Challenges and opportunities of species distribution modelling of terrestrial arthropod predators

Stefano Mammola[1,2,*], Julien Pétillon[3], Axel Hacala[3], Jérémy Monsimet[4], Sapho-Lou Marti[3], Pedro Cardoso[1], Denis Lafage[3,5]

[1] Laboratory for Integrative Biodiversity Research (LIBRe), Finnish Museum of Natural History (LUOMUS), University of Helsinki, Helsinki, Finland

[2] Molecular Ecology Group (MEG), Water Research Institute (RSA), National Research Council (CNR), Verbania Pallanza, Italy

[3] UMR ECOBIO, Université de Rennes 1, Rennes, France

[4] Inland Norway University of Applied Sciences, Campus Evenstad, Koppang, Norway

[5] Department of Environmental and Life Sciences/Biology, Karlstad University, Karlstad, Sweden

* corresponding author: stefano.mammola@helsinki.fi ; stefano.mammola@cnr.it

Author contribution statement

SM, JP, and DL conceived the study. SM, AC, S-LM, and DL mined and extracted information from papers. DL curated data, performed analyses, and prepared figures. SM and DL wrote the first draft of the paper, with substantial inputs by JP, JM, and PC. All authors contributed critically to the paper by means of discussions, comments, and additions to the text.

Conflict of interest statement

None declared

ABSTRACT

Species distribution models (SDMs) are emerging as essential tools in the equipment of many ecologists; they are useful in exploring species distributions in space and time and in answering an assortment of questions related to historical biogeography, climate change biology and conservation biology. Given that arthropod distributions are strongly influenced by microclimatic conditions and microhabitat structure, they should be an ideal candidate group for SDM research, especially generalist predators because they are not directly dependent on vegetation or prey types. However, most SDM studies of animals to date have focused either on broad samples of vertebrates or on arthropod species that are charismatic (e.g. butterflies) or economically important (e.g. vectors of disease, crop pests and pollinators). By means of a systematic bibliometric approach, we targeted the literature published on key terrestrial arthropod predators (ants, ground beetles and spiders), chosen as a model to explore challenges and opportunities of species distribution modelling in mega-diverse arthropod groups. We show that the use of SDMs to map the geography of terrestrial arthropod predators has been a recent phenomenon, with a near-exponential growth in the number of studies over the past 10 years and still limited collaborative networks among researchers. There is a bias in studies towards charismatic species and geographical areas that hold lower levels of diversity but greater availability of data, such as Europe and North America. To overcome some of these data limitations, we illustrate the potential of modern data sources (citizen science programmes, online databases) and new modelling approaches (ensemble of small models, modelling above the species level). Finally, we discuss areas of research where SDMs may be combined with dispersal models and increasingly available phylogenetic and functional data to obtain mechanistic descriptions of species distributions and their spatio-temporal shifts within a global change perspective.

Keyword: bibliometrics; climate change; ecological niche models; MaxEnt; niche-based models; predicted distribution; social network analysis; statistical modelling

INTRODUCTION

A mainstream topic in ecology, biogeography and conservation biology is the extent to which climatic conditions affect species performance (Colinet et al. 2015, Rezende and Bozinovic 2019), which together with geographical and historical constraints ultimately modulates species niches and observed range boundaries (Thomas 2010). Obtaining a nuanced understanding of the factors conditioning species distributions has gained new urgency amid the current climate emergency (Ripple et al. 2020), insofar as changing climatic conditions are determining fast redistributions of species along latitudinal, elevational and other spatial gradients (Chen et al. 2011, Lenoir et al. 2020). As global climate change redefine the geography of life, we are becoming spectators of a large-scale experiment of complex ecological responses (Walther et al. 2002), where interactions among previously isolated species can quickly occur (Krosby et al. 2015), invasions of novel areas by alien species are becoming routine (Hellmann et al. 2008, Liu et al. 2020) and unnoticed extinctions are potentially taking place on a daily basis (Hughes et al. 2004, Barnosky et al. 2011, Cardoso et al. 2020b). Therefore, mapping the diversity of life has never been so urgent.

Over the years, ecologists and statisticians have developed a wide range of methods for modelling the niches and distribution of species in space and time, several of which fall under the umbrella of correlative species distribution models or ecological niche models (defined in Box 1). For simplicity, we will hereafter refer to these as ‘species distribution models’ (SDMs), while redirecting the interested readers to key semantic and theoretical discussions (e.g. Peterson and Soberón, 2012; Sillero, 2011; Warren, 2012). Researchers have used SDM techniques for mapping the distribution of organisms in a variety of systems, although the number of applications across habitats and the tree of life have not been equal. For example, while the use of SDMs has grown exponentially in the terrestrial realm from the early 2000s onward (Lobo et al. 2010, Robinson et al. 2011, Araújo et al. 2019), applications in systems where three-dimensionality is an important feature – e.g. marine ecosystems (Robinson et al. 2017, Melo-Merino et al. 2020), tree canopies

(Burns et al. 2020), soils (Schröder 2008) and caves (Mammola and Leroy 2018) – have lagged behind. Also, applications of SDMs in animals have concentrated mostly on vertebrates (Titley et al. 2017), while studies on invertebrate groups remain scarcer, although recently increasing (Figure 1).

The paucity of SDM studies is possibly related to a number of arthropod-specific modelling challenges. First, arthropods often are small organisms that move in small spatial scales, strongly influenced by microclimatic conditions and microhabitat structure (Pincebourde and Woods 2020). These characteristics are hardly captured by the ubiquitous bioclimatic variables derived from remote sensing at relatively large spatial scales (e.g. Potter et al., 2013). Second, arthropods often have short life cycles with wide population abundance fluctuations from season to season and strong metapopulation dynamics, making it difficult to determine what their real, constantly changing, range is. Thus, arthropods pose particular modelling challenges that add to the ones already present for vertebrates, but they should also offer opportunities for future SDM research as data and new methods are made available (Maino et al. 2016).

Here, we conducted a systematic mapping of the literature to synthesize trends in the use of SDMs in arthropod research. We explored these topics through the lens of the literature on dominant terrestrial arthropod predators: ants (c. 30,000 described species; Parr et al., 2017), ground beetles (c. 39,300 species; Lorenz, 2020), and spiders (c. 49,000 species; World Spider Catalog, 2020). We begin by conducting a systematic literature search focused on SDM use in our focal group. Then, to put our survey in perspective, we compare the volume of literature with that on other key terrestrial invertebrate and vertebrate groups. By means of bibliometric analyses, we explore the geography of SDM applications and networks of collaborations among researchers working on terrestrial arthropod predators. Subsequently, we review the main areas of application of SDMs in terrestrial arthropod research, highlighting which ecological factors emerged as important in driving predicted distribution patterns. Building upon this quantitative evidence, we discuss challenges and opportunities of SDM research on terrestrial arthropod predators and delineate potential future lines of enquiry as well as promising areas of research where SDMs may be combined with other modelling tools and data sources to obtain mechanistic descriptions of species distributions and their shifts within a global change perspective.

Box 1. A general definition of SDMs and their domain of applicability

As a broad and general definition, species distribution modeling implies using some statistical algorithms to explore the relationship between species occurrences (typically geo-referenced localities) and environmental variables (typically spatial rasters whose cells represent bioclimatic and other habitat and environmental conditions). Once this relationship is determined, the model is used to characterize the ecological niche of a given species by projecting a probability surface into a geographical space to represent its potential range of distribution (Guisan et al. 2017). These models can be construed using a wide range of algorithms, from simple logistic regression up to sophisticated techniques based on machine learning (Elith et al. 2011, Ryo et al. 2020) and other artificial intelligence methods (Cardoso et al. 2020a). Given the large variety of life histories and data sources, the best modelling algorithm and approach necessarily changes, with no universal best solutions (Qiao et al. 2015).

Whereas the first paper relying on species distribution modelling is now over three decades old [e.g. the first applications of the algorithm BIOCLIM can be traced back to 1986 (Booth 2018)], there has been an acceleration in the use of these tools in just the last two decades (Araújo et al., 2019; Lobo et al., 2010; Figure 1). This trend was probably due to the increase in data (Zhang 2017, Wüest et al. 2020) and easy to use, often automated, statistical packages that perform species distribution modelling (reviewed in Angelov, 2019). These methods have become popular in the toolkit of many ecologists, being useful to answer a range of questions. Not only are SDMs routinely used to describe species distributions, they have also proved important to assist and complement taxonomic studies (Rödder et al. 2010) and to set conservation agendas (Guisan et al. 2013). Furthermore, given that these models are transferable in space and time (Yates et al. 2018), they find applications in studies on climate change (Dormann 2007, Santini et al. 2020), historical biogeography (Peterson 2009), and invasion biology (Peterson 2003, Liu et al. 2020), among other topics.

METHODS

Systematic search of SDM papers and analyses

Between 20 and 24 November 2020, we searched on the Web of Science (Clarivate Analytics) for articles relying on SDMs to predict distributions of terrestrial arthropod predators (ants, ground beetles and spiders) and, for comparative purposes, other terrestrial vertebrate and invertebrate groups (Table 1). For each taxonomic group considered, we found and extracted papers using the following general query:

TS=(“family name(s)” OR “vernacular name(s)”) AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR “Niche model*” OR “Distribution model*” OR “Habitat suitability model*”)

where TS denotes a search for ‘Topic’ and the asterisk (*) is a regular expression used to match all words including that string of characters (e.g. ‘model*’ matches ‘models’, ‘modelling’, ‘modelled’, etc.). See Appendix 1 for the list of families and vernacular names.

We exported all results into the online review application Rayyan (Ouzzani et al., 2016) for title, keywords, and abstract screening, whereby we excluded by-catches of papers not actually dealing with SDMs or our model species (e.g. our search for the keyword ‘spiders’ also captured papers dealing with spider monkeys, genus *Ateles*) (Table 1). Furthermore, for ants, ground beetles and spiders, we manually inspected all papers to extract specific data. We recorded the geographical extent of each study and all the species modelled. We classified the type of predictors used, their resolution, and the SDM algorithm(s) and modelling protocol employed. Specifically, we coded the modelling protocol under three main categories: single algorithm, when studies just applied one modelling technique; ensemble of models, when the authors applied a plethora of algorithms and took the consensus between them (Araújo and New, 2007); and no-silver bullet (Qiao et al., 2015), when the authors applied a number of algorithms and chose the best performing one(s) for projecting the distribution. Finally, we read each study and summarized the key results (see Appendix 2 for a full list of extracted information).

Data analysis

We conducted analyses in R 3.6.3 (R Core Team, 2020) and visualized data using the *ggplot2* R package (Wickham, 2009) and QGIS (Open Source Geospatial Foundation Project, 2020). The complete data set and R code used for the analyses is available on GitLab (https://gitlab.com/DenisLafage/sdm_review).

We analysed bibliometric data regarding ants, ground beetles, and spiders with the *bibliometrix* R package (Aria and Cuccurullo 2017). In order to map the production of articles per country for each group, we assigned articles to a country based on the affiliations of all the authors at the time when each article was published. In order to identify the most influential papers for researchers dealing with modelling of macro-arthropod distributions, we used a weighted co-citation network. Initially introduced for bibliometric research, co-citation networks have proved useful to identify key literature items acting as bridges between disciplines (Trujillo and Long, 2018). A particular article is included in the network when it is cited by at least two papers from the dataset under study (Batagelj and Cerinšek 2013). The number of co-citations is the number of times two articles are cited together. Furthermore, we built a collaboration network to identify the existence of bridges among scientists working on ants, ground beetles and spiders.

Caveats in the interpretation of the survey

Some caveats need to be made when interpreting the results of this research weaving. Due to our search strategy in the Web of Science and selection of keywords (Appendix A), we did not capture all possible studies on SDMs dealing with our focal groups. For example, we missed some studies on taxonomy that used SDMs to assist species delimitations, since these rarely mentioned the methodology in their keywords, title or abstract. Similarly, SDMs have recently begun to be routinely used for assessing terrestrial arthropod risk of extinction against International Union for Conservation of Nature criteria (e.g. Branco et al., 2019; Fukushima et al., 2019; Seppälä et al., 2018b, 2018a, 2018c, 2018d), but most of these studies were missed for the same reason. Furthermore, for many groups, especially vertebrates, the authors may not mention the higher taxonomic ranks included in our query but exclusively the species/genus/family, which will not be captured. We also acknowledge that our search was not exhaustive since we only included articles in English (Konno et al. 2020) and we used a single database, Internet browser, and location for the search (Pozsgai et al. 2020).

As a result, our estimation of the volume of the literature on the focal groups should be taken as an approximation of the real number of studies. While we operated under the assumption that the biases were homogeneously distributed across all taxonomic groups, allowing us to compare them and to draw general inferences, still the comparison of absolute numbers of studies across taxa (Figure 1) should be taken with caution.

SURVEY RESULTS

Volume of scientific production in comparison to other groups

By comparing the volume of SDM literature on vertebrates versus invertebrates, we observed a similar exponential increase in the number of studies for both groups, with an inflection point after 2010 (Figure 1a). However, the total number of studies was greater for vertebrates (67%) than invertebrates, despite vertebrates accounting for a considerably lower number of species. Among invertebrates, the largest fraction of studies focused on Diptera (8.2%). Other well-studied groups were butterflies (6.4%), non-carabid beetles (5.8%), and Apoidea (2.5%). Our three focal groups were comparatively less studied (0.7, 1.7 and 1.9% for ground beetles, spiders and ants, respectively); all three of them began to be the focus of SDM research after 2005, with the number of studies steadily increasing from this point (Figure 1b).

Geography of scientific production

The geography of studies, as inferred from author affiliations, revealed how production of SDM papers on ants, ground beetles and spiders is mostly concentrated in North and South America and Europe (Figure 2). There were, however, some conspicuous differences among groups. For ants, modelled species are mostly in North and South America, and Europe (Appendix C), and 15 studies modelled species distribution worldwide. For spiders and ground beetles, most studies focused on European species (Appendices E and G), and only three and one studies/y, respectively, had worldwide coverage. There were considerably more ant species which have been studied with SDMs than spiders and ground beetles (Appendices D, F, H).

Most influential papers

The co-citation network allowed us to identify key articles co-cited by the studies included in our survey (Figure 3). As expected, most co-cited papers were methodological rather than arthropod-specific papers. The top-cited papers were Phillips et al. (2006) and Hijmans et al. (2005), respectively the reference for the algorithm MaxEnt and for the most widely used global climate database (WorldClim). Among the less co-cited but still influential papers, there were several references to phylogenetic methods, suggesting that a number of articles are potentially integrative research using multiple lines of evidence to deal with species delimitation (Ross et al. 2010, Ferretti et al. 2019) and historical biogeography (e.g., Solomon et al. 2008, Magalhaes et al. 2014, Planas et al. 2014, Mammola et al. 2015).

Collaboration network

Network analysis revealed highly structured collaboration hubs around the three groups of interest (Figure 4). Observed collaboration hubs were strongly bound but limited in size, with only four cases of inter-group collaborations (ants–ground beetles, ants–spiders and ground beetles–spiders). Two cases were the result of multi-taxa studies (Christman et al., 2016; Jiménez-Valverde et al., 2009) and two were related to authors involved in articles dealing with two different groups: Williams S.E. (Steiner et al. 2008, Staunton et al. 2014) and Peterson A.T. (Roura-Pascual et al. 2004, 2006, 2009, Peterson and Nakazawa 2008, Planas et al. 2014).

Paper topics

Articles dealing with ants primarily focused on climate change (33.9% of studies) and invasion biology (30.4%). Many studies often dealt with both topics simultaneously (23.2%) as these topics often go hand-in-hand, with researchers seeking to predict the future spread of alien species in climate change scenarios.

The research spectrum of articles dealing with spiders was more diversified, with studies using SDMs to explore the environmental drivers of species distribution (28.6%), to predict distributions under future climate change (33.4%), to assist species delimitation (26.8%), as well as other miscellaneous topics (19.6%). Contrary to ants, only 7.1% of studies on spiders dealt with invasion biology, probably on account of the reduced number of globally important known invasive spiders (Nentwig 2015).

Finally, the focus of articles dealing with ground beetles was almost entirely climate change (52.6%) and the drivers of species distribution (36.8%), with only two papers dealing with biological invasions.

Algorithms used

A large majority of articles used a single algorithm (for ants, ground beetles and spiders 79.6, 94.7, 80.4%, respectively) or ensemble (24.1, 5.3 and 14.3%, respectively) for modelling species distributions, whereas no silver bullet use was sporadic and only used in spider-related articles (7.1%). A total of 33 different algorithms were used in the studies we reviewed. For all taxonomic groups, MaxEnt was the most used algorithm (Figure 5), as also emphasized by the co-citation network (Figure 3). This is a recurrent pattern in the latest SDM research, as found for the research in other animal groups (e.g. bats; Razgour et al., 2016). This trend is probably due to the fact that MaxEnt is a presence-only technique, thus allowing users to overcome some of the difficulties associated with obtaining reliable absence data in the light of imperfect detection (e.g. Ward and Stanley, 2013). Moreover, MaxEnt has proved to be a robust species distribution modelling technique according to comparative studies [e.g. Elith et al., 2006 – a highly co-cited reference in our dataset as shown in Figure 3 (Phillips and Dudík 2008)].

Surprisingly, the simple technique of logistic regression (i.e. a Bernoulli generalized linear model) was sporadically used, even before MaxEnt release in 2004. This may be a true pattern but also an artefact resulting from our keyword search. In fact, the idea that logistic regressions based on distribution records are *de facto* an SDM is relatively recent (Peterson 2006, Elith and Leathwick 2009).

Variables used in the models and their importance in explaining distribution patterns

Bioclimatic variables were by far the most used predictors to model and explain species distributions (Table 2) for the three focal groups. The mean variable resolution was rather similar for ants and ground beetles [respectively 314.9 arc.sec (max = 1.4 arc.sec) and 414.7 arc.sec (max = 0.05 arc.sec)]. The mean resolution was higher for spiders (171.7 arc.sec (max = 1.4 arc.sec). This is partly due to broad availability of free high-resolution climatic variables [e.g. CHELSA (Karger et al., 2017), CliMond (Kriticos et al., 2012) and WorldClim 2 (Fick and Hijmans, 2017)], and partly reflects the true importance of climate as a limiting factor for species distribution (Muñoz and Bodensteiner 2019), especially climatic extremes (Román-Palacios and Wiens 2020). Climatic variables, in fact, were systematically selected as important in virtually all analysed studies (Table 2).

Topography, soil and land use, and habitat variables are used less often, possibly due to greater limitations in their availability (the mean resolution of these predictors in the data set is c. 4.6 km at the equator, much higher than that of bioclimatic variables). Nevertheless, when used, these non-climatic factors were often selected as important in modelling the distribution (>65% for ants and >80% for spiders and ground beetles, table 2), suggesting that fine-scale habitat structure plays a critical role for the ecology of terrestrial arthropods .

The integration of SDM use with species functional traits and ecophysiological data was scarce. For ants, 10.2% of articles used traits and 6.2% ecophysiological data. For spiders, 2.4% of articles used functional traits and none ecophysiological data. For ground beetles, no articles used functional traits or ecophysiological data. In the few instances where similar variables were considered, these were not directly incorporated as predictors in the model but rather discussed in comparison with the modelled distribution. For the three groups, between 20 and 25% of papers used phylogenies, but as previously they were not incorporated into the models.

CHALLENGES FOR SDM RESEARCH IN TERRESTRIAL INVERTEBRATES

Taxonomic and geographical biases

While SDM studies based on comprehensive samples of vertebrate species are becoming routine (e.g. Liu et al., 2020; Thuiller et al., 2019), our survey emphasizes how just a small fraction of terrestrial arthropod predators have been subjected to the attention of modellers. The total volume of SDM articles was higher for vertebrates than for arthropods, and this difference would be even greater if these numbers are relativized to the total number of known vertebrate and arthropod species. This is a typical pattern that is partly explained by the fact that there is more available information on vertebrates (e.g. distribution data; Troudet et al., 2017) and partly the result of a cognitive bias in terms of researcher' subjective preferences for certain taxa over others (Clark and May 2002) – what has been termed by entomologists 'institutional vertebratism' or 'taxonomic chauvinism' (Leather 2009a, b). The few available studies on arthropods are drops in the ocean when considering the number of described and as yet undescribed species of insects (Stork 2018) and spiders (Agnarsson et al. 2013). However, taxonomic bias towards certain groups exists also among articles dedicated to arthropods (e.g. Cardoso, 2012; Leandro et al., 2017). For example, butterflies are among the most studied in SDM studies, which once again may be due to a greater availability of information (Thomas 2005, van Swaay et al. 2008, Brereton et al. 2011), and which in turn might be driven by aesthetic characteristics. Other well-studied groups are those relevant from an economic point of view, such as vectors of diseases (Diptera, 8.9%), crop pests (other beetles, 6.6%) and pollinators (Apoidea, 3.2%).

As for our focal groups, we found that despite spiders and ground beetles outnumbering ants in terms of described species, the number of species studied was considerably higher for ants. This may be linked to the topic of articles, with most papers focusing on one of the numerous invasive ant species – it is likely that a few globally relevant invasive ant species (e.g. Argentine ant, fire ant) allow myrmecologists to obtain research funding, thus attracting most research attention

(Holway et al. 2002, Silverman and Brightwell 2008).

Inevitably, the few studies on ants, ground beetles and spiders have often been opportunistic, largely reflecting the specific interests of the few authors who have ventured to explore the potential of SDMs in terrestrial arthropod research (Figure 4). For example, this is evident when looking at a sample of papers on spiders – most studies focused on large-sized, taxonomically unique, and/or charismatic species (e.g. Decae et al., 2019; Jiménez−Valverde et al., 2011; Wang et al., 2018), taxa of medical importance (Planas et al. 2014, Wang et al. 2018, Taucare-Ríos et al. 2018) or taxa inhabiting peculiar habitats that are the interest of certain authors, like caves (Mammola et al. 2018, 2019, Pavlek and Mammola 2020).

The paucity of multi-taxa studies also suggests a general lack of data for less common species. In fact, and this is true also for vertebrates, most SDM applications are set in geographical areas that hold lower levels of diversity but greater availability of data (Europe and North America; Figure 2). The scarcity of data has been pointed out as one of the key limitations to our understanding of the drivers of biodiversity change in invertebrates (Cardoso and Leather 2019), as summarized in eight so-called ‘biodiversity shortfalls’ (Cardoso et al. 2011, Hortal et al. 2015, Ficetola et al. 2019). SDMs may help us to combat some of these impediments by identifying unexplored regions of high environmental suitability for improving the geographical gaps in species distributions (i.e. tackling the Wallacean shortfall), by identifying the environmental drivers of these distributions (Hutchinsonian shortfall), and even by suggesting suitable sites for further sampling (Linnean shortfall). However, the SDM construction in itself requires robust and high-quality distribution data, creating a loop that is difficult to break.

Solutions to alleviate data limitations

Distribution data

A quick search for any bird species in the Global Biodiversity Information Facility (www.gbif.org) reminds us that it is unlikely we will ever possess for arthropods the same amount and quality of data available for vertebrates. Arthropods are simply too ubiquitous in space and time (Stork 2018), while natural scientists are simply too few (Tewksbury et al. 2014). However, some recent technical and technological advances may help us to overcome some impediments related to data limitation and getting close to the goal of modelling the distribution of arthropods with more confidence.

Foremost, there have been recent technical advances in SDMs that may be significant when it comes to modelling the distribution of poorly known taxa. The emergence of ensemble of small models has proved promising to optimize the modelling of species for which few occurrences are available; this is achieved by combining a set of small bivariate models to create a consensus model that avoids overfitting (details in Breiner et al., 2018, 2015).

Second, modelling above the species level (Smith et al. 2019), for example by integrating data from related species when their niche overlap is large (Qiao et al. 2017), may be a useful shortcut to overcome a lack of distribution data in many circumstances.

Furthermore, the information age is characterized by the emergence of a myriad of types of digital data (summarized in Jarić et al., 2020) that may help to fill distribution data gaps, especially for easy-to-identify species. It was shown that photo-sharing platforms and smartphone applications such as iNaturalist are valuable sources of species occurrences (e.g. Unger et al., 2020), even in the case of our focal groups (Wang et al. 2018, Jiménez−Valverde et al. 2019). Citizen science programmes are also a valuable source of distribution data, for example about swarms of ants (Hart et al. 2018a) or common species of spiders (Hart et al. 2018b). Recent modelling exercises based on similar alternative data sources have demonstrated their utility in obtaining realistic representations of niches and distributions for easy-to-identify arthropods (Wang et al. 2018, Peña-Aguilera et al. 2019).

Finally, the recent advances in metabarcoding and especially environmental DNA is of major interest to overcome the issue of species detectability (Muha et al. 2017) and lack of invertebrate taxonomists (Hebert and Gregory 2005). Metabarcoding consists in identifying species using small DNA sequences that are highly variable between species and weakly variable within a given species. It is the basis of the environmental DNA approach which consists in the identification of the species present in a given environment using the DNA left by individuals. Despite the many technical challenges, environmental DNA and metabarcoding face becoming standard survey tools (Deiner et al., 2017), including in our focal groups (Toju and Baba 2018, Piper et al. 2019, Kennedy et al. 2020). Their ability to provide reliable absence data and to produce a massive amount of presence data is predicted to improve the efficiency of SDMs in the near future (Muha et al. 2017). Recently, for example, the use of environmental DNA has proved useful to forecast the spread of invasive species (Zhang et al. 2020) or to monitor reintroduction programme success (Riaz et al. 2020). Large-scale projects including metabarcoding of terrestrial arthropod communities [e.g. LIFEPLAN (https://www.helsinki.fi/en/projects/lifeplan) and the Insect Biome Atlas (https://www.insectbiomeatlas.com)] are currently taking place and will provide an unprecedented data baseline for SDMs. This will likely trigger the parallel development of tools to handle the big data era (Hallgren et al. 2016).

Lack of micro-scale environmental predictors

The second impediment that limits our ability to model the distribution of terrestrial arthropods is the scarcity of relevant environmental predictors at the correct spatial resolution. Most environmental rasters used today for SDMs [e.g. CHELSA (Karger et al., 2017) and WorldClim 2 (Fick and Hijmans, 2017)] achieve a maximum resolution of 30 arcsec (cell size c. 1 km^2 at the equator), which is excellent but might not be enough in the case of invertebrates that are known to respond to microclimatic characteristics over spatial scales of millimetres to metres (Potter et al.

2013, Suggitt et al. 2018). On a positive note, gigantic leaps forward are being made in the development of microclimatic databases (e.g., Kearney et al., 2014), as well as approaches for downscaling temperature data at high resolutions from thermal images (Senior et al., 2019) or airborne light detection and ranging data (George et al. 2015). It is predicted that in the following years, the use of remote sensing derived data will become the standard for modelling and mapping the microclimate (Zellweger et al. 2019), especially in invertebrate research where the use of similar high-resolution data has already proved useful to achieve realistic conservation prioritization (e.g. Bombi et al., 2019).

Furthermore, our literature survey emphasizes that habitat, soil and other land use variables, although rarely used, are key features affecting the distribution of invertebrates. However, these are only rarely considered. As in the case of climatic variables, there is a general paucity of raster data at a meaningful spatial resolution. Once again, statistical downscaling and remote sensing may come to help; for example, the use of high-resolution habitat variables has proved fundamental in modelling the distribution of spiders in spatially complex alpine rocky lands (Mammola et al. 2019).

Testing new algorithms

Despite the large number of algorithms tested (33), even per article (up to 14), it remains unclear if one algorithm rather than another is more suitable for modelling the distribution of terrestrial invertebrates. At the moment, MaxEnt hegemony in SDM research is obvious in our data set. More empirical comparisons of the performance algorithms in the context of terrestrial invertebrates research would be needed (Qiao et al., 2015; Araujo et al., 2019).

OPPORTUNITIES FOR SDM RESEARCH IN TERRESTRIAL INVERTEBRATES

SDMs are often used as a simple, correlative way to estimate species ranges based on the realized niche, having large uncertainties and often over- or underfitting the real distribution. In an influential paper published 15 years ago, it was foreseen that SDMs may offer ‘more than simple habitat models’ (Guisan and Thuiller 2005), for example by tackling biotic interactions, migration processes, dispersal limitations, and (meta)population dynamics.

The challenges faced by conservation biologists today call for the development of more of these process-based models (or mechanistic models), providing causal explanations for the observed patterns (e.g. Briscoe et al., 2019). These can be defined as any model that mechanistically links model predictions and species fitness, measured either directly or indirectly using functional traits or environmental and biological (e.g. competing or mutualistic species) interactions (Kearney 2006). This idea was reinforced by a seminal paper by Kearney and Porter (2009) calling for explicitly integrating physiological data in mechanistic niche modelling, but also life history traits (including dispersal abilities, fitness, eco-physiological tolerances) and biotic interactions (competition, parasitism).

There are proportionally more such studies for plants and marine invertebrates (see e.g. Chardon et al., 2020; Webb et al., 2020) than animals, because large spatial data sets needed for integrating physiological trait variation are available (Chown and Gaston 2016). While all these applications are still rare when it comes to terrestrial arthropods (see Maino et al. 2016), recently there have been studies that have successfully addressed biotic interaction (Mammola and Isaia 2017), dispersal limitations (Monsimet et al. 2020), and metapopulations (Giezendanner et al. 2020), thereby showing promising directions for future research. Studies including probability of survival to different stresses such as cold (Cuddington et al. 2018) or desiccation (Barton et al. 2019) were also performed for particularly well-known groups like lepidopterans and pests. However, whereas mechanistic models are increasingly available, they have high data demands and thus cannot be routinely used for invertebrates (Viterbi et al. 2020), especially in terrestrial arthropods where, as previously discussed, the scarcity of data on natural history and the large number of species are a clear challenge. Some ideas towards a more mechanistic understanding of arthropod distributions are discussed in the following.

Integration of species attributes and traits in SDMs

Species traits influence the outcome of SDMs in two ways. First, they themselves influence the distribution of species. Either in the present, past or future, the ability of species to adapt to certain conditions, their history, their relation with other species or their ability to disperse, all influence species distribution and its change in time. Second, their traits may influence how complete or biased the known distribution data are and hence how adequate the modelled distributions for the different purposes are. Taking into account trait data before, during and after SDMs is therefore crucial for correct interpretation and to be aware of possible limitations.

The recent upsurge in open source trait databases and projects [ants (Parr et al., 2017), ground beetles (Homburg et al., 2014) and spiders (Lowe et al., 2020)] offers an unprecedented data baseline to integrate trait variability in modeling exercises and develop mechanistic descriptions of species distributions and their changes through time. Accordingly, the integration of correlative distribution analyses and functional approaches has recently been advocated (Mammola et al., 2019; Thuiller et al., 2009; Wittmann et al., 2016), as it would make it possible to bridge the differences in biogeography and functional ecology and move towards the novel field of 'functional biogeography' (Violle et al., 2014).

Accounting for trait variability

There are various ways to link correlative SDMs and traits (Kearney and Porter 2009). The most obvious one is a simple comparison between model outputs and trait variability, including the formulation of hypotheses about why these may concur or not. Example in invertebrates is the positive relationship between predicted habitat suitability and body size found in spiders (Mammola et al. 2019), phenotype−environment associations observed in butterflies (Zaman et al. 2019), or the use of thermal physiology tests to define thermal safe zones in ants (Coulin et al. 2019).

With the aim of obtaining more meaningful and realistic prediction of biodiversity change, recently new modelling approaches that directly incorporate phenotypic plasticity and other functional traits into correlative modelling are being scrutinized (e.g., AdaptR; Bush et al., 2016; ΔTraitSDM; Garzón et al., 2019). Following these examples, which are respectively based on dipterans and plants, and considering the recent increase in availability of traits for invertebrates, it is possible to predict a vast potential for developing trait-based SDMs in invertebrates.

Linking genetic data and distributions

SDMs have been criticized, among other things, for not taking into account heterogeneity in the genetic structure of populations within the species range (Hampe and Petit 2005, Smith et al. 2019). Indeed, SDMs generally assume uniformity of responses to climate but local adaptations and intraspecific variations have been documented (e.g. Franken et al., 2018; Hereford, 2009). Several recent studies have demonstrated that genetically informed SDMs improve climate change predictions because they incorporate possible local adaptations (Marcer et al. 2016, Ikeda et al. 2017). Instead of building SDMs based on species occurrence defined using standard taxonomy, one can model the distribution of each genetic unit of the population. The identification of these units can be achieved using traditional molecular markers such as amplified fragment-length polymorphisms, micro-satellites, and even Single Nucleotide polymorphisms (see below). For example, in their study, Marcer et al. (2016) built SDMs for each haplotype *Arabidopsis thaliana* (Brassicacae) and found that even though most haplotypes distribution ranges will shrank with global climate change, two of them will expand. Some authors also advocate the use of genetic data because it allows production of real absence data (absence of a given genetic cluster), making it possible to fit logistic regressions and incorporation of endogenous spatial autocorrelation (Gotelli and Colwell 2011). The recent advances in high-throughput sequencing techniques allow ecologists to collect single nucleotide polymorphism data (Peterson et al., 2012) for cluster identification at reasonable costs. Single nucleotide polymorphism provides fine-scale resolution of population genetic structure, which can then be incorporated into SDMs. To our knowledge this has rarely been done on animal populations (but see Razgour et al., 2018) and has never been done on terrestrial arthropod species.

Accounting for dispersal

Using a correlative approach makes the inclusion of complex processes like dispersal more difficult. While the inclusion of dispersal can improve model fit (Dormann 2007), dispersal processes are rarely accounted for in the studies on arthropods. The only such studies either considered dispersal via the calibration area (Anderson and Raza 2010, Barve et al. 2011), by including a buffer of reachable areas around presences based on species-specific dispersal abilities [e.g. long-distance dispersal via ballooning for spiders (Mammola and Isaia 2017)], or relied on more sophisticated approaches based on kernel distribution [e.g. model of butterfly accounting for both demography and dispersal via a kernel distribution (Singer et al. 2018)].

In general, these are rough estimations, given that dispersal is a complex phenomenon that is not trivial to integrate into SDMs (Thuiller et al. 2013). Indeed, dispersal is characterized by three phases (Clobert et al. 2009), i.e. departure, transfer and settlement, that the model should theoretically account for. Moreover, several factors can influence each of these phases (e.g. inbreeding, temperature development, body condition or starvation; on spiders see Bonte et al., 2009, 2008b, 2008a), often acting synergistically.

Different methods, with varying complexity levels, have been developed to integrate dispersal into SDMs. The use of mechanistic models, which is highly data demanding, is not easily feasible for mega-diverse arthropod groups. Moreover, while models that include dispersal often outperform simpler models, an increase of the model complexity usually increases the uncertainty of the prediction (Zurell et al. 2016). Integration of dispersal is thus still largely ignored in SDMs, not only in the case of arthropods, but even in most studies on vertebrates and plants. It should be considered, for example in conservation or invasive species studies, to improve predictions in time by predicting the range of potentially suitable habitat that can be reached.

CONCLUSIONS

Efforts to map the diversity of invertebrate life have been mostly concentrated in the last 10 years, emphasizing how more and more entomologists and other scientists are beginning to incorporate SDMs into their research. In the light of our ignorance about the diversity, distribution and life history of most arthropods, these versatile tools are proving useful to fill some major knowledge gaps regarding arthropod diversity. The importance of similar endeavours becomes apparent when considering the accumulating evidence about the silent extinctions of invertebrates (e.g. Cardoso et al., 2020; Eisenhauer et al., 2019), the limited conservation efforts that are directed towards them (e.g. Cardoso, 2012; Mammides, 2019; Mammola et al., 2020), and the calls for solutions to these problems (Harvey et al. 2020, Samways et al. 2020).

Apart from the conservation implications of using SDMs to map arthropod diversity, we have shown how terrestrial arthropods may provide opportunities for advancing SDM research. Given that terrestrial arthropod distributions are strongly influenced by microclimatic conditions and microhabitat structure, they represent ideal candidates for testing novel modelling approaches. So far this potential is still largely unexploited and thus we have discussed some recent avenues of research where the integration of different data sources may lead to mechanistic descriptions of key processes associated with species distributions. We are certain that our suggestions are a drop in the ocean when compared with what is currently available in terms of modelling possibilities –

methodological advances in SDM-related theory are so quick that often it is difficult to keep pace. As brand new solutions to describe patterns and processes associated with species distribution are becoming available, we hope that this review will succeed in highlighting the potential of arthropods in SDM research and, in the future, that we will more often see them involved as protagonists in these developments.

ACKNOWLEDGEMENTS

We are grateful to Cathryn Primrose-Mathisen for proof-reading our English. Dr. Tim Bonebrake provided useful suggestions on an earlier version. SM acknowledges support from the European Commission through Horizon 2020 Marie Skłodowska-Curie Actions (MSCA) individual fellowships (Grant no. 882221). JP was supported by 'BOOST ERC' OPALE and by the SAD ‘PEPPS’ (Région Bretagne). DL was supported by the SAD ‘PEPPS’ (Région Bretagne)

SUPPLEMENTARY MATERIALS

Appendix S1. Queries on the Web of Science.

Appendix S2. List of information collected for ant, ground beetle and spider papers selected.

Figures S1–S6. Spatial distributions of studied species.

DATA AVAILABILITY

The complete dataset and R code used for the analyses is available on Gitlab (https://gitlab.com/DenisLafage/sdm_review).

LITERATURE CITED

Agnarsson, I. et al. 2013. Systematics: progress in the study of spider diversity and evolution. - Spider Res. 21st Century Trends Perspect. in press.

Anderson, R. P. and Raza, A. 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela. - J. Biogeogr. 37: 1378–1393.

Angelov, B. 2019. Review of species distribution modeling open-source software.: 5. - Res. Prepr. in press.

Araújo, M. B. et al. 2019. Standards for distribution models in biodiversity assessments. - Sci. Adv. 5: eaat4858.

Aria, M. and Cuccurullo, C. 2017. bibliometrix: An R-tool for comprehensive science mapping analysis. - J. Informetr. 11: 959–975.

Barnosky, A. D. et al. 2011. Has the Earth's sixth mass extinction already arrived? - Nature 471: 51–57.

Barton, M. G. et al. 2019. Incorporating temperature and precipitation extremes into process-based models of African lepidoptera changes the predicted distribution under climate change. - Ecol. Model. 394: 53–65.

Barve, N. et al. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. - Ecol. Model. 222: 1810–1819.

Batagelj, V. and Cerinšek, M. 2013. On bibliographic networks. - Scientometrics 96: 845–864.

Bombi, P. et al. 2019. Identifying priority sites for insect conservation in forest ecosystems at high resolution: the potential of LiDAR data. - J. Insect Conserv. 23: 689–698.

Bonte, D. et al. 2008a. Thermal conditions during juvenile development affect adult dispersal in a spider. - Proc. Natl. Acad. Sci. 105: 17000–17005.

Bonte, D. et al. 2008b. Starvation affects pre-dispersal behaviour of Erigone spiders. - Basic Appl. Ecol. 9: 308–315.

Bonte, D. et al. 2009. Repeatability of dispersal behaviour in a common dwarf spider: evidence for different mechanisms behind short- and long-distance dispersal. - Ecol. Entomol. 34: 271–276.

Booth, T. H. 2018. Why understanding the pioneering and continuing contributions of BIOCLIM to species distribution modelling is important. - Austral Ecol. 43: 852–860.

Branco, V. et al. 2019. Species conservation profiles of spiders (Araneae) endemic to mainland Portugal. - Biodivers. Data J. 7: e39315.

Breiner, F. T. et al. 2015. Overcoming limitations of modelling rare species by using ensembles of small models. - Methods Ecol. Evol. 6: 1210–1218.

Breiner, F. T. et al. 2018. Optimizing ensembles of small models for predicting the distribution of species with few occurrences. - Methods Ecol. Evol. 9: 802–808.

Brereton, T. M. et al. 2011. Developing and launching a wider countryside butterfly survey across the United Kingdom. - J. Insect Conserv. 15: 279–290.

Briscoe, N. J. et al. 2019. Forecasting species range dynamics with process-explicit models: matching methods to applications. - Ecol. Lett. 22: 1940–1956.

Burns, P. et al. 2020. Incorporating canopy structure from simulated GEDI lidar into bird species distribution models. - Environ. Res. Lett. in press.

Bush, A. et al. 2016. Incorporating evolutionary adaptation in species distribution modelling reduces projected vulnerability to climate change. - Ecol. Lett. 19: 1468–1478.

Cardoso, P. 2012. Habitats Directive species lists: urgent need of revision. - Insect Conserv. Divers. 5: 169–174.

Cardoso, P. and Leather, S. R. 2019. Predicting a global insect apocalypse. - Insect Conserv. Divers. 12: 263–267.

Cardoso, P. et al. 2011. The seven impediments in invertebrate conservation and how to overcome them. - Biol. Conserv. 144: 2647–2655.

Cardoso, P. et al. 2020a. Automated discovery of relationships, models and principles in ecology. - Front. Ecol. Evol. 8:530135: 027839.

Cardoso, P. et al. 2020b. Scientists' warning to humanity on insect extinctions. - Biol. Conserv. 242: 108426.

Chardon, N. I. et al. 2020. Incorporating intraspecific variation into species distribution models improves distribution predictions, but cannot predict species traits for a wide-spread plant species. - Ecography 43: 60–74.

Chen, I.-C. et al. 2011. Rapid range shifts of species associated with high levels of climate warming. - Science 333: 1024–1026.

Chown, S. L. and Gaston, K. J. 2016. Macrophysiology–progress and prospects. - Funct. Ecol. 30: 330–344.

Christman, M. C. et al. 2016. Predicting the Occurrence of Cave-Inhabiting Fauna Based on Features of the Earth Surface Environment. - PLoS ONE in press.

Clark, J. A. and May, R. M. 2002. Taxonomic Bias in Conservation Research. - Science 297: 191–192.

Clobert, J. et al. 2009. Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. - Ecol. Lett. 12: 197–209.

Colinet, H. et al. 2015. Insects in Fluctuating Thermal Environments. - Annu. Rev. Entomol. 60: 123–140.

Coulin, C. et al. 2019. Linking thermo-tolerances of the highly invasive ant, Wasmannia auropunctata, to its current and potential distribution. - Biol. Invasions 21: 3491–3504.

Cuddington, K. et al. 2018. Probability of emerald ash borer impact for Canadian cities and North America: a mechanistic model. - Biol. Invasions 20: 2661–2677.

Decae, A. et al. 2019. Systematics, ecology and distribution of the mygalomorph spider genus Cteniza Latreille, 1829 (Araneae, Mygalomorphae, Ctenizidae). - Zootaxa 4550: 499–524.

Deiner, K. et al. 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. - Mol. Ecol. 26: 5872–5895.

Dormann, C. F. 2007. Promising the future? Global change projections of species distributions. - Basic Appl. Ecol. 8: 387–397.

Eisenhauer, N. et al. 2019. Recognizing the quiet extinction of invertebrates. - Nat. Commun. 10: 50.

Elith, J. and Leathwick, J. R. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. - Annu. Rev. Ecol. Evol. Syst. 40: 677–697.

Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. - Ecography 29: 129–151.

Elith, J. et al. 2011. A statistical explanation of MaxEnt for ecologists. - Divers. Distrib. 17: 43–57.

Ferretti, N. E. et al. 2019. An integrative approach unveils speciation within the threatened spider Calathotarsus simoni (Araneae: Mygalomorphae: Migidae). - Syst. Biodivers. 17: 439–457.

Ficetola, G. F. et al. 2019. The Racovitzan impediment and the hidden biodiversity of unexplored environments. - Conserv. Biol. 33: 214–216.

Fick, S. E. and Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. - Int. J. Climatol. 37: 4302–4315.

Franken, O. et al. 2018. Heated communities: large inter- and intraspecific variation in heat tolerance across trophic levels of a soil arthropod community. - Oecologia 186: 311–322.

Fukushima, C. S. et al. 2019. Species conservation profiles of tarantula spiders (Araneae, Theraphosidae) listed on CITES. - Biodivers. Data J. in press.

Garzón, M. B. et al. 2019. ΔTraitSDMs: species distribution models that account for local adaptation and phenotypic plasticity. - New Phytol. 222: 1757–1765.

George, A. D. et al. 2015. Using LiDAR and remote microclimate loggers to downscale near-surface air temperatures for site-level studies. - Remote Sens. Lett. 6: 924–932.

Giezendanner, J. et al. 2020. Earth and field observations underpin metapopulation dynamics in complex landscapes: Near-term study on carabids. - Proc. Natl. Acad. Sci. 117: 12877–12884.

Gotelli, N. J. and Colwell, R. K. 2011. Estimating species richness. - In: Magurran A.E., M., B. J. (ed), Frontiers in Measuring Biodiversity. pp. 39–54.

Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. - Ecol. Lett. 8: 993–1009.

Guisan, A. et al. 2013. Predicting species distributions for conservation decisions. - Ecol. Lett. 16: 1424–1435.

Guisan, A. et al. 2017. Habitat Suitability and Distribution Models: With Applications in R. - Cambridge University Press.

Hallgren, W. et al. 2016. The Biodiversity and Climate Change Virtual Laboratory: Where ecology meets big data. - Environ. Model. Softw. 76: 182–186.

Hamilton, C. A. et al. 2016. Taxonomic revision of the tarantula genus Aphonopelma Pocock, 1901 (Araneae, Mygalomorphae, Theraphosidae) within the United States. - ZooKeys: 1–340.

Hampe, A. and Petit, R. J. 2005. Conserving biodiversity under climate change: the rear edge matters. - Ecol. Lett. 8: 461–467.

Hart, H. G. et al. 2018a. The spatial distribution and environmental triggers of ant mating flights: using citizen-science data to reveal national patterns. - Ecography 41: 877–888.

Hart, H. G. et al. 2018b. Spatiotemporal Variation in House Spider Phenology at a National Scale Using Citizen Science. - Arachnology 17: 331–334.

Harvey, J. A. et al. 2020. International scientists formulate a roadmap for insect conservation and recovery. - Nat. Ecol. Evol. 4: 174–176.

Hebert, P. D. N. and Gregory, T. R. 2005. The promise of DNA barcoding for taxonomy. - Syst. Biol. 54: 852–859.

Hellmann, J. J. et al. 2008. Five Potential Consequences of Climate Change for Invasive Species. - Conserv. Biol. 22: 534–543.

Hereford, J. 2009. A Quantitative Survey of Local Adaptation and Fitness Trade−Offs. - Am. Nat. 173: 579–588.

Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. - Int. J. Climatol. 25: 1965–1978.

Holway, D. A. et al. 2002. The Causes and Consequences of Ant Invasions. - Annu. Rev. Ecol. Syst. 33: 181–233.

Homburg, K. et al. 2014. Carabids.org – a dynamic online database of ground beetle species traits (Coleoptera, Carabidae). - Insect Conserv. Divers. 7: 195–205.

Hortal, J. et al. 2015. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. - Annu. Rev. Ecol. Evol. Syst. 46: 523–549.

Hughes, M. et al. 2004. The Development of a GIS-based Inventory of Standing Waters in Great Britain together with a Risk-based Prioritisation Protocol. - Water Air Soil Pollut. Focus 4: 73–84.

Ikeda, D. H. et al. 2017. Genetically informed ecological niche models improve climate change predictions. - Glob. Change Biol. 23: 164–176.

Jarić, I. et al. 2020. iEcology: Harnessing Large Online Resources to Generate Ecological Insights. - Trends Ecol. Evol. 35: 630–639.

Jiménez-Valverde, A. et al. 2009. Species Distribution Models Do Not Account for Abundance: The Case of Arthropods on Terceira Island. - Ann. Zool. Fenn. 46: 451–464.

Jiménez−Valverde, A. et al. 2011. Environmental suitability of new reported localities of the funnelweb spider Macrothele calpeiana: an assessment using potential distribution modelling with presence-only techniques. - J. Biogeogr. 38: 1213–1223.

Jiménez−Valverde, A. et al. 2019. Photo−sharing platforms key for characterising niche and distribution in poorly studied taxa (S Leather and A Stewart, Eds.). - Insect Conserv. Divers. 12: 389–403.

Karger, D. N. et al. 2017. Climatologies at high resolution for the earth's land surface areas. - Sci. Data 4: 170122.

Kearney, M. 2006. Habitat, environment and niche: what are we modelling? - Oikos 115: 186–191.

Kearney, M. and Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. - Ecol. Lett. 12: 334–350.

Kearney, M. R. et al. 2014. microclim: Global estimates of hourly microclimate based on long-term monthly climate averages. - Sci. Data 1: 140006.

Kennedy, S. R. et al. 2020. High-throughput sequencing for community analysis: the promise of DNA barcoding to uncover diversity, relatedness, abundances and interactions in spider communities. - Dev. Genes Evol. in press.

Konno, K. et al. 2020. Ignoring non-English-language studies may bias ecological meta-analyses. - Ecol. Evol. in press.

Kriticos, D. J. et al. 2012. CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. - Methods Ecol. Evol. 3: 53–64.

Krosby, M. et al. 2015. Climate-induced range overlap among closely related species. - Nat. Clim. Change 5: 883–886.

Leandro, C. et al. 2017. Bias and perspectives in insect conservation: A European scale analysis. - Biol. Conserv. 215: 213–224.

Leather, S. R. 2009a. Taxonomic chauvinism threatens the future of entomology. - Biologist 56: 10–13.

Leather, S. R. 2009b. Institutional vertebratism threatens UK food security. - Trends Ecol. Evol. 24: 413–414.

Lenoir, J. et al. 2020. Species better track climate warming in the oceans than on land. - Nat. Ecol. Evol.: 1–16.

Liu, X. et al. 2020. Animal invaders threaten protected areas worldwide. - Nat. Commun. 11: 2892.

Lobo, J. M. et al. 2010. The uncertain nature of absences and their importance in species distribution modelling. - Ecography 33: 103–114.

Lowe, E. J. et al. 2020. Towards establishment of a centralized spider traits database. - J. Arachnol. in press.

Magalhaes, I. L. F. et al. 2014. Strong spatial structure, P liocene diversification and cryptic diversity in the N eotropical dry forest spider S icarius cariri. - Mol. Ecol. 23: 5323–5336.

Maino, J. L. et al. 2016. Mechanistic models for predicting insect responses to climate change. - Curr. Opin. Insect Sci. 17: 81–86.

Mammides, C. 2019. European Union's conservation efforts are taxonomically biased. - Biodivers. Conserv. 28: 1291–1296.

Mammola, S. and Isaia, M. 2017. Rapid poleward distributional shifts in the European cave-dwelling Meta spiders under the influence of competition dynamics. - J. Biogeogr. 44: 2789–2797.

Mammola, S. and Leroy, B. 2018. Applying species distribution models to caves and other subterranean habitats. - Ecography 41: 1194–1208.

Mammola, S. et al. 2015. Alpine endemic spiders shed light on the origin and evolution of subterranean species. - PeerJ 3: e1384.

Mammola, S. et al. 2018. Climate change may drive cave spiders to extinction. - Ecography 41: 233–243.

Mammola, S. et al. 2019. Associations between habitat quality, body size and reproductive fitness in the alpine endemic spider Vesubia jugorum. - Glob. Ecol. Biogeogr. 28: 1325–1335.

Mammola, S. et al. 2020. Towards a taxonomically unbiased EU Biodiversity Strategy for 2030. - Proc. R. Soc. B Biol. Sci. in press.

Marcer, A. et al. 2016. Tackling intraspecific genetic structure in distribution models better reflects species geographical range. - Ecol. Evol. 6: 2084–2097.

Melo-Merino, S. M. et al. 2020. Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. - Ecol. Model. 415: 108837.

Monsimet, J. et al. 2020. Explicit integration of dispersal-related metrics improves predictions of SDM in predatory arthropods. - Sci. Rep. in press.

Muha, T. P. et al. 2017. Using Environmental DNA to Improve Species Distribution Models for Freshwater Invaders. - Front. Ecol. Evol. in press.

Muñoz, M. M. and Bodensteiner, B. L. 2019. Janzen's Hypothesis Meets the Bogert Effect: Connecting Climate Variation, Thermoregulatory Behavior, and Rates of Physiological Evolution. - Integr. Org. Biol. in press.

Nentwig, W. 2015. Introduction, establishment rate, pathways and impact of spiders alien to Europe. - Biol. Invasions 17: 2757–2778.

Parr, C. L. et al. 2017. GlobalAnts: a new database on the geography of ant traits (Hymenoptera: Formicidae). - Insect Conserv. Divers. 10: 5–20.

Pavlek, M. and Mammola, S. 2020. Niche-based processes explaining the distributions of closely related subterranean spiders. - J. Biogeogr. in press.

Peña-Aguilera, P. et al. 2019. Niche segregation in Iberian Argiope species. - J. Arachnol. 47: 37–44.

Peterson, A. T. 2003. Predicting the Geography of Species' Invasions via Ecological Niche Modeling. - Q. Rev. Biol. 78: 419–433.

Peterson, A. T. 2006. Uses and requirements of ecological niche models and related distributional models. in press.

Peterson, A. T. 2009. Perspective: Phylogeography is not enough: The need for multiple lines of evidence. - Front. Biogeogr. in press.

Peterson, A. T. and Nakazawa, Y. 2008. Environmental data sets matter in ecological niche modelling: an example with Solenopsis invicta and Solenopsis richteri. - Glob. Ecol. Biogeogr. 17: 135–144.

Peterson, A. T. and Soberón, J. 2012. Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right. - Nat. Conserv. 10: 102–107.

Phillips, S. J. and Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. - Ecography 31: 161–175.

Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. - Ecol. Model. 190: 231–259.

Pincebourde, S. and Woods, H. A. 2020. There is plenty of room at the bottom: microclimates drive insect vulnerability to climate change. - Curr. Opin. Insect Sci. in press.

Piper, A. M. et al. 2019. Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. - GigaScience in press.

Planas, E. et al. 2014. Ecological niche and phylogeography elucidate complex biogeographic patterns in Loxosceles rufescens (Araneae, Sicariidae) in the Mediterranean Basin. - BMC Evol. Biol. 14: 195.

Potter, K. A. et al. 2013. Microclimatic challenges in global change biology. - Glob. Change Biol. 19: 2932–2939.

Pozsgai, G. et al. 2020. A comparative analysis reveals irreproducibility in searches of scientific literature.

Qiao, H. et al. 2015. No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. - Methods Ecol. Evol. 6:

1126–1136.

Qiao, H. et al. 2017. Using data from related species to overcome spatial sampling bias and associated limitations in ecological niche modelling. - Methods Ecol. Evol. 8: 1804–1812.

Razgour, O. et al. 2016. Painting maps with bats: species distribution modelling in bat research and conservation. - Hystrix in press.

Razgour, O. et al. 2018. An integrated framework to identify wildlife populations under threat from climate change. - Mol. Ecol. Resour. 18: 18–31.

Rezende, E. L. and Bozinovic, F. 2019. Thermal performance across levels of biological organization. - Philos. Trans. R. Soc. B Biol. Sci. 374: 20180549.

Riaz, M. et al. 2020. Combining environmental DNA and species distribution modeling to evaluate reintroduction success of a freshwater fish. - Ecol. Appl. 30: e02034.

Ripple, W. J. et al. 2020. World Scientists' Warning of a Climate Emergency. - BioScience 70: 8–12.

Robinson, L. M. et al. 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. - Glob. Ecol. Biogeogr. 20: 789–802.

Robinson, N. M. et al. 2017. A Systematic Review of Marine-Based Species Distribution Models (SDMs) with Recommendations for Best Practice. - Front. Mar. Sci. in press.

Rödder, D. et al. 2010. Molecules meet macroecology—combining Species Distribution Models and phylogeographic studies. - Zootaxa 2426: 54–60.

Román-Palacios, C. and Wiens, J. J. 2020. Recent responses to climate change reveal the drivers of species extinction and survival. - Proc. Natl. Acad. Sci. 117: 4211–4217.

Ross, K. G. et al. 2010. Species Delimitation: A Case Study in a Problematic Ant Taxon. - Syst. Biol. 59: 162–184.

Roura-Pascual, N. et al. 2004. Geographical potential of Argentine ants (Linepithema humile Mayr) in the face of global climate change. - Proc. R. Soc. Lond. B Biol. Sci. 271: 2527–2535.

Roura-Pascual, N. et al. 2006. Niche Differentiation and Fine-Scale Projections for Argentine Ants Based on RemoTELy Sensed Data. - Ecol. Appl. 16: 1832–1841.

Roura-Pascual, N. et al. 2009. Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. - Biol. Invasions 11: 1017–1031.

Ryo, M. et al. 2020. Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models. - Ecography in press.

Samways, M. J. et al. 2020. Solutions for humanity on how to conserve insects. - Biol. Conserv. 242: 108427.

Santini, L. et al. 2020. Assessing the reliability of species distribution projections in climate change research. - bioRxiv: 2020.06.10.143917.

Schröder, B. 2008. Challenges of species distribution modeling belowground. - J. Plant Nutr. Soil Sci. 171: 325–337.

Senior, R. A. et al. 2019. ThermStats: An R package for quantifying surface thermal heterogeneity in assessments of microclimates. - Methods Ecol. Evol. 10: 1606–1614.

Seppälä, S. et al. 2018a. Species conservation profiles of a random sample of world spiders II: Gnaphosidae to Nemesiidae. - Biodivers. Data J. in press.

Seppälä, S. et al. 2018b. Species conservation profiles of a random sample of world spiders I: Agelenidae to Filistatidae. - Biodivers. Data J. in press.

Seppälä, S. et al. 2018c. Species conservation profiles of a random sample of world spiders III: Oecobiidae to Salticidae. - Biodivers. Data J. in press.

Seppälä, S. et al. 2018d. Species conservation profiles of a random sample of world spiders IV: Scytodidae to Zoropsidae. - Biodivers. Data J. in press.

Sillero, N. 2011. What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. - Ecol. Model. 222: 1343–1346.

Silverman, J. and Brightwell, R. J. 2008. The Argentine Ant: Challenges in Managing an Invasive Unicolonial Pest. - Annu. Rev. Entomol. 53: 231–252.

Singer, A. et al. 2018. Constructing a hybrid species distribution model from standard large-scale distribution data. - Ecol. Model. 373: 39–52.

Smith, A. B. et al. 2019. Niche Estimation Above and Below the Species Level. - Trends Ecol. Evol. 34: 260–273.

Solomon, S. E. et al. 2008. Paleodistributions and comparative molecular phylogeography of leafcutter ants (Atta spp.) provide new insight into the origins of Amazonian diversity. - PloS One 3: e2738.

Staunton, K. M. et al. 2014. Projected distributions and diversity of flightless ground beetles within the Australian wet tropics and their environmental correlates (R Pradel, Ed.). - PLoS ONE 9: e88635.

Steiner, F. M. et al. 2008. Combined modelling of distribution and niche in invasion biology: a case study of two invasive Tetramorium ant species. - Divers. Distrib. 14: 538–545.

Stork, N. E. 2018. How Many Species of Insects and Other Terrestrial Arthropods Are There on Earth? - Annu. Rev. Entomol. 63: 31–45.

Suggitt, A. J. et al. 2018. Extinction risk from climate change is reduced by microclimatic buffering. - Nat. Clim. Change 8: 713–717.

Taucare-Ríos, A. et al. 2018. Matching global and regional distribution models of the recluse spider Loxosceles rufescens: to what extent do these reflect niche conservatism? - Med. Vet. Entomol. 32: 490–496.

Tewksbury, J. J. et al. 2014. Natural History’s Place in Science and Society. - BioScience 64: 300–310.

Thomas, J. A. 2005. Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. - Philos. Trans. R. Soc. B Biol. Sci. 360: 339–357.

Thomas, C. D. 2010. Climate, climate change and range boundaries. - Divers. Distrib. 16: 488–495.

Thuiller, W. et al. 2013. A road map for integrating eco-evolutionary processes into biodiversity models. - Ecol. Lett. 16: 94–105.

Thuiller, W. et al. 2019. Uncertainty in ensembles of global biodiversity scenarios. - Nat. Commun. 10: 1446.

Titley, M. A. et al. 2017. Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. - PLOS ONE 12: e0189577.

Toju, H. and Baba, Y. G. 2018. DNA metabarcoding of spiders, insects, and springtails for exploring potential linkage between above- and below-ground food webs. - Zool. Lett. 4: 4.

Unger, S. et al. 2020. iNaturalist as an engaging tool for identifying organisms in outdoor activities. - J. Biol. Educ. 0: 1–11.

van Swaay, C. A. M. et al. 2008. Butterfly monitoring in Europe: methods, applications and perspectives. - Biodivers. Conserv. 17: 3455–3469.

Viterbi, R. et al. 2020. Effects of Temperature Rise on Multi-Taxa Distributions in Mountain Ecosystems. - Diversity 12: 210.

Walther, G.-R. et al. 2002. Ecological responses to recent climate change. - Nature 416: 389–395.

Wang, Y. et al. 2018. Predicting the distribution of poorly-documented species, Northern black widow (Latrodectus variolus) and Black purse-web spider (Sphodros niger), using museum specimens and citizen science data (M Kuntner, Ed.). - PLOS ONE 13: e0201094.

Warren, D. L. 2012. In defense of ‘niche modeling.’ - Trends Ecol. Evol. 27: 497–500.

Webb, T. J. et al. 2020. Occupancy-derived thermal affinities reflect known physiological thermal limits of marine species. - Ecol. Evol. in press.

Wüest, R. O. et al. 2020. Macroecology in the age of Big Data–Where to go from here? - J. Biogeogr. 47: 1–12.

Yates, K. L. et al. 2018. Outstanding Challenges in the Transferability of Ecological Models. -

Trends Ecol. Evol. 33: 790–802.

Zaman, K. et al. 2019. Testing the role of ecological selection on colour pattern variation in the butterfly Parnassius clodius. - Mol. Ecol. 28: 5086–5102.

Zellweger, F. et al. 2019. Advances in Microclimate Ecology Arising from Remote Sensing. - Trends Ecol. Evol. 34: 327–341.

Zhang, J. 2017. Biodiversity science and macroecology in the era of big data. - Biodivers. Sci. 25: 355.

Zhang, Z. et al. 2020. Impacts of climate change on the global potential distribution of two notorious invasive crayfishes. - Freshw. Biol. 65: 353–365.

Zurell, D. et al. 2016. Benchmarking novel approaches for modelling species range dynamics. - Glob. Change Biol. 22: 2651–2664.

TABLES

Table 1: Number of articles returned by the queries on Web Of Science (WOS) and number of articles kept after title, keywords and abstract screening.

Group	# papers WOS	# papers kept
Spiders (Araneae)	74	55
Ground beetles (Carabidae)	32	24
Ants (Formicidae)	108	51
Other arachnids	37	34
Mites and ticks (Acari)	159	110
Molluscs (Gasteropoda)	164	121
Flies (Diptera)	454	320
Grasshoppers and crickets (Orthoptera)	59	34
Beetles other than Carabidae (Coleoptera)	313	183
Butterflies (Lepidoptera)	391	253
Dragonflies and damselflies (Odonata)	50	42
Bees (Apoidea)	116	81
Reptiles (Reptila)	529	347
Amphibians (Amphibia)	652	412
Mammals (Mammalia)	854	617
Birds (Aves)	1411	930

Table 2: Percentage of studies predicting distribution in the past, present and future, using different types of predictor variables and where a given predictor variable type was selected in the best models (values in brackets).

		Ants	Ground beetles	Spiders
Projection	Past	10.7	22.2	21.4
	Present	87.5	61.1	96.4
	Future	28.6	22.2	19.6
Predictor variables	Climate	92.9 (100)	77.8 (100)	98.2 (100)
	Soil	19.6 (65)	27.8 (100)	10.7 (80)
	Geology	5.4 (0)	11.1 (0)	10.7 (33.3)
	Topography	25.0 (73.6)	27.8 (100)	41.1 (80)
	Habitat	28.6 (73.9)	38.9 (100)	21.4 (80)
Mechanistic explanations	Species traits	8.9	0.0	3.6
	Ecophysiology	5.5	0.0	0.0

FIGURE CAPTIONS

Figure 1: Cumulative number of articles per year. a) Comparison between arthropods and vertebrates; b) comparison between main arthropod groups.

Figure 2: Production of studies per country for the three groups. Papers were attributed to a country based on the affiliations of the authors using the *bibliometrix* package.

Figure 3: Weighted co-citation network for the top 30 cited papers in the entire data set (ants, ground beetles and spiders). The size of the vertex is proportional to the number of articles citing a given reference. The colours of the links and vertex reflect citation clusters. The colour of the text corresponds to the paper theme.

Figure 4: Collaboration network between authors. Colours represent clusters of collaboration and pictograms the group targeted. For readability, the network is restricted to those papers with at least one author having two articles in the data set. This represents 64 articles (out of 103) and 211 authors (out of 355).

Figure 5: Number of articles using a given algorithm for species distribution models by year and group. To improve readability, only algorithms used at least five times in the entire data set are shown. ANN: artificial neural network; CT: classification tree; FDA: flexible discriminant analysis; GAM: generalized additive model; GARP: genetic algorithm for rule-set production; GBM: generalized boosting model; GLM: generalized linear model; MARS: multiple adaptive regression spline; MaxEnt: maximum entropy.

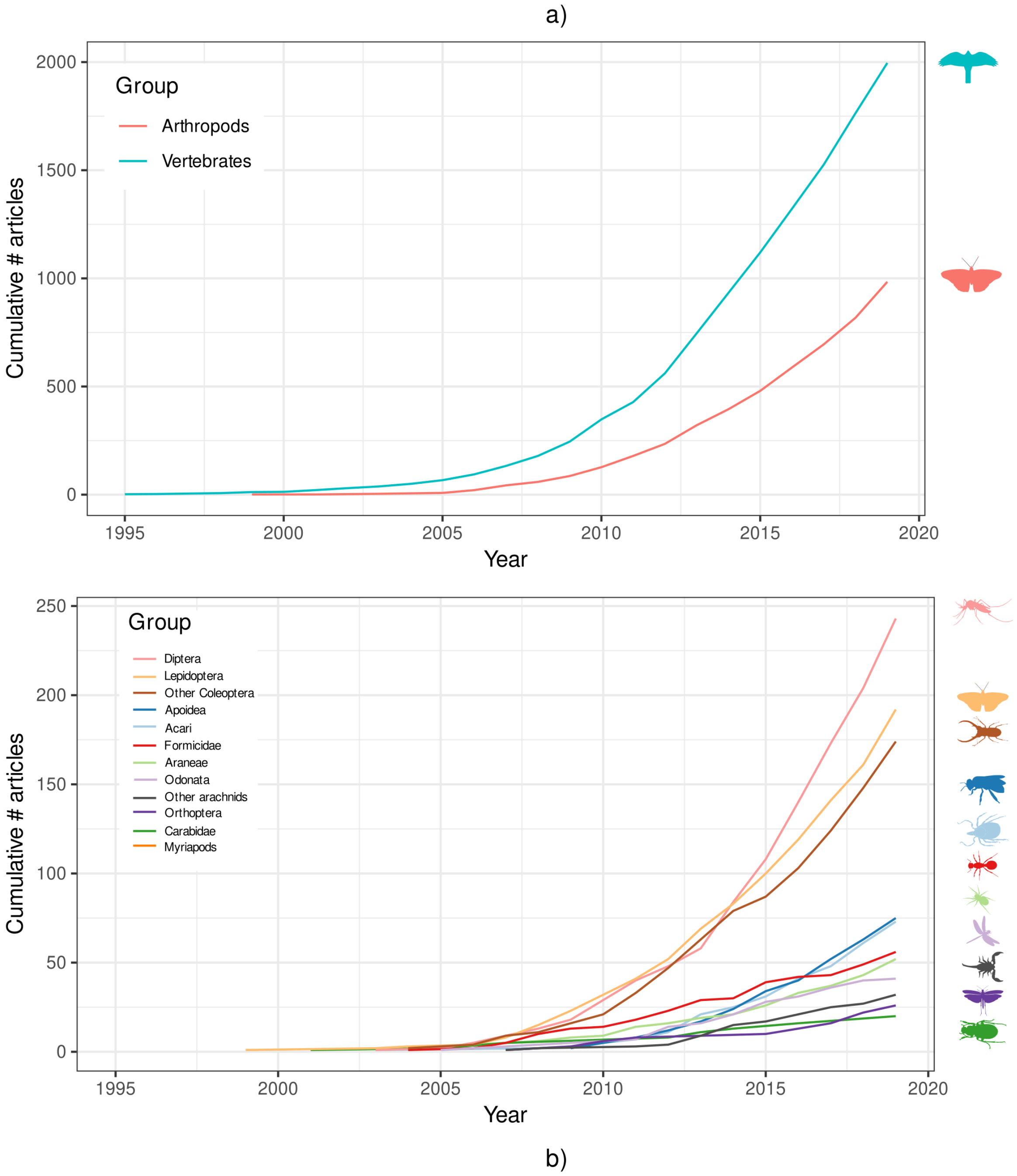
a)
Group
Arthropods
Vertebrates
Cumulative # articles
0
500
1000
1500
2000
1995
2000
2005
2010
2015
2020
Year
Group
Diptera
Lepidoptera
Other Coleoptera
Apoidea
Acari
Formicidae
Araneae
Odonata
Other arachnids
Orthoptera
Carabidae
Myriapods
Cumulative # articles
0
50
100
150
200
250
1995
2000
2005
2010
2015
2020
Year
b)

Figure 1

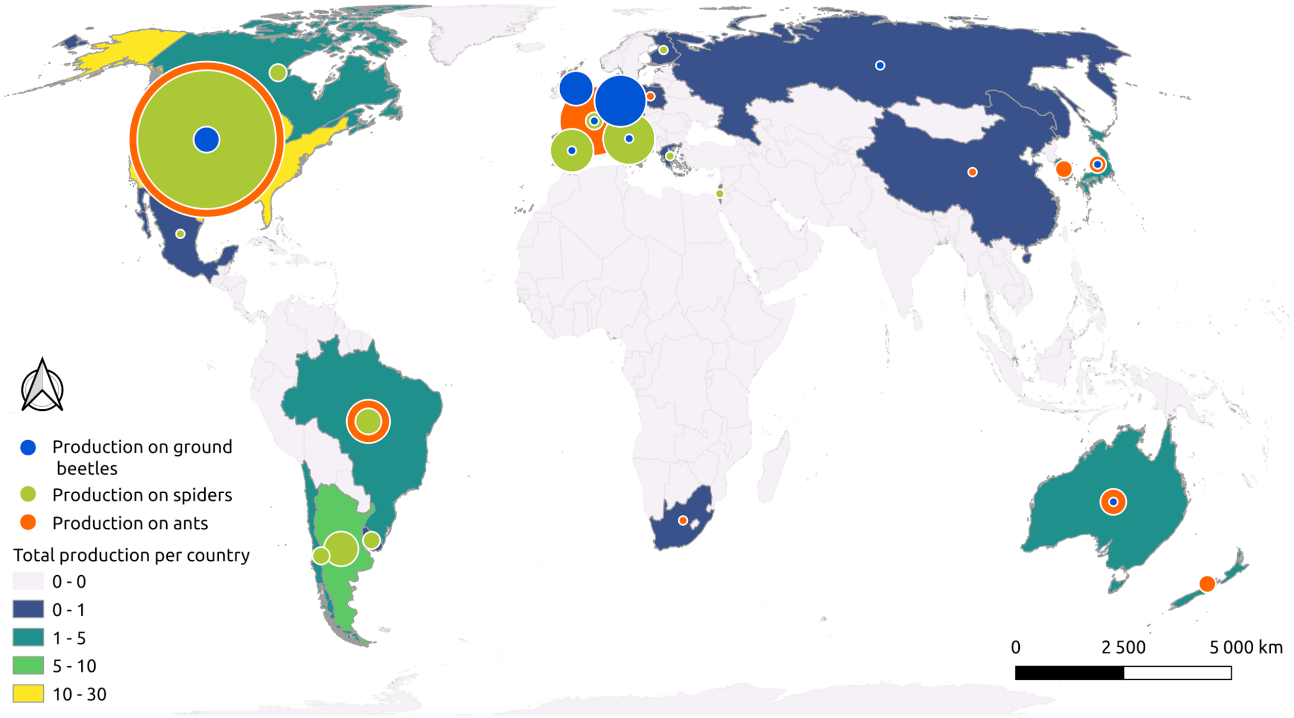
Production on ground beetles
Production on spiders
Production on ants
Total production per country
0 - 0
0 - 1
1 - 5
5 - 10
10 - 30
0
2 500
5 000 km

Figure 2

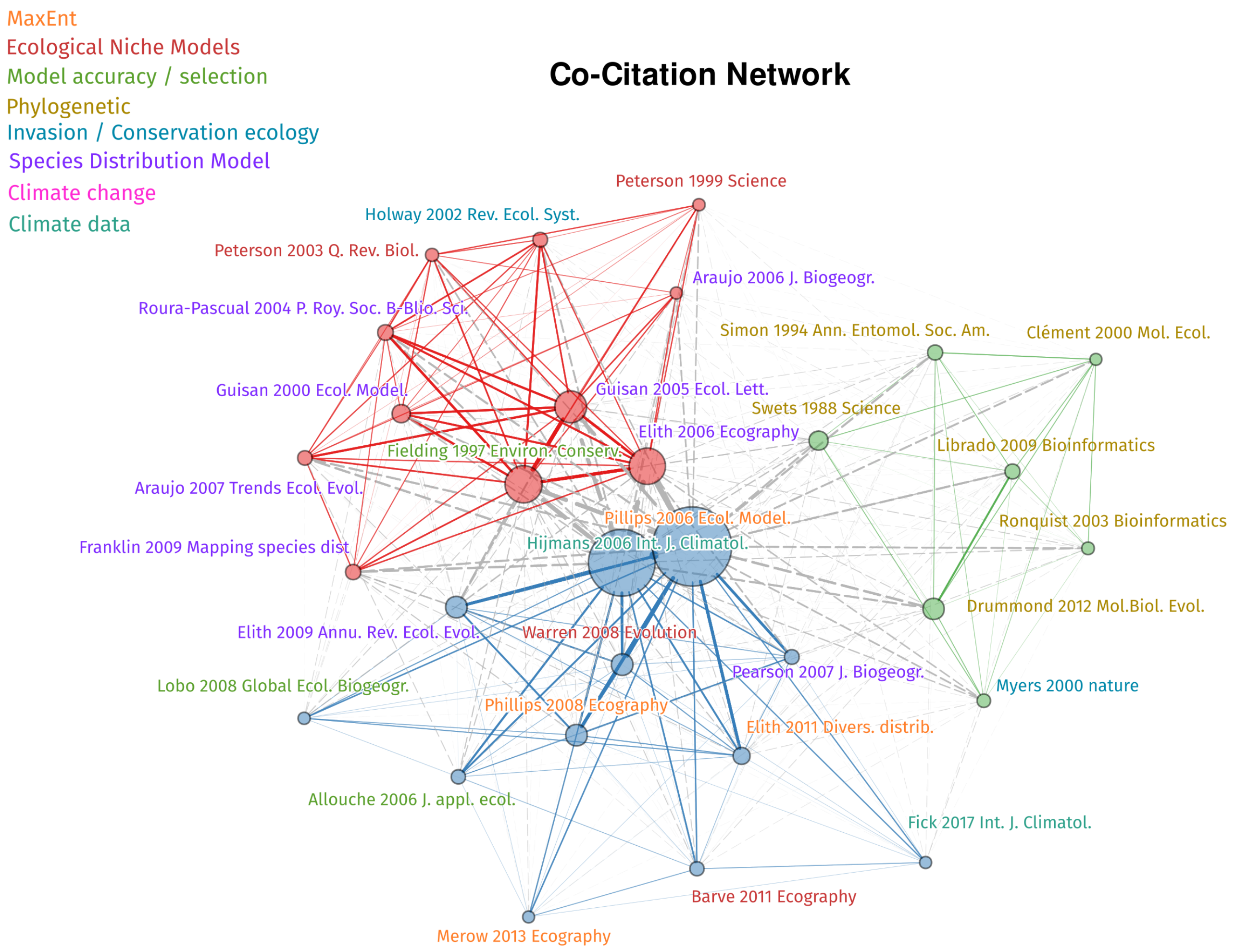
MaxEnt
Ecological Niche Models
Model accuracy / selection
Phylogenetic
Invasion / Conservation ecology
Species Distribution Model
Climate change
Climate data
Co-Citation Network
Peterson 1999 Science
Holway 2002 Rev. Ecol. Syst.
Peterson 2003 Q. Rev. Biol.
Araujo 2006 J. Biogeogr.
Roura-Pascual 2004 P. Roy. Soc. B-Blio. Sci.
Simon 1994 Ann. Entomol. Soc. Am.
Clément 2000 Mol. Ecol.
Guisan 2000 Ecol. Model.
Guisan 2005 Ecol. Lett.
Swets 1988 Science
Elith 2006 Ecography
Fielding 1997 Environ. Conserv.
Librado 2009 Bioinformatics
Araujo 2007 Trends Ecol. Evol.
Pillips 2006 Ecol. Model.
Ronquist 2003 Bioinformatics
Hijmans 2006 Int. J. Climatol.
Franklin 2009 Mapping species dist
Drummond 2012 Mol.Biol. Evol.
Elith 2009 Annu. Rev. Ecol. Evol.
Warren 2008 Evolution
Pearson 2007 J. Biogeogr.
Lobo 2008 Global Ecol. Biogeogr.
Myers 2000 nature
Phillips 2008 Ecography
Elith 2011 Divers. distrib.
Allouche 2006 J. appl. ecol.
Fick 2017 Int. J. Climatol.
Barve 2011 Ecography
Merow 2013 Ecography

Figure 3

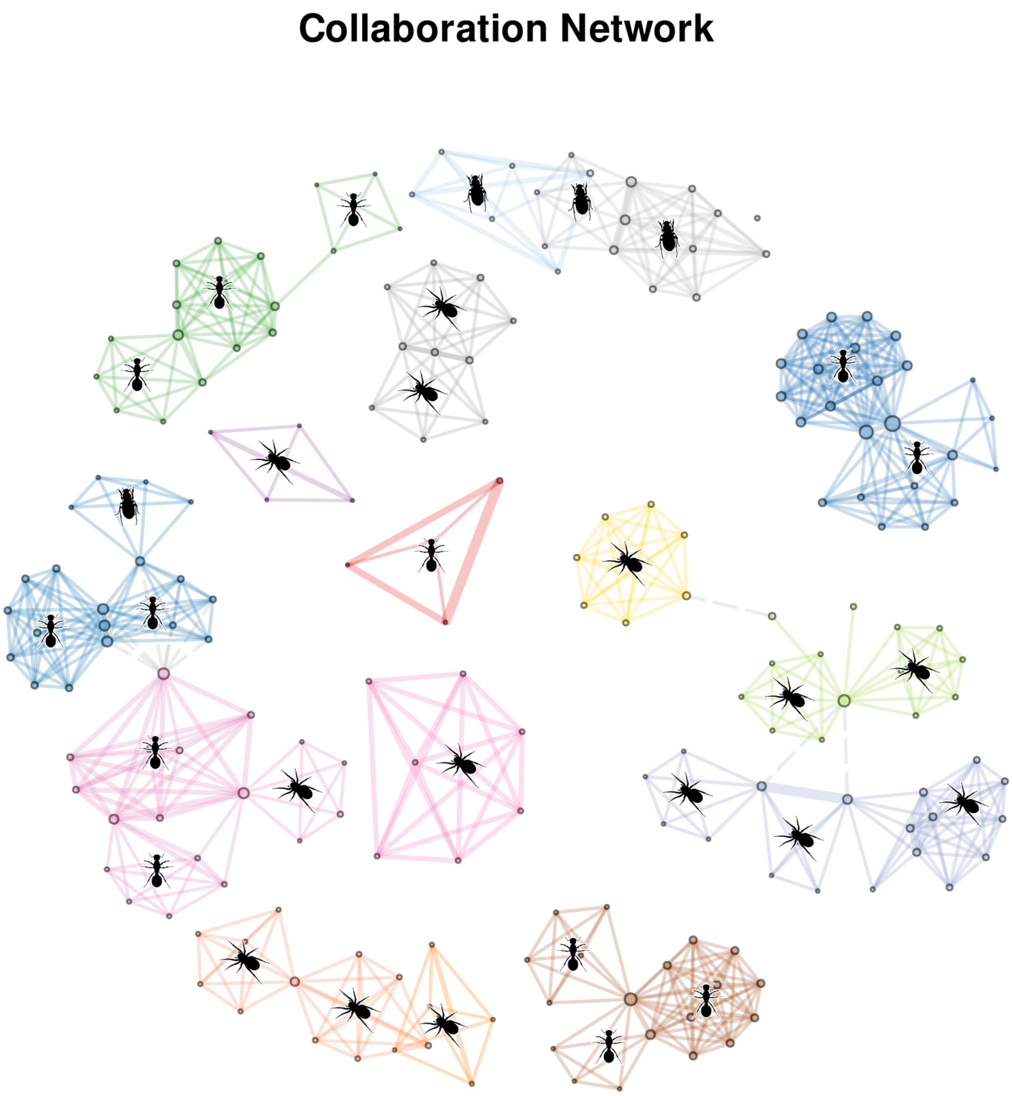

Figure 4

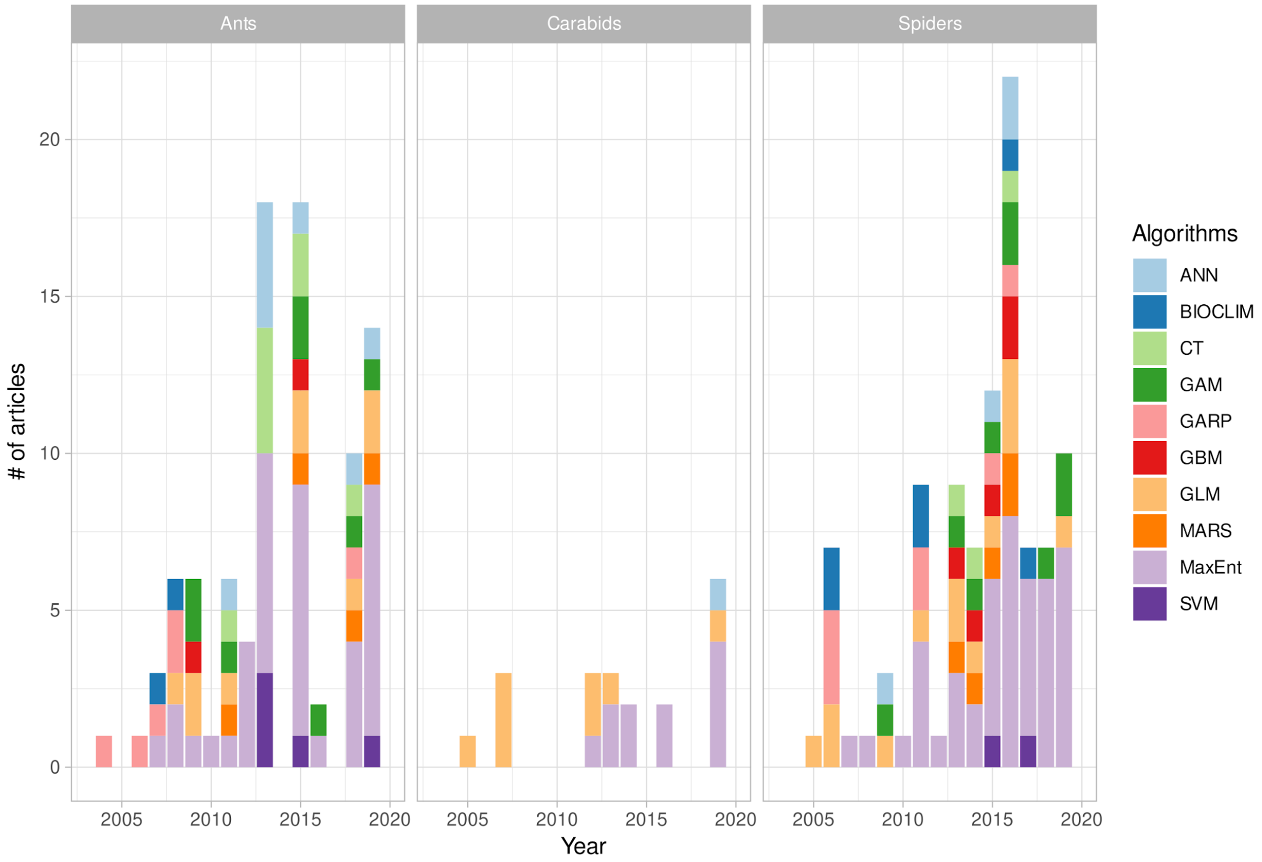
Ants
Carabids
Spiders
# of articles
0
5
10
15
20
2005
2010
2015
2020
Year
Algorithms
ANN
BIOCLIM
CT
GAM
GARP
GBM
GLM
MARS
MaxEnt
SVM

Figure 5

SUPPLEMENTARY MATERIAL FOR

Challenges and opportunities of species distribution modeling in terrestrial arthropods

Mammola et al.

Appendix S1. Queries on the Web of Science. Queries were made between 20 and 24 November 2020. TS denotes search for "Topic" and the asterisk (*) is a regular expression indicating to match all words including that string of characters (for example, "spider*" matches "spider", "spiders", "spiderling(s)", etc.).

- **Acari**: TS= ("Acari" or "tick" or "mites") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR "Niche model*" OR "Distribution model*" OR "Habitat suitability model*")

- **Amphibians**: TS= ("amphib*" or "anura" or "urodela" or "caudata" or "frog*" or "salamander*" or "newt*") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR "Niche model*" OR "Distribution model*" OR "Habitat suitability model*")

- **Ants**: TS= ("Formicidae" OR "Ants" OR "Ant") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR "Niche model*" OR "Distribution model*" OR "Habitat suitability model*")

- **Bees**: TS= ("Apoidea*" OR "bee" or "bees") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR "Niche model*" OR "Distribution model*" OR "Habitat suitability model*")

- **Birds**: TS= ("bird*" or "aves") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR "Niche model*" OR "Distribution model*" OR "Habitat suitability model*")

- **Butterflies**: TS= ("butterfl*" OR "lepidoptera*" or "moth*) AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR "Niche model*" OR "Distribution model*" OR "Habitat suitability model*")

- **Coleoptera except ground beetles**: TS= ("beetle*" OR "coleoptera" NOT "Carabidae" NOT "Carab*") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR "Niche model*" OR "Distribution model*" OR "Habitat suitability model*")

- **Diptera**: TS= ("diptera*" OR "fly" OR "flies" OR "mosquitoe*" or "midge*" or "gnats") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR "Niche model*" OR "Distribution model*" OR "Habitat suitability model*")

- **Gasteropoda**: TS= ("Gastropod*" or "snail*" or "slug*") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt"

OR “Niche model*” OR “Distribution model*” OR “Habitat suitability model*”)

- **Ground beetles**: TS= ("Carabidae*" OR "Carabid*") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR “Niche model*” OR “Distribution model*” OR “Habitat suitability model*”)

- **Mammals**: TS= ("mammal*" or "Monotremata" or "Dermoptera" or "Chiroptera" or "Primates" or "Pholidota" or "Lagomorpha" or "Rodentia" or "Carnivora" or "Tubulidentata" or "Proboscidea" or "Hyracoidea" or "Perissodactyla" or "Artiodactyla" or "Didelphimorphia" or "Scandentia" or "Paucituberculata" or "Microbiotheria" or "Dasyuromorphia" or "Peramelemorphia" or "Notoryctemorphia" or "Diprotodontia" or "Macroscelidea" or "Afrosoricida" or "Erinaceomorpha" or "Soricomorpha" or "Cingulata" or "Pilosa") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR “Niche model*” OR “Distribution model*” OR “Habitat suitability model*”)

- **Myriapods**: TS= ("Chilopod*" or "Centiped*" or "Myriapod*" or "Milliped*" OR "Diplopod*" OR Pauropod* OR "Symphyl*" OR "pseudocentiped*") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR “Niche model*” OR “Distribution model*” OR “Habitat suitability model*”)

- **Odonates**: TS= ("Odonat*" OR "dragonfl*" OR "damselfl*") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR “Niche model*” OR “Distribution model*” OR “Habitat suitability model*”)

- **Orthoptera**: TS= ("Orthoptera*" OR "Grasshopper*" or "locust*" or "cricket*") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR “Niche model*” OR “Distribution model*” OR “Habitat suitability model*”)

- **Other arachnids**: TS= ("Opilion*" OR "Scorpion*" OR "Pseudoscorpio*" OR "Amplypyg*" OR "Solifug*" OR "Palpigrad*" OR "whip scorpion*" OR "Microwhip scorpion*" OR "harvestman" OR "camel spider*" OR "wind scorpion*" OR "sun spider*" OR "solifug*") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR “Niche model*” OR “Distribution model*” OR “Habitat suitability model*”)

- **Reptila:** TS= ("reptil*" or "snake*" or "lizard*" or "turtle*" or "crocodyl*" or "squamata") AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR “Niche model*” OR “Distribution model*” OR “Habitat suitability model*”)

- **Spiders**: TS= (“Araneae” OR “Spider*”) AND TS=("Species distribution model*" OR "Ecological niche model*" OR "Bioclimatic envelope model*" OR "MaxEnt" OR “Niche model*” OR “Distribution model*” OR “Habitat suitability model*”)

Appendix S2: List of information collected for ant, ground beetle and spider papers selected.

Article information	CODE	Arbitrary code relating paper to taxa information
	TITLE	Article title
	JOURNAL	Journal in which the paper was published
	DOI	Digital Object Identifier
	AUTHORS	List of authors
	YEAR	Year of publication
	GEOGRAPHIC AREA	Area covered by the SDM
	PURPOSE	Purpose of the paper
	MODELLING_ALGORITHM	Modelling algorithm used in the SDM
	MODELLING_APPROACH	Single, ensemble, or no silver bullet approach for algorithm
	PAST_PROJECTION	Past projection of the distribution (0/1)
	PRESENT_PROJECTION	Present projection of the distribution (0/1)
	FUTURE_PROJECTION	Future projection of the distribution (0/1)
	CLIMATIC_VAR	Climatic variables used as predictors (0/1)
	GEOLOGICAL_VAR	Geological variables used as predictors (0/1)
	SOIL_VAR	Soil variables used as predictors (0/1)
	HABITAT_VAR	Habitat variables used as predictors (0/1)
	TOPOGRAPHIC_VAR	Topographic variables used as predictors (0/1)
	HISTORICAL_VAR	Historical variables used as predictors (0/1)
	BIOTIC_INTERACTIONS_VAR	Biotic interactions effect included in the SDM (0/1)
	VAR_RESOLUTION	Predictor variables resolution
	TRAITS	Species traits included in the SDM
	PHYLOGENETIC_DATA	Phylogenetic data used in the paper
	ECOPHYSIOLOGICAL_DATA	Ecophysiological variables used in the paper
	NOTES	Remarks
Taxa information	CODE	Arbitrary code relating paper to taxa information
	FAMILY	Family of the species studied
	GENUS	Genus of the species studied
	SPECIES	Species studied
	ENDEMIC	Is the species studied endemic of the area (0/1)
	ALIEN	Is the species studied considered by the authors as alien
	HABITAT	Species habitat
	PAST_SHIFT	Is the past distribution larger/shifted than the present (0/1)
	PAST_SMALLER	Is the past distribution smaller than the present (0/1)
	FUTURE_DECLINE	Is the future distribution smaller than the present (0/1)
	FUTURE_SHIFT	Is the distribution forecasted to shift compared to the present (0/1)
	FUTURE_EXPANSION	Is the future distribution larger than the present (0/1)
	CLIMATIC_VAR_SEL	Were climatic variables selected among the best predictors (0/1)
	GEOLOGICAL_VAR_SEL	Were geological variables selected among the best predictors (0/1)
	SOIL_VAR_SEL	Were soil variables selected among the best predictors (0/1)
	HABITAT_VAR_SEL	Were habitat variables selected among the best predictors (0/1)
	TOPOGRAPHIC_VAR_SEL	Were topographic variables selected among the best predictors (0/1)
	HISTORICAL_VAR_SEL	Were historical variables selected among the best predictors (0/1)
	BIOTIC_INTERACTIONS_VAR_SEL	Were biotic variables selected among the best predictors (0/1)
	NOTES	Remarks

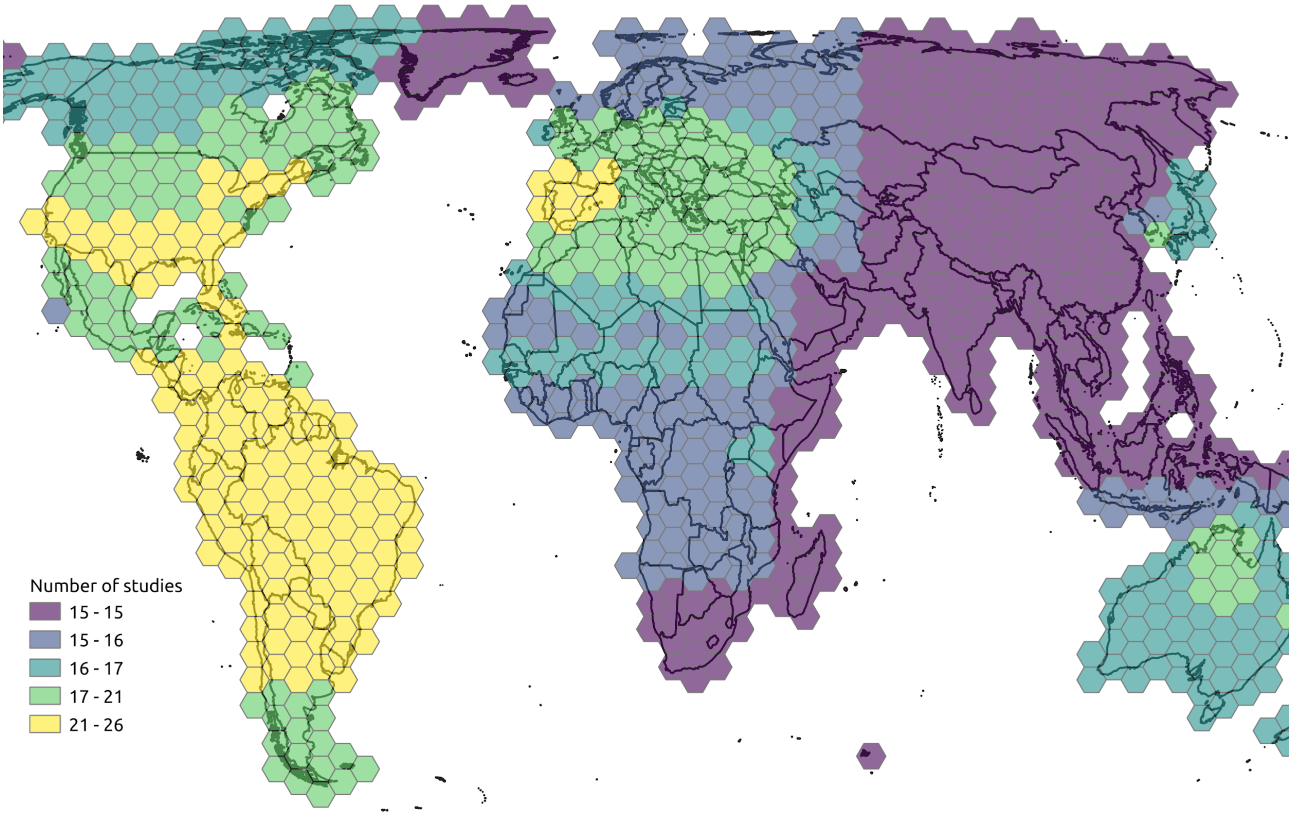

Figure S1. Distribution of the number of ant studies (for 250 km side hexagon). For each study the distribution of each species resulting from SDM were mapped on the hexagon grid.

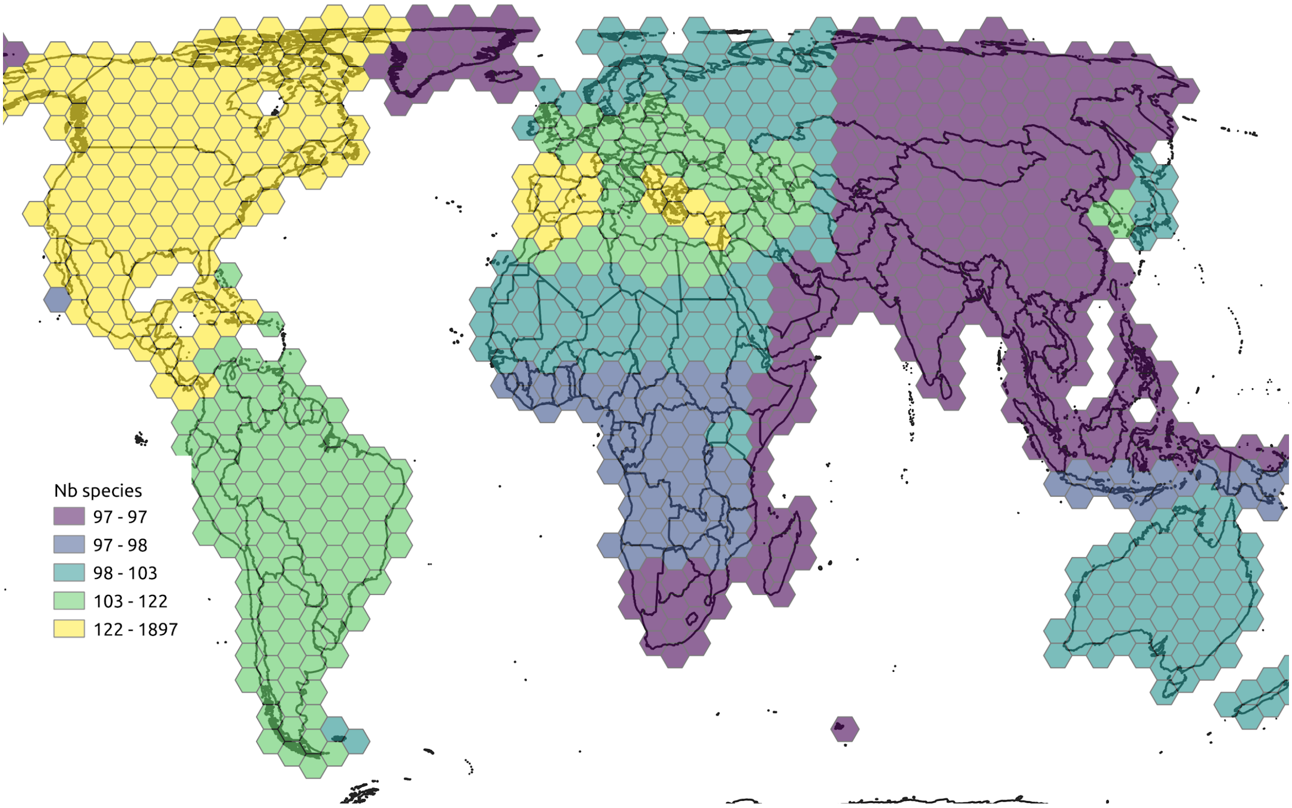

Figure S2. Distribution of the number of ant species (for 250 km side hexagon). For each study the distribution of each species resulting from SDM were mapped on the hexagon grid.

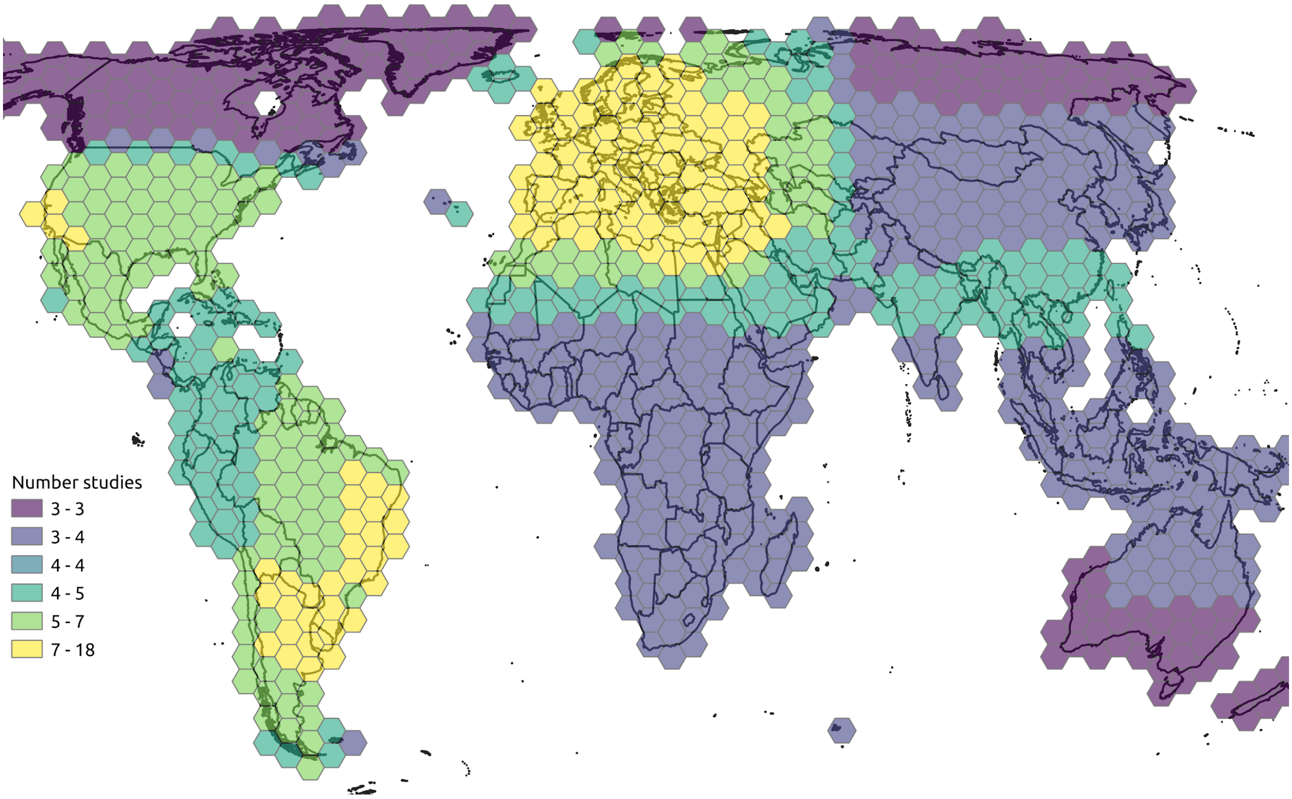

Figure S3. Distribution of the number of spider studies (for 250 km side hexagon). For each study the distribution of each species resulting from SDM were mapped on the hexagon grid.

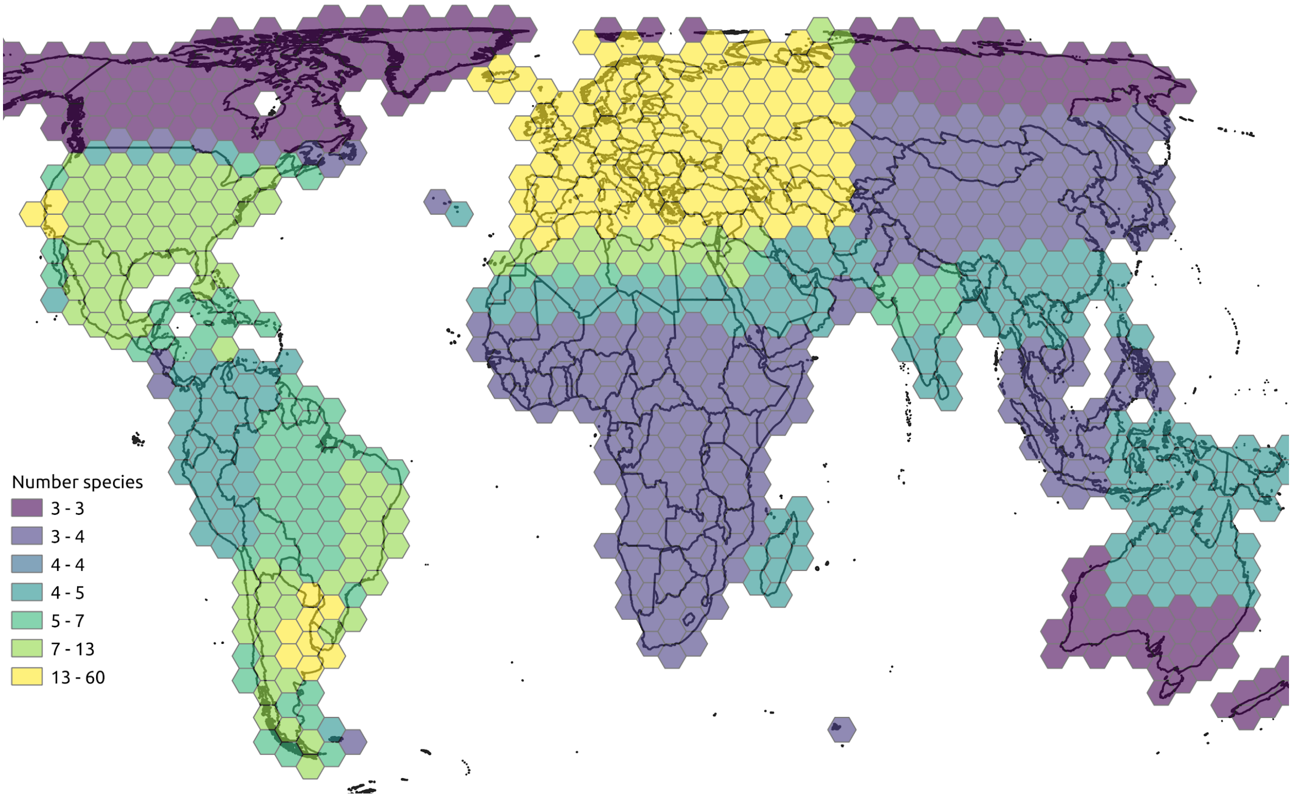

Figure S3. Distribution of the number of spider species (for 250 km side hexagon). For each study the distribution of each species resulting from SDM were mapped on the hexagon grid.

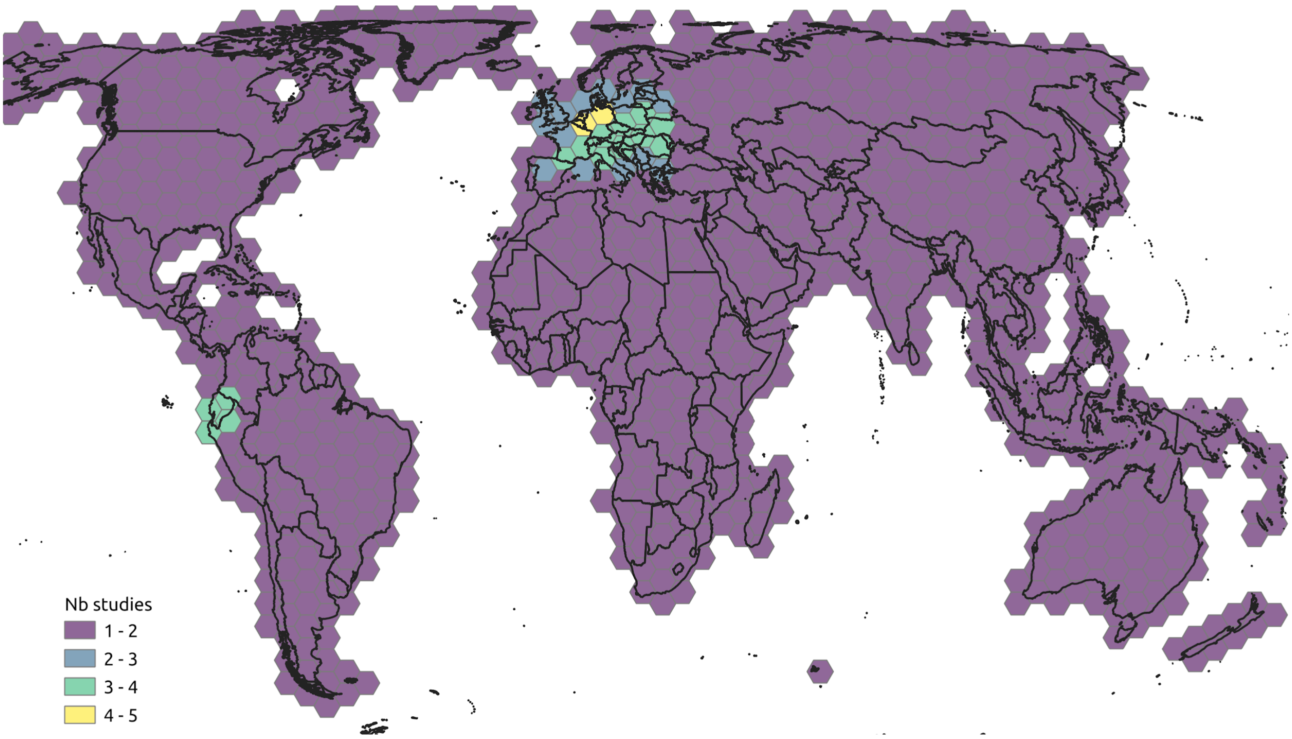

Figure S4. Distribution of the number of ground beetles studies (for 250 km side hexagon). For each study the distribution of each species resulting from SDM were mapped on the hexagon grid.

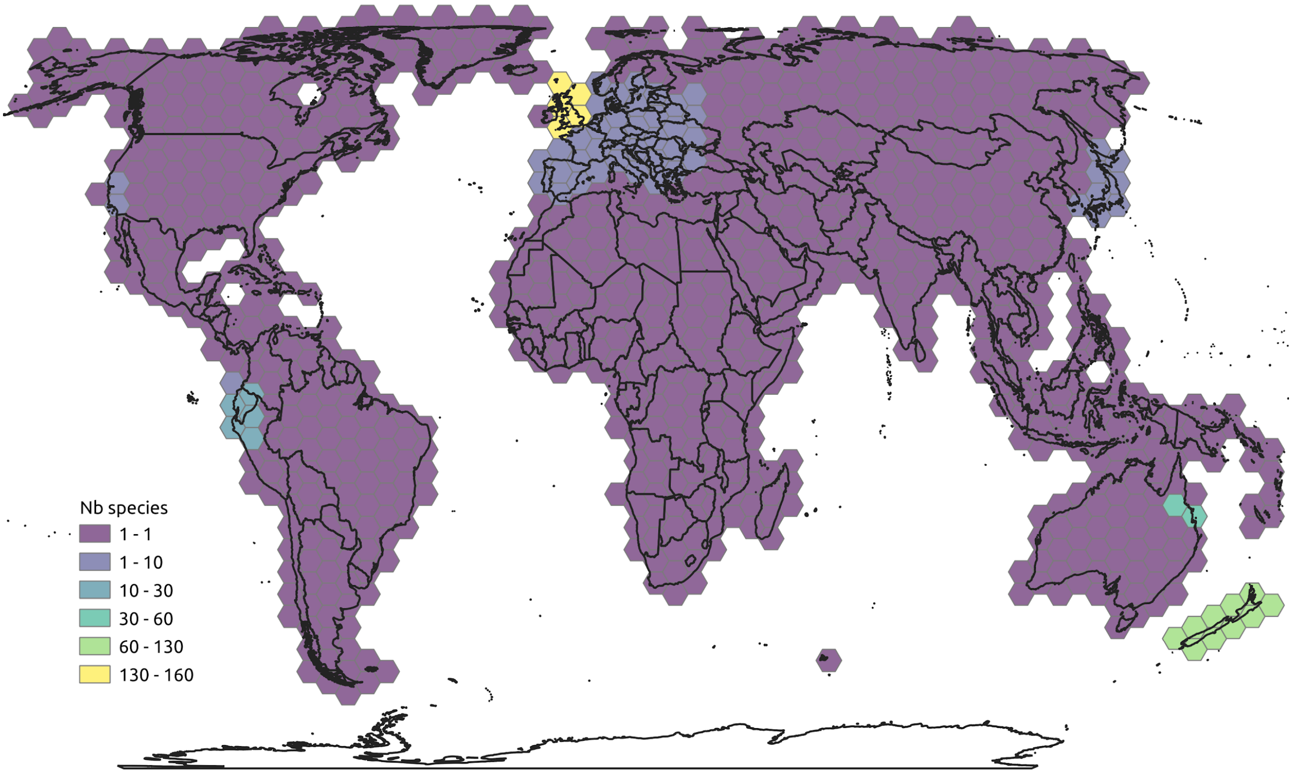

Figure S5. Distribution of the number of ground beetles species (for 250 km side hexagon). For each study the distribution of each species resulting from SDM were mapped on the hexagon grid.