Biological Conservation 227 (2018) 167–180

Contents lists available at ScienceDirect

Biological Conservation

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

From general research questions to specific answers: Underspecificity as a source of uncertainty in biological conservation

Susana Suárez-Seoane[a,*], Jose Manuel Álvarez-Martínez[b], Carlos Palacín[c], Juan Carlos Alonso[c]

[a] Área de Ecología, Departamento de Biodiversidad y Gestión Ambiental, Facultad de Ciencias Biológicas y Ambientales, Universidad de León, Campus de Vegazana s/n, 24071 León, Spain

[b] Environmental Hydraulics Institute "IH Cantabria", Universidad de Cantabria, Parque Científico y Tecnológico de Cantabria, C/Isabel Torres, 15, 39011 Santander, Spain

[c] Museo Nacional de Ciencias Naturales (CSIC), José Gutiérrez Abascal 2, 28006 Madrid, Spain

ARTICLE INFO

Keywords:
Breeding success
Breeding site selection
Great Bustard
Species distribution modelling

ABSTRACT

Species distribution modelling may support ecologists in conservation decision-making. However, the applicability of management recommendations depends on the uncertainty associated to the modelling process. A key source of uncertainty is the underspecificity of the research question. Modelling specific questions is straightforward since they drive clearly the methodological choices about input data and model building. Nevertheless, when the research questions remain underspecific, modellers must choose among a wide spectrum of choices, with each decision sequence driving to a different outcome that explain partially the target question. We show how the underspecificity associated to a general research question about Great Bustard breeding success at geographic scale drives to multiple decision choices, leads to a variety of model outcomes and hampers the identification of specific conservation actions. We ran generalised linear models using multi-model inference on a set of databases built according to specific sequences of methodological choices. Then, we evaluated variations in model performance, complexity (parsimony) and nature of predictors, as well as averaged model predictions and spatial congruence among model outputs. Deviance and parsimony varied widely (11.46% to 83.33% and 7 to 18, respectively), as did model averaged mean predictions in occupied areas, contributing predictors and spatial congruence among outputs ($r_{Pearson} = 0.44 \pm 0.23$ for models calibrated in occupied areas; 0.48 ± 0.06 for models calibrated in potential/accessible areas). We recommend to carefully fix research questions and associated methodological options through collaborative working frameworks to conceptualize modelling approaches and, thus, to mitigate problems arising from underspecificity and other forms of uncertainty in conservation applications.

1. Introduction

Species distribution models (SDM) are correlative approaches that allow for the estimation of species' ecological requirements at multiple scales in the framework of the ecological niche theory (Austin, 2002, 2007; Kearney, 2006; Peterson, 2006; Jiménez-Valverde et al., 2008; Warren, 2012). When SDM are based on static limiting conditions (scenopoetic variables) that may control species' ecophysiology and drive its occurrence at large scale, these models can be interpreted as describing the Grinnellian niche. When SDM are based on biotic interactions and resource consumer dynamics (bionomic variables) that determine species abundance and reproductive rates operating at more detailed scales, they can be interpreted as describing the Eltonian niche (Soberon, 2007; Tingley et al., 2009; Peterson et al., 2012; Alvarez-Martínez et al., 2015).

SDM are, therefore, potentially useful tools for ecologists and land managers dealing with processes of decision-making addressed to biological conservation actions (Suárez-Seoane et al., 2002; Franklin, 2010a). However, their applicability decreases if they fail to describe the ecological system under study across different scenarios (Elith et al., 2002; Dawson et al., 2011; Guisan et al., 2013). This problem is inherently linked to the uncertain nature of ecological processes, but it might also depend on the methods and decision-making applied by researchers. In this sense, scientists must cope with the uncertainties derived from incompatibilities between ecological background, input data and statistical methods (Austin, 2007) that can permeate model results and management recommendations. Authors such as Harwood and Stokes (2003) and Ascough II et al. (2008) have argued that the

* Corresponding author.
E-mail address: s.seoane@unileon.es (S. Suárez-Seoane).

https://doi.org/10.1016/j.biocon.2018.09.005
Received 23 April 2018; Received in revised form 1 August 2018; Accepted 5 September 2018

failure of ecologists to evaluate accurately the uncertainties associated with their advice (e.g., integrated assessment models, optimization algorithms and multi-criteria decision analysis tools) diminishes their influence in decision-making. Therefore, it is compulsory to understand the practical consequences of these uncertainties, which may be exacerbated when the system under study is complex and changing (Álvarez-Martínez et al., 2010; Polasky et al., 2011).

Random uncertainty refers to the inherent variability of a given system and is typically named as variability, irreducible uncertainty, inherent uncertainty or stochastic uncertainty. Besides this uncertainty due to chance, three main human dependant sources of uncertainty can affect soundly SDM predictions: epistemic, linguistic and human decision-making (Regan et al., 2002; Elith et al., 2002; Kujala et al., 2013). Epistemic uncertainty is due to a lack of knowledge about the state of the system that is being modelled. It is associated to technical decisions, as well as systematic and measurement errors in data collection, such as those related to location (e.g., spatial accuracy and representativeness of the species' range requirements; Thuiller et al., 2004; Randin et al., 2006; Menke et al., 2009), shape (e.g., areal unit for which data are collected, which may lead to the modifiable areal unit problem MAUP; Openshaw and Taylor, 1981; Wong, 2004) or nature of input data (e.g., species data can be continuous, counts or binary and predictor values can be collected from different sources; Synes and Osborne, 2011). Linguistic uncertainty originates when language used for stating research questions or defining terms is underspecific, vague, ambiguous or context-dependent. Underspecificity arises when the research question is excessively general and portraits a lack of clarity. For example, the question “which environmental factors determine temporal changes in biological fitness?” is underspecific because it does not clarify whether we are interested in the assessment of factors behind inter or interannual variability in fitness, neither if we are looking for the factors that determine the mean or the variation of fitness values for a particular time span, the biotic or abiotic factors driving the temporal changes and so on. On the opposite, “is annual maximum temperature behind interannual variation of mean breeding success?” or “does spring cumulative rainfall determine weekly variations in breeding success during the reproductive season?” are both specific research questions that can be nested within the former underspecific question. Vagueness refers to the possibility of borderline cases due to categorical classifications of data (Regan et al., 2002). It arises when terms are defined using arbitrarily sharp boundaries (e.g., “high fitness”, “optimal habitat” or “viable population”) that may alter drastically the output (Bull et al., 2016). Ambiguity originates because some words may have more than one meaning (e.g., “fitness” can be defined either as the variation in reproductive success or as the genotype's rate of increase in future generations; Michod, 2000). Context-dependence is problematic when the framework of the question at hand has not been correctly specified, which may affect its interpretation. The lack of a coherent understanding of the context under which to answer ecological questions may be associated with large variability in the research predictions (Lajoie and Vellend, 2015). For instance, different answers should be expected if we explore the fitness variations in Great Bustard populations in Spain or Russia due to different biotic (e.g. interspecific relationships) and abiotic (e.g. climate constraints) contexts. Finally, uncertainty derived from human decision-making may arises from human beliefs, values, preferences, choices and actions, as it is the case of scale preferences (e.g. spatial resolution) or subjective choice of predictors to be included in a model. However, the best recognized type is subjective judgement, which is frequently associated to scenario planning or interpretation of model results. This problem is especially relevant when data are scarce and error prone. The standard way of dealing with it is to assign a level of certainty to the target event in the form of a subjective probability (Regan et al., 2002). For instance, we might assign a probability of 0.8 to the event “the mean breeding success of Great Bustard in Spain during the last five years was 15%”. There are different techniques within decision science that can help to address this area of uncertainty. Among them, structured decision making (Gregory and Keeney, 2002) and adaptive management (Runge, 2011) are particularly relevant for applying formal decision-analysis tools in natural resource management decisions.

In this context, the underspecificity associated to the statement of the research question emerges as a critical and under evaluated issue that can be particularly relevant in conservation applications. Despite wide assessment of epistemic uncertainty in SDM approaches (e.g., Elith et al., 2002; Thuiller, 2003; Pearson et al., 2006; Convertino et al., 2012), the role of underspecificity still remains underexplored, as it occurs with other forms of linguistic and human decision-making uncertainty. In fact, according to Kujala et al. (2013), from the set of papers considering uncertainty that were published during the period 1945–2011 with an ecological scope, only 1.5% mentioned specifically the underspecificity, as did the 1% of the papers with a conservation scope. Therefore, underspecificity should be deeper explored, as it may have serious implications in model interpretation and subsequent applicability. The specific definition of the research question is crucial since it frames the problem and drives the methodological choices to be done, structuring the modelling approach and the nature of the requested data (Kuhnert et al., 2010). Modelling specific questions, with a low level of underspecificity, is straightforward since they drive clearly the methodological choices to be done through the development of the modelling approach. However, research questions might be underspecific due to a lack of theoretical knowledge or because scientists do not consider explicitly the complexity and dynamics of the ecological systems under evaluation. General questions, with a high level of underspecificity, are associated to broader environmental responses and result much more difficult to capture in a single model. This is problematic since scientists are required to make specific decisions about input data and methods, which leaves open a wide spectrum of possibilities and, therefore, underspecificity issues. Each sequence of decisions will determine the nature of the outputs, the model performance and the chance of error given the different types of data and approaches, as well as the different levels of risk aversion (Barry and Elith, 2006). Practitioners must be aware of the implications of their decisions since each possible model outcome would explain only partially the target question, with a wide array of different outputs.

In this work, we aim to show how the underspecificity associated to a general research question drives to multiple modelling decisions, leads to a variety of model outcomes and, therefore, hampers the identification of specific management choices. We formulate an underspecific research question about Great Bustard (*Otis tarda*) breeding success in Spain during the last two decades (period 1987–2010) and evaluate differences in explained deviance, averaged mean predicted values across distributional areas, model complexity, nature of predictors (scenopoetic and bionomic) and spatial prediction patterns obtained when applying different sequences of methodological choices (with different ecological meaning) in a SDM framework. The evaluation of this ecological trait remains a challenge because of complex trade-offs between individual life traits of Great Bustards (quality, age or experience; Lescroël et al., 2009), social relationships (e.g., reproductive skew in social species; Johnstone, 2000; Ryder et al., 2009) and environmental constraints of distribution (widely changing across space and time; Barbraud and Weimerskirch, 2005). We intend to draw attention to the risks of generalising the outcomes obtained by applying specific methodological choices when modelling underspecific questions, highlighting the importance of carefully specifying the ecological question that one aims to disentangle.

2. Methods

2.1. Data on Great Bustard breeding success in Spain

Great Bustard is a globally threatened lekking bird species with a population severely fragmented throughout the Paleartic (BirdLife