Ecological Modelling 323 (2016) 87–95

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Simple or complex: Relative impact of data availability and model purpose on the choice of model types for population viability analyses

Viktoriia Radchuk[a,*], Steffen Oppel[b], Jürgen Groeneveld[c], Volker Grimm[c,d,e], Nicolas Schtickzelle[a]

[a] Earth and Life Institute, Université catholique de Louvain, Croix du Sud 4, L7.07.04, 1348 Louvain-la-Neuve, Belgium
[b] RSPB Centre for Conservation Science, Royal Society for the Protection of Birds, The Lodge, Sandy, Bedfordshire SG19 2DL, United Kingdom
[c] Department of Ecological Modelling, Helmholtz Centre for Environmental Research – UFZ, Permoserstr. 15, 04318 Leipzig, Germany
[d] Institute for Biochemistry and Biology, University of Potsdam, Maulbeerallee 2, 14469 Potsdam, Germany
[e] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

ARTICLE INFO

Article history:
Received 21 July 2015
Received in revised form 30 October 2015
Accepted 6 November 2015
Available online 9 January 2016

Keywords:
Model complexity
Individual-based model
Population-based model
Matrix model
Structured population model
Stage-based model

ABSTRACT

Population viability analysis (PVA) models are used to estimate population extinction risk under different scenarios. Both simple and complex PVA models are developed and have their specific pros and cons; the question therefore arises whether we always use the most appropriate model type. Generally, the specific purpose of a model and the availability of data are listed as determining the choice of model type, but this has not been formally tested yet. We quantified the relative importance of model purpose and nine metrics of data availability and resolution for the choice of a PVA model type, while controlling for effects of the different life histories of the modelled species. We evaluated 37 model pairs: each consisting of a generally simpler, population-based model (PBM) and a more complex, individual-based model (IBM) developed for the same species. The choice of model type was primarily affected by the availability and resolution of demographic, dispersal and spatial data. Low-resolution data resulted in the development of less complex models. Model purpose did not affect the choice of the model type. We confirm the general assumption that poor data availability is the main reason for the wide use of simpler models, which may have limited predictive power for population responses to changing environmental conditions. Conservation biology is a crisis discipline where researchers learned to work with the data at hand. However, for threatened and poorly-known species, there is no short-cut when developing either a PBM or an IBM: investments to collect appropriately detailed data are required to ensure PVA models can assess extinction risk under complex environmental conditions.

1. Introduction

Population viability analysis (PVA) is an important tool used in conservation biology to assess the viability of populations and rank alternative management scenarios (Beissinger and Westphal, 1998; Reed et al., 2002). Population viability analysis is a generic label affixed to a variety of modelling techniques differing in their complexity, approach, and type of data used (Morris and Doak, 2002) including: patch-occupancy models (POMs, Hanski, 1997, 1994), matrix projection models (Caswell, 2002), structured population models (SPMs, Akçakaya and Sjögren-Gulve, 2000; Schtickzelle and Baguette, 2009), and individual-based models (IBMs, DeAngelis and Gross, 1992; Grimm and Railsback, 2005). Individual-based models themselves cover a wide spectrum of models, ranging from simpler IBMs driven by demographic rates (e.g. Grimm et al., 2003) to more complex models driven by the adaptive behaviour of individuals (Stillman et al., 2015). Generally, the type of PVA model used is believed to be determined by three main factors: the life history of the species in consideration, the specific model purpose, and constraints of data availability and resolution (Akçakaya and Sjögren-Gulve, 2000; Boyce, 1992; DeAngelis and Rose, 1992; Grimm and Railsback, 2005).

The choice of a given model type used for a PVA first depends on the life history of the species in consideration (Boyce, 1992;

* Corresponding author. Present address: Evolutionary Genetics Department, Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Straße 17, 10315 Berlin, Germany. Tel.: +49 15756038711.
E-mail addresses: radchuk@izw-berlin.de (V. Radchuk), steffen.oppel@gmail.com (S. Oppel), juergen.groeneveld@ufz.de (J. Groeneveld), volker.grimm@ufz.de (V. Grimm), nicolas.schtickzelle@uclouvain.be (N. Schtickzelle).

http://dx.doi.org/10.1016/j.ecolmodel.2015.11.022
0304-3800/

Vucetich and Creel, 1999; Wiegand et al., 1998). As an example, population dynamics will not be modelled the same way for an insect species with an r-selected strategy and abundant populations, and for a rare large mammal, characterized by a k-selected strategy and complex social or territorial behaviour. The latter species is believed to benefit more from incorporation of behaviour in the model, leading to the development of more complex models, such as IBMs (Grimm and Railsback, 2005). Second, the specific purpose of the model guides the choice of the model type. For example, models depicting generic phenomena and those predicting species responses to new environmental conditions are likely to strongly differ in model structure (e.g. Grimm and Railsback, 2005). Last but not least, the availability and resolution of data necessary for modelling certain processes can often impose limitations on the model type that can be developed and parameterized (Boyce, 1992; DeAngelis and Rose, 1992; Reed et al., 2002). Even if model aim or species life history suggest the development of a more complex model, the lack of data or their coarse resolution can force the adoption of less complex model types.

Although these three factors are widely recognized to affect model type choice, in particular its structure and complexity (Akçakaya and Sjögren-Gulve, 2000; Boyce, 1992; Grimm and Railsback, 2005), very few studies explicitly justified why they chose a specific model type over others (Pe'er et al., 2013). Thus, there is no clear understanding of how the species' life history, data at hand (availability and resolution), and model aim can affect the choice of a specific model type for PVA in practice. Given that data scarcity is probably a major limitation in assessing population viability, it is striking that no systematic approach has been undertaken to study the constraints posed by data restrictions on model type choice (Brook et al., 2002; Fieberg and Ellner, 2000; Reed et al., 2002; Thomas et al., 2002). Such an understanding would aid to expedite the decision making in the era of the sixth species mass extinction (Barnosky et al., 2011) by providing a more streamlined and efficient procedure of PVA design and implementation (Pe'er et al., 2013).

In this study, we quantified the relative importance of specific model purpose and data availability and resolution (hereafter abbreviated to "data availability") for the choice of PVA model type. We used 37 pairs of PVA models identified from the literature, each consisting of an individual-based model (IBM) and a population-based model (PBM) developed for the same species. This 'paired' design effectively controls for the effect of species life history and therefore allows us to objectively assess the impact of data availability and model purpose on the choice of model type. We first extracted from each PVA study the information on the model purpose and a series of factors describing data availability about the species and its environment. We then elicited identical information via a questionnaire sent to the authors of the 74 PVA models to verify that the published information indeed reflected what data were available to modellers at the time of model development. We then quantified the importance of data availability and model purpose on the choice of model type.

2. Methods

2.1. Selecting published PVA models in a 'paired' design

We searched the ISI "Web of Knowledge" for papers reporting PVA models constructed for the same animal species but differing in their complexity: IBMs vs. PBMs (including structured population models, projection matrix models and patch occupancy models). Throughout this manuscript we consider IBMs to be generally more complex than PBMs, although we acknowledge that both PBMs and IBMs can range in their complexity. We used the following search keys: (1) for PVA: "(population viability analysis) OR (vulnerability analysis) OR (population dynamics model)"; combined with either (2) for IBMs: "(individual based model) OR (behavioural model) OR (mechanistic model)"; or (3) for PBMs: "(structured population model) OR (matrix model) OR (Leslie model)". The literature search (conducted on 29.07.2014) yielded 542 papers, all classified as either IBM or PBM. From these papers, we identified 37 pairs of PVA models (Table 1), i.e. when at least one IBM and one PBM were developed for the same species. When more than one PVA of a certain type was developed for the same species (17 out of 37 species, 46%), we retained only the model of each type that was published first. Such focus on the earliest studies is believed to better reflect the general lack of data, which is characteristic for species for which a PVA is typically performed (Sitas et al., 2009). However, to assess how this decision to use only the papers published first affected our conclusions, we also ran the analyses with all available PVA studies for the 37 species.

2.2. Assessing data availability

From each PVA model, we extracted a series of nine factors to describe the availability of data covering all main model components that we identify as relevant to PVA model structure (Fig. 1 and Table 2), grouped into factors describing the species and factors describing its environment (Beissinger and Westphal, 1998). We distinguish three factors for species data: *demography* (i.e. vital rates, specifics of sociality or territoriality if relevant), *genetics* (e.g. inbreeding depression, heterozygosity, etc.), and *dispersal* (e.g. immigration, emigration, movement rules); three factors for environment data: *climate* (covering both the short-term weather component, and the longer-term climatic component), *biotic interactions* (e.g. the influence of competition, predation or parasitism on survival and/or reproduction), and *landscape*, further subdivided into *space representation* (i.e. location and configuration of territories/populations) and *habitat heterogeneity* (i.e. habitat quality, here defined as any measure of structural and/or functional habitat heterogeneity that was considered in the study); and two factors about the impact of the environment on species demography: *effect of climate on vital rates* (i.e. quantification of how weather factors affect survival and/or reproduction), and *effect of habitat on vital rates* (i.e. quantification of how habitat affects survival and/or reproduction). These nine factors characterize both the availability and, whenever possible, the resolution of relevant data (see Table 2 for details).

Information about data availability extracted from published studies does not necessarily reflect the true data availability at the time of model development. To obtain a more accurate understanding of what data were available at the time of model development, we contacted the authors of the 74 case studies by e-mail and asked them to fill in a questionnaire (see Supplementary material). Authors were asked to give their assessment about the availability of data for each of the nine factors (listed in Table 2) at the time when the model was developed. Additionally, we asked authors to assess how the choice of the model type they used may have been affected by their expertise and familiarity with a certain type of PVA. We obtained information on data availability for 38 models (a 51% response rate). Because not all of these models represented paired studies, we acquired author-based estimates for data availability and resolution for 12 pairs of published PVA studies.

We then assessed whether we could glean data availability from the published model structure. To that end, we used Kendall's rank correlation (a modified version for data with ties based on normal approximation with continuity correction) between the values for each of the nine factors reflecting data availability as extracted directly from published studies and those obtained from questionnaire responses. We considered Kendall's rank correlation more appropriate than Spearman rank correlation because most of our