

Preface

Species distribution modelling has become one of the most widely used tools in ecology, conservation biology and wildlife management. While methods for species distribution modelling are continually being developed, it is fair to say that the field itself is well established. Thousands of research papers have developed and applied statistical methods to map how the occurrence or abundance of species depends on environmental and spatial predictors. These methods have been summarised in several influential reviews and books, some of which are part of the Ecology, Biodiversity and Conservation series of Cambridge University Press (Franklin 2009; Guisan et al. 2017). However, the largest body of species distribution modelling literature concerns single-species models in which the response variable is the occurrence or abundance of a focal species. Compared to single-species distribution modelling, the methodological advances in multiple-species distribution modelling have lagged behind. When applying single-species models to data on multiple species, a separate model needs to be developed and validated for each species, making it challenging to model real communities consisting of many species. This is particularly difficult with regard to rare species, which are inherently common in most ecological communities. Furthermore, species do not live in isolation from each other, and thus viewing a community as a set of species that respond individualistically to environmental variation represents a major simplification. From the perspective of assembly theory in community ecology, biotic filtering is ignored when treating each species independently.

While species distribution modelling is routinely applied in single-species studies, the reasons outlined above make it less ideal for modelling species-rich ecological communities. Instead, the most widely applied methods in community ecology are ordination-based methods. Ordination methods were developed to enable the patterns in community composition to be summarised along spatial and environmental gradients. This is done by simplifying the high-dimensional structure of

community data into few axes that explain the dominating part of the variation. While ordinations are very powerful for summarising the patterns in complex community data, they have limitations as well. Most importantly, ordination methods have been criticised for being of descriptive rather than predictive nature.

Both species distribution models and ordination methods are used to achieve the same general aim, namely to better understand the drivers controlling biodiversity across environmental gradients, space and time. Consequently, there is no reason why these two methodological fields should continue to develop independently of each other; rather, they could each learn from each other and eventually merge to combine their strengths. In recent years, statistical ecologists have taken the first steps in this direction, by developing the so-called joint-species distribution modelling (JSDM) approaches. JSDMs build more heavily on single-species distribution models than on ordinations, as they involve a single-species distribution model for each of the species comprising the community. However, they are not a mere collection of single-species models: the species are not modelled independently of each other, but jointly, as with ordination methods. The 'joint' aspect of JSDM relates to both environmental filtering and biotic filtering. The responses of the species to environmental variation (i.e. environmental filtering) are assumed to have a joint structure that can depend on e.g. species' traits or phylogenetic relationships. This is achieved by a hierarchical model structure that involves both community-level and species-level parameters. The species' responses to each other (i.e. biotic filtering) are modelled through residual association matrices that describe the co-occurrence or co-abundance patterns that are not explained by environmental filtering.

The first JSDM approaches modelled species associations separately for each pair of species, and were thus feasible only for communities with few species (the number of pairs of species – and hence model parameters – becomes otherwise too large to be estimated). To overcome this limitation, the next generation of JSDMs applied latent variable approaches, making it possible to estimate association matrices also for communities with many species. This is where joint species distribution models have approached ordination methods. Namely, the latent variable approach is used to reduce the high dimensionality of community data. In fact, it can actually be viewed as model-based ordination. Therefore, JSDMs involve both species-specific distribution models and

ordinations in their machinery, bringing these two fields closer to each other.

JSDM is currently one of the fastest developing fields in statistical ecology. While several kinds of JSDMs have already been implemented and successfully applied, the field is still in its infancy, especially compared to single-species modelling and ordination techniques. Consequently, the field of JSDM is currently experiencing much turbulence, with new approaches emerging at a fast rate and parallel developments of related approaches being simultaneously undertaken by different research groups. Some of these approaches may prove to be viable in the long run, while others may become superseded by improved approaches. While the ongoing rapid turnover of JSDM provides exciting possibilities, it also makes it difficult for their users to keep track of the pros and cons of the different approaches, and to gain an adequate understanding of their underlying assumptions and limitations. For these reasons, we considered it timely to devote an entire book to joint species distribution modelling, as this provided the possibility to present the conceptual, statistical and implementation aspects of JSDM in a much more profound and collective way than would be possible in focal research papers or software tutorials.

While several JSDM approaches and software implementations have been developed over the past decade, this book develops the argument of joint species distribution modelling from the point of view of one specific framework, namely Hierarchical Modelling of Species Communities (HMSC). However, as many of the existing JSDM approaches are closely related and can be considered as different branches of the same tree, we hope that this book will help deepen the readers understanding of the fundamentals of JSDM in general. In addition to presenting the conceptual, theoretical and statistical foundations of JSDM, this book also provides 'hands on' examples of how JSDM can be applied in practice. To this end, we build heavily on the R-package *Hmsc*; its use is demonstrated through R-scripts, and it has also been used to generate the majority of the figures/illustrations. Furthermore, we note that writing this book motivated us to implement some new features and extensions of HMSC, so some of the material here has not been published yet in research papers. We hope that the many R-scripts presented in this book (and the related online resources at www.helsinki.fi/en/researchgroups/statistical-ecology/hmsc) will provide a convenient starting point for a reader who wishes to apply JSDM for his or her own purposes.