Austral ECOLOGY A Journal of ecology in the Southern Hemisphere

Austral Ecology (2018) **43**, 852–860

Why understanding the pioneering and continuing contributions of BIOCLIM to species distribution modelling is important

TREVOR H. BOOTH*
CSIRO Land and Water, GPO Box 1700, Canberra, ACT 2601, Australia (Email: Trevor.Booth@csiro.au)

Abstract Modern species distribution modelling (SDM) began in 1984 with the launch of the BIOCLIM program in Australia. With more than 900 papers mentioning 'species distribution model' published around the world to date, SDM is one of the most active areas of global ecology. Three books published in 2009, 2011 and 2017 have reviewed SDM, and the closely related areas of ecological niche modelling and habitat suitability modelling. All three books provide excellent introductions to these topics, but give very little information on the role that BIOCLIM played in laying the foundation for these research areas. Understanding the history of BIOCLIM is vital because it was the first package to implement the basic SDM process in an easy-to-use integrated system. It provided what are still the most commonly used set of 19 bioclimatic variables and contributed to the development of the interpolation routines used to prepare the most commonly used source of bioclimatic data (WorldClim). Early BIOCLIM studies investigated important issues such as ecological niche, invasion risk, conservation planning and impacts of climate change. Although all three books acknowledge that the BIOCLIM package was important in early SDM research, they all deal with the pioneering work very briefly and omit important details which are described here. Virtually all current SDM studies owe something to the pioneering BIOCLIM work, but this is rarely acknowledged.

Key words: biological conservation, climate change, ecological niche modelling, habitat suitability modelling, species distribution model.

INTRODUCTION

Species distribution modelling (SDM) packages take in geocoded data describing species distributions, relate these to environmental space and map likely species distributions under past, current or possible future conditions in geographic space (see Fig. 1.1 in Franklin (2009) for an illustration of this generic correlative process). In the case of BIOCLIM, interpolation relationships provide monthly mean estimates for maximum temperature, minimum temperature and precipitation for each location for the species of interest. Values are calculated for bioclimatic indices, such as mean temperature of the warmest quarter and mean precipitation of the driest quarter. The characteristics of the species distribution are described in terms of the range (or percentile range) of these variables. Locations in a regular grid are then checked and mapped to show the extent to which they satisfy this description (i.e. core, marginal or fail). Nix (1986) and Busby (1991a,b) described early versions of BIOCLIM. Xu and Hutchinson (2011) have described the current version and the User's Guide is downloadable from the internet.

Species distribution modelling along with the closely related areas of ecological niche modelling (ENM) and habitat suitability modelling (HSM) have enjoyed an enormous rise in usage since the BIOCLIM package became available in Australia in 1984. Booth *et al.* (2014) reviewed early applications of BIOCLIM and cited 20 Australian studies from the years 1984–1991. A search in the Web of Science for publications to date mentioning "species distribution model" yielded more than 900 references from around the world, with 186 being published in 2017 alone.

Three books have reviewed SDM/ENM/HSM studies, hereafter all referred to simply as SDM studies (Franklin 2009; Peterson *et al.* 2011; Guisan *et al.* 2017). The first two books have been well-received and recommended by reviewers (see, for example, Friedman 2011; Soberon 2011; Pearce 2012; Warren 2012). The highly cited review of Guisan and Thuiller (2005) (>2500 citations in Web of Science to date) suggests that the Guisan *et al.* (2017) book will be similarly well-received. However, all these books and the Guisan and Thuiller (2005) review deal with the pioneering SDM studies, in particular BIOCLIM, only very briefly.

*Corresponding author.
Accepted for publication May 2018.

doi:10.1111/aec.12628

For example, the main description of BIOCLIM in Franklin (2009) is on pages 182–183, with some brief comments elsewhere including on page 78 in relation to BIOCLIM's environmental variables. The Peterson *et al.* (2011) book describes BIOCLIM mainly on pages 103–105, again with some brief comments on other pages. The main BIOCLIM description in the Guisan *et al.* (2017) book is on pages 156–159. In total the equivalent of only about ten complete pages of the combined 1096 pages of the three books are used to describe BIOCLIM and its variables.

The purpose of this paper is to examine how the three books describe BIOCLIM and to clarify some issues with a view to helping current researchers consider issues, such as the 'equilibrium assumption' and variable selection. It is not suggested that much more lengthy descriptions of BIOCLIM would have been desirable in the three books. However, some key points about the early work, of which all current SDM users should be aware, could have been more clearly made. These are described in the following paragraph.

Modern SDM studies began with BIOCLIM (the first SDM package) which became available in January 1984. Improved thin-plate spline climatic interpolation methods, later used to create the WorldClim database (Hijmans *et al.* 2005), were a key part of the BIOCLIM developments. The first version of BIOCLIM used 12 climatic variables, later increased to 19 in 1996 and 35 in 1999. SDM research areas for which Guisan and Thuiller (2005) provide relatively recent examples were all first explored using BIOCLIM in the period 1984–1991 (see Booth *et al.* 2014 for details). Although more sophisticated SDM methods such as Maxent (Phillips *et al.* 2006) have been developed, BIOCLIM continues to be used particularly in training and for baseline comparisons. The 19 BIOCLIM variables available from the WorldClim database (Hijmans *et al.* 2005) are still used in many current SDM analyses. To date several hundred publications have used the BIOCLIM SDM package or its associated variables.

The following sections examine the extent to which the three books do or do not recognize these key points. Where appropriate, the consequences of omissions or misleading information for current SDM users are considered. Some of the key features of BIOCLIM are outlined in Table 1.

Table 1. Key features of BIOCLIM – the first SDM package

- Reliable climatic interpolation – thin-plate splines (see Hutchinson 1991, 1995, 2013)
- Interpolated estimates for actual locations – not climate data from nearest grid point
- Ease of use – requires only geocoded occurrence data
- Continental-scale coverage – Australia-wide in 1984 (later extended to other regions e.g. Africa, see Xu & Hutchinson 2011)
- High resolution – in comparison to preceding studies e.g. Box (1981)
- Bioclimatic variables – wide use of 19 BIOCLIM variable set e.g. WorldClim (Hijmans *et al.* 2005)
- Only requires presence data – not absence or pseudo-absence data as needed by MaxEnt (Phillips *et al.* 2006)
- Transparency of method – explicit, quantitative and easily understood
- Clear links to ecological theory – Evelyn Hutchinson's (1957) niche concepts (see Booth *et al.* 1988)
- Integrated mapping – with links to the MAPROJ package Hutchinson (1981)
- Error checking – cumulative plots help to identify anomalous outliers"

KEY POINTS

Modern SDM studies began with BIOCLIM (the first SDM package) which became available in January 1984

BIOCLIM was developed by Nix (concept), Busby (coding with input also from Ms J.P. McMahon) and Hutchinson (interpolation) (Nix 1986; Busby 1991a, b; Hutchinson 1995). All three books mention that BIOCLIM was the first, or one of the first, SDM packages. The actual comments are the "original climate modelling system" (Franklin 2009, p. 5), "one of the earliest methods" (Peterson *et al.* 2011, p. 103) and "the first model to predict the geographic distribution of a species" (Guisan *et al.* 2017, p. 156). However, none of the books clearly state when the package became available. Guisan *et al.* (2017, p. 156) wrote that BIOCLIM had become available "more than 25 years ago", though in 2017 it was actually 33 years earlier. As described in a following section the first eight years of BIOCLIM applications (1984–1991) were important as several key SDM areas were investigated for the first time (Booth *et al.* 2014).

Surprisingly, considering the early and continuing relevance of the BIOCLIM package and its variables, Guisan *et al.* (2017, p. 20) make no mention of it in their 'short (non-exhaustive) history of HSMs (adapted from Guisan & Thuiller 2005)'. They do mention several interesting early SDM applications, but none of these methods provided generic packages that were taken up by other users in the early days of SDM.

In contrast, BIOCLIM provided an integrated SDM package for the whole of Australia, which could be applied to any plant or animal species for which geocoded (i.e. latitude, longitude and elevation) distributional data were available. The ease of use of BIOCLIM was critically important in

 doi:10.1111/aec.12628

establishing SDM studies. The first publications using BIOCLIM appeared in 1985 (Booth 1985; Prendergast & Hattersley 1985) and BIOCLIM was soon applied to a wide range of ecological problems (see Booth *et al.* 2014). In later years, the package and/or its related set of variables were used in hundreds of publications, so it can reasonably be described as the start of modern SDM studies.

The failure to specify when BIOCLIM became available in the three books has led to some confusion over when modern SDM studies began. For example, in an encyclopaedia of geography Miller and Holloway (2017) wrote that correlative SDM techniques have been used for "approximately 20 years". If scientists do not know when research began in their area of interest this increases the risk of them lacking relevant information.

For example, it is instructive to consider SDM studies of climate change impacts on forests. The first two SDM climate change studies of any sort were published in 1988, but are not mentioned in the three books being considered here or in highly cited relevant SDM reviews, such as Pearson and Dawson (2003), Guisan and Thuiller (2005) or Elith and Leathwick (2009). The climate change study of Busby (1988) included a BIOCLIM examination of likely impacts on the natural distribution of *Nothofagus cunninghamii* (Hook.) Oerst. (Myrtle or Southern Beech, now known as *Lophozonia cunninghamii* (Hook f.) Heenan & Smissen) in Australia, while the other (Booth & McMurtrie 1988) looked at likely impacts on plantations of *Pinus radiata* D. Don. (Monterey Pine) across Australia.

Busby (1988) stated what has become known as the 'equilibrium assumption' at the start of his discussion noting that "the primary assumption of the BIOCLIM system is that entities can only colonise and survive in areas with climates fitting within their present climate profile". Ironically, the Booth and McMurtrie (1988) paper showed how widely *P. radiata* is grown in Australia, and thus provided some evidence why this assumption should be viewed with caution for long-lived tree species under climate change. *P. radiata* is an endangered species in its natural distribution of just 5300 ha in California, but is the most widely planted softwood species in the southern hemisphere with over 4 M ha of plantations demonstrating its considerable climatic adaptability (Mead 2013).

Recently Dyderski *et al.* (2018) have collated a list of 124 papers published between 1996 and 2017 related to SDM studies of climate change impacts on forests (Fig. 1). It is safe to assume that most of these authors were unaware of the pioneering work in this area, as only one paper (Mellert *et al.* 2015) cited one of the early BIOCLIM climate change studies. Dyderski *et al.* (2018) commented that the 124 studies "were usually limited to one country or region within a country. Only 18 of the studies were conducted at continental or broader spatial extents". Almost all of the 124 studies followed the approach of Busby (1988) in simply analysing tree species current natural distributions, applying climate change scenarios and projecting changes in species distributions.

However, early BIOCLIM studies had demonstrated that tree species can often grow successfully in trials under climatic conditions somewhat different

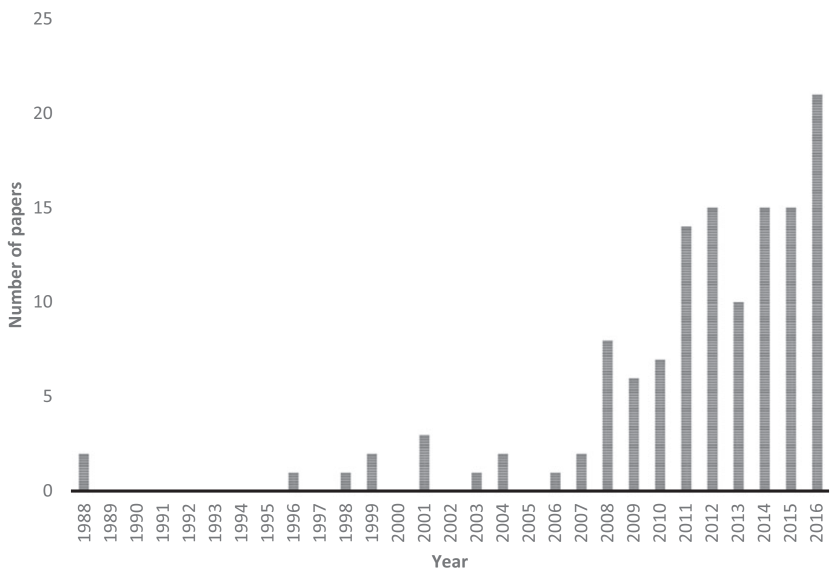

Fig. 1. Published papers describing SDM climate change studies in forests. Data shown include the Busby (1988) and Booth and McMurtrie (1988) papers and then the frequency of occurrence for the papers collated by Dyderski *et al.* (2018) for complete years. The sole observation from 2017 was omitted.

doi:10.1111/aec.12628

from those of their natural distribution (Booth *et al.* 1988). Competitive processes for trees in natural stands occur largely during establishment. It is reasonable to assume that under climate change a long-lived tree species, already well-established at a particular site, may be able to display some of the climatic adaptability that it shows in trials outside its natural distribution (Booth 1991, 2017). One of the papers listed by Dyderski *et al.* (2018) that did make efforts to collate data from outside, as well as within, natural distributions was that of McKenney *et al.* (2007). The senior author had worked closely with the BIOCLIM team and one of the team (M.F. Hutchinson) was a co-author.

Improved thin-plate spline climatic interpolation methods, later used to create the WorldClim database, were a key part of the BIOCLIM developments

Understanding that the development of improved climatic interpolation methods were a vital part of the BIOCLIM initiative is essential to appreciate its pioneering significance and continuing relevance. Of the three books only Franklin (2009) clearly described how the improved climatic interpolation methods developed by Michael Hutchinson were "integral to the BIOCLIM system" (p. 78). Peterson *et al.* (2011, p. 92) cited the Hijmans *et al.* (2005) reference that describes the WorldClim global climatic database, and noted their importance for niche modelling. However, they appeared to be unaware of the link between the 19 bioclimatic variables available from WorldClim and BIOCLIM. Guisan *et al.* (2017, p. 63) mention the use of thin-plate splines in the development of WorldClim, and cite Hutchinson (1995). However, they also did not appear to appreciate the close links between the development of the ANUSPLIN interpolation package (Hutchinson 2013) and BIOCLIM (see Booth *et al.* 2014 for details).

In their description of BIOCLIM Guisan *et al.* (2017, p. 156) suggest that the package simply extended the approaches pioneered by the life zone concept of Holdridge (1947) and the macroclimate study of Box (1981). Although the use of simple climatic ranges to define the groups is a point of similarity, the differences are great. Only BIOCLIM takes in locational data for particular species, determines their environmental requirements and maps likely areas of occurrence. In other words it was the first package to provide the three key stages of the basic SDM approach illustrated in Figure 1.1 in Franklin (2009) and Figure 4.1 in Guisan *et al.* (2017).

The use of improved methods of climatic interpolation was crucial to the important advance that BIOCLIM provided. Box (1981) based his analysis on climatic data estimated for only 1225 locations across the whole world. In contrast, BIOCLIM could reliably estimate climatic conditions for any location in Australia. The thin-plate spline programs, now known as ANUSPLIN (Hutchinson 2013), built on the work of Wahba and Wendelberger (1980). They include a generalized cross-validation. This effectively uses all the observations to fit the surface and all the observations to check the surface. Errors are estimated for every observation, so regions and periods requiring more data points can be readily identified. A key aspect in the accuracy of these surfaces was the incorporation of an appropriately scaled dependence on land elevation, as later demonstrated by Hutchinson (1995). The programs of what has become the ANUSPLIN package were originally developed in part for the BIOCLIM initiative (Hutchinson & Bischof 1983; Hutchinson *et al.* 1984; Hutchinson 1991).

Guisan *et al.* (2017, p. 63) describe the use of climatic data from the nearest grid point to estimate conditions at particular occurrence locations. This is appropriate for many current users, as they may not have access to interpolation relationships. However, it is remarkable that most current SDM analyses continue to use bioclimatic estimates from the nearest grid point. More than 34 years ago BIOCLIM incorporated interpolation relationships for the whole of Australia, an area comparable to the contiguous United States. Climatic estimates are made by BIOCLIM for the actual locations of occurrences rather than taken from the nearest grid point. It is not a great leap to imagine future SDM studies using detailed topography to provide improved estimates of environmental conditions that take account of variations in factors such as solar radiation, frost and soil water at specific locations.

The first version of BIOCLIM used 12 bioclimatic variables, later increased to 19 in 1996 and 35 in 1999

BIOCLIM has used various numbers of variables ranging from 12 in the earliest version to 35 in the current version. Just three of the initial 12 variables were the same as those used in the set of eight variables applied by Box (1981). Booth (1985) and Nix (1986) provide a list of the initial 12 BIOCLIM variables, which were derived from mean monthly values of daily maximum temperature and daily minimum temperature, as well as mean monthly values of precipitation. The derived variables, such as mean minimum temperature of the coldest month and mean precipitation of the driest quarter were calculated to provide meaningful variables independent of particular months, which could be used for any location in the world. Xu and Hutchinson (2011, p. 44) provide a list of the 35

doi:10.1111/aec.12628

variables, which made use of additional monthly mean values for Class A pan evaporation and solar radiation.

Of the three SDM books only Franklin (2009, p. 78) makes any specific mention of the number of variables used by the BIOCLIM package, and then only the 16 variables that had replaced the initial 12 by 1991 (see Busby 1991a,b). Most confusingly Guisan *et al.* (2017) describe a BIOCLIM analysis (p. 156–158) using only four variables, but do not make it clear that this was a greatly simplified example. At the other extreme they comment that a BIOCLIM model using 100 variables "might prove highly accurate in defining the current extent of a given species, but it would expectedly perform relatively poorly when used to project the distribution of the species in space and time.".

Peterson *et al.* (2011, p. 105) note that BIOCLIM cannot handle interactions between variables. This is true, but individual BIOCLIM variables can be based on complex calculations of interactions between environmental factors. For example, the 35 variable version includes eight moisture index values calculated from water balance equations based on weekly estimates of precipitation and evaporation (see section 6.2 of the ANUCLIM manual (Xu & Hutchinson 2011) for a description of the method).

In their Figure 7.7 (p. 130) Peterson *et al.* (2011) present the graph from the highly cited (>3500 citations in Web of Science) paper of Elith *et al.* (2006) which compared different methods for modelling species distributions. This showed BIOCLIM performing poorly in comparison with other methods. As it uses only simple ranges BIOCLIM is sensitive to the number of variables used. This is reflected in the increase in number of variables from 12 to 35 from 1984 to 1999. It is worth noting that these comparative analyses were based on a range of between 11–13 variables, with a maximum of only 10 climatic variables (see Table 1 in Elith *et al.* 2006) i.e. below the number of variables used in even the earliest version of BIOCLIM and only 28 per cent of the bioclimatic variables used by the then current version.

The advantages of some of the more recent analysis methods, such as Maxent, are acknowledged here and include the ability to perform well with relatively few variables. However, readers of the SDM books may not appreciate that limitations of BIOCLIM would be exaggerated by the use of relatively few variables by Elith *et al.* (2006) and particularly by Guisan *et al.* (2017).

SDM research areas for which Guisan and Thuiller (2005) provide relatively recent examples were all first explored using BIOCLIM

All three books cover papers describing early BIOCLIM applications very sparsely. For example, Guisan *et al.* (2017, p. 50) list 'some possible uses and applications of HSMs' in their Table 4.1). Eight of these types (1 and 4–10) are the same as example application areas described in the highly cited Guisan and Thuiller (2005) paper. Booth *et al.* (2014) showed how examples of all the areas for which Guisan and Thuiller (2005) provide later examples had all been first explored in BIOCLIM studies published between 1985 and 1991, but none are included in Table 4.1. Guisan *et al.* (2017) noted that it would be impossible to list all published examples, so below the table they listed some review papers and complementary books that also largely ignore the early BIOCLIM studies. Unfortunately, they do not list the Booth *et al.* (2014) review that describes the earliest examples in several SDM application areas there, though it is cited later in the book (p. 156).

It is not suggested that Guisan *et al.* (2017) should have cited all the early SDM papers in their Table 4.1. However, it would have been desirable to have included at least a couple. For example, as climate change studies are such an important part of current SDM studies, the first SDM climate change study (Busby 1988) might have been listed. Similarly, the first paper to relate Hutchinson's (1957) niche ideas to SDM studies and to show how both the realized niche and at least parts of the fundamental niche could be assessed (Booth *et al.* 1988) might also have been worth including. This was arguably the first ENM paper, as it recognized two types of niche and showed how they could be assessed, thus satisfying the ENM definition of Peterson *et al.* (2011; p. 271).

A consequence of the sparse coverage of the earliest BIOCLIM studies in the three books is that some important early observations have been overlooked. For instance, Guisan *et al.* (2017, p. 313) note that "very few studies have so far attempted to quantify the difference between the fundamental and realized niche...because this question is extremely difficult to assess from empirical data on species distributions and experimental *in situ* and *ex situ* studies are also needed to explore this issue.". Yet the desirability of considering results from trials outside natural distributions, as well as data from natural distributions, was well-understood in early SDM forestry studies (Booth *et al.* 1988; Booth 1991). Some recent papers have questioned the 'equilibrium assumption' that analyses of species natural distributions alone provide a reliable indication of the climatic requirements of long-lived species under climate change (see, for example, Early & Sax 2014 and the review by Booth 2017).

While more sophisticated SDM methods have been developed BIOCLIM is still widely used

It is not disputed that more modern SDM methods, such as Maxent (Phillips *et al.* 2006) have advantages

doi:10.1111/aec.12628

in comparison to BIOCLIM. However, the BIOCLIM package was included in various SDM and GIS packages (e.g. IDRISI and ARCGIS) and is still used. For example, it has been included in the Biodiversity and Climate Change Virtual Laboratory (BCCVL) system (Hallgren *et al.* 2016; www.bccvl.org.au), which is closely integrated with the Atlas of Living Australia (Belbin & Williams 2015; ala.org.au). BIOCLIM has advantages particularly for teaching purposes in being based on easily understandable processes closely related to the niche concepts of Hutchinson (1957).

Hijmans and Elith (2011) used BIOCLIM as one of the examples in their manual on using the R statistical computing package for SDM (R Development Core Team 2016). Guisan *et al.* (2017) based their book on the use of R scripts and mention that BIOCLIM has value for providing baseline predictions for comparison with other more sophisticated analyses. Similarly, Kass *et al.* (2018) have chosen to support BIOCLIM as well as Maxent in their Wallace system, which provides a graphical user interface for SDM in R.

The 19 BIOCLIM variables are still used in many current SDM analyses

The set of 19 bioclimatic variables was first adopted for BIOCLIM by the team of Nix, Busby and Hutchinson in about 1996 (see, for example, the application by Lindenmayer *et al.* 1996). The set of variables was adopted for use in the WorldClim database (Hijmans *et al.* 2005) and has been very widely used in SDM analyses. For example, Booth *et al.* (2014) noted that more than 75% of recent published Maxent analyses of terrestrial ecosystems used some or all of the 19 BIOCLIM variables available from WorldClim. Current SDM papers seldom acknowledge the origin of these 19 variables with the BIOCLIM package. Similarly, none of the three books clearly relate the origin of the 19 variables to the BIOCLIM package. WorldClim had only become available relatively recently when the Franklin (2009) book was written, so this omission was not surprising. Peterson *et al.* (2011) cite Hijmans *et al.* (2005) in their chapter six on environmental data, but do not mention the 19 variables or their relation to BIOCLIM. Guisan *et al.* (2017) refer to "19 so-called bioclimatic variables (bioclim)" (p. 63), but elsewhere refer to the SDM package as "BIOCLIM".

The lack of knowledge in the SDM community about the origins of the set of 19 variables is best illustrated by two recent papers which focussed particularly on variables used in SDM studies. Bradie and Leung (2017) mention 19 "BioClim" variables available from WorldClim and note their use in >1000 models. Similarly, Title and Bemmels (2017) refer to "the 19 bioclimatic variables from WorldClim" as "the *bioclim* model". However, neither paper provides any reference to the BIOCLIM SDM.

The aim of the Title and Bemmels (2017) paper was to extend the set of 19 variables. They briefly mentioned the CliMond dataset (Kriticos *et al.* 2012) and wrote that though only available at 10 and 30 arc-minutes it is "the only existing multi-variable dataset that is truly complementary to WorldClim in its breadth, application and accessibility". However, they did not mention that the BIOCLIM variables included in CliMond are based on the extended set of 35 BIOCLIM variables. This now also includes a further five variables based on a principal component analysis (PCA) of the 35 BIOCLIM variables, which account for more than 90% of the variability in the data (Kriticos *et al.* 2014).

To date several hundred publications have used the BIOCLIM SDM package or the BIOCLIM variables

Although the three books mention that BIOCLIM was the first or one of the first SDM packages they do not describe its importance in the years from 1984–2006. Elith *et al.* (2006) recognized BIOCLIM as one of only three "well-established modelling methods" for SDM studies.

The BIOCLIM package or its associated variables are used in many SDM studies, but BIOCLIM reviews such as Busby (1991a,b) or Booth *et al.* (2014) are cited in few of these papers. For example, of the 124 papers listed by Dyderski *et al.* (2018) one mentions Busby (1991a), four cite Busby (1991b) and one cites Booth *et al.* (2014).

Values for the set of 19 variables developed for BIOCLIM in 1996 are usually downloaded for current studies from the WorldClim system (Hijmans *et al.* 2005). They are sometimes referred to as "Bioclim variables", though their origin is seldom acknowledged.

CONCLUSIONS

It should be emphasized that the three SDM books discussed here all make very useful contributions and are all highly recommended. However, the descriptions of BIOCLIM in each of the three books are very brief and omit key points made in the paragraph in the introduction.

It is therefore not surprising that the pioneering work on which much of modern SDM studies are based goes largely unacknowledged in most current SDM papers. Rather than indicating the origins of

 doi:10.1111/aec.12628

SDM studies many current papers begin by saying that there has been an enormous increase in SDM studies in recent years. They tend to cite reviews such as that by Guisan and Thuiller (2005) and/or the Franklin (2009) and Peterson *et al.* (2011) books. It is likely that the Guisan *et al.* (2017) book will be a key reference in future. Researchers preparing SDM papers might reasonably assume that these publications adequately describe the pioneering SDM research and its continuing relevance, but they do not. The main omissions are failures to mention when BIOCLIM became available (i.e. January 1984), the range of variables used by BIOCLIM, the key SDM application areas explored from 1984–1991, the importance of the package from 1984–2006, and its continuing significance (both in terms of the 19 variables now commonly used and actual data prepared by the WorldClim team (Hijmans *et al.* 2005) using the climatic interpolation programs developed in part for BIOCLIM).

The authors of the three books all faced a great problem with the vast increase in the number of SDM studies in recent years and the need to summarize hundreds of scientific papers. The two early summaries of BIOCLIM studies also had limitations being, respectively, very brief (Busby (1991a) was just one and a half pages long) and not easily accessible (Busby (1991b) was in a conference volume). Early BIOCLIM publications were not easy to locate as no standard term such as "species distribution modelling" had then emerged and even the BIOCLIM package was initially called the "Bioclimate Prediction System".

It was particularly unfortunate that the Guisan and Thuiller (2005) review made no mention of BIOCLIM in its text, though reference to it and the Busby (1991a) review was included in a table. The Guisan and Thuiller (2005) review aimed to provide "an overview of recent advances" in SDM and it fulfilled that aim very well. However, as shown here it does not cover pioneering SDM work adequately, yet it is often cited at the beginning of SDM papers.

There are good reasons why the pioneering BIOCLIM research and its continuing significance should be acknowledged. At least three types of problem have arisen as a result of the widespread lack of awareness of early SDM studies. The first type relates to the 'equilibrium assumption' which has been generally applied in climate change studies. The need to assess species climatic requirements beyond those of the realized niche, especially for long-lived species such as trees, was well-understood in early BIOCLIM studies (Booth *et al.* 1988; Booth 1991). It has taken about 25 years for this concern to be re-discovered and in the meantime numerous potentially misleading analyses involving trees or forest-dependent species have been completed.

The second type of problem relates to ignoring the potential of commercial forestry trials to inform studies of species and provenances climatic requirements. The three SDM books all consider the use of information from species invasions, but curiously ignore commercial forestry species elimination trials, which are specifically designed to determine species' climatic and edaphic requirements. Invasion data can be useful, but have several limitations. Has the invading species reached equilibrium? Which species is it competing with in the invaded area? In contrast, forestry species elimination trials take unimproved material from known locations within natural distributions and test them at known locations outside natural distributions. Foresters do everything possible to reduce competition with other species. In many cases they may test a number of provenances of the same species from different locations within the natural distribution (see, for example, Eldridge *et al.* 1996). It would be hard to imagine trials better suited to informing SDM climate change studies, but they are almost totally ignored in the three SDM books.

The third problem relates to users not appreciating that the 19 variables available from WorldClim were originally created for BIOCLIM. If Title and Bemmels (2017) had known that the set of 19 BIOCLIM variables had been extended to 35 variables they could have tested their proposed variables against the existing extended set. They could also have made use of the existing PCA analysis of the extended BIOCLIM dataset to select the most useful additional variables.

Current researchers would benefit from improved knowledge of the pioneering BIOCLIM research. This should assist some of them to make better decisions about the use of particular methods and variables. It would be good to see at least a few current SDM papers acknowledge some of the key points made here in the paragraph in the introductory section. Just the addition of a single phrase at the start of some current papers noting that "modern SDM studies began when the BIOCLIM package became available in 1984" might cause some researchers to examine the pioneering studies.

ACKNOWLEDGEMENTS

I am particularly grateful to my colleagues Henry Nix, John Busby and Michael Hutchinson, for their excellent work in making BIOCLIM available for users such as myself. The late June McMahon provided important input to the development of the program. Henry Nix, Stephen Roxburgh and Peter Szabo, as well as the anonymous referees, provided very helpful comments on earlier drafts of this paper.

doi:10.1111/aec.12628

REFERENCES

Belbin L. & Williams K. J. (2015) Towards a national bio-environmental data facility: Experiences from the Atlas of Living Australia. *Int. J. Geogr. Inf. Syst.* **30,** 108–25.

Booth T. H. (1985) A new method to assist species selection. *Commonw. Forest. Rev.* **64,** 241–50.

Booth T. H. (1991) A global climatological audit to assist conservation and sustainable development. Proceedings of 10th World Forestry Congress, 2, 65–70.

Booth T. H. (2017) Assessing species climatic requirements beyond the realized niche: Some lessons mainly from tree species distribution modelling. *Clim. Change.* **145,** 259–71.

Booth T. H. & McMurtrie R. E. (1988) Climatic change and *Pinus radiata* plantations in Australia. In: *Greenhouse: Planning for Climate Change* (ed G. I. Pearman), pp. 534–45. CSIRO, Melbourne and E.J. Brill, Leiden.

Booth T. H., Nix H. A., Hutchinson M. F. & Jovanovic T. (1988) Niche analysis and tree species introduction. *Forest Ecol. Manage.* **23,** 47–59.

Booth T. H., Nix H. A., Busby J. R. & Hutchinson M. F. (2014) BIOCLIM: The first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Divers. Distrib.* **20,** 1–9.

Box E. O. (1981) *Macroclimate and Plant Forms: An Introduction to Predictive Modeling in Phytogeography*. Junk, The Hague.

Bradie J. & Leung B. (2017) A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. *J. Biogeogr.* **44,** 1344–61.

Busby J. R. (1988) Potential implications of climate change on Australia's flora and fauna. In: *Greenhouse: Planning for Climate Change* (ed G. I. Pearman), pp. 387–98, CSIRO, Melbourne and E.J. Brill, Leiden.

Busby J. R. (1991a) BIOCLIM – a bioclimate analysis and prediction system. *Plant Protect. Quarterly* **6,** 8–9.

Busby J. R. (1991b) BIOCLIM – a bioclimate analysis and prediction system. In: *Nature Conservation: Cost Effective Biological Surveys and Data Analysis* (eds C. R. Margules & M. P. Austin) pp. 64–8. CSIRO, Melbourne.

Dyderski M. K., Paz S., Frelich L. E. & Jagodzinski A. M. (2018) How much does climate change threaten European forest tree species distributions? *Global Change Biol.* **24,** 1150–63.

Early R. & Sax D. F. (2014) Climatic niche shifts between species' native and naturalized ranges raise concern for ecological forecasts during invasions and climate change. *Global Ecol. Biogeogr.* **23,** 1356–65.

Eldridge K., Davidson J., Harwood C. & van Wyk G. (1996) *Eucalypt Domestication and Breeding*. Cambridge University Press, Cambridge, UK.

Elith J. & Leathwick J. R. (2009) Species distribution models: Ecological explanation across space and time. *Ann. Rev. Ecol. Syst.* **40,** 677–97.

Elith J., Graham C. H., Anderson R. P. *et al.* (2006) Novel methods improve predictions of species' distributions from occurrence data. *Ecography* **29,** 129–51.

Franklin J. (2009) *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press, Cambridge, UK.

Friedman S. K. (2011) Review of mapping species distributions: Spatial inference and prediction by Janet Franklin. *Landscape Ecol.* **26,** 895–7.

Guisan A. & Thuiller W. (2005) Predicting species distribution: Offering more than simple habitat models. *Ecol. Lett.* **8,** 993–1009.

Guisan A., Thuiller W. & Zimmermann N. E. (2017) *Habitat Suitability and Distribution Models with Applications in R*. Cambridge University Press, Cambridge, UK.

Hallgren W., Beaumont L., Bowness A. *et al.* (2016) The biodiversity and climate change virtual laboratory: Where ecology meets big data. *Environ. Modell. Softw.* **76,** 182–6.

Hijmans R. J. & Elith J. (2011) *Species Distribution Modelling with R*. University of Crete, Crete, Greece. 79 p.

Hijmans R. J., Cameron S. E., Parra J. L., Jones P. G. & Jarvis A. (2005) Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25,** 1965–78.

Holdridge L. R. (1947) *Life Zone Ecology*. Tropical Science Center, San Jose, Costa Rica.

Hutchinson G. E. (1957) Concluding remarks. *Cold Spring Harb. Sym.* **22,** 415–27.

Hutchinson M. F. (1981) *MAPROJ – A computer map projection system*. CSIRO Division of Land Use Research. Technical Paper No. 39. CSIRO, Melbourne.

Hutchinson M. F. (1991) The application of thin plate smoothing splines to continent-wide data assimilation. In: *Data Assimilation Systems* (ed J. D. Jasper) pp. 104–13. Bureau of Meteorology Research Report No. 27, Melbourne.

Hutchinson M. F. (1995) Interpolating mean rainfall using thin plate smoothing splines. *Int. J. Geogr. Inf. Syst.* **9,** 385–403.

Hutchinson M. F. (2013) *ANUSPLIN version 4.4*. Centre for Resource and Environmental Studies, Australian National University. Available at: fennerschool.anu.edu.au/files/anusplin44.pdf (accessed 19 December 2017)

Hutchinson M. F. & Bischof R. J. (1983) A new method for estimating the spatial distribution of mean seasonal and annual rainfall applied to the Hunter Valley, New South Wales. *Aust. Meteorol. Mag.* **31,** 179–84.

Hutchinson M. F., Booth T. H., McMahon M. F. & Nix H. A. (1984) Estimating monthly mean values of daily solar radiation for Australia. *Sol. Energy* **32,** 277–90.

Kass J. M., Vilela B., Aiello-Lammens M., Muscarella R., Merow C. & Anderson R. P. (2018) Wallace: A flexible platform for reproducible modelling of species niches and distributions built for community expansion. *Methods Ecol. Evol.* **9,** 1151–6.

Kriticos D. J., Webber B. L., Leriche A. *et al.* (2012) CliMond: Global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods Ecol. Evol.* **3,** 53–64.

Kriticos D. J., Jarošik V. & Ota N. (2014) Extending the suite of bioclim variables: A proposed registry system and case study using principal components analysis. *Methods Ecol. Evol.* **5,** 956–60.

Lindenmayer D. B., Mackey B. G. & Nix H. A. (1996) The bioclimatic domains of four species of commercially important eucalypts from south-eastern Australia. *Aust. Forestry* **59,** 74–89.

McKenney D. W., Pedlar J. H., Lawrence K., Campbell K. & Hutchinson M. F. (2007) Potential impacts of climate change on the distribution of North American trees. *Bioscience* **57,** 939–48.

Mead D. J. (2013) Sustainable management of *Pinus radiata* plantations. FAO Forestry Paper No. 170. Rome, FAO.

Mellert K. H., Deffner V., Kuchenhoff H. & Kolling C. (2015) Modeling sensitivity to climate change and estimating the uncertainty of its impact: A probabilistic concept for risk assessment in forestry. *Ecol. Model.* **316,** 211–6.

Miller J. A. & Holloway P. (2017) Niche theory and models. In: *The International Encyclopedia of Geography* (ed D. Richardson) pp. 1–10. John Wiley & Sons, Hoboken, NJ.

 doi:10.1111/aec.12628

Nix H. A. (1986) A biogeographic analysis of Australian elapid snakes. In *Atlas of Elapid Snakes of Australia: Australian Flora and Fauna Series 7* (ed R. Longmore), pp. 4–15. Bureau of Flora and Fauna, Canberra.

Pearce I. S. (2012) Review of ecological niches and geographic distributions by Peterson A.T. et al. *Q Rev. Biol.*, **87,** 376–7.

Pearson R. G. & Dawson T. P. (2003) Predicting the impacts of climate change on the distribution of species: Are bioclimate models useful?. *Global Ecol Biogeogr.* **12,** 361–72.

Peterson A. T., Soberon J., Pearson R. G., Anderson R. P., Martınez-Meyer E., Nakamura M. & Araujo M. B. (2011) *Ecological Niches and Geographic Distributions*. Monographs in Population Biology No. 49. Princeton University Press, Princeton, NJ.

Phillips S. J., Anderson R. P. & Schapire R. E. (2006) Maximum entropy modelling of species geographic distributions. *Ecol. Model.* **190,** 231–59.

Prendergast H. D. V. & Hattersley P. W. (1985) Distribution and cytology of Australian Neurachne and its allies (Poaceae), a group containing C3, C4 and C3-C4 intermediate species. *Aust. J. Bot.* **33,** 317–36.

R Development Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Soberon J. (2011) Review of mapping species distributions: Spatial inference and prediction by Janet Franklin. *Q. Rev. Biol.* **86,** 219–20.

Title P. O. & Bemmels J. B. (2017) ENVIREM: An expanded set of bioclimatic and topographic variables increase flexibility and improves performance of ecological niche modeling. *Ecography* **41,** 291–307.

Wahba G. & Wendelberger J. (1980) Some new mathematical methods for variational objective analysis using splines and cross validation. *Mon. Weather Rev.* **108,** 1122–43.

Warren D. L. (2012) Review of ecological niches and geographic distributions by Peterson A.T. et al. *Ecology*, **93,** 2126–7.

Xu T. & Hutchinson M. F. (2011) *ANUCLIM Version 6.1 User Guide*. Fenner School of Environment and Society, Australian National University. 90 p. fennerschool.anu.edu.au/files/anuclim61.pdf (Accessed 15 April 2018)

doi:10.1111/aec.12628