Preface

The presence or absence of a species from a collection of sampling units is a basic concept that is widely used in wildlife and ecological studies. When we first became involved in this topic in the late 1990's–early 2000's, we were more familiar with problems of estimating abundance, birth rates, survival probabilities and related demographic parameters, primarily from capture–recapture data, and did not appreciate the generality of the concept of 'occupancy'. At that time, some of us were involved with the newly formed U.S. Geological Survey's Amphibian Research and Monitoring Initiative (ARMI), advising on general study design and the types of data that should be collected as part of broad-scale monitoring programs. It quickly became apparent that it was logistically impossible to estimate changes in absolute abundance across large areas over time, so one suggestion was to simply measure the presence or absence of the species at a number of ponds or wetlands, which was soon known within ARMI as the 'proportion of area occupied'. With our background in capture–recapture methods, where estimation of capture probability is the key concept for making robust inference about abundance, survival, and other parameters, questions of whether it would be reasonable to always detect the target species within an occupied sampling unit were quick to surface. The response from the biologists was a resounding "No!" Our experience made us realize that inferences about 'occupancy' would be much improved if the collected data, and analytic methods used, accounted for the imperfect detection of the species. To us, the use of repeated 'presence/absence' surveys at each sampling unit would obviously supply the type of information required to estimate detection probabilities, but at the time of our initial collaborations with ARMI, we did not have a clear idea on the best approach to efficiently analyze the resulting data, particularly if unequal numbers of surveys were conducted at the different sampling units. Our thoughts on the topic crystallized in 2000–2001, sometimes aided by a fine pint or two of Guinness, which resulted in our first paper on the general topic in 2002 (MacKenzie et al., 2002).

Soon after, we realized that the general concept of occupancy is used much more widely than simply in monitoring as a surrogate for abundance. Despite widespread acknowledgment that in many applications a species may be detected only imperfectly, we were surprised at the lack of development of appropriate analytic techniques that explicitly accounted for detection probabilities, finding only a handful of previously published methods, with all of them being developed in a monitoring context. This led us to quickly extend our original method in an attempt to develop practical approaches for analyzing occupancy data in different situations (e.g., MacKenzie et al., 2003, 2004a; Royle and Nichols, 2003; Dorazio and Royle, 2005; MacKenzie and Royle, 2005). We recognized a substantial gap in the literature that was relevant to a wide range of applications, wherever reliable estimates of occupancy or changes in occupancy were required about species that are imperfectly detected. The simultaneous development of similar methodologies by independent groups (e.g., Tyre et al., 2003; Stauffer et al., 2004), and research into the effect of imperfectly detecting the species in different contexts, e.g., habitat modeling (Tyre et al., 2003; Gu and Swihart, 2004) and metapopulation incidence functions (Moilanen, 2002), suggested that perhaps the time was right for a more comprehensive treatment of the issues associated with estimating occupancy and related parameters in wildlife and ecological studies. The first edition of this book was our first attempt at doing so.

In the preface of the first edition, we noted that part of our motivation for developing the original book was to provide an early synthesis of the topic, in the hope that it would promote use of these methods and rapid methodological development in the area of occupancy estimation and modeling. This was based on what we had observed in other areas of ecological statistics, where a synthetic treatment of a topic had been a catalyst for both rapid assimilation of methods into practice, and rapid development of extensions and new methods, often many years after the methods were originally published during which time their usage tended to languish. Such was the case for band recovery models (Brownie et al., 1978, 1985), closed capture–recapture models (Otis et al., 1978; White et al., 1982), distance sampling (Burnham et al., 1980; Buckland et al., 1993, 2001), and open capture–recapture models (Burnham et al., 1987; Pollock et al., 1990; Lebreton et al., 1992). Since the publication of the book's first edition in 2006, occupancy models have certainly become very widely used, on a diverse set of taxa, around the world. There have been novel applications of the initial methods and numerous extensions, both large and small. While correlation is not causation, we would like to think the first edition played some role in the widespread uptake of the methods.

In 2012 we began consideration of a second edition, which seemed warranted due to the amount of new material that had been developed in a few short years, both in terms of new methods and increased understanding around the application of the methods. And thus began a journey that is finally coming to completion five years later. It has been a rewarding, and tiring, project that has allowed us to learn, and relearn, about many facets of modeling patterns and dynamics of species occurrence. While we have endeavored to provide details of many of the interesting extensions and notable applications of 'occupancy modeling' of which we are aware, there are undoubtedly some works that are deserving of publication in this book that we have overlooked, not been aware of, or simply forgot about. The methods have become so widespread that it would be a Herculean effort to stay abreast of all the occupancy modeling applications and extensions. At some point we had to draw the line on including further material to finish the book. We apologize to the reader if they have been party to relevant research that we have not included.

In the second edition of this book you will find a synthesis of most of the current literature on estimating occupancy-type metrics while explicitly accounting for detection probability. We declare that one cannot make reliable inferences about such metrics when issues related to the detectability of the species are effectively ignored, and we are also of the view that detection probabilities are likely to change over times, locations and species. Therefore, appropriate field and analytic methods should be employed to account for imperfect detection at every opportunity. We have divided this book into five parts. Part I provides readers with important introductory material containing our views of modeling and the conduct of management and/or science, and of the role models play in making inference about populations; historical and potential applications of species occurrence concepts in ecology; and an introduction to some of the fundamental statistical methods used throughout this book. We strongly advise readers not to skip these introductory chapters as they provide the necessary background, upon which we build in the latter chapters. In Part II we focus on methods for examining patterns of species occurrence using, what we refer to as, single-season, or static, occupancy models. This could be in the more common case where occupancy is defined as a binary variable (e.g., species presence/absence), or where occupancy is defined with more than two possible states. We detail potential modeling approaches that can incorporate a wide-range of issues such as imperfect detection, covariates, unequal sampling effort, spatial correlation and heterogeneity. Example of the methods are provided throughout. In Part III we switch focus to investigation of species occurrence dynamics, using extensions of the models discussed in Part II to the multi-season case, where data has been collected over longer time periods. Part IV is devoted to the topic of study design, for both the single- and multi-season cases. We have an overriding adherence to the GIGO principle – Garbage In, Garbage Out. Irrespective of model complexity, if the underlying data is of insufficient quality then any inferences based on those models may be unreliable. We view the topics covered in Part IV as the most important ones covered in this book for the practical implementation of the modeling methods discussed. All inferences are conditional upon the collected data, therefore the data is at the foundation of everything we do, and sound study design is of critical importance. Finally, in Part V we detail more complex methods that build upon the modeling techniques discussed in Parts II and III, turning our attention to integrated modeling of species and habitat dynamics, multi-species and community-level applications. We end the book with some concluding comments.

Our intended audience for this book is those at the 'coal face' of wildlife and ecological research, who may not necessarily have a strong background in statistics. We have thus taken great pains to use as little statistical jargon as possible, and provide detailed descriptions of how the models are developed, applied, and interpreted throughout, together with practical examples to illustrate the methods. This is, however, a technical topic, and in places some readers may struggle to understand the techniques exactly. Naturally, material gets more complex as the book progresses, and at some stage, many readers will benefit from reading the section on vectors and matrices in the Appendix. However, we would hope that readers will at least come away with an understanding of the general concepts and what the modeling is attempting to achieve. We believe this book should be useful to biologists internationally working in government agencies, private research institutes, and universities. It may also be a useful supplemental text for graduate courses on sampling wildlife populations or wildlife biometry. Further it would be an excellent choice, and has been used, for the basis of short courses on this specific topic.

We recognize that user-friendly software is necessary if the methods discussed in this book are to be widely adopted. As such, we have developed Windows-based software specifically for their application: Program PRESENCE, which is public domain software freely available for download from https://www.mbr-pwrc.usgs.gov/software/presence.shtml. Many techniques have also been incorporated into Program MARK (http://www.cnr.colostate.edu/~gwhite/mark/mark.htm) which was originally developed for the application of capture–recapture models to data collected on marked individuals. Since the publication of the first edition we have noticed that the statistical programming environment R, has become much more mainstream, particularly at the graduate level. There are now a number of R packages available that implemented many of the methods described in this book. Copies of the files associated with many examples we provide in the book, are available for free download from DIM's company's website http://www.proteus.co.nz.

Finally, as we stressed in the preface to the first edition, we still do not view the contents of this book as the final word on occupancy estimation and modeling, but an overview of the available methods as of 2017. Without doubt there will be further extensions and novel applications of the methods from both ourselves, and many others. Such extensions will only add to the available toolbox of appropriate analytic methods. We hope that readers will find the methods detailed in the following pages as thought-provoking and useful as we have found them to be during their development over the last 15 years.

USGS Disclaimer: Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. USGS employees authored Chapters 7 and 15.