(f) MEM species richness vs S-SDM species richness

$y = 0.34\,z + 108$

Fig. 4

SESAM framework

SPECIES PRODUCTION (global pool)

All species N

WORLD:
2×10^5 plant species

Evolutionary history
&
Dispersal filtering (DF)

European Alps:
4500 species

STEPS

Regional Source Pool

1 Source Species Pool (SSP)

Local Source Pool

$N_1 = N - DF$

Local study area:
800 species

2 Habitat suitability models (S-SDM)

Abiotic Habitat Filtering (HF)

Abiotic Habitat Pool

$N_2 = N_1 - HF$

Alpine grasslands on acidic bedrock:
80 species

3 Macroecological constraints (MEM)

4 Ecological assembly rules (EAR)

Biotic Filtering (BF)

REALIZED ASSEMBLAGE
of analytical unit

$N_3 = N_2 - BF$

Extant community:
50 species