

Causes and consequences of global pollinator biodiversity change in the Anthropocene

Joseph William Millard

**A thesis submitted for the degree of Doctor of Philosophy at University
College London
(September 2021)**

Primary supervisor: Dr. Tim Newbold

Secondary supervisors: Dr. Robin Freeman, and Prof. Richard D. Gregory

I, Joseph William Millard, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Any contributions made by any coauthors I describe in the thesis outline.

Joseph William Millard, September 2021

Abstract

87.5% of wild plant species are thought to be pollinated to some extent by animals, with an estimated global value of \$235-577 billion US dollars per annum to crop pollination. Across North America and Europe numerous studies have documented declines in pollinating animals. A number of anthropogenic drivers—primarily land-use and climate change—have been associated with these losses. However, the extent to which land-use and climate act individually and synergistically to drive changes, and how this might risk the contribution pollinators make to crop pollination, is still unclear. In this thesis I explore the causes and potential consequences of pollinator biodiversity change. First, I use a set of name-entity recognition algorithms to quantify the geographic and taxonomic distribution of the animal pollination literature, confirming that although the pollination literature does over-represent the honey bees and bumblebees of North America and Europe, there is also pollination information across a range of other taxa and locations. Second, I then apply these same algorithms in combination with a manual literature check to identify a set of likely pollinating species. I then use this set of likely pollinating species and a database of local biodiversity records to model the global response of pollinator biodiversity to land-use type and intensity, showing how response differs among taxonomic groups, biodiversity metrics, and geographic regions. Third, I then use my set of likely pollinating species to investigate the interactive effects of land use and climate change on pollinator abundance, demonstrating a strongly negative synergistic interaction between climate change and land use. Using this model, estimations of global pollination dependent production, and future climate scenarios, I then project temporal and spatial changes in crop pollination risk. Fourth, I develop a new metric of biodiversity awareness using Wikipedia page views, which I use to show that public interest in pollinators has likely not increased. I conclude by discussing the core findings of my thesis in the context of the current debate around pollinator biodiversity change and conservation culturomics, with some suggestions as to how these respective fields might move forward.

Contents

Impact statement	6
Acknowledgements	8
Thesis outline.....	10
Chapter 1: Introduction	13
Chapter 2: Text-analysis reveals taxonomic and geographic disparities in animal pollination literature	26
Abstract	26
Introduction.....	27
Methods	33
Results and Discussion	38
Chapter 3: Global effects of land-use intensity on local pollinator biodiversity	61
Abstract	61
Introduction.....	62
Methods	66
Results	74
Discussion	85
Chapter 4: Worldwide vulnerability of local pollinator abundance and crop pollination to land-use and climate change	92
Abstract	92
Introduction.....	93
Methods	97
Results	105
Discussion	114
Chapter 5: The species awareness index as a conservation culturomics metric for public biodiversity awareness	120
Abstract	120
Introduction.....	121
Methods	124
Results	131
Discussion	141
Chapter 6: Discussion and synthesis.....	148
References	163
Appendix 1:	186
Appendix 2:	201
Appendix 3:	233

Appendix 4:	239
--------------------------	------------

Impact statement

Pollinator biodiversity is highly important to humanity, both through the way in which it underpins the structure of many ecosystems, and through the contribution it makes to crop production. Numerous studies have indicated that pollinator biodiversity is undergoing rapid change, driven primarily by climate change and land use, but the magnitude of change and the extent to which it differs among taxonomic groups, biodiversity metrics, and geographic regions, remains unclear. Moreover, how these changes might affect crop pollination is an ongoing question, with a large body of research indicating losses of crop yield can result from pollinator biodiversity change. The primary aim of this thesis is to further our understanding of both the causes and potential consequences of pollinator biodiversity change, in this present epoch of unprecedented human activity. I make a number of novel contributions to the field, which could potentially help to inform the future conservation of pollinator biodiversity.

In Chapter 3, I model the effect of land-use intensity on global pollinator biodiversity, using the most comprehensive space-for-time model to date, showing strong effects of land use on tropical pollinator biodiversity and a differential effect among pollinating taxa. The results of this chapter are published in *Nature Communications*, where it was selected for an editor's highlight in the category 'Ecology and Evolution'.

In Chapter 4 I further develop the models in Chapter 3, here focussing on the interactive effects of climate change and land use. Using future scenarios of climate change and estimations of global pollination dependent agriculture, I show that crop pollination risk will likely be highest in northern South America and sub-Saharan Africa, and that under RCP 8.5 rapid change in pollination production risk could occur in the 2030s. As far as I know this work represents the broadest study of the interactive effects of climate change and land use on pollinator biodiversity, and the first attempt to project how this interaction might relate to the spatial and temporal distribution of crop pollination risk. The results of this chapter are currently in preparation for submission in a high-impact journal. In addition I have also developed an

interactive web app to present the findings of this paper (https://joemillard.shinyapps.io/pollinator_dependence_visualisation/), which can be used to inform both crop pollination risk to individual countries and global crop production.

In addition to Chapters 3 and 4, I also contribute towards developing our understanding of the pollination literature, and efforts to develop online metrics for public biodiversity awareness, both of which are important in contextualising the causes and consequences of pollinator biodiversity change. For example, Chapter 2 confirms that the pollination literature is concentrated in the honey bees and bumble bees of North America and Europe, with a set of name-entity recognition tools rarely applied in ecology and conservation. The results of this chapter are published in *Ecography*, and were presented at SCCS Europe 2019, the RSPB Annual Science Meeting 2019, and the Linnean Student Conference 2019. In Chapter 5, I introduce a new metric of public biodiversity awareness using Wikipedia page views, and suggest how this might be combined with other online metrics already in existence. The results of this chapter are published in a special section of *Conservation Biology*, entitled 'Advancing Conservation Culturomics', and were presented at the BES Festival of Ecology 2020.

Acknowledgements

Wow, what a journey! Doing a PhD has been tough, but undoubtedly a hugely enjoyable and fulfilling experience. Credit for that is largely due to my team of supervisors. An especially huge thanks to Tim Newbold. Tim, you've been a guiding light for me as I start out on my scientific career; an inspiration not only in the practice of science, but also in the living of a scientist.

A massive thanks also to Robin Freeman. Rob, your seemingly endless stream of ideas on all things data and conservation is staggering. I don't doubt that you have in your mind ideas sufficient for a lifetime(s) of PhDs.

And to Richard Gregory. Richard, your incisive ability to grasp the root of a problem, and your encouragement to think carefully about uncertainty has—I like to think—kept me anchored to the bedrock of what's useful.

To the people of CBER, thank you for making CBER great—I am so grateful to have had the opportunity to work (and go to the pub) with you all. A few people I'd like to thank specifically. Residents of SDL (past and present), thanks primarily for the laughs, friendship, and heated Star Wars discussions, but also the willingness to help and support. To the Global Biodiversity Change group, thanks for all the biodiversity discussions, happy chats, BBQs, dinners, advice, and general support. I'm sure we'll all stay in touch wherever we happen to find ourselves next. To Seirian Sumner, my upgrade panel chair, thanks for your direction and clarity at a critical period of my PhD. Without your course correction, this thesis would have charted a very different path. And to the people of CBER data club and journal club, thanks for the conversations on all things data and journals. I've greatly enjoyed learning and chatting with you all, both back in the office and in lockdown.

A few other researchers I'd like to thank for their valuable contributions. To all those who provided much-needed expert animal pollinator advice (Opeyemi Adedoya, Sabrina Gavini,

Esther Kioko, Michael Kuhlmann, Jeff Ollerton, Manu Saunders, and Zong-Xin Ren), thank you. Your input was absolutely invaluable, I do hope that I get the chance to work with you all again, and perhaps someday the opportunity to meet.

To the 'Leeds peeps', thanks for always being awesome and just generally great. A particular thanks to Patrick Molgaard. Your help in taming the beast that is CLIFF-CLAVIN saved me weeks of blindly stumbling through Tomcat server tutorials. I don't have much love for Tomcat servers. I hate to think how I'd feel without your input.

To my mum and dad, thanks for always supporting and caring for me. This thesis (and anything I've ever done) I owe to how you taught me to do life. To my brothers and sisters (Tom, Sophie, Ben, Chloe, Katie, and Emily), and the whole ever-growing Millard+ family, thank you for all the happiness, and past, present, and future memories. Hopefully I get to see a whole lot more of all of you post-pandemic.

And finally to Bruna, my fiancée. Thank you so much for being you. The love and support we have for each other has carried us through *all* this. I know I couldn't have done it without you.

Thesis outline

Chapter 1: Introduction

Chapter 2: Text-analysis reveals taxonomic and geographic biases in pollinator information

As anthropogenic activity has driven changes in global biodiversity, concerns have arisen for the prospects of pollinators and the services they provide humanity. Such changes have driven a significant push for further research. Much of this growth in research is said to have occurred in the honeybees and bumblebees of North America and Europe, but the extent of this bias is unclear. In Chapter 2 I carry a text-analysis systematic review of the animal pollinator literature, using name-entity recognition algorithms to quantify the temporal, taxonomic, and geographic distribution of the animal pollinator literature. The work for this chapter is published in *Ecography* (Millard et al. 2020a), in collaboration with Dr. Robin Freeman and Dr. Tim Newbold. I conceived and designed the review (with Dr Robin Freeman and Dr Tim Newbold), carried out all the analysis, produced the figures, wrote the initial draft of the manuscript, and handled the manuscript through submission and review. Thanks also to Prof. Richard D. Gregory, Adrienne Etard, and Patrick Molgaard for suggestions, edits, and advice.

Chapter 3: Global effects of land-use intensity on local pollinator biodiversity

Anthropogenic land-use intensity is well known as a driver of pollinator biodiversity change, but at the global scale the magnitude of its effect among taxonomic groups, geographic regions, biodiversity metrics, and forms of intensity is unclear. In Chapter 3 I use the name-entity recognition algorithms of Chapter 2, in conjunction with manual checking from both myself and a panel of experts, to build a database of likely pollinating species. I then subset this set of likely pollinating species from a database of local biodiversity records (the PREDICTS database), before building a set of space-for-time models predicting pollinator biodiversity response to land-use intensity and type. The work for this chapter is published in

Nature Communications (Millard et al. 2021a), in collaboration with Dr. Charlotte L. Outhwaite, Robyn Kinnersley, Dr. Robin Freeman, Prof. Richard D. Gregory, Dr. Opeyemi Adedaja, Dr. Sabrina Gavini, Dr. Esther Kioko, Dr. Michael Kuhlmann, Prof. Jeff Ollerton, Dr. Zong-Xin Ren, and Dr. Tim Newbold. I conceived and designed the analysis (with Dr Tim Newbold), led all the analysis and figure design, wrote the initial draft of the manuscript, and handled the manuscript through submission and review. Thanks also to Dr Manu Saunders and Bruna Abreu for advice and comments, and Dr Monica Ortiz for aggregating the forest cover data in the supplementary information (Appendix 2).

Chapter 4: Worldwide vulnerability of local pollinator abundance and crop pollination to land-use and climate change

Although anthropogenic land use has been a significant driver of pollinator biodiversity change, its effect is predicted to be surpassed by climate change over the coming decades. These climate effects likely interact with the structure of the land, such that the effects of both land use and climate change together have a magnified effect on biodiversity. In the future therefore, places in which cropland and rapid climate change intersect may be most vulnerable to significant pollinator biodiversity change. For regions that experience this interactive effect in combination with a high dependence on pollination for crop production, one might expect crop pollination shortfall. In Chapter 4 I use the PREDICTS database subset from Chapter 3 to model the response of pollinator abundance to land-use type (cropland and primary vegetation) and a standardised climate anomaly. I then apply my model of insect pollinator abundance in conjunction with global crop data, ratios for crop production dependence on animal pollination, and future projections of climate change, to predict temporal and spatial patterns of future crop pollination risk. The work for this chapter was undertaken in collaboration with Dr. Abbie Chapman, Dr. Silvia Ceausu, Dr. Robin Freeman, Prof. Richard D. Gregory, Dr. Charlotte L. Outhwaite, and Dr. Tim Newbold, and is being prepared for submission. I conceived and designed the analysis (with Dr Tim Newbold), led all the analysis and figure design, and wrote the initial draft of the manuscript as a thesis chapter.

Chapter 5: The species awareness index as a conservation culturomics metric for public biodiversity awareness.

Given the recent media attention afforded to pollinator biodiversity change, and biodiversity change more generally, one might expect that public awareness of biodiversity and its threats has experienced a concomitant increase. Conservation culturomics has emerged as a means through which such awareness can be measured, with online data sources such as social media, online newspapers, and search trends used to quantify human-nature interactions. In Chapter 5 I introduce a new conservation culturomics metric for public biodiversity awareness, using Wikipedia page views and a methodology inspired by the Living Planet Index. Here I show that although awareness of biodiversity has likely marginally increased, awareness of the value of biodiversity likely has not, since both pollinating and traded species do not have significantly greater rates of change in awareness than their non-pollinating or non-traded counterparts. The work for this chapter is published in *Conservation Biology* in a special section entitled ‘*Advancing Conservation Culturomics*’ (Millard et al. 2021b), in collaboration with Dr. Robin Freeman, Prof. Richard Gregory, and Prof. Kate Jones. I conceived and designed the analysis (with Dr Robin Freeman), carried out all the analysis, produced the figures, wrote the initial draft of the manuscript, and handled the manuscript through submission and review. Thanks also to Dr. Yan Wong and Dr James Rosindell for providing taxonomic mapping data and Dr Tim Newbold, Dr Rory Gibb, and the CBER Journal Club and Data Club for discussions and comments on earlier drafts of the manuscript.

Chapter 6: Discussion and synthesis

Chapter 1: Introduction

A brief introduction to animal pollination

Pollination facilitates the reproduction of many plant species (IPBES 2016). Its importance to the continuing functioning of the biosphere, as it currently exists today, is undoubtedly highly significant. Specifically, pollination describes the transfer of pollen—the male plant gamete—from the male anther structure of the plant to the female stigma, causing fertilisation and the production of a fruit and seed (Proctor 1996). This fruit and seed forms the basis of the following generation, dispersed by either the plant itself or some animal vector (Proctor 1996).

Pollen can be transferred either among the sexual structures of a single plant, or across individual plants. When pollen is transferred from the male to female structures of the same flower, or across flowers within the same individual, it's described as self-pollination (Proctor 1996). Such a reproductive strategy is present at least to some extent in many flowering plants, with some species having evolved to reproduce entirely through self-pollination (Schoen et al. 1996). When pollen is transferred across individual plants, it's described as cross-pollination (Proctor 1996). Since self-pollination increases inbreeding and reduces genetic diversity, in many flowering plants selection has favoured the evolution of cross-pollination (Barrett 2010). In order to achieve cross-pollination however, given the distance between individual plants, many plants have evolved to recruit some external means of pollen transfer (Proctor 1996). For many flowering plants this means abiotic pollination, using either wind or water (Ackerman 2000). These flowering plants release large quantities of pollen into their immediate surroundings, with the pollen then carried in wind or water currents to neighbouring plants (Ackerman 2000). Wind pollination is typical of grass species, and more common at temperate than tropical latitudes (Ollerton et al. 2011). Water pollination is less common, but an important mechanism for many seagrass species (Ackerman 2000).

Since releasing large quantities of pollen is energetically expensive, if they exist within an environment in which they can, many flowering plants have evolved to recruit animal vectors for the transfer of pollen (Proctor 1996). Such a reproductive strategy is thought to have evolved perhaps as early as the Carboniferous (300 mya) (van der Kooi and Ollerton 2020). Since then, animal pollination has been a hugely successful evolutionary innovation for flowering plants. At tropical latitudes approximately 95% of all flowering plants are animal pollinated (Rech et al. 2016), and at the global level, 87.5% of wild plant species are thought to be pollinated to some extent by animals (Ollerton, Winfree and Tarrant 2011). These pollinators are represented across a variety of taxonomic groups, including but not limited to insects, bats, birds, lizards, and rodents (Ollerton 2017). The greatest contributors globally are the insects (Wardhaugh 2015). Insect pollinators include the bees (Anthophila), beetles (Coleoptera), butterflies and moths, (Lepidoptera), bugs (Hemiptera), flies (Diptera), and multiple wasp groups (Wardhaugh 2015). Spatial differences in pollinator importance for wild and crop plants do exist (Ollerton 2017), but of these pollinating groups, today the bees are likely the most important pollinators (Potts et al. 2010). A relatively small group at approximately 20,000 species (Michener 2000), the bees are almost completely reliant on floral resources in both the larval and adult developmental stages, meaning the frequency with which they visit flowers is relatively high (Michener 2000).

The importance of animal pollinators to people

Over roughly the past 10,000 years, humans have domesticated a number of plant species in the development of agriculture, artificially selecting desirable plant traits for large-scale production and consumption (Harlan 1992). At the global level today, by both total production and the number of crops, production of these species is in the main not reliant on animal pollination (Aizen et al. 2009). Staple crops such as wheat, rice, maize, and barley are all wind-pollinated (Klein et al. 2007). Root vegetable crops such as potatoes, carrots, turnips, and cassava are propagated from tubers of previous generations, often only requiring self or cross pollination for the production of new crop varieties (Aizen et al. 2009). Indeed, owing to

the contribution of wind, self-pollination, and propagation, only 5-8% of global crop production (by tonnage) is said to be dependent on animal pollinators (Aizen et al. 2009).

Given the relatively low contribution animal pollination makes to global crop production, it is tempting to conclude that pollinators are unimportant to humanity. However, for a number of reasons this conclusion is misguided. First, although the extent to which crop production is reliant on animal pollination is relatively low, the importance of pollinators in maintaining the terrestrial ecosystems in which we have evolved is not (Ollerton 2021). With the instantaneous disappearance of animal pollinators, one would expect the environment in which we live to change abruptly, with unknown consequences for our own future (Potts et al. 2016). Second, although total crop production dependent on animal pollinators is relatively low, this production forms an important nutritional component in the diets of many people. Wind-pollinated crop species more often produce a high-calorie crop (Richards 2001; Prescott-Allen and Prescott-Allen 1990; Ghazoul 2005); a core dietary component, but not sufficient alone for a healthy diet. Animal-pollinated fruits, on the other hand, tend to provide a higher density of many essential vitamins and minerals (Eilers et al. 2011). Third, although total crop production has low dependence on animal pollination, for some crops animal pollination is essential (Klein et al. 2007). For example, production of atemoya, Brazil nuts, cantaloupes, cocoa, kiwis, macadamia nuts, passion fruit, pawpaw (*Asimina* fruit), rowanberries, sapodilla, squashes and pumpkins, vanilla and watermelons is wholly reliant on animal pollination (Klein et al. 2007). Since the spatial distribution of these crops is not even, there is therefore variation in the extent to which agricultural regions are dependent on animal pollination. Many tropical crops, for example, have a higher reliance on animal pollination than crops grown elsewhere (Klein et al. 2007). Reliance in the tropics is such that, via the production of cocoa alone, the national economies of countries such as Ivory Coast and Ghana are in theory highly vulnerable to changes in animal pollinator biodiversity (Schroth et al. 2016).

Animal pollinator biodiversity change in the Anthropocene

Pollinators exist within a world that's experiencing an abrupt—at least on geological time-scales—period of environmental change. A period of change driven by the activities of humanity on Earth, often referred to as the Anthropocene (Schroth et al. 2016). Typically, these human activities are associated with the Industrial Revolution and the Great Acceleration (Lewis and Maslin 2015), but likely go back much further (Mottl et al. 2021). Indeed, current anthropogenic impacts are now seen in the context of a period of disturbance going back thousands of years (Mottl et al. 2021). Since the Great Acceleration however, effects on the Earth have increased rapidly (Steffen et al. 2015). Over the past ~40 years, average global temperature has increased at a rate of 0.2°C per decade (IPCC 2018). This upward trend is projected to continue and accelerate, with a potential 4°C rise by 2100 (IPCC 2018). (Models also indicate precipitation will experience regional changes, leading to novel conditions of both temperature and aridity (IPCC 2018). Anthropogenic land use has also grown significantly, with more than 75% of the terrestrial world now showing evidence of historical or current transformation (Ellis and Ramankutty 2008), and just over 50% currently used by humans (Le B. Hooke, et al. 2012). And land-use will continue to grow, particularly in the tropics where agriculture is expected to expand quickly over the coming decades (Laurance et al. 2014).

The consequences of human activity for biodiversity have been great. The clearest effects are seen in species extinction rates, currently thought to be ~100-1000 times the background rate (Pimm et al. 1995), which has been said could lead to a sixth mass extinction (Ceballos et al. 2015). Specifically, 28% of all assessed species are currently threatened with extinction (IUCN 2021), and at least 778 animal species have gone extinct since 1500 (IUCN 2021). Most likely this total is an underestimate, given the difficulties associated with recording a species as extinct (Dirzo et al. 2014), and the many species we know very little or nothing about (Pimm et al. 1995). Local change is more complex, owing to the multiple ways in which local biodiversity can be measured, and the complexities associated with taxonomic, spatial, and

temporal scale (Gonzalez et al. 2016). Space-for-time models of local biodiversity indicate that anthropogenic land use tends to reduce site-level species richness and total abundance, and homogenise biodiversity between sites (Newbold et al. 2018; Newbold et al. 2015a). Assemblage trends indicate that local diversity has experienced no net change and community composition a reduction, both since the late 1800s (Dornelas et al. 2014). Vertebrate population trends specifically have tended to decrease on average since the 1970s (McRae et al. 2017), but again there is variation in the direction and magnitude of change among taxa and geographic regions (Leung et al. 2020)

Animal pollinators have not escaped the effects of the Anthropocene. A number of recent studies have found evidence of change in pollinator biodiversity, typically referred to in the literature as “pollinator decline” (Lever et al. 2014; Garibaldi et al. 2009; Gallai et al. 2009), and in the media as an “ecological Armageddon” (Embury-Dennis 2017). For the evidence currently available, the term “pollinator decline” is well-founded, at least for average change in abundance, species richness, and biomass (Ollerton 2017; see below for a summary of the evidence). But there are still quite profound differences among taxonomic groups and geographic regions (Powney et al. 2019), and data still tends to be biased towards select taxonomic groups in North America and Europe (Saunders et al. 2020). For “pollination crises”, it’s an open question as to whether declines yet warrant this term (Ghazoul 2005; Ghazoul 2015), or if the term “crisis” could still be warranted if global pollinator biodiversity change is found to be highly heterogeneous. Indeed, for the services pollinators provide in the form of crop pollination and ecosystem functioning, a negative net-change in some metric of biodiversity doesn’t necessarily have to mean a negative impact on people.

One of the core distinctions for pollinator biodiversity change is between domesticated and wild pollinators (Ollerton 2021; Potts et al. 2010), both in terms of the difference in change and the extent to which we can be confident that our evidence of change is representative. Domesticated pollinators refer to those that humans deliberately rear for the production of honey and the pollination of crops, and wild pollinators those that have no history of deliberate

and direct human use (IPBES 2016). For domesticated pollinators, that typically means the western honey bee (*Apis mellifera*; Linnaeus, 1758) (Potts et al. 2016). The western honey bee is a highly important generalist crop pollinator, providing much of the pollination service in many crop species (Rucker et al. 2012). The convenience of the western honey bee as a pollinator is such that hives are brought onto cropland in large numbers to meet pollination demand and maximise crop yield (Degrandi-Hoffman et al. 2019). For example, in the USA, hives are shipped into California in February for almond pollination, then onto Oregon and Washington for apple, cherry, and pear pollination, and then by May-June for pollination of crops such as strawberries and raspberries (Rucker et al. 2012). The importance of the honey bee to agriculture means it has reasonably good estimates of biodiversity change, typically inferred from the total number of hives (Aizen and Harder 2009; vanEngelsdorp and Meixner 2010). Total hive numbers indicate that the western honey bee is not globally declining (Aizen and Harder 2009), despite fears that colony collapse disorder—a quite sudden breakdown of the colony driven by a number of interacting factors (van Engelsdorp et al. 2009)—in European and North American colonies could become more widespread (Williams et al. 2010). In fact, the total number of honey bee hives increased by ~45% between 1961 and 2007 (Aizen and Harder 2009). There is again heterogeneity in this change, with decreases in some parts of Europe and North America, and large gains in China (Aizen and Harder 2009), but taken together it appears that honey bee populations are not currently of global concern.

Unlike domesticated pollinators, the extent to which wild pollinator biodiversity has changed is less clear, but from research that does exist it is reasonable to conclude that wild pollinators have undergone more extensive change than domesticated pollinators, and in many cases decline (Powney et al. 2019; Burkle et al. 2013; Biesmeijer et al. 2006; Hallmann et al. 2017). Uncertainty in the existence and magnitude of pollinator declines stems in part from the much greater range of taxonomic groups involved (Wardhaugh 2015; Ollerton 2017), making it both a bigger task to evidence change in all groups, and far more likely that there will be differences in the direction and magnitude of change. Moreover, it's also true that wild insect pollinators

tend to experience significant inter-annual population fluctuations, meaning detecting real population change requires research intensive, long-term time series (Roubik 2001). Nevertheless, despite these uncertainties and difficulties, the drivers of wild pollinator biodiversity change are at least now well-known. Change in land cover, chemical application (e.g. pesticides, fertilisers, herbicides, and fungicides), disease, domesticated pollinator management, the introduction of invasive species, and climate change have all been linked to wild pollinator biodiversity change to varying degrees (IPBES, 2016).

Evidence for wild pollinator decline has been found in a number of groups, and continues to grow. Pollinator declines in Europe and North America are some of the best studied. 3 bumblebee species in the UK, and at least 4 species across 11 European countries, have become extinct (Kosior et al. 2007; Goulson et al. 2008). In Britain and the Netherlands, wild bee species richness declined significantly in ~52% and ~67% of spatial cells between a set of observations pre-1980 and a set of observations post-1980 (Biesmeijer et al. 2006). In the UK, such richness declines have occurred in conjunction with bee and flower-visiting wasp extinctions, which began increasing rapidly in the 1920s following a period of agricultural expansion (Ollerton et al. 2014). Also in the UK, butterflies have experienced significant net losses, with a median change in occupancy of -13% (Thomas et al. 2004). For UK pollinators more broadly, occupancy models suggest a net loss of wild pollinator range, with the most significant losses seen in rare species (Powney et al. 2019). Declines have also been reported for biomass, particularly in Germany where flying insect biomass was found to have declined by 75% between 1989 and 2016 (Hallmann et al. 2017). Such declines are concerning, but apparently do not represent a consistent trend for all regions and taxonomic groups, given that moth biomass change in the UK was found to have no difference in mean biomass between the 1960s and 2010s (Macgregor et al. 2019). In North America, change in plant-pollinator interaction networks between 1888 and 2010 was associated with a significant reduction in bee species richness (Burkle et al. 2013), and a study of historical bumblebee records found significant reductions in range and relative abundance (Cameron et al. 2011). Bumblebees in

particular have been associated with strong climatic effects, with frequency of warmer temperatures predicting extinction risk and change in local species richness (Soroye et al. 2020). Given the importance of insect pollinators to crop pollination (Potts et al., 2010), insects have typically been the focal point of study, but vertebrate pollinators have also been found to have declined. Indeed, bird and mammal pollinators have on average moved 1 IUCN category towards extinction every 2.5 years, although on average this is still less threatened than non-pollinators (Regan et al. 2015).

Outside western Europe and North America long-term data on pollinator biodiversity change are fewer, but for studies that do exist declines also represent the average trend. In Argentina significant reductions in range for native bumblebees have occurred, driven in part by competition from *Bombus terrestris* (Linnaeus, 1758), a species introduced in much of the world for crop pollination (Morales et al. 2013). A similar process of bumblebee homogenisation has occurred historically in a number of other locations, including but not limited to Japan and Tasmania (Matsumura et al. 2004; McQuillan and Hingston 1999). In China, wild bee abundance, diversity, and occupancy have fallen in a number of agricultural localities (Teichroew et al. 2017; Williams and Osborne 2009). In South Africa, declines in some pollinating bees have been inferred from reduced orchid seed set (Pauw 2007). Outside North America and Europe, hoverfly, butterfly, and vertebrate pollinator trends represent a significant gap, perhaps with the exception of bat and bird island extinctions (Ollerton 2017). Such island extinctions have been reported for birds in Hawaii, and two bat species in New Zealand (Ollerton 2017).

Average trends in biodiversity change hide variation among taxonomic groups, with some groups of wild pollinators experiencing increases in biodiversity (Powney et al. 2019; Biesmeijer et al. 2006). Hoverflies, for example, experienced no significant change in richness in Britain between pre- and post- 1980, and a significantly increasing trend in the Netherlands (Biesmeijer et al. 2006). Dominant crop pollinators also appear to be more resilient (e.g.

Bombus terrestris, *Osmia Bicornis*) with an increase in occupancy from 1980-2013 (Powney et al. 2019).

The drivers of pollinator biodiversity change

Of the threats described by IPBES (2016), wild pollinator biodiversity change has primarily been linked to land use, land-use intensity (including the effects of pesticide application), climate change (Kerr et al. 2015; Kerr 2001; Goulson et al. 2015), and the interaction between land-use and climate change (Dalsgaard 2020). Here I will briefly review the effects of land-use, land-use intensity, and the interaction between land-use and climate change. I will not discuss climate change alone in detail, since this is not directly relevant to the aims of this thesis. I will also not discuss the effects of anthropogenic activity on domesticated pollinators (e.g. colony collapse disorder), since these have been studied extensively elsewhere (e.g. van Engelsdorp et al. 2009) and are again not directly relevant to the aims of this thesis (i.e. large-scale, cross-taxa effects on wild pollinators).

A number of studies have documented a negative impact of anthropogenic land use on wild pollinator populations (Tscharntke et al. 2002; Kennedy et al. 2013; IPBES 2016; Goulson et al. 2015; Klein et al. 2003; Ricketts et al. 2008). Such effects have typically been associated with reduced floral availability on managed land (Goulson et al. 2015). Modern agricultural practice often means monoculture cropland and infrequent crop-rotation, resulting in land which provides a lower diversity of the floral resources required by pollinating species (Goulson et al. 2005). Fragmentation of the land through disturbance will also be detrimental to some species (Tews et al. 2004). Patchy habitat isolates populations through acting as a barrier to migration (Saunders et al. 1991), leaving pollinators less able to exploit neighbouring floral resources.

Some studies have shown positive effects of land-use change, finding that moderate anthropogenic pressure and the creation of more open and diversified habitats can also benefit some pollinators (Winfrey et al. 2011). Such effects likely relate to two key factors: first, that

at low levels of intensity, human activity can increase niche diversity by making habitats more heterogeneous (Baldock et al. 2015; Tschardt et al. 2002; Tews et al. 2004; Hagen and Kraemer 2010; Winfree et al. 2011); and second, that some species happen to be better able to adapt to the environments humans create, with these species tending to move into anthropogenic habitats (McKinney 2006). The conjunction of these two points means increased pollinator biodiversity for some groups on low-intensity managed land. For example, some bees, butterflies, syrphid flies, and nectarivorous pollinating birds have been found to favour open, less forested areas of semi-natural grassland or agroforestry (Deans et al. 2007; Kuussaari et al. 2007; Michener 2000; Tschardt et al. 2008). Globally however, for many pollinators it's unclear the extent to which the direction and magnitude of response to land use differs among biodiversity metrics, taxonomic groups, scale, and geographic regions.

Climate change has been highlighted in the literature as a concern for the current and future prospects of pollinating animals. IPBES (2016) summarised the effects of climate on pollinators at three levels: spatial range shifts, changes in abundance, and shifts in seasonal and daily activity patterns. These changes are well-known in butterflies (Chen et al. 2011; Parmesan 2007) and bumblebees (Kerr et al. 2015), but less well understood in hoverflies, bats and birds (see IPBES (2016) for a thorough review of the evidence). Climate change is also thought to interact multiplicatively with land-use, such that the effects of climate change and land use are greater together than individually (Oliver and Morecroft 2014). Such interactive effects are thought to occur through two means: first, through one driver altering the magnitude of another; and second, through the effect of one driver on biodiversity being moderated by change in another (Oliver et al. 2014). Recent research has begun to investigate these potential interactions broadly (Mantyka-Pringle et al. 2011; Spooner et al., 2018), and in the context of pollination services specifically (Marshall et al. 2017), but our understanding is far from complete (see Chapter 4).

The extent to which a species is impacted by environmental change will depend on the traits of that species (Marini et al. 2014; Vanbergen 2014), since traits are indicative of a species'

ability to adapt to a new environment. Responses to environmental change will also be experienced indirectly, through effects at the level of interactions among species (Lever et al. 2014). Such effects refer to concepts of trophic cascade and co-extinction, both of which don't directly relate to individual response characteristics, but rather the strength of the connection between one species and another (Colwell et al. 2012; Polis et al. 2000). At the trait level, body size, dietary and habitat specialism, mobility (wing length), tongue length, nesting behaviour, and sociality have all been highlighted as important traits in determining pollinator response to environmental change (Winfree et al. 2011; Goulson et al. 2015; Öckinger et al. 2010; De Palma et al. 2015; Newbold et al. 2013; Newbold et al. 2014b; Burivalova et al. 2015; Montero-Castaño and Vilà 2012). However, there is an ongoing challenge associated with broad cross-taxa trait-based predictions (Bartomeus et al. 2017). Indeed, for these broader studies, with the exception of specialism and body size, traits appear to weakly predict overall response to land-use (Bartomeus et al. 2017). Although trait-based research represents an interesting area of work, in the remainder of this thesis I do not consider the predictive nature of these traits specifically, meaning I will not discuss them in further detail here. I will return to them in Chapter 3, since across taxonomic groups traits likely still underpin response to anthropogenic activity.

Public awareness of pollinator biodiversity change

In tackling pollinator biodiversity change, public awareness is a key factor. Through high-carbon lifestyles and dietary choices, and more indirectly through the way in which the public votes, the choices people make continue to have a significant impact on biodiversity change (McCrea et al. 2015; Machovina et al. 2015; Garnett 2013). Admittedly the link between awareness and behaviour can be weak, but nevertheless awareness is likely still important to some extent in leveraging behaviour change (Asvatourian et al. 2018). Aichi Biodiversity Target 1 encapsulates this need, aiming that “by 2020, at the latest, people are aware of the values of biodiversity and the steps they can take to conserve and use it sustainably” (CBD 2011). However, although important to recognise, quantifying progress on this target at the

global level is difficult. In part the problem has been the logistical and financial difficulties in surveying the opinions of large groups of people globally (Leadley et al. 2013), and the biases associated with such an approach (Mcowen et al. 2016).

One emerging option for the quantification of biodiversity awareness, and pollinator awareness more specifically, is in the field of conservation culturomics (Ladle et al. 2016). Coined by Michel et al. (2011) in a study in which the authors built a dataset comprised of over 5 million books, culturomics refers to the process of extracting insights from large bodies of text. Later Ladle et al. (2016) introduced the term to conservation, suggesting that culturomics could be used to quantify changes in public biodiversity interest. The digital age of the internet is largely driving interest, with Google Trends, Google Books, Twitter, Weibo, and website crawling all having been suggested as potential data sources (Ladle et al. 2016). Recently efforts have been made to calculate changes in biodiversity awareness at the global scale using Twitter, online newspapers, and Google Trends (Cooper et al. 2019). However, this study focused on awareness of conservation issues rather than species themselves, meaning there is still not a means through which to quantify awareness of biodiversity itself. For pollinators therefore, in the first instance we need a metric with which to quantify change in awareness for all species, which we can then use to focus in on changes in pollinator awareness.

Thesis aim

The main aim of my thesis is to understand better the causes of global pollinator biodiversity change, and the potential consequences for future crop pollination risk. To this end, my thesis will address four key research questions, with each question addressed in one of four core data chapters: 1) how disparate temporally, taxonomically, and geographically is the information available in the animal pollination literature, and what approaches can we use to utilise it better? 2) to what extent is land-use type and intensity associated with changes in local pollinator biodiversity (species richness, total abundance, and diversity) at the global

scale? 3) to what extent are interactive effects between land-use type (cropland and primary vegetation) and climate change associated with changes in local pollinator abundance, and how might these interactions relate to potential crop pollination risk in the future? and 4) using Wikipedia page view data, how might we measure changes in public awareness of biodiversity itself, and what can we learn from these metrics about change in awareness of pollinators and the value of biodiversity?

Chapter 2: Text-analysis reveals taxonomic and geographic disparities in animal pollination literature

Abstract

Ecological systematic reviews and meta-analyses have significantly increased our understanding of global biodiversity decline. However, for some ecological groups, incomplete and biased datasets have hindered our ability to construct robust, predictive models. One such group consists of the animal pollinators. Approximately 88% of wild plant species are thought to be pollinated by animals, with an estimated annual value of \$230-410 billion dollars. Here I apply text-analysis to quantify the taxonomic and geographical distribution of the animal pollinator literature, both temporally and spatially. I show that the publication of pollinator literature increased rapidly in the 1980s and 1990s. Taxonomically, I show that the distribution of pollinator literature is concentrated in the honey bees (*Apis*) and bumble bees (*Bombus*), and geographically in North America and Europe. At least 25% of pollination-related abstracts mention a species of honey bee and at least 20% a species of bumble bee, and approximately 46% of abstracts are focussed on either North America (32%) or Europe (14%). Although these results indicate strong taxonomic and geographic biases in the pollinator literature, a large number of studies outside North America and Europe do exist. I then discuss how text-analysis could be used to shorten the literature search for ecological systematic reviews and meta-analyses, and to address more applied questions related to pollinator biodiversity, such as the identification of likely interacting plant-pollinator pairs and the number of pollinating species.

Introduction

The number of publications and journals in the academic sciences is vast and continuing to increase (Ferreira et al. 2015). The field of ecology and biodiversity is no exception. Between 1990 and 2014, the total number of ecological research articles increased more than tenfold, from fewer than 10,000 in 1989 to at least 125,000 in 2014 (Nunez-Mir et al. 2016). In conjunction with this increase, digitisation of the literature, indexing tools (such as Scopus, Web of Science, and Google Scholar), and the research structures of systematic review and meta-analysis have all become standard practice (Lortie 2014, Gurevitch et al. 2018). Understanding of global biodiversity decline in particular has reaped the benefits of these changes (Loh et al. 2005, Butchart et al. 2010, Pereira et al. 2010, Tittensor et al. 2014, Newbold et al. 2015a). However, for some important ecological and taxonomic groups, incomplete and biased datasets have hindered our ability to construct robust, predictive models (De Palma et al. 2016, Bartomeus et al. 2018).

One such important ecological group consists of the animal pollinators, animals that act as a vector for the transfer of pollen from the male to the female reproductive parts of a flowering plant, causing fertilisation and the production of a fruit and seed (Proctor et al. 1996). Animal pollination is highly important, especially in tropical humid and warm environments where approximately 95% of flowering plant species are animal pollinated (Rech et al. 2016). Globally, approximately 88% of wild plant species are thought to be pollinated by animals (Ollerton, Winfree and Tarrant 2011), providing an ecosystem service valued at \$230-410 billion dollars per annum (Lautenbach et al. 2012). Although disputed by some on the basis of taxonomic and geographic biases in the data used (Ghazoul 2005, Ghazoul 2015), many papers have suggested that pollinators are declining in the face of several environmental pressures (Biesmeijer et al. 2006, Steffan-Dewenter and Westphal 2007, Potts et al. 2010b, Winfree, Bartomeus and Cariveau 2011, Goulson et al. 2015, Woodcock et al. 2016).

IPBES (2016) summarised the anthropogenic threats to pollinators as change in land cover, chemical application (pesticides, fertilisers, herbicides, and fungicides), disease, pollinator management, the introduction of invasive species, and climate change. Through the interacting effect of these threats, populations of wild invertebrate pollinators have declined (see Ollerton (2017) for a summary of the evidence for pollinator decline), although we know little about the status of wild pollinators outside North America and Europe (IPBES 2016). In Britain and the Netherlands, wild bee species richness has declined over ~50-70% of the total land area (Biesmeijer et al. 2006). Bumble bee declines are some of the best studied, with at least 3 bumble bee species having gone extinct in the UK, and at least 4 species across 11 European countries (Kosior et al. 2007, Goulson, Lye and Darvill 2008). Regional colony losses in European (1985-2005) and USA (1947-2005) honey bees have also been well documented (Stokstad 2007, Potts et al. 2010a), at 25% and 60% respectively, despite a global increase in managed colonies (Aizen and Harder 2009).

Predictive models are important in understanding pollinator biodiversity change, but are a challenge to implement robustly. This difficulty is in part driven by the geographical and taxonomic distribution of available pollinator biodiversity data (De Palma et al. 2016). Pollinators are represented across a variety of taxonomic groups, including bats, birds, and multiple insect taxa, but many of the key syntheses of pollinator decline have been restricted to the bees of North America and Europe (Winfree, Bartomeus and Cariveau 2011, Ghazoul 2015, Goulson et al. 2015, De Palma et al. 2016), but see Regan et al. (2015) for a global study on the status of mammal and bird pollinators. Although widely accepted, the degree of this bias and the extent to which it might influence biodiversity models is uncertain (Ghazoul 2005, Ghazoul 2015, De Palma et al. 2016, Ollerton 2017). Some studies have made progress towards quantifying the geographical or taxonomic distribution of the animal pollination literature (Archer et al. 2014, Ollerton 2017), but the way in which taxonomy interacts with spatial distribution globally has not to our knowledge been the subject of a thorough review. This lack of research is in part a symptom of article indexing tools, which despite their

contribution, still have significant limitations, a problem not confined to the animal pollination literature (Westgate et al. 2018b, Westgate and Lindenmayer 2016, Westgate et al. 2015). Indexing search tools such as Scopus do have functions to account for differences in spelling (fuzzy-matching), and variable suffixes for the same family of words (stemming), but searching for geographical and taxonomic names and identifying overall text topic, is only possible through discrete search terms and phrases. As a result, returning literature fully representative of a particular theme, geographical region, or taxonomic group is difficult to accomplish, given the semantic ambiguity of search terms across academic fields (Westgate and Lindenmayer 2016, Roll, Correia and Berger-Tal 2017). In the context of the animal pollination literature, better tools for extracting pollinator information could be used as the basis for more taxonomically and geographically representative meta-analyses and systematic reviews, in turn increasing the robustness of synthetic analyses.

Text-analysis could help to mitigate the problem of biased and incomplete pollinator response data. Also often called text-mining, text-analysis refers to the automated extraction of information from large volumes of text (Cohen and Hunter 2008), most notably across multiple documents (Griffiths et al. 2004, Grimmer and Stewart 2013, Westgate et al. 2015). Given the very large numbers of published papers containing potentially useful information, such technologies are invaluable in automatically drawing together results across lots of studies (Grimmer and Stewart 2013), thereby reducing the duration of the “synthesis gap”, or in other words the lag between the practice of science and the synthesis of evidence (Westgate et al. 2018b). Text-analysis tools can be used to optimise the systematic review and meta-analysis literature search path. For example, topic categorisation algorithms can be used to allocate articles automatically to particular fields of study, enabling the curator to discard articles of low relevance (O'Mara-Eves et al. 2015, Westgate 2018a). In the context of pollinator data, the application of such tools could increase recall of the relevant literature and decrease the effort required, in turn reducing data biases in systematic reviews and meta-analyses. These tools could be particularly beneficial for pollinator literature, given the effect of taxonomic and

geographic biases on predictions of pollinator response to human land-use (De Palma et al. 2016). Here I briefly discuss two tools that could be particularly useful in the context of pollination ecology: geoparsing and taxonomic entity extraction.

Geoparsing allows place names in text to be identified, resolved, and assigned geographical coordinates (Leidner and Lieberman 2011). Geoparsing can therefore be broken into two steps: firstly, the identification of geographical mentions (known as toponyms); and secondly, the resolution of mentions as the most likely physical coordinates (D'Ignazio et al. 2014). The first step is a key obstacle; the problem being semantic (D'Ignazio et al. 2014). Identical words can be used to describe both place and non-place information, interpretable only in the context the term is written (Leidner and Lieberman 2011). For example, the words “Rio” and “Alexandria” could be used to describe both a geographic location and the name of a person. High performance machine learning algorithms will therefore attempt to resolve locations through contextual information (Leidner and Lieberman 2011, Gritta et al. 2018). CLIFF-CLAVIN is one such tool (D'Ignazio et al. 2014, Gritta et al. 2018). An open-source geoparser, CLIFF-CLAVIN was developed for extracting geographical information from news articles (D'Ignazio et al. 2014). CLIFF-CLAVIN also has an implementation of focus, meaning it attempts to resolve the primary country location of a given piece of text, even when the country is not mentioned (D'Ignazio et al. 2014). CLIFF-CLAVIN estimates focus on the basis of the most frequently mentioned country, and in the absence of country mentions, the frequency of specific locations within countries (D'Ignazio et al. 2014). CLIFF-CLAVIN will attempt to find geographical locations from the local to continental level. For example, CLIFF-CLAVIN is able to find correctly the records “Krakatoa”, “Sumatra”, “Indonesia”, and “Asia”. Although still in the early stages of development, and requiring significant improvements in accuracy and speed (Gritta et al. 2018), geoparsers have previously been used to identify the main geographical location of news reports (Imani et al. 2017), to geotag museum specimens (Beaman and Conn 2003), and to digitise historical maps (Chiang 2017).

Taxonomic entity extraction refers to the identification of taxonomic names (in theory of any taxonomic rank) from blocks of text (Sarkar 2007). Such algorithms tend to use taxonomic dictionary string matches, rule-based inference, and machine learning, either independently or in combination (Akella, Norton and Miller 2012). Dictionary match algorithms search for each word (unigram) and pair of words (bigram) in a taxonomic database such as NameBank (Leary et al. 2007), returning a record if the strings match. Similarly, rule-based inference searches for regular expressions indicating a form often associated with a species record, such as bigram capitalisation and abbreviation. Machine learning approaches identify text likely to represent a taxonomic record, inferring from both context and string structure (Akella, Norton and Miller 2012). The R package ‘taxize’ has implementations for two of these algorithm types in the function `scrapenames`: dictionary string match (Taxonfinder) and machine-learning (Neti Neti) (Chamberlain and Szöcs 2013). Scrapenames will search for strings resembling taxonomic records at any taxonomic rank, including abbreviated records, hybrids, and higher taxa. For example, scrapenames is able to correctly find the records “*A. manicatum*”, “*Apidae*”, and “*Viburnum macrocephalum* f. *Keteleeri*”. Many authors have emphasised the value of extracting taxonomic information (Sarkar 2007, Guralnick and Hill 2009, Parr et al. 2012, Thessen, Cui and Mozzherin 2012), and others its associated methodological difficulties (Correia et al. 2018), but to my knowledge few studies have explored potential applications.

Here I demonstrate how CLIFF-CLAVIN and taxize can be used in combination to quantify the taxonomic and geographical distribution of the animal pollinator literature. This analysis builds on the reviews of Archer et al. (2014) and Ollerton (2017), introducing new text-analysis methods, examining temporal changes in pollinator publications, and investigating the interaction between the taxonomic and geographical distributions of pollinator studies. I reveal disparities in pollinator literature, reinforcing the problem of biases in the context of pollinator biodiversity modelling. Finally, in exploring future directions, I discuss how these tools may be used to decrease data biases in biodiversity meta-analyses and systematic reviews. I

summarise by emphasising two different although related points: firstly, the geographical and taxonomic distribution of the animal pollination literature is indeed highly concentrated in North America and Europe in the honey bees and bumble bees, although many studies do exist for other species and geographic regions; and secondly, the development of text-analysis tools shows significant promise in optimizing the search for information on animal pollination, both through capturing data on underrepresented regions and taxa, and through speeding up the search process.

Methods

I scraped the pollination literature for mentions of animal species and location data to investigate the taxonomic, geographical, and temporal distribution of studies on animal pollination. I considered any primary research article published in English returned through a search for the term “pollinat*” in Scopus, that mentioned an animal species in the abstract. Animal species scraping and geographical entity extraction were accomplished through a methodology built on the ‘taxize’ R package and the geoparser CLIFF-CLAVIN. My rationale for applying this semi-automatic approach, rather than manually checking all abstracts, was that identifying all Latin binomials and geographic locations would not be feasible given the volume of text. I first describe the methodology applied (see Appendix 1 for additional validation), before discussing change over time, the taxonomic breakdown of the animal pollinator literature, overall geographical distribution of information, and finally geographical distribution for individual taxonomic groups.

Taxonomic extraction

I queried Scopus using the stemmed term “pollinat*” (28/03/18 - 29/03/18), before subsetting for primary research articles in English (Appendix 1, Figure S1.1). Duplicated records were filtered out by removing duplicated titles. Any records without titles were also removed. I then retained any papers with abstracts that mentioned a taxonomic name, applying in conjunction both the Neti Neti and Taxonfinder algorithms implemented in the package ‘taxize’ (Appendix 1, Figure S1.1). Taxonfinder represents a dictionary match algorithm, searching against dictionaries of multiple taxonomic levels (built from NameBank¹) for potential taxonomic records. The Neti Neti algorithm applies a machine learning approach to extract strings deemed likely to be taxonomic records.

¹ <http://ubio.org/index.php?pagename=namebank>

I then carried out a series of data-cleaning steps to identify Latin binomial animal species within our initial scrape. I chose to use the Catalogue of Life (COL) in validating species records as animal species, due to its greater coverage (84% of all described species: Roskov et al. 2017). Animal species were validated by performing character string matches against the Latin binomials of a Metazoan subset of the COL (Appendix 1, Figure S1.1). Matches with the COL were carried out at two levels (1 and 2). Level 1 represented direct string matches, and level 2 any matches with an abbreviated record. Within level 1, I distinguished a series of sub-levels, reflecting the approach used to resolve a record as an animal species: level 1a represents any direct match with the original string, level 1b any direct match following the removal of punctuation, and level 1c any direct match following the removal of punctuation and the string “spp”. An abbreviated record refers to an abbreviated genus and full species picked up by taxize. For example, *Apis mellifera* would be abbreviated in the form *A. mellifera*. I also encountered problems regarding accepted and synonymous species names. In attempting to mitigate this issue, I substituted any record picked up by the COL as a synonym with its corresponding accepted name. For any further analysis, I then worked only from accepted names.

For all studies related to pollination that mention an animal species, I initially calculated change over time in publication of pollination related studies mentioning an animal species (‘Temporal distribution’). I then investigated change over time in mentions of the animal genera *Apis*, *Bombus*, and all other pollination-related animal genera (henceforth “other genera”). I opted to cluster all other genera to test the hypothesis that publications concerning *Apis* and *Bombus* species are largely responsible for the rapid increase in pollination-related papers.

I also calculated the frequency of taxonomic mentions at the level of genera (‘Taxonomic distribution’, Appendix 1, Figure S1.2). I opted to investigate the frequency of mentions at the level of genera given the hypothesized dominance of studies on *Apis* and *Bombus* species. I included only those genus names associated with a Latin binomial, given the increased ambiguity with just genus names. For example, *Prunella* is both a genus of plants and a genus

of birds. I presented taxonomic mentions as a raw count rather than a proportion since mentions of different genera are not necessarily independent of one another: one study abstract may mention multiple genera (see Millard et al. 2020b for the full list of pollination related animal genera).

Geoparsing

Following the verification of animal species, I then anchored each abstract mentioning an animal species to a geographic location, using an approach called geoparsing. Geoparsing refers to the resolution of ambiguous free-text place name descriptions as specific geographic coordinates. Not all abstracts will mention a location, but I assume that those that did were representative of the geographic distribution of the animal pollination literature as a whole.

I chose to use the open-source geoparser CLIFF-CLAVIN, due to the high accuracy of its focus implementation, relative to commercial geoparsers such as Yahoo Placespotter and OpenCalais. The main focal country for a given text will henceforth be referred to as “major” mentions, and any specific locations found in an abstract as “minor” mentions. “Minor” mentions can therefore be of any geographical scale, from the continental to the local level. I used vagrant—a software tool for leveraging virtual environments—and the GitHub repository CLIFF-up to host CLIFF-CLAVIN (<https://github.com/ahalterman/CLIFF-up>).

After geoparsing the pollination-related abstracts, I carried out a series of verification steps to improve the quality of the data. First, I plotted the data on a global map to check for any unusual-looking patterns, which revealed that continental “minor” mentions were distorting the apparent geographic distribution of studies. For example, the continental “minor” mention ‘Europe’ appeared in a number of abstracts, which CLIFF-CLAVIN had assigned to a single coordinate in central Europe. I was also not interested in oceanic “minor” mentions, since these would not relate to the study of terrestrial animal pollinators. Before proceeding with any further analysis, I therefore removed any continental or oceanic “minor” mentions. Second, in initial runs of CLIFF-CLAVIN I also noticed that the geoparser was picking up geographic

information associated with copyright details, typically included at the end of the Scopus abstract following a copyright symbol. I therefore removed any characters following the copyright, before rerunning the geoparser. Third, after removing any low-resolution geographic locations, I then visually inspected the whole raw dataset, searching for any place names that were either questionable or clearly wrong. For example, place names that had been incorrectly disambiguated by CLIFF-CLAVIN, such as “Ivory” and “Hay Meadows”, as well as locations that seem overly specific or strange, such as “Blue Ridge Parkway Milepost 234 ca. 886 m”. Fourth, given CLIFF-CLAVIN was originally trained on news articles, I was aware that performance might be reduced when applied to academic texts, particularly those mentioning Latin binomial species names. For example, the “Linnaeus Terrace”, a rock terrace in the Antarctic, was incorrectly identified from Linnaeus, the species authority, while taxonomic names such as “Peia” and “Pavonia” were also incorrectly identified as place names. I therefore also manually inspected any locations that could have been referred to in text as a species or genera, and removed any that we identified as mistakes. After the removal of text following the copyright symbol, CLIFF-CLAVIN identified geographic locations in 2087/3974 (53%) of the pollination-related abstracts containing an animal species (Appendix 1, Figure S1.1). After further verification of the geoparsed data, geographic locations were identified in 2072 abstracts (Appendix 1, Figure S1.1), meaning 2072/2087 (~99%) of those abstracts with a location contained a usable sub-continental geographical location.

I calculated a study count through counting the number of “major” mentions coordinates within each set of country polygons, and a study density by dividing this value by the area of those polygons (Appendix 1, Figure S1.3). Beforehand I removed duplicated study-country combinations for both “minor” and “major” mentions, accounting for abstracts mentioning a given location more than once. Given focus describes the algorithm’s estimation of the main geographic focus of an abstract, I reasoned that “major” mentions would provide an indication of primary study location. Due to the highly right-skewed distribution of the country study

counts, I $\log_{10}$ transformed the values. “Minor” mentions were also plotted onto this map, with the size of each point representing the number of unique study-location combinations.

For abstracts mentioning an animal species and a geographic location, I examined the way in which taxonomy and geographic location interact. I assumed that all geographic locations mentioned within a given abstract related directly to all animal species mentioned within that same abstract. Consistent with our investigation of overall taxonomic distribution, I examined taxonomy-geography interaction for Latin binomial species at the level of genera.

Results and Discussion

Temporal distribution

Over time, the number of studies on pollination has increased substantially, with a particularly rapid increase beginning in the mid-1990s and 2000s (Figure 2.1), occurring in conjunction with widespread incidences of colony collapse disorder (CCD) in the early 2000s (van Engelsdorp et al. 2008, Genersch et al. 2010). Much of this increase can be attributed to studies of *Apis* and *Bombus* species (Figure 2.2, Appendix 1, Table S1.4). From 1980-2017, the number of studies mentioning *Apis* and *Bombus* species increased non-linearly. The rapid increase for *Apis* coincided with the introduction of the parasite *Varroa destructor* to the United States in the 1980s (Oldroyd 1999), and for *Bombus* with the first commercialisation of *Bombus* pollination in the late 1980s (Velthuis and van Doorn 2006). Over this period, the general trend in publication number for other pollination-related genera increased marginally, with a slight non-linear increase from the year 2000. Given that *Apis* and *Bombus* are often referred to by their common name, it is likely that I underestimate the disparity in publication rate between *Apis*, *Bombus*, and other genera (see 'Limitations' and Appendix 1, Figure S1.5).

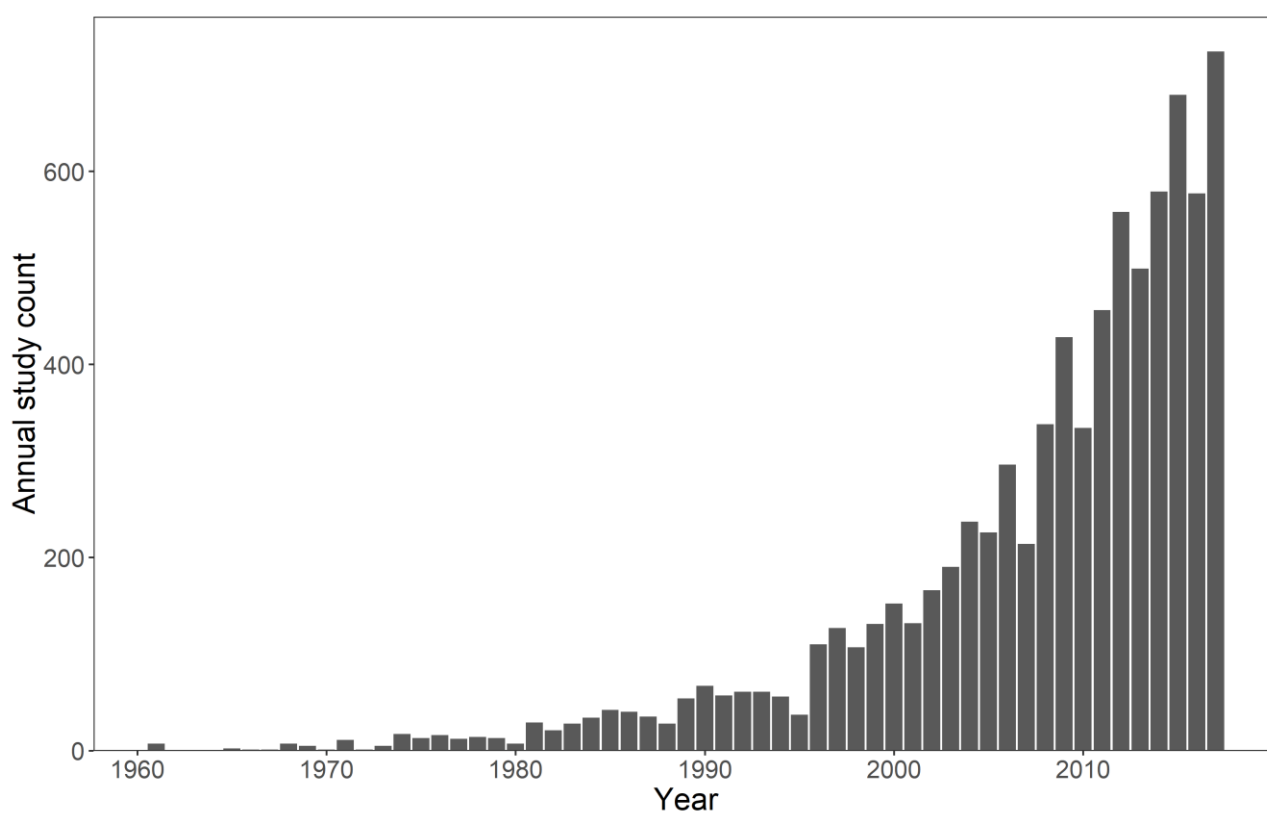


Figure 2.1 Annual study count for pollination-related studies mentioning an animal species, years 1961-2017.

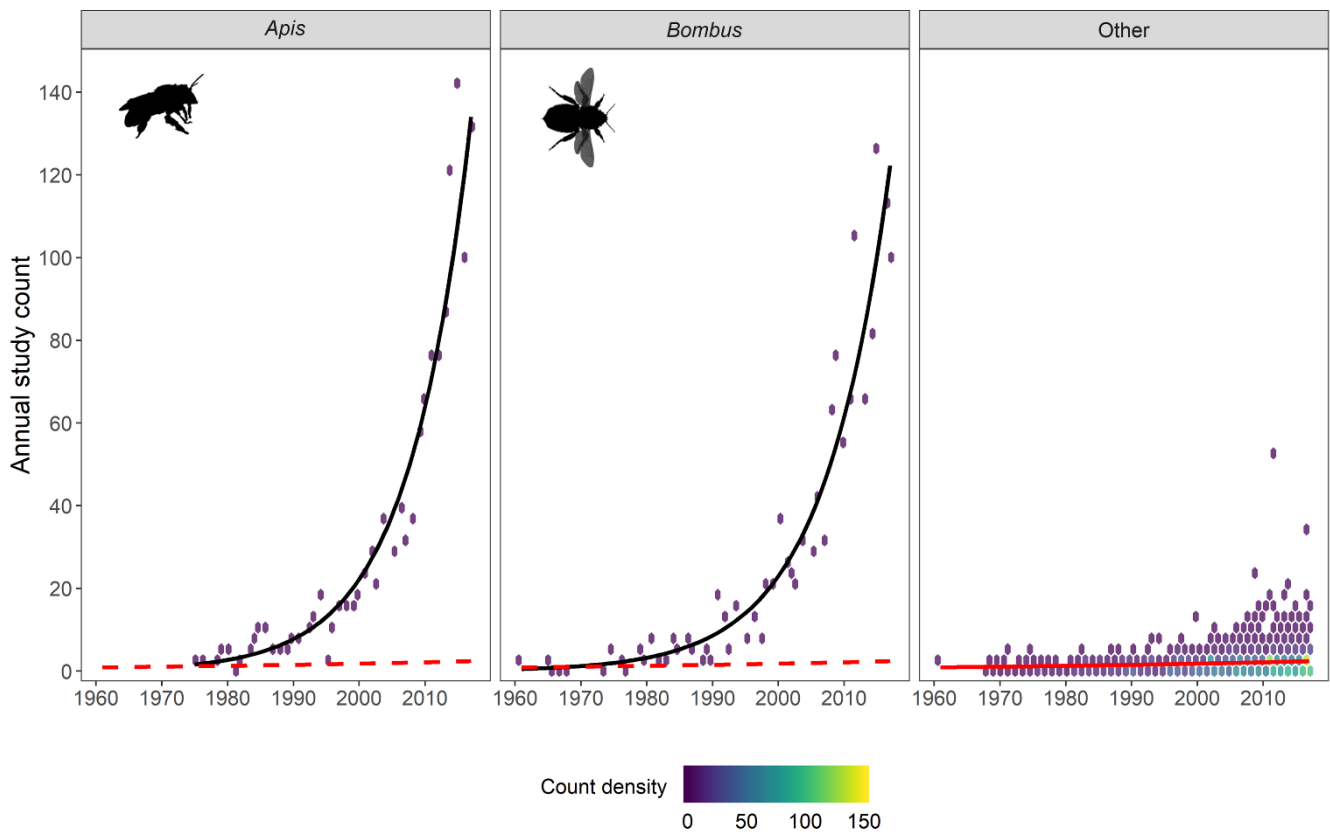


Figure 2.2 Annual study count for *Apis*, *Bombus*, and all other genera 1961-2017. Black lines represent the fit of a generalized linear model with Poisson errors for *Apis* and *Bombus*, relating study count to year. Red lines represent the output of the same model for all other genera, presented in dotted form in the *Apis* and *Bombus* facets. Counts have been binned as a density to represent multiple counts at the same study-year combination, from dark blue to yellow (150 studies). The model for both *Apis* and *Bombus* deviates from all other pollination-related animal genera in the 1980s, with a rapid and non-linear increase.

Taxonomic distribution

Of the abstracts related to pollination mentioning an animal species (3974), the Hymenoptera were overwhelmingly the most frequently mentioned of all taxonomic orders, in approximately 65% of all abstracts (Figure 2.3). Of the 13 most frequently mentioned genera, 11 were hymenopteran genera, all but one (*Ceratosolen*) of which were bees. This is to be expected, since bees are regarded as the most important pollinating group (Potts et al. 2010b, Ollerton 2017). Unsurprisingly, *Apis* and *Bombus* figure highly. Approximately 1/4 of abstracts mention a species of honey bee (*Apis*) and 1/5 a species of bumble bee (*Bombus*) (Figure 2.3), which is likely an underestimate given we do not consider common names (see ‘Limitations’ and Appendix 1, Figure S1.5). The disparity for *Apis* and *Bombus* is consistent with more anecdotal descriptions in the literature, describing honey bees and bumble bees as the main study groups (i.e. Ghazoul 2015). This taxonomic pattern probably to a large extent reflects commercial value: both *Apis* and *Bombus* are economically important commercial pollinators (Goulson 2003, Klein et al. 2007), with bumble bees in particular providing a unique contribution in the form of buzz-pollination (Goulson 2003). Moreover, 4 out of 13 other top-mentioned genera (*Osmia*, *Megachile*, *Melipona*, *Trigona*) are also managed commercially to some extent, either for pollination services or honey production (Vit, Medina and Eunice Enríquez 2004, National Research Council 2007, Velthuis and van Doorn 2006).

More generally, the insects are dominant in the pollination literature. Five of the top eight orders are insects (in decreasing order: Hymenoptera, Lepidoptera, Diptera, Coleoptera, and Hemiptera), the first four of which are well-known pollinating groups. The greater number of studies mentioning lepidopteran than dipteran species probably reflects a bias in study effort. Although flies are likely the second most important pollinators today—behind only the Hymenoptera (Ssymank et al. 2008)—and evolved as one of the first angiosperm pollinators (Endress 2001), lepidopteran flower-visitors are often deemed more conspicuous and attractive (New 2004), making them likely study candidates. The smaller number of studies mentioning beetles is more likely a true reflection of pollination importance. Although beetles

are important ecosystem service providers on the whole (Noriega et al. 2018) and evolved as some of the earliest gymnosperm pollinators (Labandeira, Kvaček and Mostovski 2007, Ollerton 2017), modern beetles are widely recognised as less important pollinators relative to the Diptera, Hymenoptera, and Lepidoptera (Ollerton 2017). A surprising result was the appearance of the moth *Manduca* in the top 13 genera, given the reputation of the moths as understudied relative to other lepidopterans (Hahn and Brühl 2016). Most likely this is an artefact of model taxa rather than pollinator importance. Although *Manduca* species can be agricultural pests and pollinators, with the larval stage feeding on a variety of plant species in the Solanaceae family (Kessler and Baldwin 2002), and the adult stage a generalist nectar feeder (Raguso and Willis 2002), mentions are driven by *Manduca sexta* (Linnaeus, 1763), an important model species for molecular and genetic studies (Riddiford et al. 2003). The hemipterans are represented primarily by aphid genera: ~1/4 of the hemipteran genera in the pollination literature are aphids. Despite being flower visitors, hemipterans more often feed on plant stem sap, making them incidental pollinators (Wardhaugh 2015). Broadly, the extent to which absolute distribution of mentions might predict pollinator importance for the five insect orders is an interesting point. Although the Hymenoptera are likely the most important pollinating order, it seems unlikely that this would be by a factor of ~7 globally. Probably there will be a signal of importance, but confounded by geography and study biases.

Vertebrates are also mentioned relatively frequently in pollination studies. Three vertebrate orders fall in the top eight: two avian, Apodiformes and Passeriformes; and one mammalian, Chiroptera. The Apodiformes are entirely represented by the Trochilidae (hummingbirds), a well-known nectar-feeding (and thus pollinating) family. The Passeriformes are represented more diversely, with approximately 75% of identified species coming from six nectar-feeding families: the sun birds (Nectarinidae), honey-eaters (Meliphagidae), Icteridae, honey-creepers (Thraupidae), white-eyes (Zosteropidae), and sugar birds (Promeropidae) (Proctor *et al.* 1996). Many species of bats are known to feed on fruit or nectar (Fleming, Geiselman and Kress 2009). The bat genus *Glossophaga*, the only vertebrate genus falling in the top 13, is a

common lowland nectar-feeding group distributed in central and South America (Fleming, Geiselman and Kress 2009).

Interestingly, some groups are associated with pollination through their nature as pollinator parasites and predators. For example, the Mesostigmata (an order containing the *Varroa* mites, 41 abstracts) and Araneae (spiders, 27 abstracts) are primarily parasites and predators respectively. *Varroa* mites parasitize honey bees, and are implicated in colony collapse disorder (van Engelsdorp et al. 2009). Araneae, such as crab spiders, prey on pollinators through hiding on the flower and ambushing at visitation (Dukas and Morse 2003).

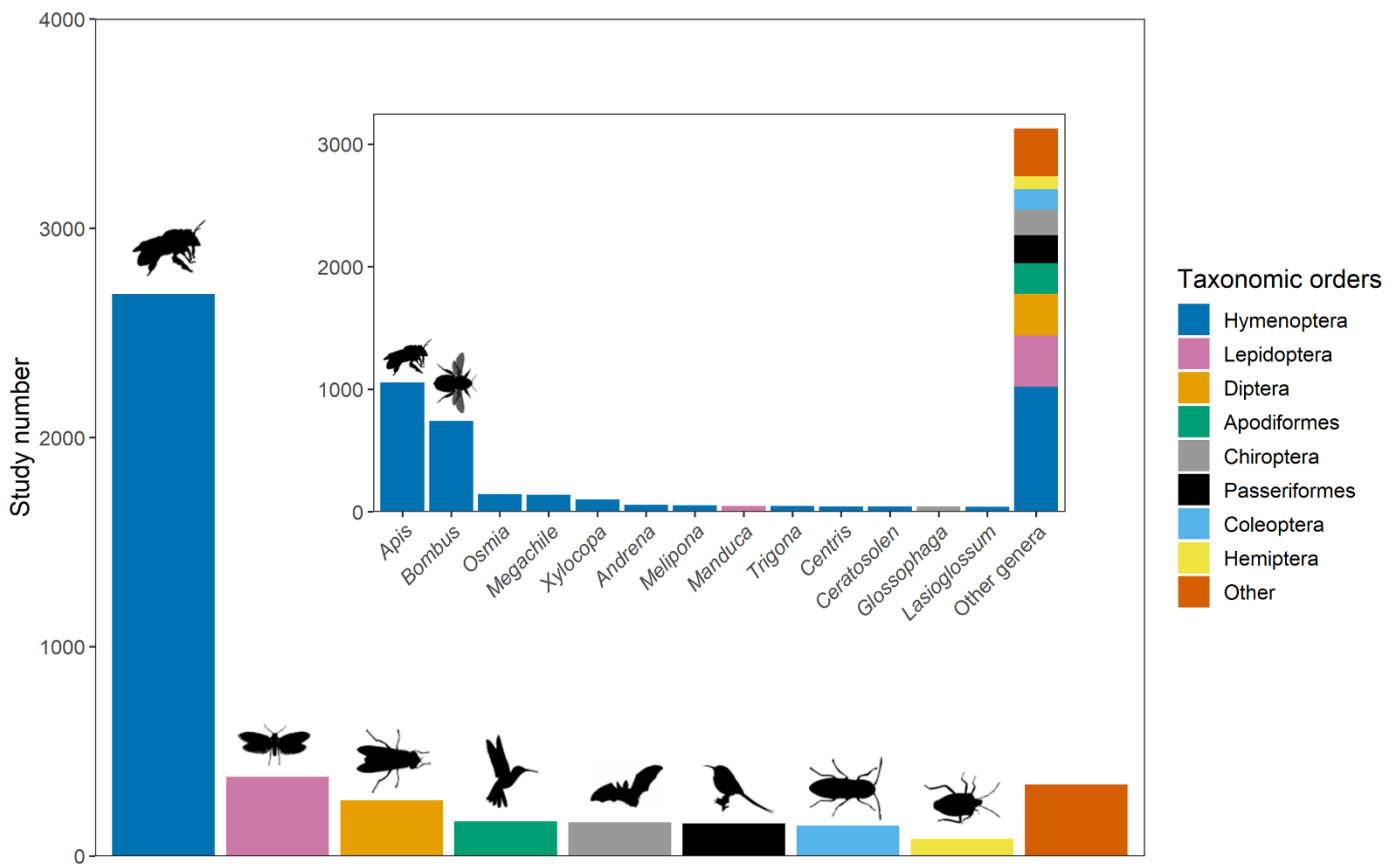


Figure 2.3 Order-level distribution of animal species in the pollination literature across 3974 studies. Given a study may mention multiple genera or orders, each bar is not independent, meaning study count values will not sum to the total of 3974. “Other” orders are represented by 55 orders, with 29/55 of these appearing in 2 or fewer abstracts. Inset shows genus-level distribution of animal species in the pollination literature. “Other genera” are represented by 1000 genera across the Hymenoptera, Lepidoptera, Diptera, Apodiformes, Chiroptera, Passeriformes, Coleoptera, Hemiptera, and “other” orders. Colours are the same in the main panel and the inset.

Geographical distribution

I also investigated the geographical distribution of the animal pollinator literature, inferred by extracting place names with the geoparser CLIFF-CLAVIN. The top five countries for animal pollination studies are the United States, Brazil, Australia, Canada, and China (Figure 2.4), together representing ~50% of all studies. Previous systematic reviews of the distribution of pollinator data identified Australia, Brazil, the United States, Germany, and Spain, as the top five contributors (Archer et al. 2014). Germany is notable by its absence in our analysis. However, exact character string matches of the term “Germany” with each of the abstracts indicate that Germany is indeed less strongly associated with the animal pollination literature (Appendix 1, Figure S1.9). Potentially lower representation of Germany is explained by the confounding effect of study subject. Archer et al. (2014) found that German studies were frequently represented among studies of pollinator perturbation, but relatively infrequently among general pollination studies. Habitat perturbation studies represented a relatively small percentage of the pollination-related papers analysed here (Appendix 1, Figure S1.7), while general pollination studies accounted for a much higher proportion. In general, my results suggest that overall pollinator information is less restricted to western Europe and North America than was previously thought (Mayer et al. 2011), although I recognise that my analysis likely underestimates geographic disparities (see ‘Limitations’). Indeed, only three European countries appear in the top 15 (United Kingdom, Spain, and Greece). However, although study count in European countries is relatively low, density is higher since European countries tend to have small areas (Appendix 1, Figure S1.3).

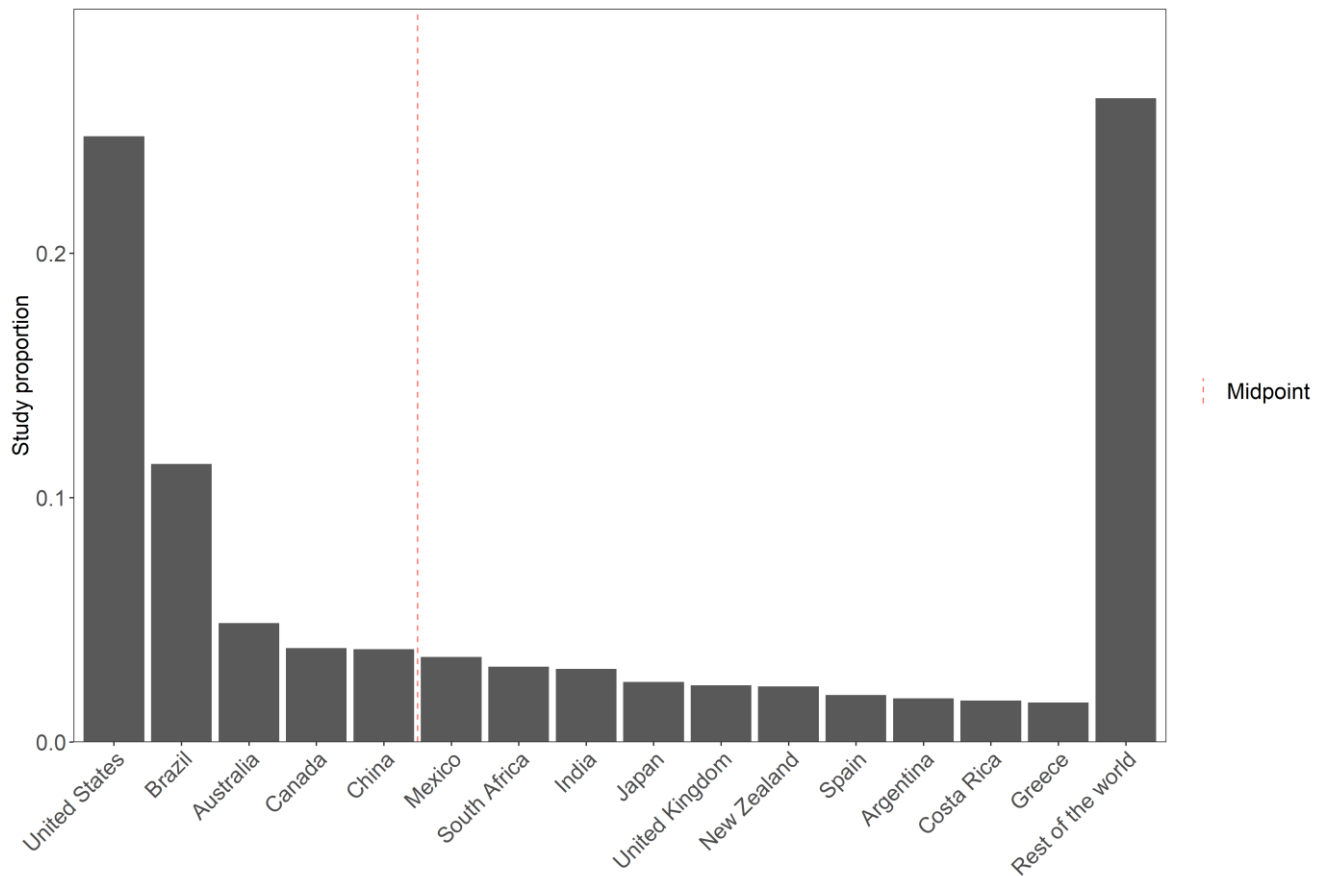


Figure 2.4 Distribution of the animal pollinator literature among countries. The “major” focus of each abstract, as resolved by CLIFF-CLAVIN, was used as an indicator of the geographical location likely representing the main study area. The red dotted line represents the study proportion midpoint. Half of all animal pollinator related studies fall in only five countries. “Rest of the world” is represented by 238 countries.

The global distribution of study counts reveals geographic disparities in animal pollination literature (Figure 2.5). Study counts are particularly low across large regions of Africa, with the exception of Kenya, South Africa, and Madagascar. Central Asia is also underrepresented, with no studies returned for any of Afghanistan, Turkmenistan, Uzbekistan, Kazakhstan, and Tajikistan. It is probable that some Central Asian pollination studies were published in Russian, meaning they were missed in the initial download. Interestingly, and as you might expect, the geographical distribution of animal pollinator-related studies to some extent reflects the global crop production, as shown by the high research effort in Eastern Brazil, India, Europe, and North America (Potts et al. 2016). Indeed, the largely unproductive region of North Africa has low study density with the exception of the Nile Delta, a fertile region of the Sahara Desert (Elbasiouny et al. 2014).

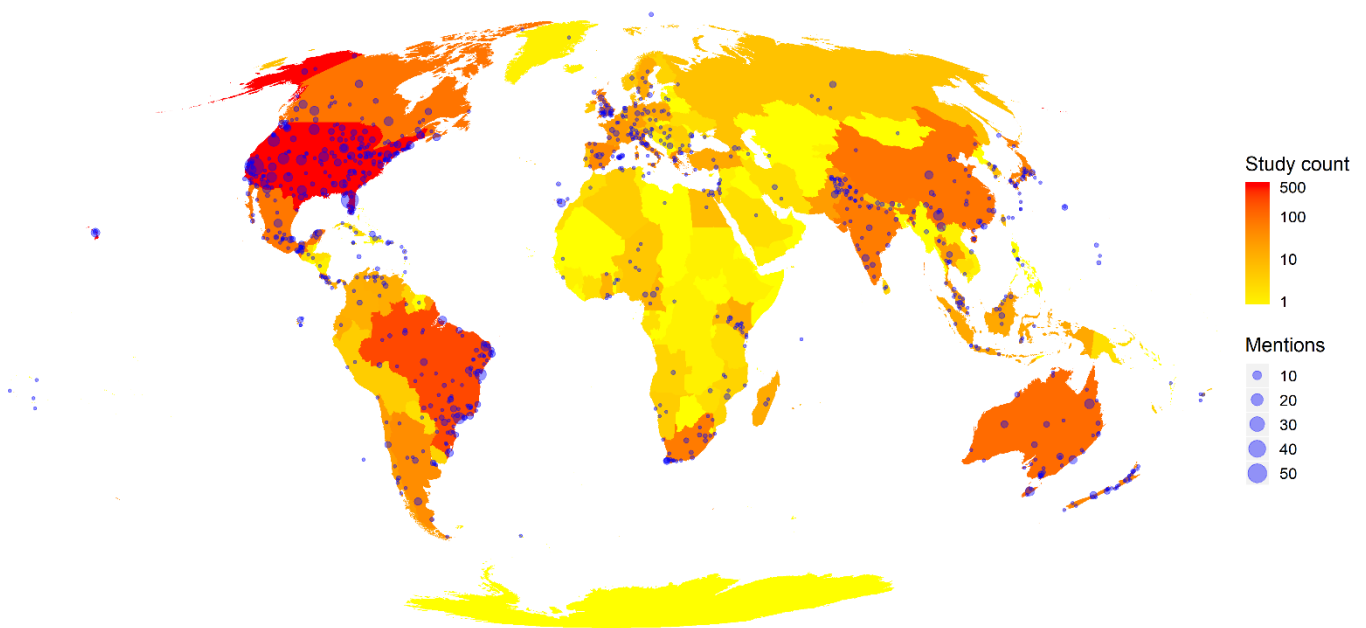


Figure 2.5 Global study counts for animal pollinator related studies, aggregated at country level. Study counts were derived from the number of abstracts with their “major” focus in each country. All oceanic and otherwise obviously incorrect mentions, as well as mentions that could only be resolved to a unit larger than a country, were removed. Study counts were $\log_{10}$ -transformed. Partially transparent blue points (“minor” mentions) represent the number of unique abstracts in which CLIFF-CLAVIN resolved that location. “minor” mentions include all specific geographical locations geoparsed by CLIFF-CLAVIN, with the exception of continents, oceans, and incorrectly geoparsed locations.

Interactions between taxonomy and geography

I also investigated the geographical distribution of pollination studies across taxa (Figure 2.6). I assumed that a taxonomic record in an abstract was related to any geographical location in the same abstract, and then plotted all unique abstract-species-location combinations (Figure 2.6) for the top five taxonomic orders. My analysis shows clear spatial patterns in pollination studies for different taxa, which probably reflects some combination of the actual distribution of pollinator species, study biases, and methodology-induced biases (see ‘Limitations’ for more details). Given the inaccuracies inherent in CLIFF-CLAVIN, and the problem in assuming all taxonomic names are associated with all locations in an abstract, my results should be interpreted with caution.

Genus-level study distributions for the Hymenoptera are associated with North America, South America, and Europe, reflecting some signal of actual distribution, although in my analysis Africa is conspicuous by its absence. My analysis indicates *Apis* has the largest study distribution, associated with North America, Europe, South and South-East Asia, Australia, and eastern South America, with some “minor” mentions in Kenya, South Africa, and Ghana. These results are consistent with the almost-global distribution of *Apis* (Han, Wallberg and Webster 2012), an important pollinating genus non-native to large portions of its current range (Whitfield et al. 2006). *Bombus* also appears to be associated with a global study distribution, albeit reduced in Africa, Central and South-East Asia, and Australia. This is to some extent concordant with the actual distribution of *Bombus* as a genus of the temperate regions, with anthropogenic introductions to New Zealand and Tasmania in the late 1800s and early 1990s respectively (Semmens, Turner and Buttermore 1993, Velthuis and van Doorn 2006). However, its association with studies in Africa and Australia is surprising, given it has no known distribution in either region. I found that all three *Bombus* abstracts associated with Africa, and all three with mainland Australia were false positives. Although each mentions a species of *Bombus*, the locations were inaccurate (either being mistakenly georeferenced taxonomic names or incorrect identification of a location). In all other top hymenopteran

genera, my results show some signal of actual distribution: *Centris* and *Melipona* are found naturally in the Neotropic and Nearctic realms, *Trigona* in the Neotropic and Indo-Australian, *Andrena* and *Osmia* in the Holarctic and North America, and *Megachile* the Western Hemisphere and Palearctic (Michener 2007).

Order-level trends likely indicate spatial patterns of pollinator importance, in part confounded by geographical study biases. For example, study records for dipterans are absent from Brazil—in a region highly populated with hymenopteran studies—but concentrated in Europe and North America. This would suggest some signal of lower ecological importance for dipterans in Brazil relative to hymenopterans. Indeed, previous studies have suggested an opposing latitudinal relationship for dipteran and hymenopteran pollination importance, with fly visitation decreasing at low latitude and hymenopteran increasing at low latitude (Ssymank et al. 2008). Similarly, chiropteran studies are concentrated in Central America, and the Apodiformes in South America, both regions within part of their respective native distribution (Fleming and Muchhala 2008). Some localities however are again conspicuous by their absence. Although hummingbirds went extinct in Africa in the Miocene, nectar-feeding fruit bats do occur in Africa and much of the tropics (Fleming and Muchhala 2008), which appears not to be represented in my analysis. Potentially also my outputs are influenced by the taxonomic spread of vertebrate pollinators in the Old and New World. The New World vertebrate pollinators are more diverse, but this diversity is concentrated in the hummingbirds and leaf-nosed bats, whereas in the Old World diversity is represented across multiple avian orders (Fleming and Muchhala 2008). For example, Old World pollinators include nectar-feeding Psittaciformes and Passeriformes, both of which are not considered major pollinators in the New World (Fleming and Muchhala 2008).

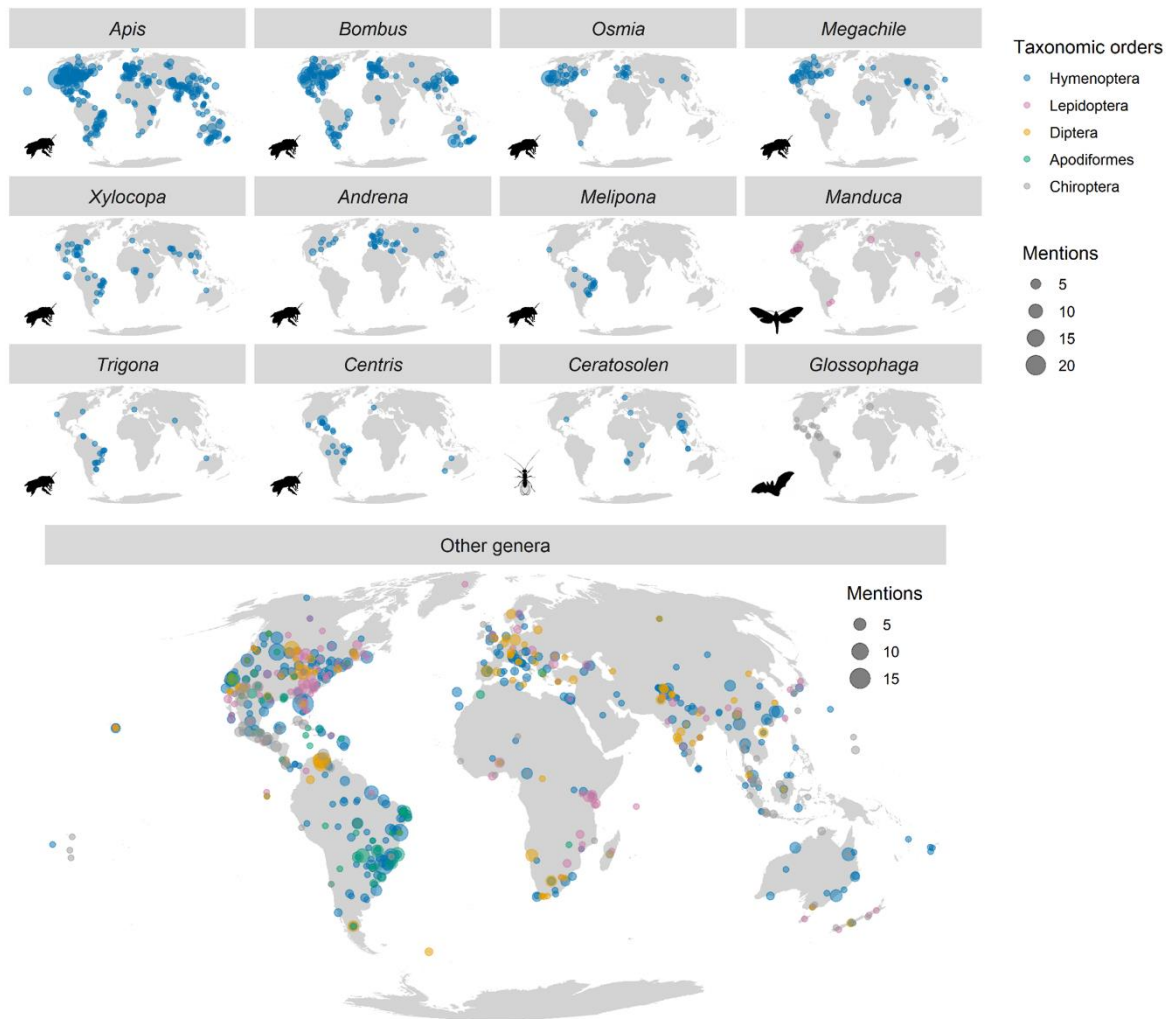


Figure 2.6 Distribution of the animal pollinator literature broken down into taxonomic groups, for the top 12 genera and top 5 orders. Taxonomic orders are indicated here by fill colour, consistent across top and bottom panels: Hymenoptera (blue), Lepidoptera (pink), Diptera (orange), Apodiformes (green), and Chiroptera (grey). Point size represents the frequency of unique abstract-genera-location combinations. “Other genera” here consists of 1001 genera across the Hymenoptera, Lepidoptera, Diptera, Apodiformes, and Chiroptera.

Limitations

Although taxonomic and geographical entity extraction are exciting developments, and show promise in both synthesising systematic review findings and prioritising the search path, these technologies are not without their limitations.

I acknowledge that applying a single academic indexing tool (Scopus) will mean my initial search likely overlooked relevant literature. Notably, relative to indexing tools such as Web of Science (WoS) or Scopus, Google Scholar has been shown to return a greater proportion of the relevant literature for a given search term (Beckmann and von Wehrden 2012). Both Scopus and Web of Science have also been shown to exhibit geographical and language biases (Mongeon and Paul-Hus 2016), which could potentially influence my outputs. However, I reasoned that although an academic indexer might favour some geographical regions, taxonomic groups, or years over another, and Google Scholar would give greater coverage, an academic indexer would return a text corpus I could be confident had undergone peer-review (Mongeon and Paul-Hus 2016). Given the underlying motive—to identify studies more efficiently for systematic review and meta-analysis—I decided that as a proof of concept, there was greater value in excluding the grey literature. In minimising the potential for biases, I opted for Scopus over WoS given its greater coverage (Mongeon and Paul-Hus 2016), and a single indexer to minimise the potential for duplication, a non-trivial risk in systematic reviews (Rathbone et al. 2015).

Potentially, my restriction to abstracts from English-language articles biased the outputs of my results. Articles of any language can be indexed in Scopus, with the caveat that an English version of the abstract must be included (Scopus 2018). Choice of language should not significantly change my outputs, for two key reasons: firstly, English is the dominant language of the scientific literature (Tardy 2004, Hamel 2008), independent of the nationality of researcher, meaning articles published in English are representative of a geographical distribution greater than just native English speaking countries; and secondly, options for other

languages within Scopus are minimal—meaning their inclusion would likely not significantly influence the distribution of our results—with the greatest contributors after English (for the term *pollinat**, returned on 16/08/18) being Portuguese (~1.7%), Chinese (~1.6%), Spanish (~0.9%), German (~0.5%), and Russian (~0.4%). However, I cannot exclude the possibility that the exclusion of some key languages may have biased the outputs, particularly given that 35.6% of the biodiversity conservation literature is not written in English (Amano *et al.* 2016).

Biases may also have been introduced by the taxonomic entity extraction algorithms, Neti Neti and Taxonfinder. Because the scrape for taxonomic information only picks up Latin binomial names, any species more often referred to by its common name, or any species more often referred to in the abstract through a higher taxonomic level, will likely be underrepresented. This may be the case for the western honey bee, *Apis mellifera*, which is often referred to by its common name, and possibly also for bumble bees. I briefly explored this limitation through investigating the frequency of common names for each of the top 4 genera (*Apis*, *Bombus*, *Osmia*, and *Megachile*), finding that the taxonomic disparity between honey bees, bumble bees and other taxa is likely even greater than suggested by our results, potentially by a factor of ~2 (Appendix 1, Figure S1.5). Given the strong association of *Apis* and *Bombus* studies with North America, Europe, and East Asia, I also expect that our analysis underestimates geographic disparity. Groups such as hummingbirds, which are more often mentioned in the abstract without an accompanying Latin binomial species name, may also be underrepresented. I explored this limitation through investigating the frequency of family names for 5 families with well-known common names (fig wasps, hawk-moths, hoverflies, hummingbirds, and leaf-nosed bats), selected from each of the top 5 orders (Hymenoptera, Lepidoptera, Diptera, Apodiformes, and Chiroptera). I found that three of these families (hummingbirds, fig wasps, and hoverflies) were likely under-represented by considering only Latin binomials (Appendix 1, Figure S1.6).

My findings were also influenced by the approach used to verify animal species records. In particular, in counting only Latin binomials I will have missed records. However, I reasoned

that the unambiguity of the Latin binomial would help to reduce noise. Moreover, I assumed that, with the possible exception of taxa referred to by widely accepted common names, the frequency of mentions for the full species record would likely correlate with higher taxonomic levels (Correia et al. 2017).

Spelling mistakes and failure to resolve as an accepted name are two different although closely-related limitations. Failure to resolve as an accepted name could potentially be caused by a spelling mistake in a synonym or accepted name, or by that record being absent from the COL as either an accepted name or synonym. Although I investigated fuzzy-matching for non-matched records, I opted not to include this implementation here because taxonomic name resolution became more ambiguous as a result. Fuzzy-matching returned multiple potential matches for a given record, requiring significant input to exclude false positives. Moreover, spelling mistakes would only be problematic for my conclusions if unevenly distributed among taxonomic groups, which is unlikely to be the case.

CLIFF-CLAVIN may also have introduced geographical biases in the distribution of our outputs, through the way in which it is trained and its probabilistic nature. Given that CLIFF-CLAVIN is trained on news articles, its effectiveness on academic texts is unclear. During my analysis, I noticed that CLIFF-CLAVIN would occasionally mistake Latin taxonomic entities for geographic locations. For example, the genus *Pavonia* was mistaken for a geographical location. There may be instances in which the algorithm's training interacts with taxonomy to bias our outputs. Relatedly, since CLIFF-CLAVIN is trained on news outlets based primarily in the US, US-based studies may be overestimated (Imani et al. 2017). However, my results for the representation of US pollination studies are consistent with Archer et al. (2014), which applied a different methodology. The probabilistic nature of CLIFF-CLAVIN may also have influenced my results to a small degree (indeed, running the algorithm a second time led to a reduced number of “minor” mentions in Brazil).

Future directions

Here I have used two text-analysis tools to quantify the geographical and taxonomic distribution of the animal pollinator literature. I showed that the literature is heavily concentrated in the honey bees and bumble bees of North America, albeit less biased than some authors have implied (Mayer et al. 2011).

The skewed taxonomic and geographical distribution of pollinator literature is a problem for the robustness of animal pollinator biodiversity models. Unfortunately, solving this problem is hard. Well-designed, long-term, and resource-intensive studies on little known taxonomic and geographical regions are needed. However, such studies are logistically difficult, expensive, and may not be achievable in time to inform important decisions. Another option—although not mutually exclusive—could be mitigating the problem through fully engaging with the available literature. Here I explore how this could be achieved by using the same text-analysis tools to yield a more representative and comprehensive set of studies for systematic reviews and meta-analyses. I briefly describe the conventional literature search path, as used for example in systematic reviews, before introducing a new search process.

The conventional literature search process for a systematic review can be conceptualised as three key phases, with the first two concerning literature retrieval (Figure 2.7): the search-term phase, in which key words in an online database are optimised to return literature deemed representative of the given research question; and the manual filtering phase, in which each article is assessed according to a series of specific criteria, and then excluded if it is deemed irrelevant. This manual filtering phase can be long and labour-intensive (Haddaway and Westgate 2018); some authors will assess >10,000 papers (Lavoie, Verbeek and Pahwa 2014, O'Mara-Eves et al. 2015), with one review as high as 800,000 papers (Shemilt et al. 2014). The manual filtering phase is followed by a third phase: the appraisal of each selected article, in which data are extracted to quantify the main findings of the study (Pullin and Stewart

2005). Although article appraisal can in part be addressed through text analysis (Lajeunesse 2016), here I focus on optimising the manual filtering and data-extraction phases.

Text-analysis has been introduced as a useful tool in optimising the literature filtering phase, but uptake in ecology is still low (Westgate et al. 2015). This is in part a symptom of unintuitive text-analysis tools, and insufficient technical skills required to use them (Westgate 2018a). However, arguably the bigger barrier is the lack of understanding as to how text-analysis tools relate practically to the literature search process. Although text-analysis approaches in ecology have advanced (Nunez-Mir et al. 2016, Westgate et al. 2018b, Roll et al. 2017), as far as I know there are no clear recommendations as to how ecological researchers should implement these approaches in the literature search. Here, taking inspiration from the 'revtools' package (Westgate 2018a), I propose a text-analysis search path in the context of the systematic review (Figure 2.7). The technology is available for this path, but not yet the specific intuitive tools or validation in the context of pollination ecology. My proposed synthesis path can be conceptualised as five key phases: search term, topic similarity, taxonomic and geographic identification, manual filtering, and appraisal. Below I briefly describe each of the first three modified phases:

1. Initial search terms should be used to return a broad body of literature for a given field. Fewer and less specific search terms should be used across multiple databases, aiming to return all of the relevant literature irrespective of a potentially high false positive rate. This less restrictive initial search will require less researcher input, thus reducing the time required. Such an approach will also reduce the likelihood of overlooking relevant literature through overly specific search terms.
2. The key filtering step should be shifted downstream to a text-analysis filter. Topic-clustering algorithms should be used to exclude irrelevant articles. For example, in the context of the potential pollination literature returned by Scopus, topic-clustering can be used to exclude papers on the flower-pollination algorithm, an area of computer science unrelated to pollination ecology. Topic clustering is more reproducible than

database search terms, less subject to researcher methodology, more representative of overall content, and not subject to differences across database.

3. Taxonomic and geographical entity extraction algorithms should be used to indicate the geography and taxonomy of each study. Taxonomic group and geography can be used to prioritise for underrepresented taxa or regions, as well as to identify likely literature for regional or taxonomic systematic reviews.

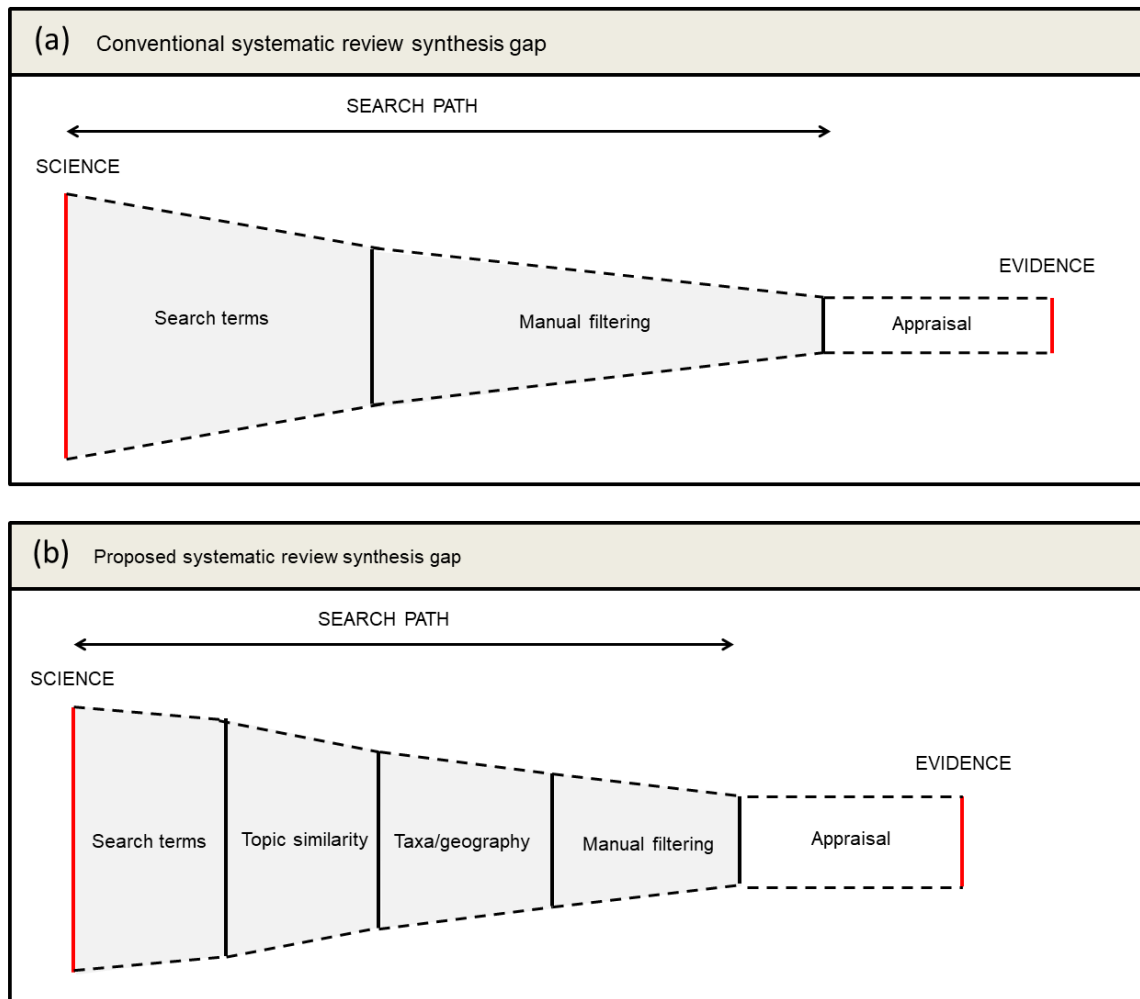


Figure 2.7 Reducing the length of the “synthesis gap” for ecological systematic reviews. The conventional synthesis gap (a) proceeds through three primary steps: the search term phase, manual filtering, and appraisal. My proposed, more efficient synthesis gap (b) proceeds through five steps: a less-exclusionary search term phase; text-analysis prioritisation (topic similarity and taxonomic/geographic filtering or prioritization); manual filtering, and appraisal, with both the search term and manual filtering stages shortened. Solid lines represent the research volume between each stage, which in my proposed process is greater, owing to the increased taxonomic and geographic representativeness. Dotted lines represent the change in research volume for each selection step. Red lines represent the beginning and ends of the synthesis gap, from the practice of science to the synthesis of evidence.

Text-analysis tools could also be applied to answer other research questions in the field of animal pollination, such as the identification of likely plant-pollinator interactions and the estimation of the number of pollinating species. Here I briefly explore how these research questions might be addressed.

Likely plant-pollinator interactions could be identified through investigating the strength of animal-plant associations across the pollination literature. For example, a plant and animal species frequently occurring in the same abstract would imply a closely-related pair, while an animal-plant combination occurring infrequently would imply a weak or non-existent interaction. These animal-plant networks could then be validated with observational plant-pollinator interaction data, using an approach similar to Tamaddoni-Nezhad et al. (2013), who showed that text-derived food web networks approximated empirical data. Such a database of likely interacting plant-pollinators could be invaluable for pollinator biodiversity modelling. For example, we might be able to predict better the likelihood of co-extinctions, as well as improve estimations of global and regional pollinator importance.

I also envisage that an accurate lower bound for the number of pollinating species could be estimated through text-analysis. This would build on the work of Ollerton (2017) and Wardhaugh (2015), who estimated there to be ~350,000 described species of animal pollinator, based on the number of species in key pollinating groups. I propose that within taxonomic groups, the number of pollinating species could be estimated through quantifying the rate at which unique animal species accumulate as pollination text volume increases. This would be analogous to the species-effort rarefaction curve; a mathematical relationship between pollinator text volume (effort) and species number, revealing how the number of unique pollinators increases as a function of research volume (effort). Such work would likely require scraping of full articles, rather than just abstracts as presented here. Full-text scraping presents additional problems regarding rapid article download and permissions, file format,

text volume quantity, but the benefits relative to abstract scraping are known (Westergaard et al. 2018). Similar methodologies have been applied to quantify the number of species on Earth, both for the flowering plants (Joppa, Roberts and Pimm 2011) and for all Eukaryotes (Mora et al. 2011). In the context of pollinator biodiversity models, good estimates of the diversity and number of pollinating species are fundamental for understanding how ecosystems will respond to future environmental change (Ollerton 2017).

Summary

Here I have shown—using a novel combination of informatics tools—how text-analysis can be used to quantify the taxonomic and geographical distribution of the animal pollinator literature. In doing so I have confirmed that the literature is heavily focused on the honey bees and bumble bees of North America and Europe, although many studies also exist for other taxa and regions. This skewed taxonomic and geographical distribution likely has a large impact on the robustness of systematic reviews and meta-analyses of animal pollinator decline. I have shown how text-analysis might to some degree mitigate these data biases. Text-analysis could be used to make the literature search process more efficient, as well as increase the taxonomic and geographic representativeness of the studies fed into systematic reviews and meta-analyses. To this end, I briefly outlined a new literature search process, using an ecological systematic review as an example for how text-analysis might contribute. I have also explored some potential broader applications of text-analysis in pollination ecology, such as the identification of likely plant-pollinator interactions and the estimation of the number of pollinating species, both of which could feed into more robust systematic reviews and meta-analyses in the future. Text-analysis undoubtedly shows promise in increasing our understanding of the rapidly growing pollination ecology literature, and in turn the robustness of studies estimating pollinator decline.

Chapter 3: Global effects of land-use intensity on local pollinator biodiversity

Abstract

Pollinating species are reported to be in decline globally, with land use an important driver. However, most of the evidence on which these claims are made is patchy, and generally based on studies with low taxonomic and geographic representativeness. Further research incorporating a broader spectrum of taxa and geographies is required. Here, I model the effect of land-use type and intensity on global pollinator biodiversity, using a database of local-scale biodiversity covering 303 studies, 12,170 sites, and 4,502 unique species judged to be pollinating animals. Relative to a primary vegetation baseline, I show that low and intermediate levels of intensity can have beneficial effects on pollinator biodiversity. Within most anthropogenic land-use types however, intensity is associated with reductions in pollinator species richness and total abundance, particularly in urban (43% richness and 62% abundance reduction compared to the least intensive urban sites), plantation (38% richness reduction), and pasture (75% abundance reduction) areas. I further show that on cropland, the strongly negative response to intensity is restricted to tropical areas, and that the direction and magnitude of response differs among taxonomic groups, with significantly higher pollinator biodiversity for some taxonomic groups (e.g. flies) on intensively fertilised cropland. My findings confirm widespread effects of land-use intensity on pollinator biodiversity, most significantly in the tropics, where climate and land use are predicted to change rapidly in the coming decades.

Introduction

Pollinating species, particularly insect pollinators, are reported to be in decline, with change in land-cover, land-use intensity, and climate thought to be the primary drivers (Potts et al. 2010; Steffan-Dewenter and Westphal 2007; Woodcock et al. 2016; Winfree et al. 2011; Soroye et al. 2020; IPBES 2017; Biesmeijer et al. 2006). In the media, insect pollinator biodiversity change has been reported to constitute an “ecological Armageddon” (Embury-Dennis 2017). However, the evidence on which these claims are made is patchy and contested, based on studies with low taxonomic and geographic representativeness (Sánchez-Bayo and Wyckhuys 2019; Martin et al. 2019; Powney et al. 2019; Lister and Garcia 2018; Hallmann et al. 2017). For pollinators more broadly, declines have been reported in wild bees, honeybees, hoverflies, butterflies and moths, flower-visiting wasps, birds, and mammals (see Ollerton 2017 for a summary of the evidence), but comprehensive studies of change tend to be biased towards North America and Europe (Saunders et al. 2020), which are unlikely to be globally representative (De Palma et al. 2016). Moreover, even within well-studied taxonomic groups and regions, the magnitude and direction of change can vary depending on methodological approach, spatial scale, and metric of biodiversity change (Kunin 2019; Macgregor et al. 2019). Recent research indicates there may be pollinator information for other geographic regions and taxonomic groups, previously untapped in synthetic analyses of pollinator biodiversity change (Millard et al. 2020a). Given the value of animal pollination to the global economy, at an estimated \$235-577 billion US dollars per annum (IPBES 2017), further research considering multiple metrics of biodiversity, across a broader spectrum of taxa and geographies, is required.

The reliance of global crop production on animal pollinators makes pollinator biodiversity research highly relevant to policy-makers. More than 75% of globally important food crops are at least partially reliant on animal pollination, including fruits, vegetables, coffee, cocoa and almonds (Klein et al. 2007). Three recent policy initiatives demonstrate recognition from the international community that pollinator biodiversity change represents a significant

problem, and needs to be addressed: 1) the EU Pollinators Initiative (European Commission 2018) called for improved knowledge of declines and causes, action to tackle drivers, and raised awareness across society on the importance of pollinators; 2) the International Pollinator Initiative plan of action aims to coordinate global action for pollinator conservation (FAO 2018); and 3) more broadly, the draft post-2020 framework of the Convention on Biological Diversity (Secretariat of the Convention on Biological Diversity 2020) describes the need for the sustainable use of biodiversity to support the productivity of ecosystems.

Much of the Earth's terrestrial surface is subject to anthropogenic use. More than 75% of the terrestrial world exhibits direct evidence of historical or current transformation (Ellis and Ramankutty 2008), with just over 50% (~67 million km²) currently used by humans (Hooke et al. 2012). This area is comprised of ~44% for agriculture and forestry, and ~7% for infrastructure including urban areas (Hooke et al. 2012). Within both natural and disturbed land-use types, intensity of human use varies markedly. Broadly capturing the inputs used in managing land, high-intensity farming refers to a suite of technological practices designed—although not always successfully—to increase yield (Donald et al. 2001). Treatments of the land are often in the form of chemical applicants, such as pesticides, fungicides, herbicides, and fertilisers, as well as mechanical management (tillage). Such intensive agricultural practices are commonplace in much of the modern world (Benton et al. 2002; Benton et al. 2003).

Anthropogenic land-use and land-use intensity are interrelated drivers of pollinator biodiversity change (Tscharntke et al. 2002; Kennedy et al. 2013; IPBES 2017). Much of the research investigating land-use effects on pollinator biodiversity has demonstrated the importance of landscape-level habitat composition, often as distance to natural habitat (Ricketts et al. 2008) and distance to managed land (Klein et al. 2003; Kennedy et al. 2013), or habitat fragmentation (Xiao et al. 2016) and edge density (Martin et al. 2019). Land-use intensity effects are generally typified by chemical application (Woodcock et al. 2016; Goulson 2013; Pisa et al. 2015; Sanchez-Bayo and Goka 2014; Pilling and Jepson 1993).

Pesticides such as neonicotinoids have been a focal point of study, given their association with declining bee populations (Goulson 2013; Pisa et al. 2015; Woodcock et al. 2016), honey bee health (Sanchez-Bayo and Goka 2014), and bumblebee behaviour (Crall et al. 2018). Other chemical inputs such as fungicides and herbicides have also been subject to investigation, tending to have indirect effects on pollinator biodiversity by increasing pesticide toxicity (Schmuck et al. 2003; Pilling and Jepson 1993) and reducing floral diversity (Morandin and Winston 2005). Similar indirect effects have also been shown for fertiliser application. For example, nitrogen-based fertilisers reduce plant species diversity (Ridding et al. 2020), and dispense with the requirement for clover field crop rotation, further reducing floral availability for pollinators (Goulson et al. 2008).

Pollinator response to landscape-level land use is mixed, with the magnitude and direction of change differing among taxonomic groups. For example, some bees, butterflies, syrphid flies, and nectarivorous pollinating birds have been found to favour open, intermediate-level forested areas of semi-natural grassland or agroforestry (Michener 2007; Deans et al. 2007; Kuussaari et al. 2007; Tscharntke et al. 2008). Similarly, species rich and abundant wild bee communities have been found in urban environments, indicating that for some species anthropogenic activity can be beneficial (Hall et al. 2017). For both open forested and urban areas, benefits to pollinators can in part be attributed to floral availability (Goulson et al. 2008; Hall et al. 2017). More broadly however, differences in pollinator response are often attributed to traits (Öckinger et al. 2010; Burivalova et al. 2015; Montero-Castaño and Vilà 2012; De Palma et al. 2015), such as dietary specialism, mobility, and nesting behaviour (Aguirre-Gutiérrez et al. 2016; Shuler et al. 2005). Trait data are not available for many pollinating species, but given that phylogeny to some extent predicts traits, one would expect broad differences in response among taxonomic groups (Cusser et al. 2018).

Differences in pollinator response to intensity are also likely between tropical and non-tropical regions. There are a number of reasons why this is the case. First, temperate non-tropical regions have a longer history of agricultural activity, which may have acted to filter

more sensitive species (Balmford 1996), meaning more recent shifts towards intensive agriculture may have a smaller effect. Second, with the exception of high latitude Arctic pollinators (Høye et al. 2013), tropical biodiversity has been reported to be more sensitive to the effects of climate change (Barlow et al. 2018), which may magnify the effect of land-use (Williams et al. 2020b). In terms of insect pollinators specifically, tropical insects are thought to exist closer to their thermal tolerance limits, meaning small magnitude changes in temperature have a disproportionate effect on biodiversity (Deutsch, et al. 2008). Third, functional specialisation tends to be higher in tropical pollination systems (i.e. there is a narrower breadth of visitors to a flower across broad taxonomic levels), which may also relate to community sensitivity to land-use change. Although recent research has addressed patterns of overall biodiversity change between geographical zones (Blowes et al. 2019; Newbold et al. 2020), for pollinating taxa the extent to which response to land-use intensity differs between tropical and non-tropical regions is unclear.

I present the most comprehensive global space-for-time synthesis of pollinator responses to land-use intensity. For the first time, I test for global differences in responses among land-use types, taxonomic groups, geographic regions, and biodiversity metrics. I do so using two global compilations of data: 1) The PREDICTS database, a global compilation of site-level ecological survey data across different land uses and land-use intensities (Hudson et al. 2017), which was originally built to model the overall effect of human-land use on global terrestrial biodiversity (e.g. Newbold 2015a); and 2) a new database of animal species judged to be animal pollinators (see Millard et al. 2020a and 'Methods'). My final dataset included 3,862 invertebrate and 640 vertebrate species identified as potential pollinators, across 303 studies and 12,170 sites, primarily across North and South America, Europe, and Africa. I hypothesised that land-use intensity would decrease site-level biodiversity (species richness, Simpson diversity, and total abundance) for pollinating species overall, but that response would differ between taxonomic orders, and would be more negative in the tropical zone than elsewhere. Specifically, I set out to answer three questions related to land-use

intensity and global pollinator biodiversity: 1) What are the overall effects of land-use intensity on pollinator biodiversity across all land-use types? Then focusing on croplands, for which there is the most extensive data: 2) How does the effect of land-use intensity on pollinator biodiversity differ between tropical and non-tropical areas? and 3) How does pollinator response to land-use intensity vary among taxa?

Methods

Pollinator dataset construction

I built an animal pollinator dataset through a semi-automatic approach, combining an automatic text-analysis method (see Millard et al. 2020a) with manual inspection of the automated output. Here I describe the full methodology used to derive this dataset.

I first created a list of possible pollinating animal genera through automatic text analysis of the pollination literature. I used an initial automated search to avoid biasing towards well-known pollinators, and to markedly reduce the input required in searching. I considered the pollination literature to be any primary research article published in English returned through a search for the term “pollinat*” in Scopus, and which mentioned an animal species in the abstract. I considered a possible pollinating genus to be any animal genus appearing as part of a Latin binomial in a pollination-related abstract returned from Scopus. Genus scraping was accomplished using the *Taxonfinder* and *Neti Neti* algorithms implemented in the ‘taxize’ R package (CRAN 2018), with animal species confirmed through a series of character string matches to the Catalogue of Life (see Millard et al. 2020a for a detailed methodology).

For each possible pollinating genus, I then read the abstracts in which these animals appeared, searching for evidence confirming that genus as pollinating. For any situation in which the abstract was inconclusive, I also searched the full text of the paper. For each confirmed pollinating genus, I then assigned a level of confidence between 1 and 4 based on the type of evidence, following Ollerton and Liede (1997): 1) experimental evidence

confirming pollination; 2) evidence of pollen carrying; 3) evidence of nectar/pollen feeding; 4) evidence of non-destructive/non-predatory flower visitation. I read abstracts for each genus searching for the highest level of evidence, either until I could be sure that the confidence value should be 1, or I ran out of abstracts for that genus. Non-destructive flower visitor refers to any animal which visits a flower without causing damage to the plant. This meant the exclusion of ants, which are typically referred to as poor pollen vectors, given that they damage pollen through secretions from the meta-pleural gland (Dutton and Frederickson 2012). Non-predatory flower visitor refers to any animal which visits for some purpose other than predation. This meant the exclusion of animals such as crab spiders, which predate on pollinators during visitation, and therefore contribute minimally or negatively to pollination (Dukas and Morse 2003). I did not classify broad statements as evidence for pollination—for example, one study stated that *Phylidonyris novaehollandiae* (Latham, 1790) is a “key pollinator” (Myers et al. 2012)—unless it was associated with specific evidence reinforcing that statement, or some claim that pollination in that genus is “well-known” or “widely acknowledged”.

Given that I only had direct evidence for a sample of all pollinating genera, I then searched for higher-level groups of likely pollinators. From the confirmed pollinators in the original list of genera, I identified all unique families with at least one pollinator. For each family, I assessed the breadth of evidence for pollination through consulting the abstracts and taxonomic group reference books. For any family with evidence of pollination across multiple branches of that family, and no evidence of any species definitely not pollinating, I assumed that the whole family is pollinating. If unable to extrapolate across the whole family, I then searched progressively lower taxonomic groups (i.e. subfamily, tribe, subtribe), searching for the point at which we could be relatively confident that the entire group contributes to pollination. If unable to extrapolate for a given group, I kept only the genera with direct evidence. For example, within the family Macroscelididae (elephant shrews), I found only

one genus (*Elephantulus*) with pollination evidence, and no evidence across the rest of the family, meaning I kept only that genus.

To compile the final list of pollinators, I merged all genera identified directly as pollinators, and then all taxonomic groups identified indirectly, with all biodiversity records in the PREDICTS database. Any record in the PREDICTS database not for a pollinating genus or extrapolated taxonomic group was thereby filtered out. As a result, specific sites or studies were only kept if they were represented by at least 1 pollinator record. I merged direct evidence (i.e. confidence levels 1-4) first to pick up each species record at its highest level of confidence, and to ensure that each species record is assigned only one confidence value. PREDICTS does not record additional taxonomic ranks between family and genus, so for any species extrapolated at a taxonomic level below the level of family (i.e. subfamily, tribe, subtribe), I consulted compiled genera lists for each group, using taxonomic references and Wikispecies (see 'Supplementary Data 2' in Millard et al. 2021a for the list of taxonomic references), and then filtered these genera from PREDICTS. As an additional check of our final list of likely pollinating species, I sought the opinion of experts in pollination ecology. Initially I shortlisted a set of pollination experts, using my knowledge of the pollination literature and contacts made through conferences. From this initial shortlist, I then selected a group of individuals such that I had expert opinion across a breadth of geographic regions and taxonomic groups. From this abridged list, I then sent lists of the likely pollinating genera in PREDICTS to 7 experts (OA, SG, EK, MK, JO, Z-XR, MS; see Appendix 2, Table S2.15), and removed any taxa identified as highly unlikely pollinators.

Effect of land use and land-use intensity on global pollinator biodiversity

I used the PREDICTS database to model responses of animal pollinators to land-use type and intensity (Hudson et al. 2017). The PREDICTS database is structured such that each site is nested at a series of levels (Appendix 2, Figure S2.13), allowing one to account for variation owing to study methodology. The database contains variables for land-use intensity (minimal, low, and high) and land-use type (primary vegetation, mature secondary

vegetation, intermediate secondary vegetation, young secondary vegetation, plantation, pasture, cropland, and urban). Land-use intensity for each land-use type is defined according to a series of variables, such as fertiliser and pesticide application, mechanisation, and hunting (see Newbold et al. 2015a for more details).

After merging the PREDICTS database with my set of likely pollinating species I performed a series of data-processing steps. I removed any sites for which land-use type and land-use intensity was unknown. I also removed sites in secondary vegetation at an unknown stage of recovery. I combined the factors for land-use intensity and type to create a single variable (henceforth referred to as LUI), following the methodology of De Palma et al. (2016). After combining land-use intensity and type, I then removed the class “Mature secondary vegetation-Intense use”, which was represented by only 5 sites, and “Intermediate secondary vegetation-Intense use”, which was represented by only 23 sites. After removing these factors, site representation was ≥ 43 sites for all land use type and land-use intensity combinations (see Appendix 2, Table S2.19). I then calculated site-level species richness (the number of uniquely named species sampled at a site), Chao1-estimated species richness (the number of species at a site controlled for abundance; Chao et al. 2005), total abundance (the sum of all species sampled abundances at a site), and the Simpson diversity index (the reciprocal of the sum of squared proportional abundances for all species sampled at a site). Sampling effort was accounted for by dividing the abundance values for each measurement by the sampling effort (rescaled to a max value of 1 for each study) for that record, as in De Palma et al. (2016). For any subsequent analyses I worked only from the sampling effort adjusted measurements. Given sampling effort adjustments, and that raw abundances were in some cases measured as densities, many total abundances will be non-integer values.

I built generalised linear mixed-effects models with a Poisson error distribution for species richness and Chao1-estimated species richness (Chao et al. 2005), and linear mixed-effects models for Simpson diversity and total abundance. In an initial set of models all biodiversity

metrics were fitted as a function of land-use intensity, land-use type, and their interaction, for all likely pollinators in PREDICTS (see Appendix 2, Table S2.17). I then built a set of models predicting each biodiversity metric as a function of LUI, for the same set of pollinators. I did not use a generalised model with Poisson errors for total abundance or Simpson diversity because most recorded measurements are not integer counts of individuals. Instead I log_e-transformed all total abundance and Simpson diversity values (adding one because of zero values) to normalise the model residuals. Due to the nested nature of the database (see Appendix 2, Figure S2.13 and Hudson et al. (2017)), I included a random intercept of study identity to account for variation in sampling methods, sampling effort and broad geographical differences among studies, and a random intercept of spatial block within study to account for the spatial structuring of sites. An additional (observation-level) random intercept of site identity was included in the species richness model, to control for the over-dispersion present in species richness estimates (Rigby et al. 2008). Random-effects structures were selected to minimise AIC values. I checked for overdispersion in the species richness models using the function `GLMEROverdispersion` in the R package `StatisticalModels` (Newbold 2015b). I compared each model against an intercept-only model, and discarded any main model for which AIC was greater than the null model (see Appendix 2, Table S2.18 for pseudo R squared values for all significant models).

I carried out a series of additional validation analyses for our set of LUI models. 1) I checked for study-level spatial autocorrelation in the residuals of any significant model, using the Moran's I test (Appendix 2, Figure S2.5). 2) I checked the extent to which a negative binomial zero-inflated model for total abundance would have differed from a linear model approach (Appendix 2, Figure S2.1). 3) I checked the extent to which the fixed effects would have differed if we had fit a model with climatic variables as potentially confounding covariates (Appendix 2, Figure S2.3), including both the maximum temperature of the hottest month and the total precipitation of the wettest month—both over the 12 months previous to the end data of each sample—which have previously been indicated as important biological

variables (Williams et al. 2020a). 4) I jack-knifed the sites for each significant model by continental region to check the extent to which geographic biases influenced our predictions (Appendix 2, Figure S2.2). 5) I checked the extent to which an abundance-controlled estimate of species richness (Chao1-estimated species richness; Chao et al. 2005) would have differed from species richness alone (Appendix 2, Figure S2.4).

Effect of land-use intensity on cropland pollinator biodiversity

I focused on cropland in our remaining analyses, given the importance of animal pollination to crop production. I built 3 models for all potential pollinating species, modelling each of three biodiversity metrics (species richness, total abundance, and Simpson diversity) in cropland as a function of land-use intensity (minimal, low, high), geographical zone (temperate/tropical), and their interaction. I included minimally used primary vegetation in these models as a baseline. Given that the structure of this baseline differs among sites—particularly between tropical and non-tropical areas (see Appendix 2, Figures S2.11 and S2.12)—and that this may affect our predictions for pollinator biodiversity, I also built a set of models with a high ($\geq 60\%$ cover) and low ($\leq 40\%$ cover) forest cover baseline, using Hansen (2013) forest cover data (Appendix 2, Figure S2.6). I also carried out two additional validation analyses. 1) I checked whether unequal site number between our tropical and non-tropical data predicted the size of our 95% confidence intervals. Specifically, I resampled 1,000 sites from each of the tropical and non-tropical sites a total of 100 times, and then for each group of 2,000 (tropical and non-tropical) fitted total abundance as a function of land-use intensity, geographical zone, and their interaction. I then plotted the distribution of the size of the 95% confidence intervals for all models (Appendix 2, Figure S2.7). 2) I checked whether response to land-use intensity between the tropics and non-tropics would have been the same if we had analysed only the main crop pollinating groups (Appendix 2, Figure S2.8; i.e. bees, wasps, beetles, thrips, flies, birds, and bats; see Ollerton 2021).

I also built 3 models for a vertebrate and invertebrate cropland subset of the database, modelling the same biodiversity metrics as a function of land-use intensity, taxonomic order, and their interaction, again including minimally used primary vegetation as a baseline. My taxonomic subset included the better-sampled invertebrate orders Hymenoptera, Lepidoptera, Diptera, and Coleoptera, and the vertebrate orders Apodiformes and Passeriformes, and represented 3006 sites in total (see Appendix 2, Table S2.25). For both my geographical zone and taxonomic order models, I selected from the same set of random-effects structures (as in the main models), aiming to minimise AIC values. I tested each model against an intercept-only model and a model with one fixed effect for land-use intensity, and discarded any main model for which AIC was greater than the null model (see Appendix 2, Table S2.18).

I also explored the effect of a continuous variable describing a specific aspect of land-use intensity (fertiliser application rate) on pollinator biodiversity, specifically for cropland. I used Earthstat fertiliser data at a spatial scale of 5 x 5 minutes—equivalent to 10 × 10 km at the equator—largely for the years 1999-2000 (Mueller et al. 2012; West et al. 2014), aggregated as the total application in kg per hectare for nitrogen, phosphorous, and potassium on 17 major crops (see Appendix 2, Table S2.34 for the full list crops). I aggregated the Earthstat fertiliser data by summing the per hectare application rate rasters for all crop/fertiliser combinations, and then extracting the summed fertiliser values at each site (see Appendix 2, Figure S2.9, note that site-level geographic distribution is lesser relative to our overall pollinator biodiversity models). Given that the spatial scale of this aggregated application rate data is greater than that of specific sites, my fertiliser metric refers to application rate in the surrounding landscape, rather than at that specific site. I chose to use fertiliser data given its availability at the global scale, reasoning that its application would both drive change itself, and broadly act as a surrogate for intensity. I built models for all potential pollinating species, modelling each of three biodiversity metrics (species richness, total abundance, and Simpson diversity) in cropland as a function of log10(fertiliser application

rate + 1), geographical zone (temperate/tropical), and their interaction. I also built models for the invertebrate and vertebrate subset (Hymenoptera, Lepidoptera, Diptera, Coleoptera, Apodiformes, and Passeriformes), modelling each of three biodiversity metrics as a function of total fertiliser application rate, taxonomic order, and their interaction. I compared each model against an intercept model and a model with one fixed effect for total fertiliser application rate, and excluded any main model for which AIC was greater than the null model. As a supplement to our total fertiliser application rate analyses, I also used PEST-CHEMGRIDS to build an analogous set of total pesticide application rate models (Figure 3.6; Maggi et al. 2019). PEST-CHEMGRIDS represents 20 of the most common pesticides for 6 individual crops and 4 aggregated crop groups, again at a scale of 5 x 5 minutes (Appendix 2, Figure S2.10). All analysis and data processing were carried out in R v.4.0.3 (R Core Team 2020).

Results

Pollinator dataset

I identified 1,013 possible pollinating genera across 3,974 abstracts in the initial automatic search of the pollination literature. After reading the abstracts associated with each genus, I confirmed 545 genera as likely pollinators at confidence levels 1-4. These 545 genera represented 141 unique families, of which 46 families, 10 subfamilies, and 5 tribes were judged to consist entirely of pollinators. Whilst consulting literature prioritised by the automatic search, I also identified an additional 51 genera with direct pollination evidence, which I assigned a confidence level between 1 and 4, and 18 additional families with extrapolated evidence. Filtering all expert-assessed pollinators from the PREDICTS database returned records for 4502 species in total, sampled at 12,170 sites (see Millard *et al.* 2021a 'Data availability' for the final list of pollinating species in PREDICTS). After selecting only sites in which land-use intensity and type were recorded in the PREDICTS database, a total 8,639 sites remained. The number of sites was highest in Europe (26.2%), North America (24.4%), and Africa (20%), and lowest in South America & the Caribbean (12.3%), Oceania (15.2%), and Asia (8.6%) (Figure 3.1).

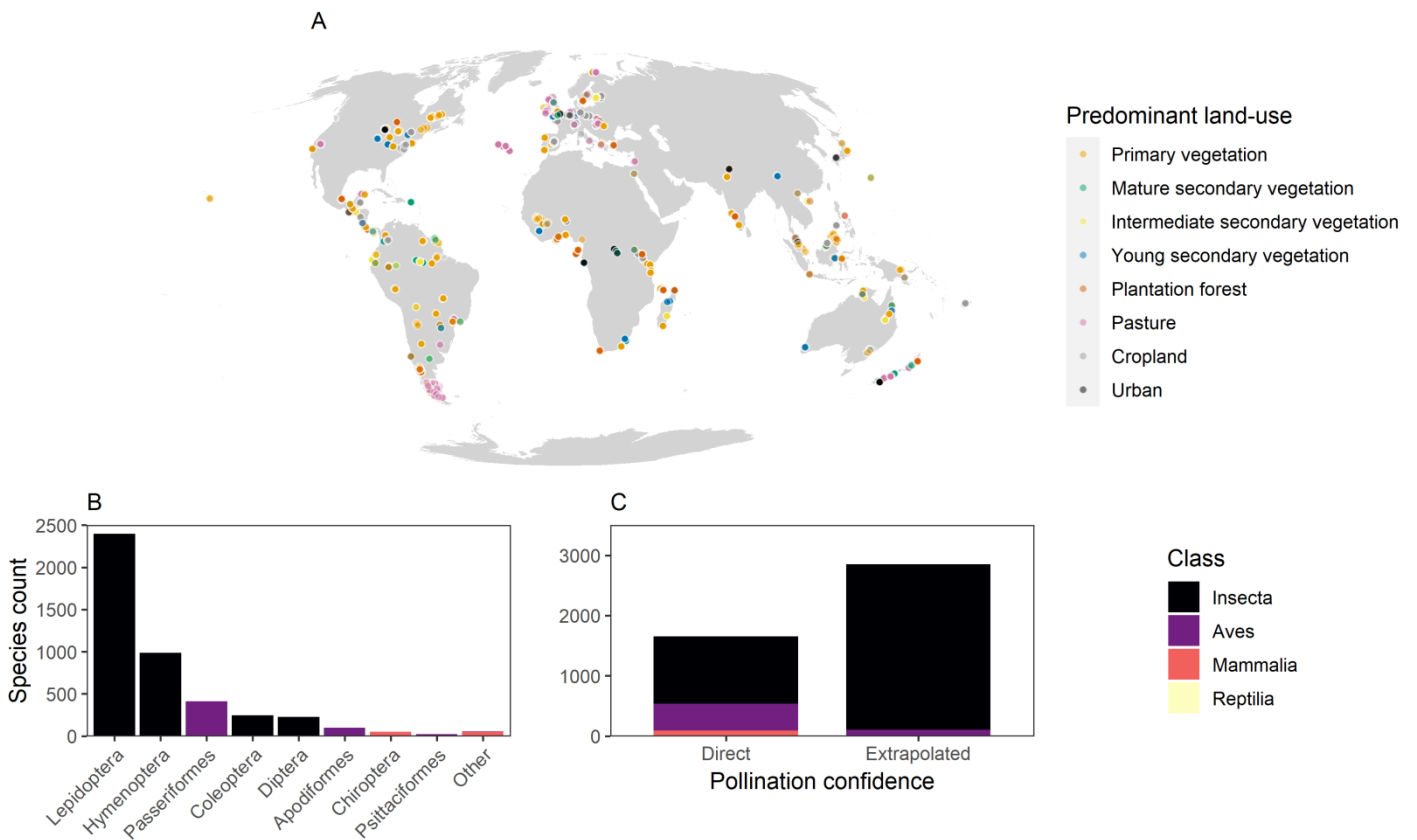


Figure 3.1 Data on species in the PREDICTS database identified as likely pollinators, after automatic text-mining, manual filtering, and expert consultation. **A)** The global distribution of PREDICTS sites containing likely pollinating species, for which both the land-use type and intensity of that site are known ($n_{\text{sites}} = 8639$). **B)** The taxonomic distribution of likely pollinating species in PREDICTS for all sites ($n_{\text{species}} = 4502$). The number of species indicated here will be an underestimate of the number of pollinating species in PREDICTS, since this figure only includes records for which there is a full scientific binomial in the database. Some biodiversity records in the PREDICTS database are recorded above the level of species. **C)** The source of information (direct evidence at the genus level or extrapolated to groups based on information for groups at higher taxonomic levels) for pollinators in PREDICTS, broken down by taxonomic class. In both B and C, there are four taxonomic Classes: Insecta (black), Aves (purple), Mammalia (red), and Reptilia (yellow). The reptiles are represented by only 5 species with ‘Direct’ confidence (see Appendix 2, Table S2.16).

Effect of land use and land-use intensity on global pollinator biodiversity

With human land-use type, land-use intensity, and their interaction fitted separately, increasing land-use intensity from minimal to intense use was associated with a significant change in pollinator biodiversity (species richness, $F = 9.4384$; total abundance, $F = 4.8075$, $p < 0.01$; Simpson diversity, $F = 11.6691$, $p < 0.01$; Figure 3.2; see Appendix 2, Table S2.17 for ANOVA tables of land-use intensity and type fitted separately). Land-use type was also a significant predictor (species richness, $F = 8.9440$; total abundance, $F = 8.0346$, $p < 0.01$; Simpson diversity, $F = 4.4150$, $p < 0.01$; see Appendix 2, Table S2.17), although declines occur more strongly within a land-use type as opposed to among land-use types (note that p values are not included for species richness since deriving these from F values for mixed effects generalised linear models is problematic).

Fitting human land-use type and intensity as a combined 'LUI' variable, relative to the primary vegetation minimal use baseline, biodiversity was often higher at low intensity (Figure 3.2). Indeed, with the exception of cropland and young secondary vegetation, all land-use types had species richness and total abundance significantly greater than the baseline, for at least one of low or intermediate intensity (Figure 3.2). Effects of land-use intensity were strongest in urban areas, with a 43% reduction for species richness and 62% for total abundance, between minimal and intense use. Plantation forest also experienced strong declines, decreasing by 38% for species richness. For anthropogenic land uses the weakest effects of land-use intensity were seen in pasture and cropland. Species richness did not decline significantly for pasture—although there was a 75% decline for total abundance—or for cropland for both total abundance and species richness. Young secondary vegetation did not significantly differ for total abundance, but for species richness declined between minimal and high intensity by 16%. All other secondary-vegetation types (mature and intermediate secondary vegetation) did not show significant differences in pollinator biodiversity among intensity levels (Figure 3.2). The AIC value for my Simpson

diversity LUI model was greater than the intercept-only model, meaning it was excluded from further analysis (see Appendix 2, Table S2.18). A zero-inflated negative binomial model for total abundance did not markedly change my predictions (Appendix 2, Figure S2.1). Similarly, neither did jack-knifing total abundance and species richness by continent (Appendix 2, Figure S2.2), including environmental covariates (Appendix 2, Figure S2.3), or controlling for abundance in our measure of species richness (Figure S2.4). There was significant spatial-autocorrelation in the residuals of only a small proportion of studies (2.33% of species richness studies and 4.65% of total abundance studies; Appendix 2, Figure S2.5).

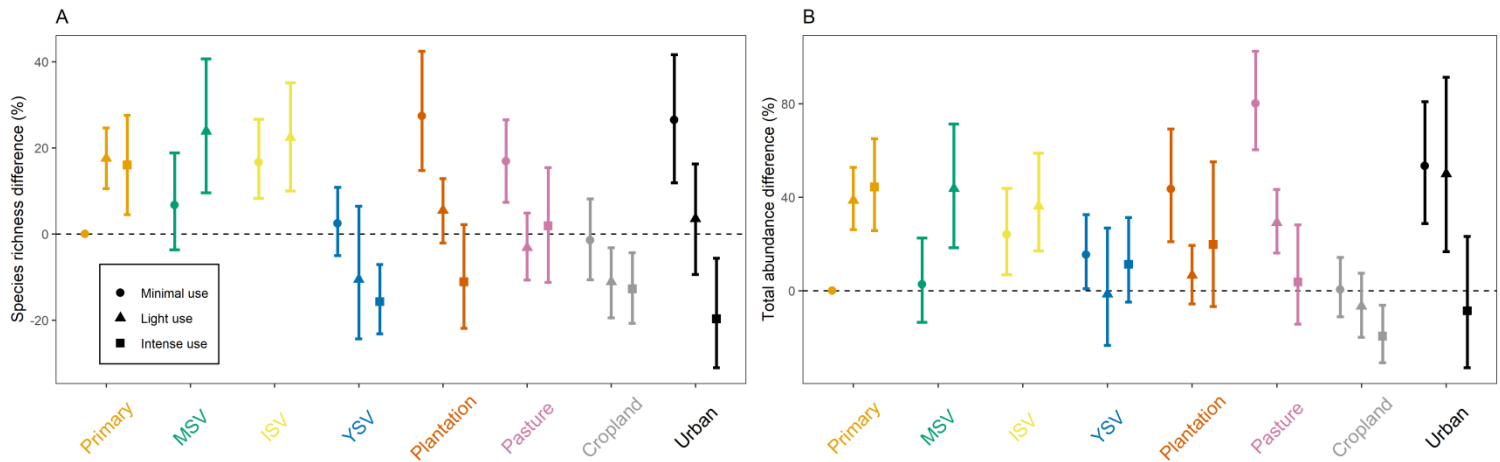


Figure 3.2 Responses of pollinator biodiversity to LUI (a combined variable of land-use type and intensity). Each panel represents a linear or generalised linear mixed-effects model for: A) species richness; and B) total abundance. I excluded Simpson diversity here since AIC was greater for the main model than the intercept-only model. Colours represent land-use type: orange (primary vegetation, Primary), green (mature secondary vegetation, MSV), yellow (intermediate secondary vegetation, ISV), blue (young secondary vegetation, YSV), dark orange (plantation forest, Plantation), pink (Pasture), grey (Cropland), and black (Urban), and point shape represents land-use intensity: circle (minimal use), triangles (light use), and squares (intense use). Effect sizes were adjusted to a percentage by drawing fixed effects 1,000 times based on the variance-covariance matrix, expressing each fixed effect within each random draw as a percentage of the baseline (primary vegetation minimal use), and then calculating the median value (shown as points) and 2.5th and 97.5th percentiles (shown as error bars). See Appendix 2, Table S2.19 for the number of sites and Tables S2.20 and S2.21 for the model summaries.

Effect of land-use intensity on pollinator biodiversity within croplands

Land-use intensity had a divergent effect on cropland pollinator biodiversity between the non-tropical and tropical geographical zones (Figure 3.3). In the non-tropical zone, species richness and total abundance did not differ significantly among cropland intensity classes, and were significantly higher in minimal-intensity cropland compared to the primary-vegetation baseline. In contrast, in the tropical zone, species richness and total abundance decreased between primary vegetation and high-intensity cropland by 44% and 49% respectively. Forest cover of the baseline did effect change relative to cropland, with relatively bigger declines from a low forest cover baseline, although the relative difference within cropland remains largely unchanged (Appendix 2, Figure S2.6. Greater variation in non-tropical areas is not predicted by high sample size (Appendix 2, Figure S2.7), and response for the main crop pollinators is likely consistent with all pollinators (Appendix 2, Figure S2.8). The AIC value for my Simpson diversity zone model was greater than the intercept-only model, meaning it was excluded from further analysis (see Appendix 2, Table S2.18). Response to total fertiliser application rate between the non-tropical and tropical zones was also insignificant for all of species richness, total abundance, and Simpson diversity (Appendix 2, Table S2.18), meaning it was excluded from further analysis.

Increasing land-use intensity in croplands had varying effects among taxa (Figure 3.4). Relative to primary vegetation, abundance declines at high intensity for the invertebrate pollinators were greater than 70% for all orders, and as high as 80% for the Lepidoptera and Diptera. The most consistent invertebrate declines were in the Lepidoptera, exhibiting a negative response across a gradient of intensity for species richness, total abundance, and Simpson diversity. For flies on the other hand, relative to minimally used cropland, intermediate levels of intensity were associated with higher species richness and total abundance. For the vertebrates, the Apodiformes exhibited a strong negative response to land-use intensity, declining by at least 20% from medium-intensity cropland to primary

vegetation for all three metrics (high-intensity cropland was not sampled for this taxonomic group). The Passeriformes experienced a significant reduction from the baseline to high intensity, of 30% for total abundance, 36% for species richness, and 26% for Simpson diversity.

Response to total fertiliser application rate in surrounding cropland landscape differed strongly in magnitude and direction (Figure 3.5). Both Hymenoptera and Lepidoptera showed a strong negative response to increasing fertiliser application rate for both species richness and total abundance. In particular, an increase of 1000kg/ha in fertiliser application rate was associated with a reduction of 44% in hymenopteran total abundance, whereas lepidopteran abundance fell 50% over the same range. Dipteran richness and abundance, on the other hand, increased markedly by 760% and 374% respectively. Coleopteran response to total fertiliser application rate was insignificant for all of species richness, total abundance, and Simpson diversity. For the vertebrates, the Apodiformes increased by 163% for species richness, whereas the Passeriformes experienced no marked change for any of species richness, total abundance, or Simpson diversity. Although the AIC value for my Simpson diversity model was less than the null model (Appendix 2, Table S2.18), all interactions between total fertiliser application rate and taxonomic order for Simpson diversity were insignificant (Figure 3.5). On the whole the direction and magnitude of the response for pesticide application rate is similar to that of fertiliser application rate (Figure 3.6).

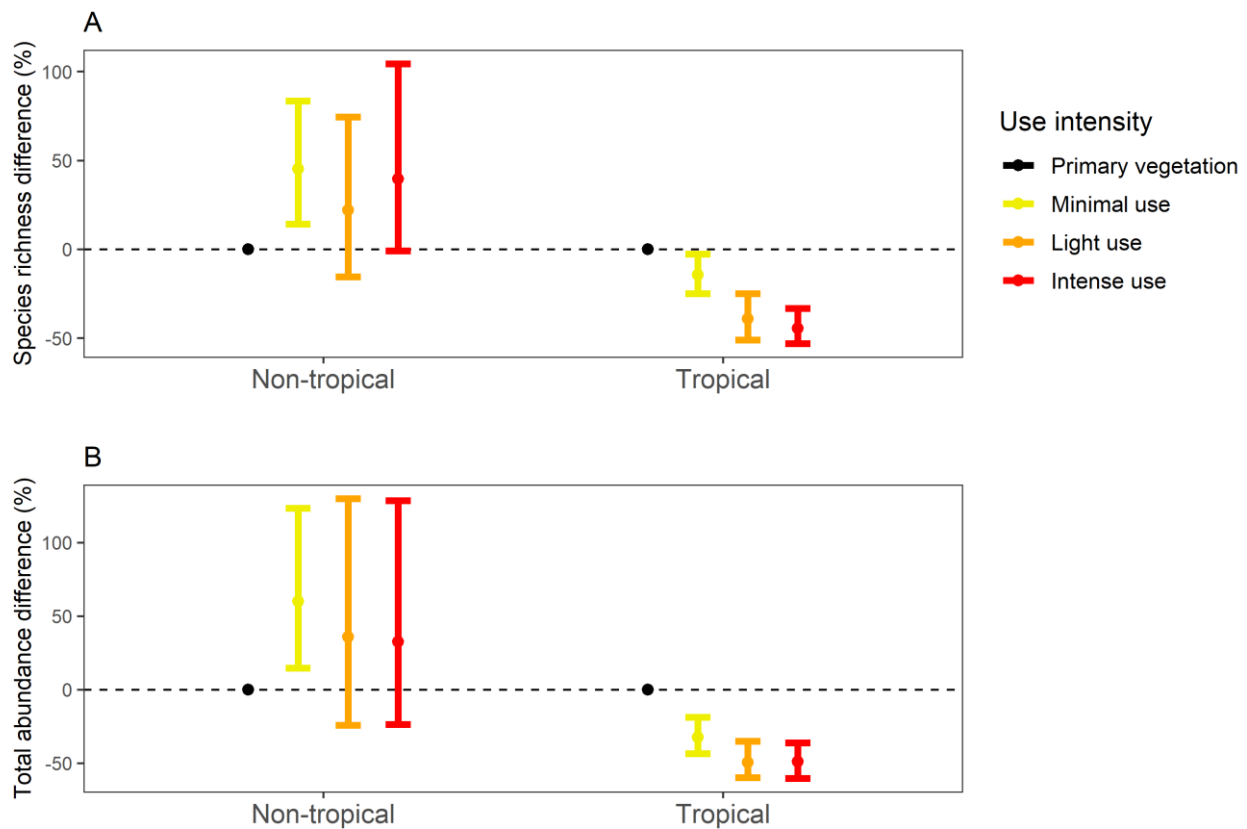


Figure 3.3 Response of pollinators to land-use intensity on cropland, for non-tropical and tropical sites. Each panel represents a linear or generalised linear mixed-effects model for a given biodiversity metric: A, species richness; and B, total abundance. I excluded Simpson diversity here since AIC was greater for the main model than the intercept-only model. Colours represent the land-use intensity level, with primary vegetation (minimal use) as the reference factor: black (primary vegetation, minimal use); yellow (cropland, minimal use), orange (cropland, light use), and red (cropland, intense use). Effect sizes were adjusted to a percentage by sampling fixed effects 1,000 times based on the variance-covariance matrix, expressing each fixed effect as a percentage of the value in primary vegetation for that geographical zone, and then calculating the median value (shown as points), and 2.5th and 97.5th percentiles (shown as error bars). See Appendix 2, Table S2.22 for the number of sites and Tables S2.23 and S2.24 for the model summaries.

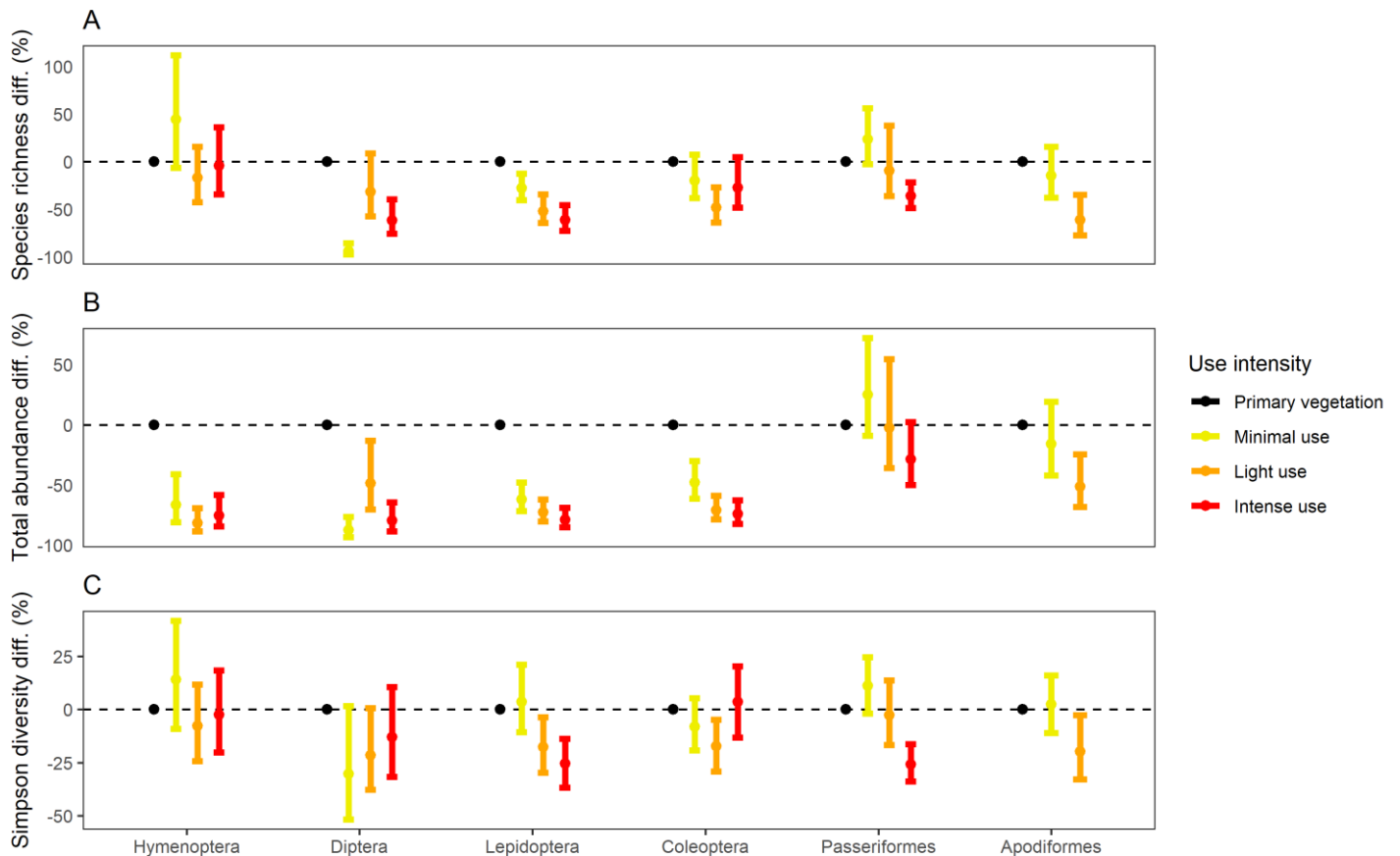


Figure 3.4 Response of different pollinator groups to land-use intensity in cropland. Each panel represents a linear or generalised linear mixed-effects model for a given biodiversity metric: A, species richness; B, total abundance; and C, Simpson diversity. Colours represent the land-use intensity level, with primary vegetation (minimal use) as a reference factor: black (primary vegetation, minimal use); yellow (cropland, minimal use), orange (cropland, light use), and red (cropland, intense use). Effect sizes were adjusted to a percentage by sampling fixed effects 1,000 times based on the variance-covariance matrix, expressing each fixed effect within each random sample as a percentage of the value in primary vegetation for that taxonomic order, and then calculating the median value (shown as points), and 2.5th and 97.5th percentiles (shown as error bars). See Appendix 2, Table S2.25 for the number of sites and Tables S2.26, S2.27, and S2.28 for the model summaries.

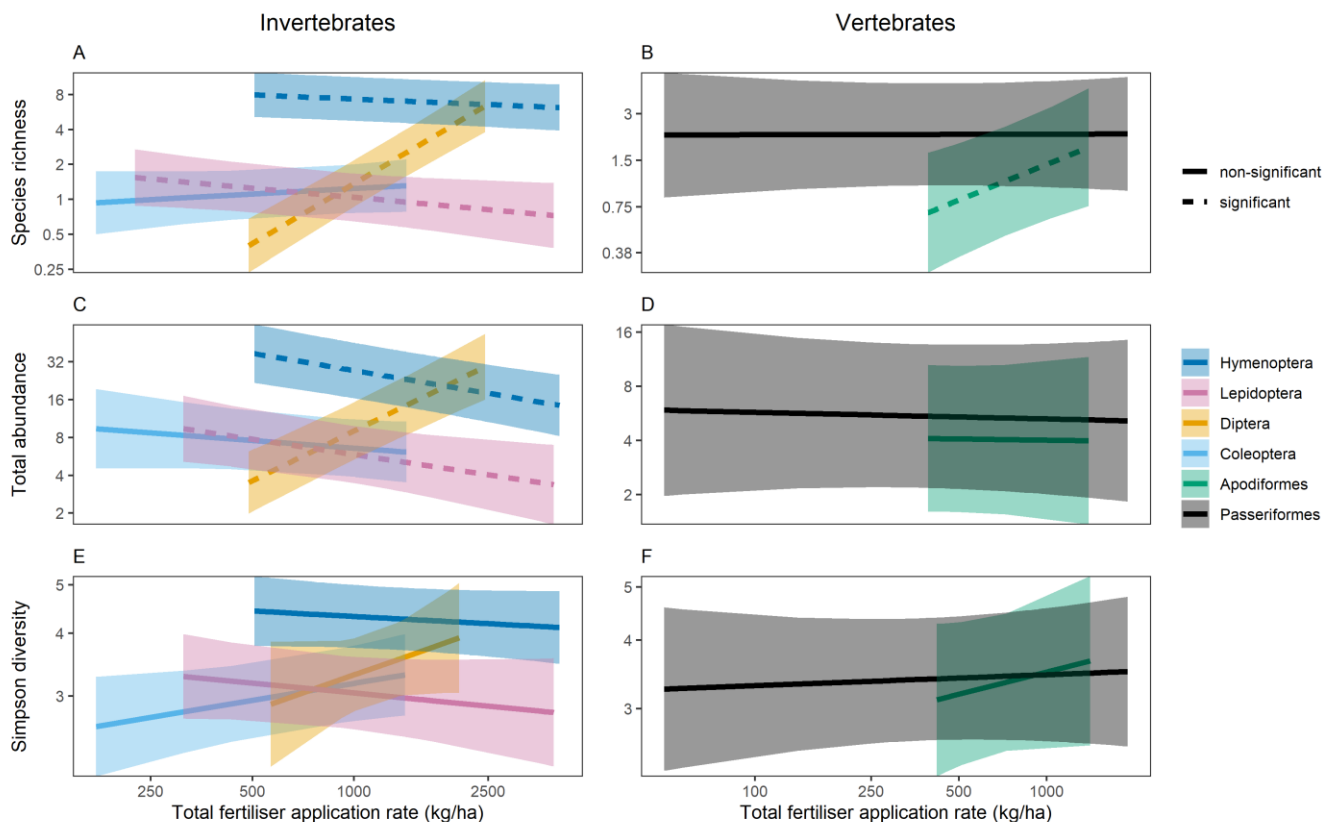


Figure 3.5 Response of pollinator biodiversity in cropland sites to total fertiliser application rate in the landscape (kg/ha) (note that each metric is plotted on a log scale), predicted across 95% of the range of sampled fertiliser values for each taxonomic order. Each panel represents a linear or generalised linear mixed-effects model for a given biodiversity metric for four invertebrate orders: A, C, and E, invertebrate species richness, total abundance, and Simpson diversity respectively; and B, D, and F vertebrate species richness, total abundance, and Simpson diversity respectively. Coloured lines represent median fitted estimates for each taxonomic order, with shading representing 95% confidence intervals: light blue (Coleoptera), light orange (Diptera), dark blue (Hymenoptera), pink (Lepidoptera), green (Apodiformes), and black (Passeriformes). Dashed lines represent significant interactions between taxonomic order and total fertiliser application rate, and solid lines non-significant interactions. See Appendix 2, Table S2.29 for the number of sites, Tables S2.30, S2.31, and S2.32 for the model summaries, and Appendix 2, Figure S2.10 for the global distribution of total fertiliser application rate.

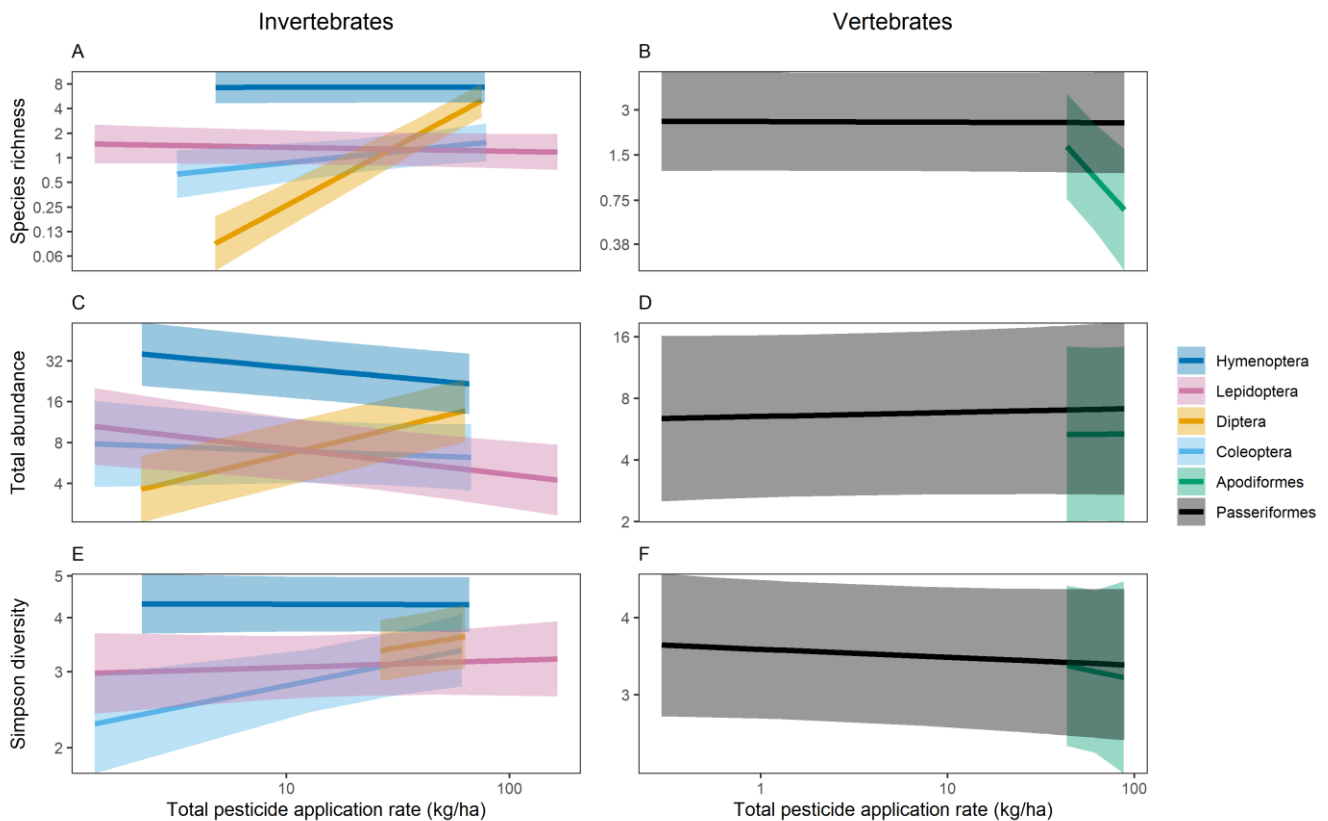


Figure 3.6 Response of pollinator biodiversity in cropland sites to total pesticide application rate in the landscape (kg/ha) (note that each metric is plotted on an absolute value log scale), predicted across 95% of the range of pesticide values for each taxonomic order. Each panel represents a linear or generalised linear mixed-effects model for a given biodiversity metric for four invertebrate orders: A, C, and E, invertebrate species richness, total abundance, and Simpson diversity respectively; and B, D, and F vertebrate species richness, total abundance, and Simpson diversity respectively. Coloured lines represent median fitted estimates for each taxonomic order, with shading representing 95% confidence intervals: light blue (Coleoptera), light orange (Diptera), dark blue (Hymenoptera), pink (Lepidoptera), green (Apodiformes), and black (Passeriformes). Each biodiversity metric is predicted across 95% of the range of its values, meaning each response is not estimated across the same range of total pesticide application rate.

Discussion

Using a space-for-time approach, here I showed that land-use intensity is associated with significant changes (both positive and negative) in pollinator species richness, total abundance, and Simpson diversity, particularly for invertebrate pollinators. This study represents, as far as I know, the largest global analysis of the response of animal pollinator biodiversity to land-use type and intensity, and the first to consider large-scale differences in responses among taxa, geographic regions, and biodiversity metrics. My results are consistent with previous analyses showing reductions in overall pollinator biodiversity at high land-use intensity (Clough et al. 2014), and increases at low-intermediate intensity (Lazaro et al. 2016). In particular, low-intensity urban sites (villages and green spaces; Hudson et al. 2017) have higher pollinator biodiversity than the primary vegetation baseline, but at high intensity urban pollinator species richness is significantly lower than the baseline (although this was not the case for abundance). This is concordant with previous research demonstrating that urban areas can support species-rich and abundant pollinator populations (Hall et al. 2017; Baldock et al. 2019). I also highlight that strong negative responses to land-use intensity within croplands are largely restricted to the tropics, with no apparent effect (and even increases at low intensity) in non-tropical cropland. This is an important result, given the dominance of animal pollinated plants in tropical environments (Ollerton et al. 2011; Rech et al. 2016), and that rapid agricultural expansion is predicted to occur in the tropics in the near future (Hurt et al. 2011). Furthermore, I show pronounced differences in response among taxonomic groups, consistent with time-series studies showing differential trends among UK invertebrate taxa (Outhwaite et al. 2020). Pollinator biodiversity change resulting from land-use intensity may have consequences for pollination (Rader et al. 2014) and crop yields (Woodcock et al. 2019), especially in the tropics: Although the abundance of some pollinating groups (i.e. flies) is greater on intensively fertilised cropland, increases may not compensate for overall losses across other pollinating groups.

Differences in response to intensity between tropical and non-tropical areas are likely driven by the interacting effects of historical land-use and climate sensitivity, which differ between the temperate and tropical zones. Non-tropical temperate regions have a long history of anthropogenic land-use, which has likely filtered out many sensitive species (Balmford 1996), meaning that contemporary differences in land use and land-use intensity may be weakly associated with pollinator biodiversity. Indeed, historical land-use has been shown to be strongly associated with current species richness and abundance of insect pollinators (Cusser et al. 2018; Cusser et al. 2015; Bommarco et al. 2014), and may contribute towards an overall increase in pollinator biodiversity at low land-use intensity (Tscharntke et al. 2005). The tropical zone, on the other hand, has a shorter history of intensive agricultural land-use (DeFries and Rosenzweig 2010), meaning recent intensification has stronger effects on pollinator community composition. Tropical biodiversity is also thought to be more sensitive to the effects of climate change (Deutsch, et al. 2008), which may be exacerbating the effects of land use (Williams et al. 2020a). Further research is required to tease out the relative contribution of historical land-use and climate change to tropical pollinator biodiversity.

Pollinating insects across multiple geographic locations and taxonomic groups have been reported to have declined, particularly for biomass and abundance (Seibold et al. 2019; Hallmann et al. 2017). In my space-for-time analysis, I found significantly lower abundance on high-intensity cropland—relative to primary vegetation—for all insect orders, especially for the hymenopterans, lepidopterans, and coleopterans. I also found significant reductions in insect abundance and richness in response to increasing fertiliser application rate on croplands, particularly for the hymenopterans and lepidopterans. Sensitivity to land-use for butterflies has previously been attributed to dietary specialism: relative to flies, butterflies are known to be dietary specialists as larvae, making a reduction in lepidopteran species richness and diversity likely when plant species richness is reduced (Weiner et al. 2011; Weiner et al. 2014; Winfree et al. 2011; Scoble 1992), which is known to occur at high

fertiliser application rates (Woodcock et al. 2016; Jones and Chapman 2011; Kleijn et al. 2009; Ridding et al. 2020). Previous research in the temperate zone has indicated beetle flower visitors are sensitive to land-use change (Weiner et al. 2011; Weiner et al. 2014). Although I found a decrease in coleopteran abundance relative to the primary vegetation baseline, response to fertiliser application rate was mixed and insignificant. It may be that fertiliser application buffers against the more negative effects of associated intensity, since some pollinating coleopterans are known to have a larval preference for fertile soils (Parker and Howard 2001). For the hymenopterans, sensitivity to land-use has been reported previously, particularly for solitary bees (Rader et al. 2014; Gathmann and Tschamntke 2002) which are on average highly specialised (Weiner et al. 2011).

Dipteran abundance and species richness increased with fertiliser application rate, concordant with previous studies demonstrating increased dipteran biodiversity on managed land (Kühnel and Blüthgen 2015; Jauker et al. 2009; Biesmeijer et al. 2006; Weiner et al. 2011; Weiner et al. 2014). Dipteran resilience to land use and land-use intensity has been attributed to a number of traits, including low dietary specialisation on floral resources (Weiner et al. 2014; Weiner et al. 2011), high mobility (Haenke et al. 2009), absence of parental care (Jauker et al. 2009), and larval preference for agricultural habitats (Jauker et al. 2009). Syrphid fly larval development in agricultural land is particularly of note. Semi-aquatic syrphid larvae are known to favour eutrophic or manure-contaminated habitats (Speight 2017), which is consistent with the strong positive gradient for fly richness and abundance in response to increasing fertiliser application rate. In contrast, the response of fly abundance and species richness to my overall measure of land-use intensity was negative, suggesting that fertiliser application does not sufficiently capture all aspects of land-use intensity on cropland. Neonicotinoids such as imidacloprid, for example, have adverse effects on flies, given their association with declines in insects in general (Wood and Goulson 2017; Goulson 2013), and visitation rate in flies specifically (Easton and Goulson 2013). I used the PEST-CHEMGRIDS global estimation of pesticide application rate (Maggi

et al. 2019) to build additional models of pollinator response between tropical and non-tropical croplands, finding that the response for flies to pesticide application rate is in fact similar to fertiliser (Appendix 2, Figures S2.10 and S2.11), the cause of which is unclear. Further analysis is required to tease out the relative effects of fertiliser and pesticide application at the global scale.

Compared to invertebrates, vertebrate pollinators appear to be less sensitive to the effects of land-use intensity, particularly with respect to change in total abundance. Relatively higher resilience to land-use intensity has been found for vertebrate pollinators in tropical forests (Tscharntke et al. 2008), and suggested in broad comparisons across taxa (Winfree et al. 2011). However, although previous work found that vertebrate resilience diminishes after controlling for study design (Winfree et al. 2011), my results suggest that vertebrate pollinators are indeed less sensitive than invertebrate pollinators to increasing land-use intensity. Such relative vertebrate resilience likely relates to body size and mobility (Henle et al. 2004; Bommarco et al. 2010), both of which are typically greater in vertebrate pollinators.

Change in global pollinator biodiversity resulting from land-use intensity may have consequences for crop pollination. Since the loss in pollination service provided by sensitive crop-pollinating taxa (Hymenoptera) will likely not be offset by gains in more resilient taxa (Passeriformes), the service will experience a net loss. Four lines of evidence indicate losses in sensitive taxa will not be buffered by gains in more resilient taxa: first, relative to bees, those species which are increasing (i.e. dipterans) are known to contribute less to crop pollination (Ollerton 2021), and do not necessarily pollinate the same set of crops; second, response to land-use intensity of the main crop pollinating groups (bees, flies, beetles, wasps, thrips, birds, and bats) appears consistent with responses across all pollinators (Appendix 2, Figure S2.8); third, there is evidence from multiple historical localities that significant pollinator deficits can result from the losses associated with intensive agriculture (Watanabe 1994; Kevan 1977; Stephen 1955), irrespective of differences among taxonomic groups; and fourth, significant crop-yield reduction has been linked to overall changes in

pollinator biodiversity in multiple studies (Kremen et al. 2002; Garibaldi et al. 2013; Garratt et al. 2014; Blitzer et al. 2016; Holzschuh et al. 2012; Klein, et al. 2003), again irrespective of taxonomic differences.

My analysis of pollinator biodiversity change is subject to limitations. First, the nature of our study as a space-for-time analysis means we may overlook extinction-debt effects (De Palma et al. 2018). Such effects can be controlled for by assessing change over time at a specific location or region (Seibold et al. 2019; Hallmann et al. 2017). However, at the global scale, given that long-term studies are lacking, space-for-time analyses represent a necessary alternative (Purvis et al. 2018). Second, most of the results I present here are relative to a baseline of primary vegetation with minimal human use, which inevitably varies in nature, especially between tropical and non-tropical sites (Appendix 2, Figures S10 and S11). In particular, we might expect non-tropical primary habitats to be more open than tropical primary habitats, which is likely to be more favourable to pollinators, and thus might partly explain the absence of responses to land-use intensity in non-tropical areas. Indeed, I show that the margin of reduction from a low forest-cover baseline is greater than from a high forest-cover baseline (Appendix 2, Figure S2.6). Nevertheless, the overall responses for tropical and non-tropical areas remain unchanged. Third, I analyse raw species richness which I recognise may be confounded by abundance. However, since my validation model predicting estimated species richness (using the Chao1 estimator) does not differ markedly from raw richness, a confounding effect of abundance is unlikely (Appendix 2, Figure S2.4). Fourth, my dataset of pollinator biodiversity responses is spatially biased towards the non-tropics, particularly Europe and North America. Given that tropical pollinators are affected more negatively, my overall results therefore likely underestimate the impact of land-use intensity. However, my continental jack-knife for my overall LUI models showed that the exclusion of any of the Americas, Europe, or Africa (the continents for which we have the most sites) did not markedly influence our predictions. Moreover, my additional validation analysis in which I re-sampled 1000 sites from each of the tropics and non-tropics would

indicate that greater variation in non-tropical regions is likely not predicted by greater sample size (Appendix 2, Figure S2.7). Fifth, total fertiliser application rate was estimated at a relatively coarse spatial scale, which for some pollinators—particularly those such as insect pollinators which respond strongly to more localised change—will not be meaningful. Since spatial scale is known to predict pollinator response to land-use intensity (Brittain et al. 2010), I could infer a different response if more localised fertiliser estimates were available globally. Sixth, I recognise that evidence for a given species will not always be representative of all species in a whole genus, or all life-history stages within a species. For example, the species *Crociodura cyanea* (Duvernoy, 1838; Common name: Reddish-gray musk shrew; Order: Eulipotyphla; Family: Soricidae) has been found to feed on and carry pollen, but most other *Crociodurans* are insectivorous (Melidonis et al. 2015). As a result, pollination confidence for many species in the genus *Crociodura* will be less than the genus-level evidence would imply. In the main, I assumed evidence for a single species would be representative of whole genera, given the association between phylogeny and traits (Cusser et al. 2018). I also reasoned that for some insect groups, searching at the species level would be ineffective, given the large number of species with little evidence. Seventh, species confirmed as pollinating one flowering plant will not necessarily make an important contribution to all flowering plants, or to the pollination of crops.

Anthropogenic activity has significantly altered the biosphere. Such changes have had, and will continue to have, profound consequences for animal pollinator biodiversity. Here I demonstrated significant pollinator biodiversity change in response to land-use intensity, with both negative and positive effects. Within an anthropogenic land-use type, I showed that intensity often decreases pollinator biodiversity. But relative to the primary vegetation minimal-use baseline, pollinator biodiversity was often greater at low and intermediate levels of intensity, suggesting that some level of disturbance can be beneficial. Across taxa within croplands, I again showed a mixed response to intensity, varying according to both facet of intensity and taxonomic group. Specifically in the tropics however, pollinators appear highly

sensitive to land-use intensity; a situation that may worsen as intensive agriculture continues to expand. Moreover, it is also likely that climate change will drive further changes in tropical pollinator biodiversity, particularly for insects which respond strongly to changes in ambient temperature (Soroye et al. 2020; Kühnel and Blüthgen 2015). Further research is required to resolve better the way in which these threats interact at the global level. For crop pollination services in the tropics, the repercussions of land-use and climate change could be great, with a growing body of evidence indicating high wild pollinator biodiversity is required to sustain productive yields (Winfree et al. 2018). Although the complexities of this relationship are not yet fully understood, there is sufficient evidence to suggest that pollination shortfalls in the tropics could result from continuing anthropogenic intensification and expansion.

Chapter 4: Worldwide vulnerability of local pollinator abundance and crop pollination to land-use and climate change

Abstract

Global climate and agricultural extent are predicted to change rapidly over the coming decades, with potential consequences for ecosystem services. For pollinators, it is unclear how the combined effects of land-use and climate change will drive shifts in biodiversity and the service they provide in the form of crop pollination. Here I model the response of pollinating insect and vertebrate abundance to agricultural land use, historical climate change, and their interaction. Given future scenarios of climate change (RCP 2.6, 6.0, and 8.5), and the global distribution of crops that depend on animal pollination, I then use my models to explore potential changes in global crop pollination risk. I show that insect pollinators experience strongly negative declines where cropland coincides with high levels of historical climate change. Under RCP 8.5—a high emissions future climate scenario—my total crop production risk index is predicted to increase by ~70% between 2016 and 2048. Again under RCP 8.5, the rate of change in total crop production risk could be up to ~25% higher than that of RCP 2.6 and 6.0, with abrupt periods of increased exposure in the 2030s. I further show that crop pollination risk is highest, and predicted to increase most rapidly, in regions of sub-Saharan Africa and northern South America, primarily for cocoa, mango, pumpkin, melon, watermelon, and coffee. This study provides the first global framework for predicting current and future crop pollination risk from land-use and climate change. Mitigating this risk will be a key challenge of the 21st Century.

Introduction

Global climate and agriculture are predicted to change rapidly in the coming years (IPCC 2018; Laurance et al. 2014). In 2015, the international community agreed on a set of policies aiming to keep change in climate to below 1.5°C, but recent research indicates this threshold will likely be exceeded by 2050 (IPCC 2018). Over the same period agriculture is expected to expand significantly, meeting an increasing global demand for both food and bioenergy (Laurance et al. 2014). The effects of future agriculture and climate change on the biosphere could be great. Although there is a debate regarding the magnitude of local change (Dornelas et al. 2014), on average agricultural activity has been shown to reduce and homogenise local biodiversity (Newbold et al. 2015a, 2018), such that agriculture likely represents the single greatest current threat to terrestrial mammals and birds (Joppa et al. 2016). Climate change on the other hand has been associated with elevational, latitudinal, and phenological shifts (Bellard et al. 2012), as well as predicted abrupt future changes in the species composition of communities (Trisos et al. 2020).

The individual effects of climate change and agricultural land use on biodiversity are great, but in combination they have been shown to have synergistic interactive effects (Frishkoff et al. 2016; Hendershot et al. 2020; Oliver et al. 2016; Williams et al. 2020b). Such interactive effects act through either ‘chain effects’ or ‘modification effects’ (Didham et al. 2007; Oliver & Morecroft 2014). Chain effects occur when one driver alters the magnitude of another. For example, areas of agricultural land use tend to be warmer and drier (Frishkoff et al. 2018), meaning on average degree of climate change is magnified on anthropogenic land. Modification effects refer to situations where the effect of one driver on biodiversity is moderated by change in another (Didham et al. 2007; Oliver et al. 2014). For example, biodiversity responses to extreme climatic events are modified by the structure of the landscape, with woodland patches reducing insect susceptibility to droughts (Oliver et al. 2012).

The interactive effects of climate change and agricultural land use on biodiversity have been the focus of a growing number of studies, but there are still at least two big research gaps. First, for important ecosystem service providers such as pollinators, it's unclear how climate change and agricultural land use might interact globally to drive biodiversity change (although there have been localised studies e.g. Oliver et al. 2012; Zaragoza-Trello et al. 2021). Given the sensitivity of insects to changes in microclimate (Boggs 2016) one might expect that, relative to vertebrate pollinators, insect pollinators will respond strongly to the interactive effect of climate change and land-use. Within the insects, one might also expect that pollinators will respond more strongly than non-pollinators, since insect pollinators tend to be associated with a set of traits that predict high sensitivity (e.g. high dietary specialism) (Winfrey et al. 2011). Second, given the uncertainty around interactive effects on service providers, there is an ongoing question as to the extent to which the actual service to people may be affected, particularly with regards to the temporal and spatial scale at which consequences might play out. Since the interactive effects of land use and climate change on biodiversity are non-linear, one potential scenario is that future anthropogenic activity leads to rapid biodiversity change, with knock-on socio-economic effects that ripple out from agricultural areas that are highly dependent on animal pollinators (Garibaldi et al. 2016; Silva et al. 2021). Insufficient data for pollinating insects has previously made addressing both the first and second of these research gaps difficult, but appropriate datasets are now more widely available (Millard et al. 2021a).

Crop pollination is one service that could be affected following rapid change in agricultural land use and climate. For example, as agricultural regions with high dependence on animal pollination begin to be exposed to novel extreme temperatures, and pollinators experience an associated change in biodiversity, these localities may be at sudden risk of pollination shortfall. Shortfalls will likely be best predicted by changes in pollinator abundance, given that a large quantity of the pollination service tends to be provided by a relatively small group of dominant species (Kleijn et al. 2015). Shortfalls will also relate to measures of taxonomic and functional biodiversity, given the connection between these measures and crop yield (Dainese et al.

2019; Woodcock et al. 2019), but since abundance has a consistent relationship with crop pollination service (Woodcock et al. 2019), here I focus on change in pollinator abundance as a measure of risk.

Given the uneven global distribution of crops dependent on animal pollination, and the spatial heterogeneity of climate change and agricultural activity, the likelihood of future crop pollination shortfall will not fall evenly. This is for three primary reasons. First, tropical agriculture is more dependent on animal pollination, with animal pollinators an essential requirement for the production of tropical crops such as cocoa, vanilla, Brazil nuts, Macadamia nuts, and kiwifruit (Klein et al. 2007). Temperate agriculture does benefit to some extent from animal pollination (e.g. apples, pears, cherries, apricots, raspberries; see Klein et al. 2007), but the proportion of agricultural production dependent on pollination is markedly lower (Aizen et al. 2019). Second, climate change is occurring fastest at high latitudes (IPCC, 2018), but the greatest rate of exposure to novel extreme temperatures is taking place in the tropics (IPCC 2018). Although Arctic pollinators have been shown to respond strongly to high latitude warming (Schmidt et al. 2016), a growing body of evidence indicates that novel temperatures are more important to insect pollinator biodiversity change in the tropics, since novel temperatures act to push insects beyond their upper thermal limits (Boggs 2016). Third, during this century the tropical regions of sub-Saharan Africa and South America will experience the greatest relative growth in intensive agriculture (Laurance et al. 2014), which will likely lead to particularly strong changes in pollinator biodiversity (Millard et al. 2021a). In the tropics, then, there is a conjunction of high pollination dependence, high vulnerability to climate change, and high potential for agricultural expansion, meaning if significant shortfalls are to occur anywhere, it is likely that they will occur at tropical latitudes.

Global production of crops dependent on animal pollination today exists within a network of international trade, meaning the consequences of pollination shortfall will not necessarily be localised (Silva et al. 2021). For localities producing animal-pollinated crops for local consumption, any shortfall resulting from pollinator loss will play out at a local scale (Garibaldi

et al. 2016). But for localities that ordinarily export large quantities of pollination-dependent crop production to multiple countries, one might expect fluctuations in global supply and consequent volatility of crop price. China, for example, is one of the main global exporters of crops dependent on animal pollination (Teichroew et al. 2017), by both the quantity of exports and the number of importing countries (Silva et al. 2021). As a result, one might expect crop pollination shortfalls in China to have consequences that extend beyond the country itself (Teichroew et al. 2017). Predicting when these rippling effects may occur, however, is complicated. One indicator might be high reductions in pollinator abundance in regions that produce a high quantity of animal pollination dependence crops, and the future year at which this occurs most rapidly. Some objective measure incorporating these factors would be a significant step towards an index of global risk to crop pollination shortfall.

Here, I model the effect on pollinator abundance of agricultural land use and its interaction with historical climate change, before exploring potential risk to the production of crops that depend on animal pollination. I construct my initial models using the PREDICTS database (Hudson et al. 2017) and a set of likely pollinating and non-pollinating species derived previously (Millard et al. 2021a). I then use future scenarios of climate change (Frieler et al. 2017), a set of rasters for global crop production (Monfreda et al. 2008), and estimates of the proportional dependence of different crops on animal pollination (Klein et al. 2007), to explore potential scenarios of change in crop pollination risk. Specifically, I address two core questions: 1) What is the difference in response to the interactive effects of agricultural land use and climate change between those species that pollinate (insects and vertebrates) and those that do not? And 2) given predicted changes in abundance of insect pollinators on cropland (based on scenarios of future climate change), and the global distribution of crops production that depends on animal pollination, where and when might we expect a risk of shortfalls for crop pollination in the future? I hypothesise that the interactive effect of climate change and agricultural land use will be stronger in insect pollinators relative to vertebrate pollinators, and that this interactive effect will be such that decline in abundance is greater on

cropland than primary vegetation. I also hypothesise that crop pollination vulnerability to future climate change will be greatest in the tropics, and that rate of change in crop pollination risk will experience abrupt increases as regions of high pollination-dependent crop production are exposed to novel extreme temperatures.

Methods

Pollinator dataset construction

I used the PREDICTs database, and a set of likely pollinating species identified in PREDICTS (see Chapter 3 for a detailed background on the development of this PREDICTS database subset), to model the response of local pollinator abundance to the interactive effect of climate change and agricultural land use. PREDICTS is a global database of local biodiversity records, with a hierarchical structure such that each record is nested according to a series of levels (i.e. ‘source’, ‘study’, ‘block’, and ‘site’; see Chapter 3 Appendix 2 Figure S2.14, and Hudson et al. 2017). Each record in PREDICTS is associated with a land use type (primary vegetation, mature secondary vegetation, intermediate secondary vegetation, young secondary vegetation, plantation, pasture, cropland, and urban) and land use intensity (high, low and minimal), meaning change in biodiversity can be predicted as a function of anthropogenic land-use disturbance (e.g. Newbold et al. 2015a; also see Chapter 3 ‘Methods’). Land-use type categories in PREDICTS are defined according to the Representative Concentration Pathway land-use estimates (Hurtt et al. 2011), plus plantation and a gradation of stages according to age within secondary vegetation (Hudson et al. 2014). Land-use intensity for each land-use type is defined according to a series of variables, such as fertiliser and pesticide application, mechanisation, and hunting (see Newbold et al. 2015a for more details).

I identified pollinating species in PREDICTS through a semi-automatic approach combining text-mining, manual inspection, and expert consultation. I first used the stemmed term ‘pollinat*’ on Scopus to return all abstracts for English language primary research papers. From this set of abstracts, I then used a set of name-entity recognition algorithms to extract

all animal species Latin binomial names (see Chapter 2 ‘Methods’; Millard et al. 2020a for details). For each animal genus returned by the name-entity recognition algorithm, I then read the corresponding abstracts searching for evidence confirming that genus as pollinating. Given my set of pollinators identified from Scopus abstracts could only ever be a sample, I then searched for higher-level taxonomic groups, aiming to identify broader groups of animals that we could be confident would pollinate (see Chapter 3 ‘Methods’ for more details). After compiling my list of pollinators from automated text-analysis and manual searching, I then consulted a group of 7 expert pollination ecologists, and removed or added any groups at their suggestion (see Chapter 3 ‘Methods’ for details).

I then built four separate datasets with different subsets of the original set of species in PREDICTS: pollinating vertebrates, non-pollinating vertebrates, pollinating insects, and non-pollinating insects. My set of non-pollinating species are not strictly a set of confirmed non-pollinators, rather a set of species not confirmed as pollinators (see Discussion for an exploration of this limitation). For each of my four data subsets, I then calculated site-level total abundance (the sum of all species sampled abundances at a site). Sampling effort was accounted for by dividing the abundance values for each measurement by the relative sampling effort among sites within a study, rescaled to a maximum value of 1 in each study, as in (Newbold et al. 2014a).

Climate change data

I used Climatic Research Unit Time Series (CRU TS) v4.03 (Harris et al. 2020) mean daily temperature estimates per month at a spatial resolution of 0.5° to calculate a global standardised temperature anomaly (STA) for the year of each PREDICTS sample, using an approach developed previously (Outhwaite et al. in review; see full details on methodology below). Although extreme temperatures have been shown to predict contemporary changes in biodiversity (Mantyka-Pringle et al. 2011), I used mean temperatures since they provide a measure of the overall change in temperature per month. Mean temperatures have been used

in a similar manner for a number of other studies on insect thermal tolerance (e.g. Deutsch et al. 2008; Johansson et al. 2020). To calculate my standardised temperature anomaly, I first calculated a 30-year baseline temperature for the years 1901-1930 as the grand mean temperature over all 360 monthly mean daily temperatures for each cell. For each PREDICTS site, I then extracted the sample end date (i.e. the last date at which that measurement was taken) for the measurements at each site, and then for the month of that end date and the 11 months previously I calculated the grand mean temperature over all monthly mean daily temperatures for that sample. I then calculated a climate anomaly for each site as the difference between the 30-year baseline mean and the mean annual temperature at the time of the sample. I then standardised this climate anomaly by dividing the anomaly at each site by the standard deviation of the baseline, calculated as the standard deviation across monthly mean daily temperatures for the same 30-year baseline period. A standardised temperature anomaly of less than 0 indicates a region that has cooled since the baseline. A value between 0 and 1 indicates a region that has warmed, but current average temperature remains within 1 standard deviation of the variability in baseline temperatures. A value greater than 1 indicates a region in which average warming is 1 standard deviation greater than the variability in the baseline (i.e. it is now experiencing high novel temperatures). Again, using CRU TS v4.03, I also calculated a global standardised temperature anomaly independent of PREDICTS sites (for an average of the 36 monthly mean temperatures in the period 2004-2006), which I mapped globally for context (Figure 4.2). I chose the period 2004-2006 to coincide with the temporal distribution of most records in the PREDICTS database (Hudson et al. 2017).

Interactive effects of land-use and climate change on pollinator biodiversity

To model the effects of land-use and climate change on pollinator abundance, I built linear mixed-effects models predicting total abundance as a function of land-use type (primary vegetation and cropland), standardised temperature anomaly, and their interaction. I did not use a generalised linear model with Poisson errors because most recorded measurements

are non-integer values (also see Chapter 3, showing that log-linear and negative binomial models for land use intensity yielded similar results). I focussed on primary vegetation and cropland given my interest in the interactive effects of land-use and climate change between natural and anthropogenic land. I chose primary vegetation since it best represents natural vegetation, and cropland since it is both an anthropogenic land-use type and most relevant to crop production that depends on animal pollination. Since the interactive effect of land-use and climate change is likely greater between natural and anthropogenic land than among levels of intensity within a land-use type (Outhwaite et al. in review), for both cropland and primary vegetation I grouped together all levels of intensity. I $\log_e$ -transformed all total abundance values (adding one because of zero values) to normalise the model residuals. Due to the nested nature of the database (Hudson et al. 2017), I included a random intercept of study identity to account for variation in sampling methods, sampling effort, and broad geographical differences among studies, and a random intercept of spatial block within study to account for the spatial structuring of sites. Random-effects structures were selected to minimise AIC values. To test how sensitive my predictions were to the inclusion of particular taxonomic groups, I ran an additional analysis in which I predicted change in total abundance on cropland following the removal of each taxonomic family (i.e. jack-knife removal with replacement).

Potential future risk to crop pollination from land-use and climate change

Given the importance of insect pollinators to global crop production, I used my model of insect pollinator abundance to predict geographic and temporal patterns of pollination shortfall risk, based on a combination of estimated crop production dependent on animal pollination, and projections of future climate change. I focus on two forms of risk: production risk, and proportional production risk. Production risk is a measure of the total crop production that could be at risk of shortfall. Proportional production risk is a measure of the crop production at risk as a proportion of the total production for a given cell (i.e. a single pixel in a gridded data layer), crop, or country. For both measures of risk, I specifically focussed on insect pollinator

abundance given that insects make a bigger contribution than vertebrates to global crop pollination (Klein et al. 2007).

My projections assume that a projected loss in pollinator abundance combined with crop dependence on pollination is a good proxy for risk to crop production. There are three core uncertainties associated with these projections (also see Discussion). First, I do not know if there is a mechanistic link between the interactive effects of land-use and climate change on pollinator abundance. I reason that a significant interactive effect is at least likely however, given prior localised studies demonstrating a synergistic effect of climate change and anthropogenic land use in insects (Outhwaite et al. in review; Zaragoza-Trello et al. 2021). Second, I do not account for changes in the distribution of crops from the effects of climate change alone. Third, it remains unclear how local abundance change will impact crop pollination, and in turn how crop pollination will relate to yield change. Given these uncertainties my projections should not be interpreted as absolute projections of yield loss, but rather as a relative measure of crop production risk. Such a measure of global risk does not yet exist, despite its marked importance to policy-makers (IPBES, 2016).

I first used a global dataset of crop production (Monfreda et al. 2008), in combination with the animal pollination dependencies reported in Klein et al. (2007), to build a map of global crop production dependent on animal pollination for the year 2000 (the date for which Monfreda et al. 2008 estimated global crop production). For each crop in Monfreda et al. (2008), I adjusted total production for the pollination dependence ratios reported in Klein et al (2007) (essential, 0.95; great, 0.65; modest/great, 0.45; modest, 0.25; little, 0.05; no increase, 0). Given some Monfreda crops are represented by multiple crops in Klein et al (2007), for each Monfreda et al (2008) crop I first calculated the mean pollination dependence ratio among Klein crops. I then adjusted each crop for its pollination dependence and summed dependence-adjusted production for all crops at each cell as

$$PollinationProd_i = \sum_{c=1}^{c=n} Prod_{ci} d_c$$

where *PollinationProd* = animal pollination dependent crop production, *Prod* = crop production, *c* = Monfreda crop category, *d* = average animal pollination dependence ratio, and *i* = a given cell. I also calculated the total production in any given cell (independent of animal pollination dependence) for the same set of crops, as

$$TotalProd_i = \sum_{c=1}^{c=n} Prod_{ci}$$

where *TotalProd* = total crop production. In any further work, I assumed that the distribution of pollination-dependent production in 2000 will be representative of that in 2050 (see 'Discussion' for an exploration of this limitation). I also did not exclude pollination-dependent production which relies on either vertebrates or non-insect invertebrates. Although there are some crops recorded in Klein et al (2007) as reliant on either vertebrate or non-insect pollinators, after reconciling these crops with Monfreda et al (2008) there are no crop groups that are exclusively dependent on vertebrates or non-insect invertebrates.

I then used ISIMIP temperature anomalies from Frieler et al (2017) to project potential future change in standardised temperature anomaly under three RCP (Representative Concentration Pathway) scenarios (8.5, 6.0, 2.6), using an ensemble mean of the climate models GFDL, HadGEM2, IPSL, and MIROC5. RCP scenarios describe how global climate might change in the future, according to the volume of greenhouse gases emitted by humans (van Vuuren et al. 2011). Each of these emission pathways are based on a set of socioeconomic assumptions (e.g. technological improvements, policy changes) about the future (van Vuuren et al. 2011). RCP 8.5 represents a worst-case high-emissions scenario, 6.0 a pathway with some degree of mitigation, and 2.6 a pathway with significant reductions in emissions (van Vuuren et al. 2011). ISIMIP temperature anomalies were added onto a historical baseline for the period 1979-2013 (inclusive), calculated as the grand mean over all

monthly mean daily CRU temperatures for each cell. For each year, the standardised temperature anomaly was projected across the terrestrial surface of the globe (for the years 2016-2048), using a 3-year rolling average to smooth change in risk over time. For each 3-year projection window, insect pollinator abundance on cropland was predicted according to the model in Figure 4.1 (top left panel) for all cells containing crop production dependent on animal pollination. These abundance values were then expressed as the proportional loss of abundance compared to the abundance expected on croplands that have experienced no warming (i.e. standardised temperature anomaly of 0) as

$$l_{ti} = 1 - \frac{a_{wti}}{a_{0i}}$$

where l = abundance loss, t = each 3-year time window, a_w = predicted abundance on cropland under projected warming, and a_0 = predicted abundance on cropland under no warming. In each cell, animal-pollination-dependent crop production was then adjusted for the percentage reduction in abundance at that time step, before summing production at risk for all cells as

$$ProdRisk_t = \sum_{i=1}^{i=n} PollinationProd_i l_{ti}$$

where $ProdRisk$ = the total crop production risk (i.e. production risk). To check the influence of single climate models on our projections, I calculated one projection as the average of all models for that RCP scenario, and a set of additional jack-knifed projections, dropping each climate model in turn. For the average of all climate models at RCP 8.5, I also checked the extent to which extrapolating abundance loss beyond the lowest fitted abundance value (i.e. greatest abundance loss) of our PREDICTS models affected my projections. Specifically, I projected change in total pollination-dependent crop production at risk when the greatest loss in predicted abundance is capped at the largest fitted value of standardised temperature anomaly (Appendix 3, Figure S3.2). For my projection in which abundance loss is not capped, I also calculated the percentage of cells where abundance loss was extrapolated beyond the maximum fitted value (Appendix 3, Figure S3.2).

To identify geographic regions in which a high proportion of crop production could be at risk under a worst-case climate scenario (i.e. RCP 8.5), I projected pollination risk for each cell as above (using the average of all four climate models), and then expressed pollination risk as a proportion of all crop production within each cell

$$PropRisk_{ti} = \frac{PollinationProd_{i}l_{ti}}{TotalProd_i}$$

where *PropRisk* = the proportion of total crop production at risk (i.e. proportional production risk). A value of 1 therefore indicates a hypothetical region in which all crop production in that cell is dependent on pollination, and predicted insect pollinator abundance loss is 100% (i.e. one would expect a very high risk of pollination shortfall). I projected this dependence risk for 2006 and 2050, and mapped these projections globally. I also plotted the change in risk for the 10 countries with the highest overall risk (calculated as the average over the whole time series), and the change in risk for the 10 countries with the highest rate of change (the difference between the lowest and highest over the whole time series). For each of these high-risk countries, I took the overall risk to be the median of all cells in that country at each time step, onto which I overlaid the 2.5th and 97.5th percentiles as a measure of the variation in risk across the country.

Using the same approach as above, I also projected proportional production risk for each of the top 20 crops by total pollination-dependent production (i.e. *PropRisk_{cti}*). Again, I took the overall risk to be the median of all cells in which that crop appears at each time step, onto which I overlaid the 2.5th and 97.5th percentiles as a measure of the variation in risk across locations where the crop is grown. I also calculated a change in total production risk for the same top 20 crops (i.e. *ProdRisk_{ct}*), which I then adjusted to a percentage of total production for each crop (i.e. *PropRisk_{ct}*). This measure therefore estimates the total proportion of global production at risk for a given crop. All analyses were carried out in R v4.0.5 (R Core Team 2020). All core analyses can be visualised interactively in a Shiny app currently hosted online (joemillard.shinyapps.io/pollinator_dependence_visualisation/).

Results

Interactive effects of agricultural land use and climate change on pollinator biodiversity

For both pollinating and non-pollinating insects, standardised temperature anomaly (STA) and land-use type had a strong interactive effect on local abundance (Figure 4.1; $F = 23.3195$, $p < 0.001$; $F = 10.5764$, $p < 0.01$). For vertebrates, there was no interactive effect for pollinators but a significant effect for non-pollinators (Figure 4.1; $F = 2.1599$, $p = 0.14$; $F = 10.6743$, $p < 0.01$). In both insect groupings, the effect of STA diverges significantly between primary vegetation and cropland (Figure 4.1), such that abundance increases with greater STA on primary vegetation, and decreases with greater STA on cropland. The interactive effect is stronger for pollinating insects than for non-pollinating insects (Figure 4.1). Relative to croplands that experienced no warming, at the most extreme level of STA (1.58) at sampled cropland sites, insect pollinators experienced declines of up to ~75%. An STA value this extreme indicates a region that has warmed such that the average temperatures occurring now would only be expected to occur in ~1/18 years under baseline conditions.

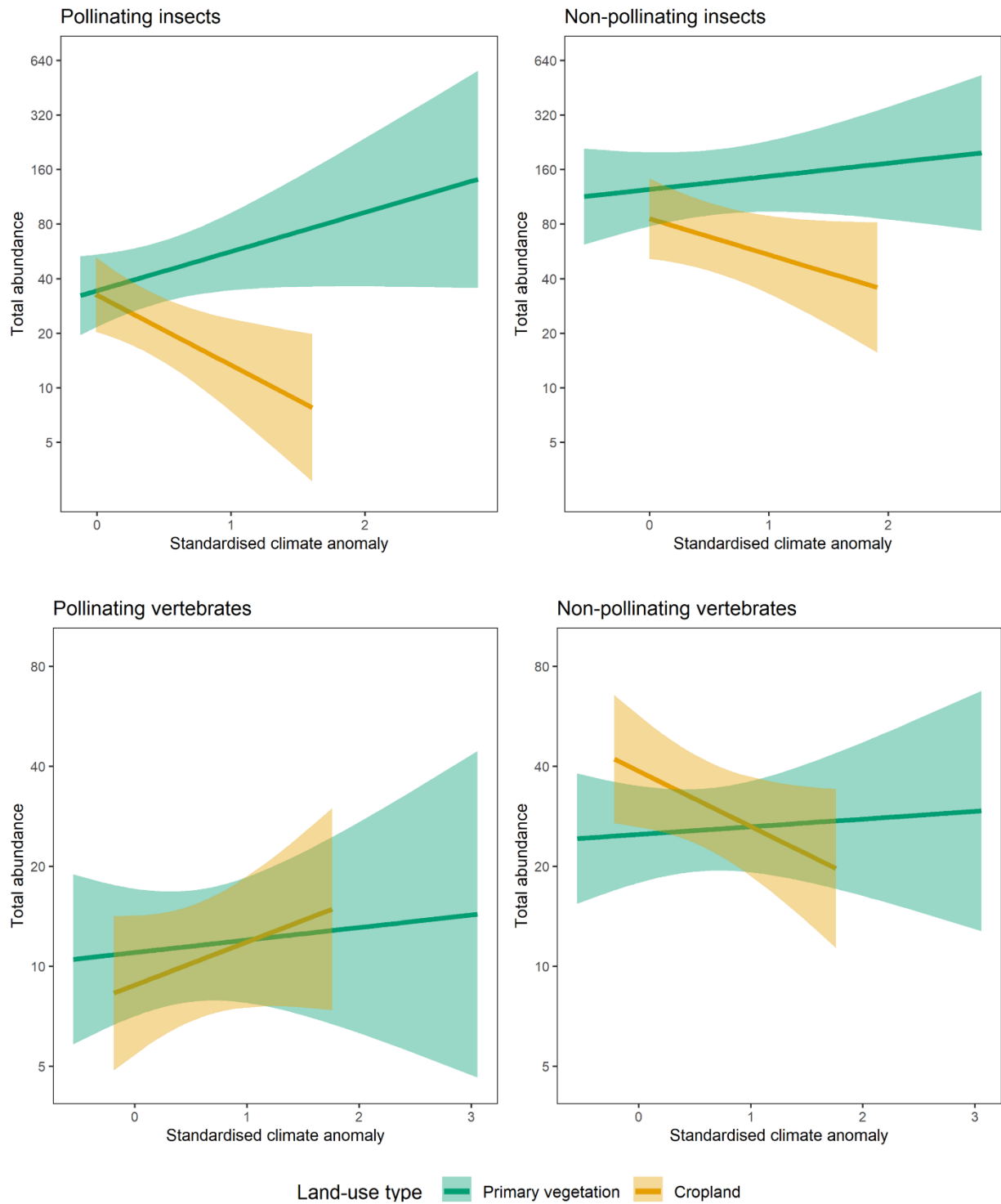


Figure 4.1 Response of pollinating and non-pollinating total abundance to standardised temperature anomaly on primary vegetation and cropland (note that abundance is plotted on a log_e scale). Each panel represents a linear mixed-effects model for pollinating or non-pollinating insects and vertebrates. Coloured lines represent mean

fitted estimates for each interaction, and shading 95% confidence intervals around that prediction: green, primary vegetation; orange, cropland.

Potential future effects of land-use and climate on crop pollination

Total production risk is predicted to increase under all climate scenarios (Figure 4.3). Rate of change in production risk is highest under RCP 8.5, rising from ~80 million to ~138 million metric tonnes by 2050. Rate of change is marginally lower under RCP 2.6 than RCP 6.0, although both are predicted to result in a production risk of ~110 million metric tonnes by 2050. Change for RCP 8.5 experiences a consistent year on year increase, with a period of increased exposure rate in the 2030s as a significant quantity of watermelon production is exposed to novel extreme temperatures (see Shiny app, https://joemillard.shinyapps.io/pollinator_dependence_visualisation/). My projections were not overly influenced by the prediction of single climate model (Figure 4.3), and they do not change markedly when abundance loss is capped at the maximum model-fitted value (Figure S3.2).

Cell-level proportional production risk (i.e. $PropRisk_{ti}$) is highest, and predicted to increase most rapidly, in regions of sub-Saharan Africa and northern South America (Figure 4.4). All 10 countries with the highest median proportional production risk, and all 10 with the highest increase in this risk, are in the tropics (Figure 4.4). Rwanda has the highest overall risk with a median index value of ~0.45 in 2020, rising to ~0.5 by 2050. Eritrea and Madagascar have the highest rates of change in this risk, increasing from ~0.27 to ~0.5 in Eritrea and ~0.02 to ~0.2 in Madagascar. In order of rate of change, median cell-level proportional production risk (i.e. $PropRisk_{cti}$) is highest for cocoa, mango, pumpkin, melon, watermelon, and coffee (Figure 4.5). Change in total proportional production risk (i.e. $PropRisk_{ct}$) is such that—according to the one in which I define risk—~60% of cocoa production and ~30% of mango production could be at risk by 2050 (Appendix 3, Figure S3.3).

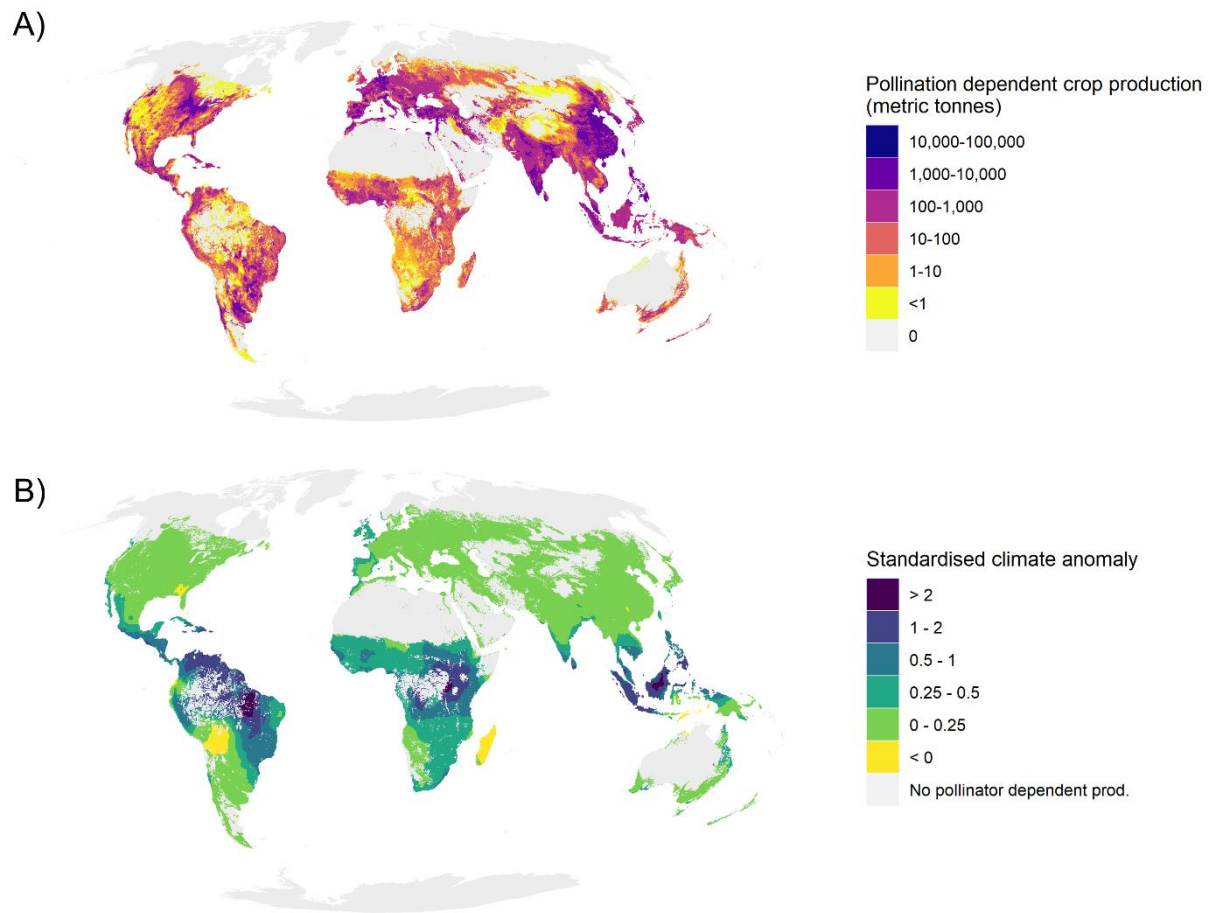


Figure 4.2 Crop production that depends on animal pollination (A), and standardised temperature anomaly (in regions of animal pollination-dependent production; B) for the years 2000 and an average of 2004-2006, respectively. For the temperature anomaly, I show an average for the years 2004-2006 to coincide with the temporal distribution of most records in the PREDICTS database, and therefore the results in Figure 4.1. A) Animal pollination-dependent production is calculated as total crop production, adjusted for the degree of dependence on animal pollination according to Klein et al (2007). B) Standardised temperature anomaly is the change in the grand mean of monthly mean daily temperatures between a baseline period (1901-1930) and 2004-2006, divided by the standard deviation across monthly mean daily temperatures, for the same 30-year baseline period.

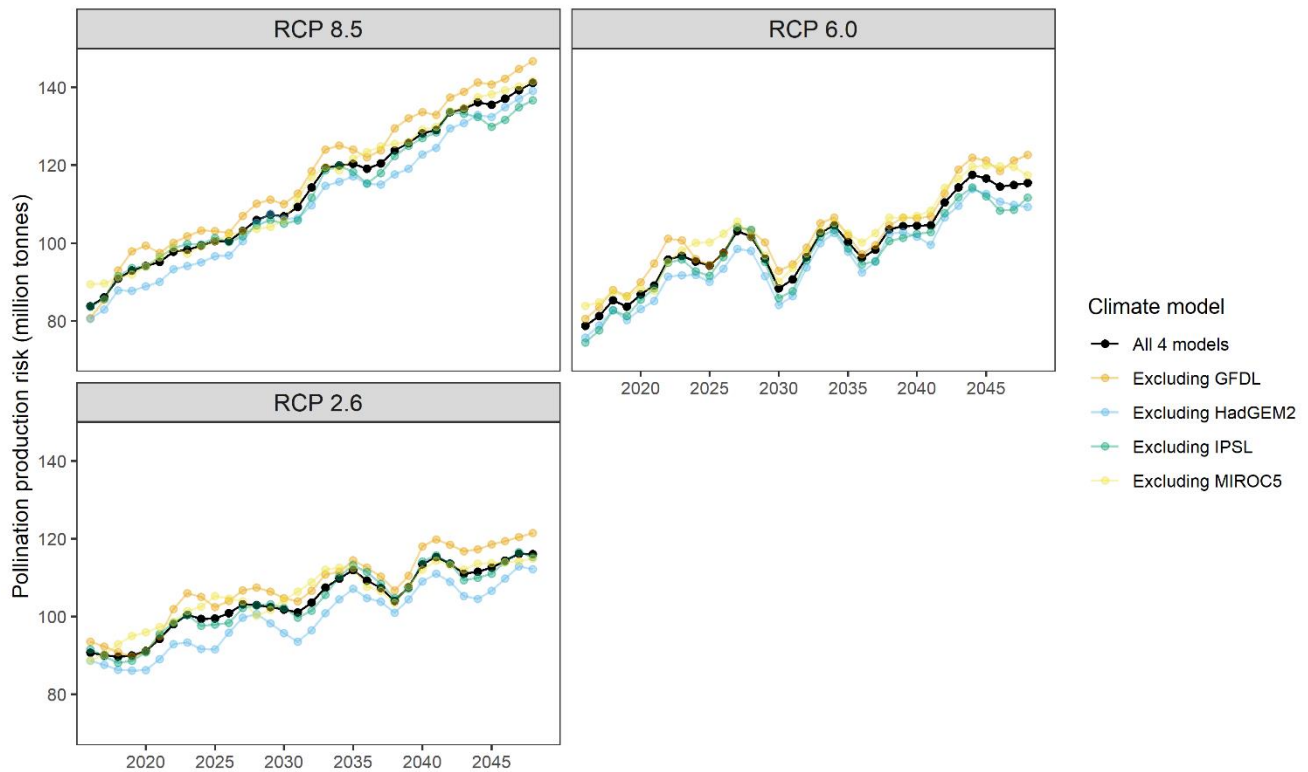


Figure 4.3 Projected change in total production risk ($ProdRisk_t$) under three RCP scenarios (8.5, 6.0, and 2.6), either as the average across all climate models, or dropping each climate model in turn. For each year into the future, the standardised temperature anomaly was projected globally for all cells of pollination-dependent production, using a 3-year rolling average. I used data on crop production from the year 2000 (the latest year when such data are available for all crops), therefore assuming that the distribution of the production of these crops does not change. For each annual projection of standardised temperature anomaly, insect pollinator abundance on cropland was predicted according to the model in Figure 4.1 (top left panel), and then expressed as proportional abundance loss compared to cropland that has experienced no warming (i.e. standardised temperature anomaly of 0). In each cell, animal pollination-dependent production was then adjusted for the percentage reduction in abundance at that cell, before summing animal-pollination-dependent production for all cells at each time step. Colours refer to the climate model excluded in that jack-knife projection with the projection for all models shown in black.

2006

2050

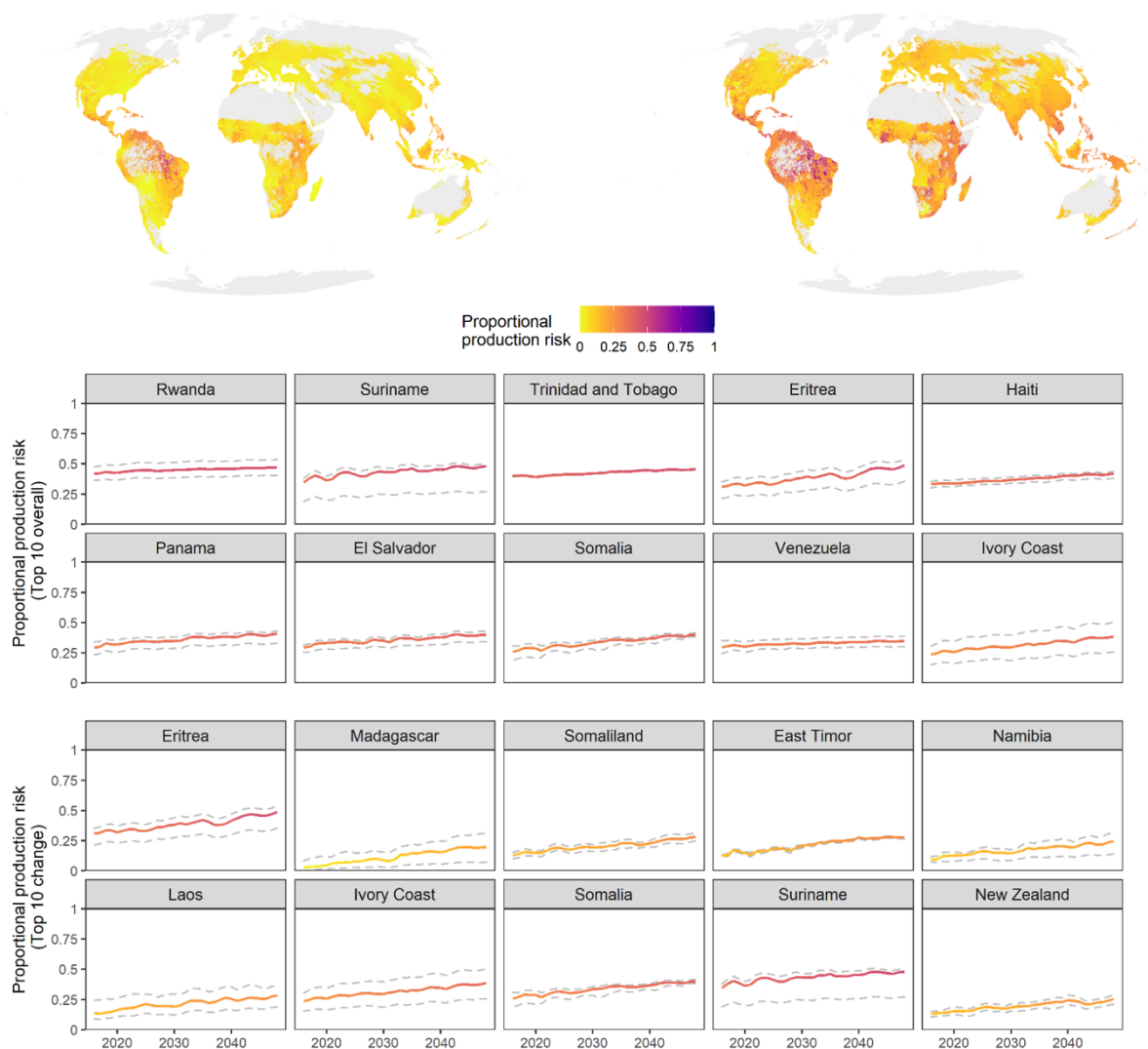


Figure 4.4 Cell-level proportional production risk (i.e. $PropRisk_{it}$) projected under RCP 8.5, from the ensemble-average of four climate models (GFDL, HadGEM2, IPSL, and MIROC5), for the whole world as well as 10 countries with the highest overall risk and change in risk. Global standardised temperature anomaly was projected to 2050 for all areas of animal-pollination-dependent crop production, using the 3-year average approach as described in Figure 4.3. I used crop production data from the year 2000 (the latest year for which such data are available for all crops), therefore assuming that the distribution of the production of these crops does not change. For each value of standardised temperature anomaly, insect pollinator abundance was predicted

according to the model in Figure 4.1, expressed as proportional loss compared to cropland that has experienced no warming (i.e. standardised temperature anomaly of 0). Animal-pollination-dependent production at each cell was then adjusted for the predicted loss of insect pollinator abundance, and then converted to a proportion of the total production at that cell. Top: global maps of proportional production risk for 2006 and 2050. A value of 1 indicates a hypothetical region in which all crop production in that cell is dependent on animal pollination, and predicted insect pollinator abundance loss is 100%. Bottom: each panel represents predicted change in proportional production risk under RCP 8.5 for a set of the most at risk countries. Coloured lines indicate the median across cells within that country, and grey dashed lines the 2.5th and 97.5th percentiles, providing an indication of variation in risk across locations within a country. The upper set of panels represents the 10 countries with the highest overall risk (average over the whole series), whereas the lower set of panels represents the 10 with the highest rate of change (difference between minimum and maximum).

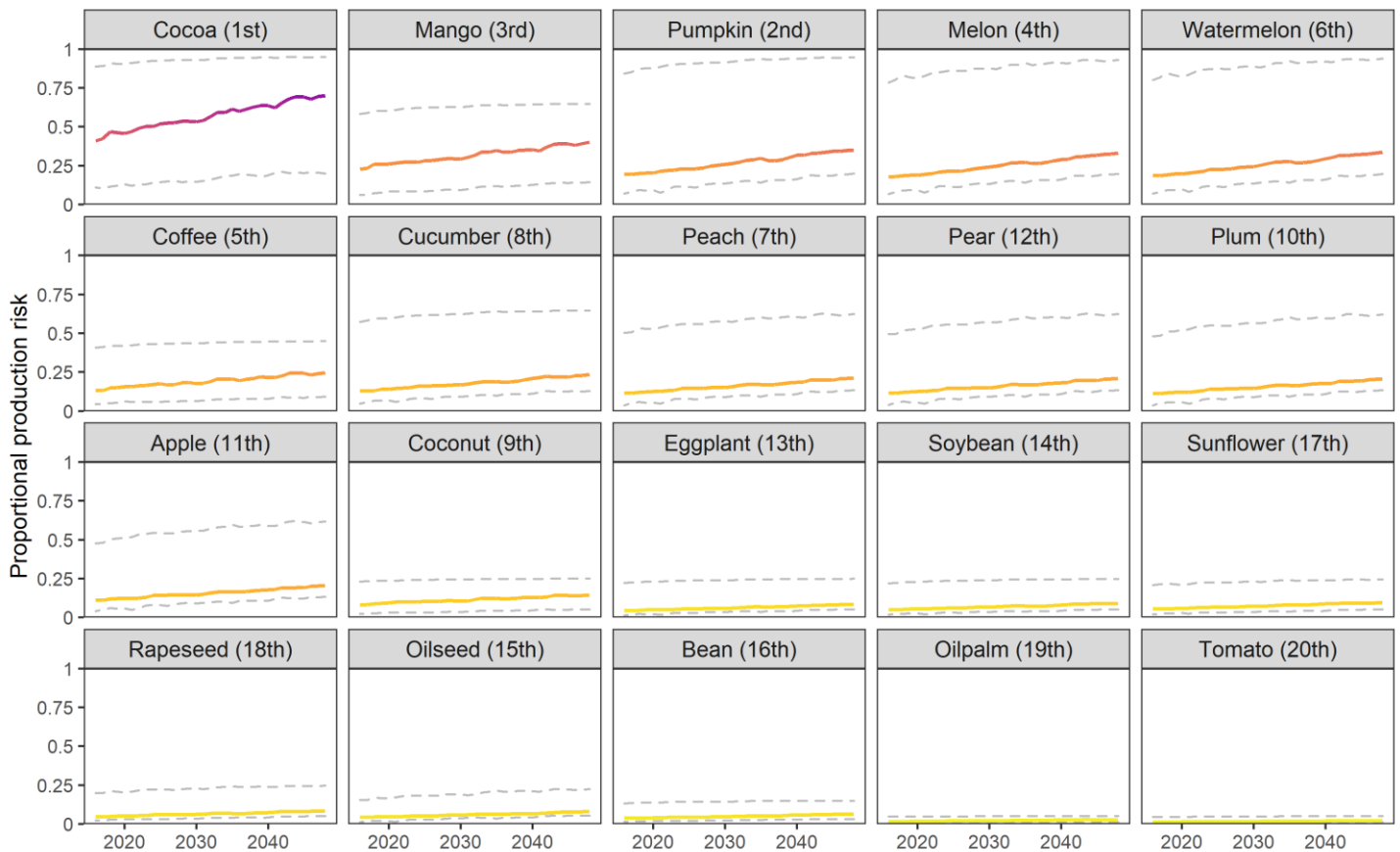


Figure 4.5 Cell-level proportional production risk (i.e. $PropRisk_{cti}$) projected under RCP 8.5, from the ensemble-average of four climate models (GFDL, HadGEM2, IPSL, and MIROC5), for the top 20 crops by total animal-pollination-dependent crop production (ordered by rate of change in risk to highlight crops of potential future concern). Global standardised temperature anomaly was projected for all areas of animal-pollination-dependent cropland to 2050, using the 3-year average approach as described in Figure 4.3. I used crop production data from the year 2000 (the latest year for which such data are available for all crops), therefore assuming that the distribution of the production of these crops does not change. For each value of standardised temperature anomaly, insect pollinator abundance was predicted according to the model in Figure 4.1, expressed as proportional loss compared to cropland that has experienced no warming (i.e. standardised temperature anomaly of 0). Animal-pollination-dependent production at each cell for each crop was then adjusted for the predicted loss of insect pollinator abundance, and then converted to a proportion of the total production for that crop at

that cell. Colours correspond to the median proportional risk for all the cells in which that crop appears, and grey dashed lines the 2.5th and 97.5th percentiles, providing an indication of variation in risk across the locations in which a crop is grown. Total proportional production risk (i.e. *PropRisk_{ct}*) rank for the final year in the series is indicated in brackets.

Discussion

Here I showed that land-use type and a standardised anomaly of temperature change have a significant interactive effect on local insect pollinator abundance, and a significant although less pronounced effect in non-pollinating insects. Change in abundance for insect pollinators on cropland was such that, relative to croplands that have not warmed (i.e. STA = 0), at an STA of ~ 1.5 insect pollinator abundance was $\sim 75\%$ lower. In contrast on primary vegetation across the same gradient of STA, insect pollinator abundance increased by $\sim 100\%$. For vertebrate pollinators and non-pollinators, these interactive effects of STA and land-use type appear to be weak or non-existent. On the basis of my predictions for insect pollinator abundance, and estimates of the global distribution of animal-pollination-dependent agriculture, I showed that the rate of change in crop production risk under RCP 8.5 could be 25% greater than under RCP 2.6 or 6.0. Under RCP 8.5, according to the way in which I define risk, total crop production risk is predicted to increase by $\sim 70\%$ between 2016 and 2048. I further showed that proportional production risk is greatest and predicted to increase most rapidly in regions of northern South America and sub-Saharan Africa. Although I cannot predict when actual pollination shortfalls will occur—due to the uncertainties associated with the link between local pollinator abundance change and crop yield—my results suggest that they will be most likely in the tropics, where pollinator biodiversity and animal-pollination-dependent crop production will be increasingly exposed to extreme novel temperatures.

Previous work has shown that insects are vulnerable to the interactive effects of land-use and climate change (Halsch et al. 2021; Oliver et al. 2016; Outhwaite et al. in review). Here I show that this effect is likely greater in insect pollinators than other insects. However, whether sensitivity to climate change and land-use interactions for insect pollinators relates to their actual reliance on floral resources, or to other correlated traits such as dispersal ability, body size, voltinism, or specialism, is unclear. Most likely some combination of both is true. For example, insect-pollinated plants have been shown to respond more strongly to warming than

wind-pollinated plants, suggesting that plant-pollinator interactions are highly sensitive to thermal changes (Fitter & Fitter, 2002). Pollinator pilosity (i.e. hairiness), on the other hand, has been shown to affect an insect's ability to adapt to changes in climate (Roquer-Beni et al. 2020), and given its nature as a trait typical of bees and hoverflies, tends to be correlated with a reliance on floral resources (Stavert et al. 2016). Reproductive behaviour change under warm temperatures may also be of relevance (Zaragoza-Trello et al. 2021). Recent work on *Osmia bicornis* (Linnaeus, 1758), a species of solitary bee, indicates that floral abundance and mean temperature interact such that sex ratios are skewed towards males when temperature is high and floral abundance is low. For this species at least, one might expect such a phenomenon to reduce local abundance (Zaragoza-Trello et al. 2021), since population growth rate is limited by the number of females (Bessa-Gomes et al. 2004).

Given that tropical agriculture has strong effects on pollinator biodiversity (Millard et al. 2021a), and is expected to increase expand rapidly in the coming decades, a significant interactive effect such that insect pollinator abundance is lower on rapidly warming croplands, is concerning. Tropical agriculture's ability to cope with pollinator losses will likely depend on the biology of specific crop-pollinator interactions. For example, tropical crops that can be pollinated by generalist domesticated pollinators will likely cope better with wild pollinator abundance decline, given the relative ease with which honey-bee hives can be transported to supplement local losses (Goodrich, 2019), although this does still come at a cost (Degrandi-Hoffman et al. 2019; Lee et al. 2019). For crops such as cocoa however, which is reliant on Ceratopogonid midges (Claus et al. 2018), the cost of pollinator abundance decline will be greater (although see limitations paragraph for the caveat of missing data for Ceratopogonid midges). Mass rearing of *Forcipomyia* midges has been proposed to buffer population size, but given their requirement for a blood-meal for the production of eggs, their low dispersal ability, and their regional specificity (Claus et al. 2018), commercialising such an approach to supplement local losses would be challenging. Recent work has also shown that hand-pollination can act to significantly increase yields and profit in cocoa, but it's unclear whether

such an approach is sustainable long-term (Toledo-Hernández et al. 2020), or whether it would be profitable in other localities as local pollinator abundance continues to change. It is therefore of particular concern that cocoa falls within a region that will experience particularly high novel temperatures, and that in the countries in which it is grown—primarily Ivory Coast and Ghana—it is so important for the livelihoods of many small-scale farmers (Claus et al. 2018; ICCO, 2014).

The difference in total pollination risk under RCP 8.5 compared with RCPs 2.6 and 6.0 further reinforces the importance of avoiding a high-warming future for Earth (Lobell et al. 2008; Thomas et al. 2004; Thuiller et al. 2005). My models project that, under RCP 8.5, crop pollination risk could change abruptly as regions of high pollination-dependent production suddenly become exposed to novel extreme temperatures. In the 2030s, for example, I predict that total production risk could increase rapidly. This may translate into a change in the availability and price of crops. However, many factors determine crop production and crop price, and so even if impacts on crop production occur, conclusive attribution to pollinator abundance changes is likely to be challenging (Khanal et al. 2018; Santeramo et al. 2018). These complications likely explain why identifying a strong effect of pollinator losses on global crop yield and price has thus far been so difficult (Aizen et al. 2008, 2009; Garibaldi et al. 2011).

There are a set of limitations associated with the approach we use in my study. First, I assume here that dynamic processes of local pollinator change can be inferred from a space-for-time model of the impacts of climate change and agricultural land use. Ideally, I would use insect trends over time, but since long-term insect studies are sparse at the global scale (see Klink et al. (2020) for one example), space-for-time inferences represent a necessary alternative (Purvis et al. 2018). Second, my set of non-pollinating species were not confirmed as such, but rather a set of species not identified as pollinators. This is not a major problem for this study, since non-pollinators were used simply as a reference group. Third, I assume that both the spatial scale and temporal scale of the temperature anomaly I use is relevant to insect

pollinators. The temperature data I use is coarse by necessity. Finer-scale temperature data, if they were available globally, may derive a different result. Moreover, I also recognise that the interactive effect of climate change and agricultural land use may differ according to the active season of insects, which is known to vary between tropical and non-tropical insects (Johansson et al. 2020). Work is currently underway investigating this effect of active seasons, indicating that the interactive effect of climate change and land use is similar when considering either only the months in which insects are most active, or all months of the year (Outhwaite et al. in review). Fourth, the crop data I used is for the year 2000. In my projections I assume that the distribution of these crops has not changed since then, and will not continue to change between now and 2050. This assumption is clearly wrong. However, although the distribution of animal-pollination-dependent crops over the next 30 years will change, it seems unlikely that the temperate and tropical zones will experience a complete reversal in pollination dependence, meaning my rates of change in total production risk and proportional production risk will likely still be informative. Moreover, long-term agricultural investment in particular geographic regions may mean time-lags between crop yield loss and the subsequent movement of agriculture from that region, and even if agriculture does move, for those people that remain local livelihoods will be impacted. Fifth, in many regions the crop data I use is highly uncertain, due to the way in which it was interpolated for regions with insufficient reporting. Such interpolation means production of some pollination-dependent crops is predicted for regions of the world in which it is highly unlikely, such as small quantities of cocoa production across large areas of Europe. In Appendix 3 (Figure S3.4) I carried out a sensitivity analysis in which I reran my production risk projections at a series of data quality levels (according to Monfreda et al (2008)). This additional analysis indicates that overall change is not sensitive to data quality. Sixth, the pollination-dependence ratios (Klein et al. 2007) I use are averaged across multiple crops to align with Monfreda et al (2008), which I do not account for in my risk projections. However, since only 6/67 of the animal pollination dependent crops in Monfreda et al (2008) have a standard deviation greater than 0 for their set of Klein et al (2007) pollination dependence ratios, accounting for this variation would have only a minimal

impact. Seventh, there is an ongoing debate as to what level of wild pollinator abundance is required to sustain yields. Here I reason that crop production risk will be predicted as a function of pollinator abundance; an assumption based on evidence from a number of studies (Blaauw & Isaacs 2014; Dainese et al. 2019; Hayter & Cresswell 2006). However, the assumption I make is that risk will be proportional to an average loss in insect pollinator abundance, which does not account for potential variation among localities/taxonomic groups, or threshold relationships between pollinator abundance and crop pollination. This is particularly relevant in the context of cocoa pollinating midges, which are absent from my set of PREDICTS pollinator data (i.e. the predictions I make are not directly representative of cocoa pollination risk). I would argue, however, that my assumption is likely still sufficient to quantify overall relative risk, since my taxonomic family jack-knife analysis shows that the negative response of pollinators to the interaction of climate change and land-use is likely consistent across families (Figure S3.5). Regardless, future work should incorporate any new empirical evidence on the relationship between local pollinator abundance, crop pollination, and yields. Future work should also associate risk for a given crop with change in only the species known to pollinate that crop. As yet the data required to make both of these developments either does not exist or is incomplete.

Climate change and anthropogenic land use have had significant effects on the biosphere. These effects will likely accelerate over the coming decades as human activity continues to increase. The consequences of such change on pollinators and the service they provide could be great. In this study I showed that climate change and agricultural land use have a strong interactive effect on insect pollinator abundance. Insect pollinators experience strongly negative declines where cropland coincides with high levels of historical climate change. I show that change in crop pollination risk could be ~25% times greater under RCP 8.5 than 2.6 or 6.0, and that the tropics will likely experience the greatest risk of future pollination shortfalls. Crop pollination risk is highest, and predicted to increase most rapidly, in regions of sub-Saharan Africa and northern South America, primarily for cocoa, mango, pumpkin, melon,

watermelon, and coffee production. These results further stress the importance of avoiding a high-warming future on Earth, and potential future risk for crop pollination in the global south (Dicks et al. 2021). The health, well-being, and livelihoods of billions of people to some extent depends upon the availability and affordability of animal pollination dependent crops (Potts et al. 2016). Climate change and agricultural land use could risk this contribution. Mitigating this risk will be a key challenge of the 21st Century.

Chapter 5: The species awareness index as a conservation culturomics metric for public biodiversity awareness

Abstract

Threats to global biodiversity are well-known, but slowing current rates of biodiversity loss remains an ongoing challenge. The Aichi Targets set out 20 goals on which the international community should act to alleviate biodiversity decline, one of which (Target 1) aimed to raise public awareness of the importance of biodiversity. Whilst conventional indicators for Target 1 are of low spatial and temporal coverage, conservation culturomics has demonstrated how biodiversity awareness can be quantified at the global scale. Following the Living Planet Index methodology, here I introduce the Species Awareness Index (SAI), a metric of change in species awareness derived from Wikipedia views. I calculated this index at the page level for 41,197 IUCN species across 10 Wikipedia languages, incorporating over 2 billion views. I then explored overall changes in species awareness, and tested the extent to which change in views is predicted by the language, taxonomic class, pollination contribution, and trade contribution of a species page. Bootstrapped indices for the page level SAI show that overall awareness of biodiversity is marginally increasing, although there are differences among taxonomic classes and languages. Among taxonomic classes, overall awareness is increasing fastest for reptiles and slowest for amphibians. Among languages, overall species awareness is increasing fastest for Japanese and slowest for Chinese and German users. Although awareness of species as a whole is increasing, and is significantly higher for traded species, over the period 2016-2020 change in awareness appears not to be strongly related to the trade of species or pollination contribution. As a data source for public biodiversity awareness, the SAI could be integrated into the Biodiversity Engagement Indicator.

Introduction

Threats to global biodiversity are well-known, but slowing current rates of biodiversity loss remains an ongoing challenge (IPBES 2019; Mace et al. 2018). One problem is the requirement for transformational behavioural and economic change (IPBES 2019), and the difficulty in leveraging this change at a global level (IPBES 2019). The Strategic Plan for Biodiversity 2011–2020, underpinned by the Aichi Targets, represents an effort to guide these changes (UNEP CBD 2010). Specifically, the Aichi Targets set out 20 goals on which the international community should act to alleviate biodiversity decline (UNEP CBD 2010). Three of the Aichi Targets have sufficient and suitable indicators (6, 9, and 11), four have intermediately sufficient indicators (4, 7, 12, and 14), ten have insufficient indicators (1, 5, 8, 10, 13, 16, 17–20), and 3 have none (Mcowen et al. 2016). Concerned with public awareness of biodiversity, Aichi Target 1 states that by 2020 the public should be aware of the value of biodiversity. Conventional indicators for Target 1 (i.e. the Biodiversity Barometer, UEBT 2019)) are of low spatial and temporal coverage (Leadley 2013, Mcowen et al. 2016), and do not incorporate awareness of biodiversity itself (i.e. species). Without robust metrics capturing evidence towards Target 1, understanding whether this target has been met is hard.

Conservation culturomics has emerged as a field concerned with digitised data and human nature interactions (Sherren et al. 2017; Ladle et al. 2016). Quantifying public awareness of biodiversity is an area of active interest. Using data sources such as Twitter, Facebook, Flickr, Wikipedia, and Google Trends, a number of researchers have shown how online data can be used to better understand how the public perceives biodiversity and environmentalism (Roberge 2014; Papworth et al. 2015; Tenkanen et al. 2017; Roll et al 2016; McCallum & Bury 2013). More recently, research has explored how online data sources can be combined to build a single indicator of biodiversity awareness. For example, Cooper et al. (2019) examined frequencies of biodiversity keywords across social media,

online newspapers, and internet searches, reasoning that relative frequencies reflect public awareness of conservation issues. A significant step forward in applying culturomic approaches to the development of indicators, Cooper et al. (2019) provided a global framework for future research. Given Cooper et al. (2019) focuses on conservation issues, one potential improvement could be to incorporate changing awareness of biodiversity itself.

Wikipedia page views represent a powerful data source for quantifying change in public awareness of biodiversity itself. Page views have previously been applied to quantify reptile public interest (Roll et al. 2016) and species phenology (Mittermeier et al. 2019). In the context of awareness, Wikipedia is valuable in that pages are linked explicitly to biodiversity across scales. Pages on Wikipedia exist for taxa at multiple taxonomic levels, Red List statuses, and ecological systems, with an unambiguous link between the taxon and page identity (Mittermeier et al. 2019). Previous research has shown Wikipedia can reveal changes in public awareness in response to natural history documentaries, demonstrating that the data source could be informative of long-term changes in awareness (Fernández-Bellon & Kane 2019). Moreover, since species characteristics provide a mechanistic link to ecosystem services, change in awareness for a particular species on Wikipedia could be used as a proxy for awareness of its contribution. For example, increasing awareness for species which contribute significantly to pollination or trade could indicate greater public awareness of biodiversity importance. For pollination specifically, such changes in awareness are particularly important, given the global economic importance and reported declines of animal pollinators (Lautenbach et al. 2012, Powney et al. 2019, Hallmann et al. 2017, IPBES 2016). Although using Wikipedia for quantifying awareness is not without its limitations and caveats (see Discussion), it provides the basis for a useful new indicator.

An awareness metric using Wikipedia page views could be thought of as analogous to The Living Planet Index (LPI). The LPI represents an aggregation of vertebrate population trends (Loh et al. 2005, Collen 2009, McRae 2017), showing an average rate of change for multiple

species populations. Treating species page-views as a population size, the LPI methodology could be similarly applied to Wikipedia to derive a rate of change for species awareness. Multiple studies have used page views or search trends to infer change in awareness of specific species (Fink et al. 2020, Fukano et al. 2020, Lenda et al. 2020, Sorian-Redondo et al. 2017, Mittermeier et al. 2019, Veríssimo et al. 2020), but as far as I know no studies have calculated such an aggregated index for overall awareness. Here, I introduce and evaluate an approach based on the frequency of views for IUCN species on Wikipedia, naming it the Species Awareness Index (SAI). I then explore variation in this metric, aiming to assess whether awareness of biodiversity has changed. Specifically, I explore the overall SAI for 41,197 IUCN species pooled, 6 distinct taxonomic classes—including the core pollinating groups (insects, birds, and mammals) and the most heavily traded vertebrates (reptiles, mammals, birds, amphibians, and ray-finned fishes)—and for each taxonomic class in each of the top 10 languages (by active user) on Wikipedia (Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian, and Spanish). I then model rate of change in the page level SAI as a function of taxonomic class, Wikipedia language, trade contribution, and pollination contribution, using a pollinator dataset derived from the academic literature through named-entity recognition. I conclude by discussing the limitations of the SAI, suggesting potential avenues for future research, and demonstrating how the SAI might be combined with other approaches for a more holistic understanding of changing biodiversity awareness.

Methods

Wikipedia data

I used the Wikipedia pageview API¹, and software written in Python, to download daily user views of IUCN species for the period 1st July 2015 – 30^{1st} March 2020 (downloaded on the 16th-21st April 2020). I downloaded views for all IUCN species with Wikipedia pages in the taxonomic groups reptiles, ray-finned fishes (Actinopterygii), mammals, birds, insects, and amphibians, from 10 Wikipedia language projects (Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian, Spanish). I retrieved my list of IUCN species on Wikipedia from OneZoom (Rosindell & Wong 2020), who used the Wikipedia API² to map between a species' Latin binomial, IUCN ID, Wikidata Q identifier³, and the main Wikipedia page name for each species in each language. Downloading views from only the main page name of each species excludes redirect views, controlling for potential variation caused by the URL used to reach a page. Each IUCN ID is unique to a species on the IUCN database, whereas each Wikidata Q identifier is unique to one page, for all the Wikipedia languages in which that page appears. Each individual Wikipedia page for a particular species in a given language I henceforth refer to as a “species page”, distinguishing from my use of the term “species” to refer to a particular species among languages.

For each species page I retrieved only user views (i.e. views for which the visitor to that page was recorded as human, excluding automated views from bots). As in Mittermeier et al. (2019), I was not able to retrieve views from before 1st July 2015, since views from before this date are not archived by Wikipedia at the pageview API. For each species page returned, I calculated the daily average views for each month, and then kept only those species pages for which the series was represented for all months (see Appendix 4, Figure S4.1 for the number of complete series). I used daily average views rather than total views

¹ <https://wikitech.wikimedia.org/wiki/Analytics/AQS/Pageviews>

² https://www.mediawiki.org/wiki/API:Main_page

³ <https://www.wikidata.org/wiki/Q43649390>

since the Wikipedia pageview API does not always return views for all days in a given month.

To account for the overall change in Wikipedia's popularity and use, I also downloaded the daily user views for a random set of 11000 pages in each language, using the Wikipedia Random API⁴ to request random pages. I then aggregated these views in the same manner as the daily average species views, and again kept only pages represented across the whole time series. From this random set of views, I then removed any page also appearing in the set of species pages for that language. I initially sampled 11,000 pages to maximise the number of remaining pages after removing incomplete series and species pages.

Pollinator and wildlife trade datasets

To explore how species awareness varied with pollination contribution, I built a list of animal pollinators with an approach combining text-analysis and manual inspection of the pollination literature (see Appendix 4 and Millard et al. (2020a) for a detailed methodology). I also used the list of traded vertebrate species released in Scheffers et al. (2019) and FAO fisheries statistics (FAO, 2020) to compile a dataset of traded mammals, birds, squamate reptiles, and harvested ray-finned fish. I then retrieved the Wikidata Q ID for each of these traded species using the Wikipedia API, which I merged onto each species page. I henceforth refer to any species that pollinates as providing a “pollination contribution”, and any species in either Scheffers et al. (2019) or the FAO statistics as “traded”.

Calculating absolute awareness of biodiversity

Before calculating the SAI I briefly explored absolute awareness of biodiversity among taxonomic classes, pollination contribution, and trade contribution. I defined “absolute awareness” as the total views for a species page on Wikipedia in the period 1st July 2015 – 30^{1st} March 2020. I joined the total views for each species page onto the taxonomic class,

⁴ <https://www.mediawiki.org/wiki/API:Random>

trade contribution, and pollination contribution of that species, and then built two generalised linear mixed-effects model: 1) modelling $\log_{10}$ total article views as a function of taxonomic class, trade contribution (Y/N), the interaction of class and trade, and a random effect for language; and 2) modelling $\log_{10}$ total article views as a function of taxonomic class, pollination contribution (Y/N), the interaction of class and pollination, and a random effect for language. Rather than attempting to find the most parsimonious model, I present full model predicted values, with AIC values for these and a set of candidate null models included in Appendix 4 (Tables S4.23 and S4.24). In Appendix 4 (Figures S4.11 and S4.12) I also present boxplots for the distribution of total views among taxa for each language.

Deriving the SAI

The SAI is a new measurement of change in species awareness, calculated at the species page level from the rate of change in daily average Wikipedia views per month. The SAI could act as a counterpart to the Biodiversity Engagement indicator, a metric of biodiversity awareness derived from change in interest on Twitter, online newspapers, and Google Trends (see ‘Discussion’ for more details). Since the SAI measures the rate of change in views within a species page, species are weighted equally irrespective of their popularity, meaning highly viewed species do not dominate the SAI. In the remainder of this paper I use the term ‘SAI’ or ‘Species Awareness Index’ to refer to the overall change in awareness for a given species page, species, or group of species on Wikipedia. Specifically, I use the term “species page SAI” to refer to rate of change at the page level, the term “species SAI” to refer to the average of all species page SAIs for a unique species among languages, and “overall SAI” to refer to a bootstrapped group of species SAIs (see Figure 5.1). I also use the term “average monthly rate of change in the species page SAI” to refer to the average rate of change for a single species page across a given time period. All of the above are distinct from absolute interest in a given species or group of species (i.e. the total Wikipedia views over the whole time series).

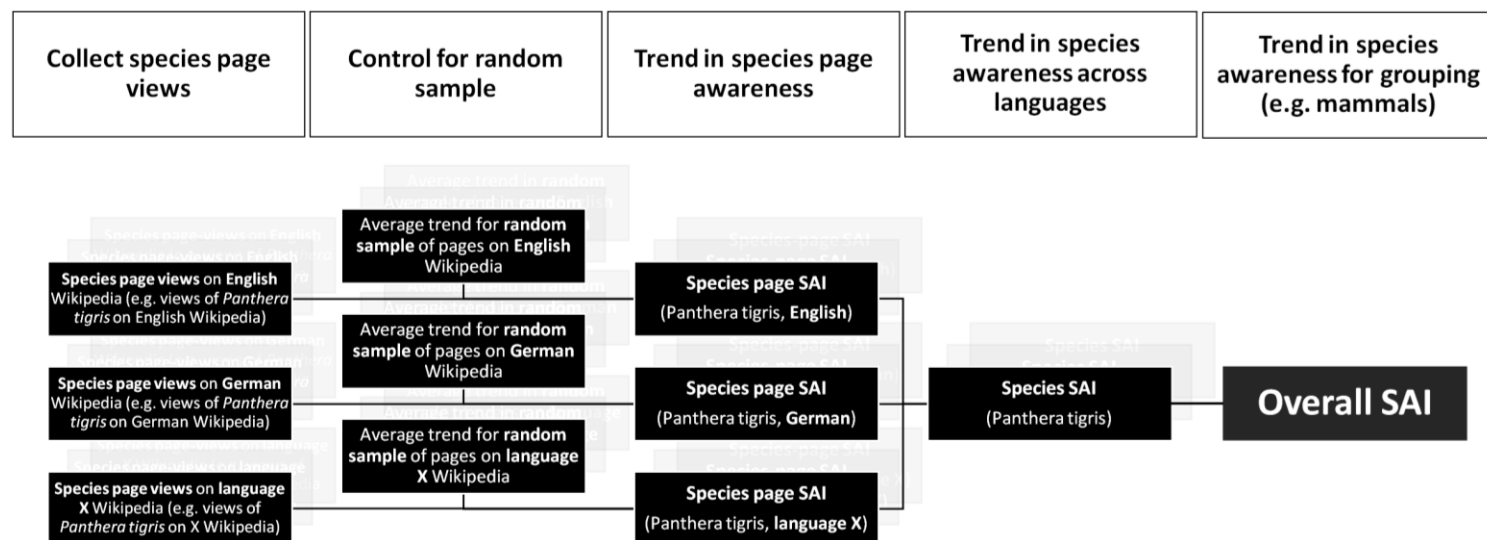


Figure 5.1 A schematic describing how the species page, species, and overall SAI were derived using Wikipedia views. The *species page SAI* represents the random adjusted trend for a given species in a given language, the *species SAI* is the average of species page SAIs for a single species across languages, and the *overall SAI* is a group of bootstrapped species SAIs.

To construct the species page SAI, I used the R package “rpi” to calculate an index of change over time for each species in 6 taxonomic groups (amphibians, birds, insects, mammals, ray-finned fish, and reptiles) on 10 Wikipedia languages (Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian, Spanish). The “rpi” package applies a generalised additive model (GAM) to smooth the daily average species page view trends, using $k = N/2$ for the degrees of freedom parameter, following Collen (2009). In ‘rpi’, these smoothed values are then used to calculate a rate of change in views for a species page article, as

$$\lambda_{st} = \log_{10}\left(\frac{N_t}{N_{(t-1)}}\right)$$

where λ_s = the rate of change in a species page, N = the smoothed number of daily average page views per month, and t = month.

To account for the overall change in popularity of Wikipedia itself over the same time period, I adjusted the rate of change for each species page using the rate of change in a random set of complete series Wikipedia pages (see Appendix 4 for the number of complete series, Figures S4.1 and S4.4). For each species page, this adjustment was made with a random set of pages in the Wikipedia language of that species page. For example, the Wikipedia page for *Panthera tigris* (Linnaeus, 1758) in the English language would be adjusted for a set of random pages in the English Wikipedia, whereas the page for *Panthera tigris* in French would be adjusted for a set of random pages in the French Wikipedia. To do so, I firstly calculated the rate of change for each random page in each language using ‘rpi’, as in species pages. I then used a bootstrap resampling approach to calculate the average rate of change for all random pages in a given language at each timestep. The average rate of change in the random pages (λ_{rt}) was calculated by bootstrapping the monthly rates of change 1000 times, and then extracting the bootstrapped mean. At each timestep, I then adjusted the species page rate of change by subtracting the monthly bootstrap estimated random rate of change (λ_{rt}) as

$$\lambda_{at} = \lambda_{st} - \lambda_{rt}$$

where λ = the rate of change, t = month, r = the bootstrapped random trend for a given language, s = the species page trend for that same language, and a = the adjusted species trend.

For each species page the SAI is then

$$I_{at} = I_{a(t-1)} * 10^{\lambda_{at}},$$

$$I_{a0} = 1$$

where I_{at} = the species page SAI at time t .

To account for differences in the tortuosity of trends among Wikipedia languages (see Appendix 4, Figure S4.7), I also smoothed the species page SAI in each Wikipedia language using a loess regression (span = 0.3), before transforming the smoothed species page SAI back into a rate of change.

After smoothing the species page SAI as above, I then calculated a species SAI for each species (across languages) by averaging rates of change at each time step across all languages. For example, the species *Panthera tigris* has the unique Wikidata ID 'Q19939', meaning the average rate of change in SAI for all species pages (irrespective of language) identified as 'Q19939' provides the overall rate of change for the species *Panthera tigris*.

I then calculated an overall SAI combining all species across 10 Wikipedia languages by averaging rates of change across all species SAIs. Bootstrap confidence intervals were calculated by taking the 2.5th and 97.5th percentiles of 1000 bootstrapped indices at each timestep. To check the extent to which single languages influence the overall SAI, I then jack-knifed the overall SAI for language, and removed any languages with a marked effect on the overall trend (see Appendix 4, Figure S4.6).

Using the same approach as above, I also calculated an overall SAI for each taxonomic class for all languages combined, and each taxonomic class in each language. For each taxonomic class I again averaged the loess smoothed rate of change in species page SAI among languages, and then bootstrapped the species rate of change in SAI at each time step for each taxonomic class, as above. To check the extent to which single languages influence class level trends, I again jack-knifed the overall SAI for language, and removed any languages with a marked effect on the overall trend (see Appendix 4, Figure S4.8). To calculate an overall SAI in each taxonomic class in each language, I bootstrapped the rate of change in species page SAI for the set of species pages in a given class-language combination.

Modelling average monthly rate of change in the SAI

After calculating the SAI for all species pages on Wikipedia, I then calculated an average monthly rate of change in each smoothed species page SAI for the period January 2016 – January 2020. This average monthly rate of change was calculated across complete yearly periods to control for the effect of seasonality. To robustly explore whether change in awareness differs for various groups, I constructed 1 linear model and 2 mixed effects linear models fitting average monthly rate of change in species page SAI: 1) a linear model for average rate of change in species page SAI as a function of taxonomic class, language, and their interaction; 2) a mixed effects model for average rate of change in species page SAI as a function of taxonomic class, pollination contribution (Y/N), their interaction, and a random effect for language; and 3) a mixed effects models for average rate of change in species page SAI as a function of taxonomic class, traded status, and a random effect for language. Rather than attempting to find the most parsimonious model, I present full model predicted values, with AIC values for these and a set of candidate null models included in Appendix 4 (Tables S4.25, S4.26, and S4.27).

Results

Wikipedia view dataset

Before removing incomplete series, my initial Wikipedia dataset included ~2.23 billion page views for IUCN species across the Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian, and Spanish Wikipedias. These views were represented across 41,197 IUCN species, over a period of 1735 days between the 1st July 2015 and 31st March 2020. Views for each language varied from ~24.92 million views in the Arabic Wikipedia to ~1.08 billion views in the English Wikipedia (Figure 5.2). For all languages unique species number was highest in the ray-finned fishes at 13,571 and lowest in the insects at 2,743 (Figure 5.2, see Appendix 4, Figure S4.1 for a full language breakdown). After subsetting for series represented for every month, the proportion of complete series was lowest in the Arabic Wikipedia, specifically the ray-finned fishes (~35%) and the reptiles (~38%). Most taxonomic classes for most languages had complete series in at least 80% of the species in that grouping (Appendix 4, Figure S4.1).

After removing pages also present in the species set, my set of random views consisted of ~2.82 billion views across 113,622 random pages (Appendix 4, Figure S4.3), again for the same 1735 day period. The total number of random views was highest for the English Wikipedia at ~629.85 million views, and lowest in the Arabic Wikipedia at ~87.94 million views (Appendix 4, Figure S4.3). After subsetting for only random pages represented for all months, total random pages varied from 3486 in the Arabic Wikipedia to 9174 in the Japanese Wikipedia (Appendix 4, Figure S4.4).

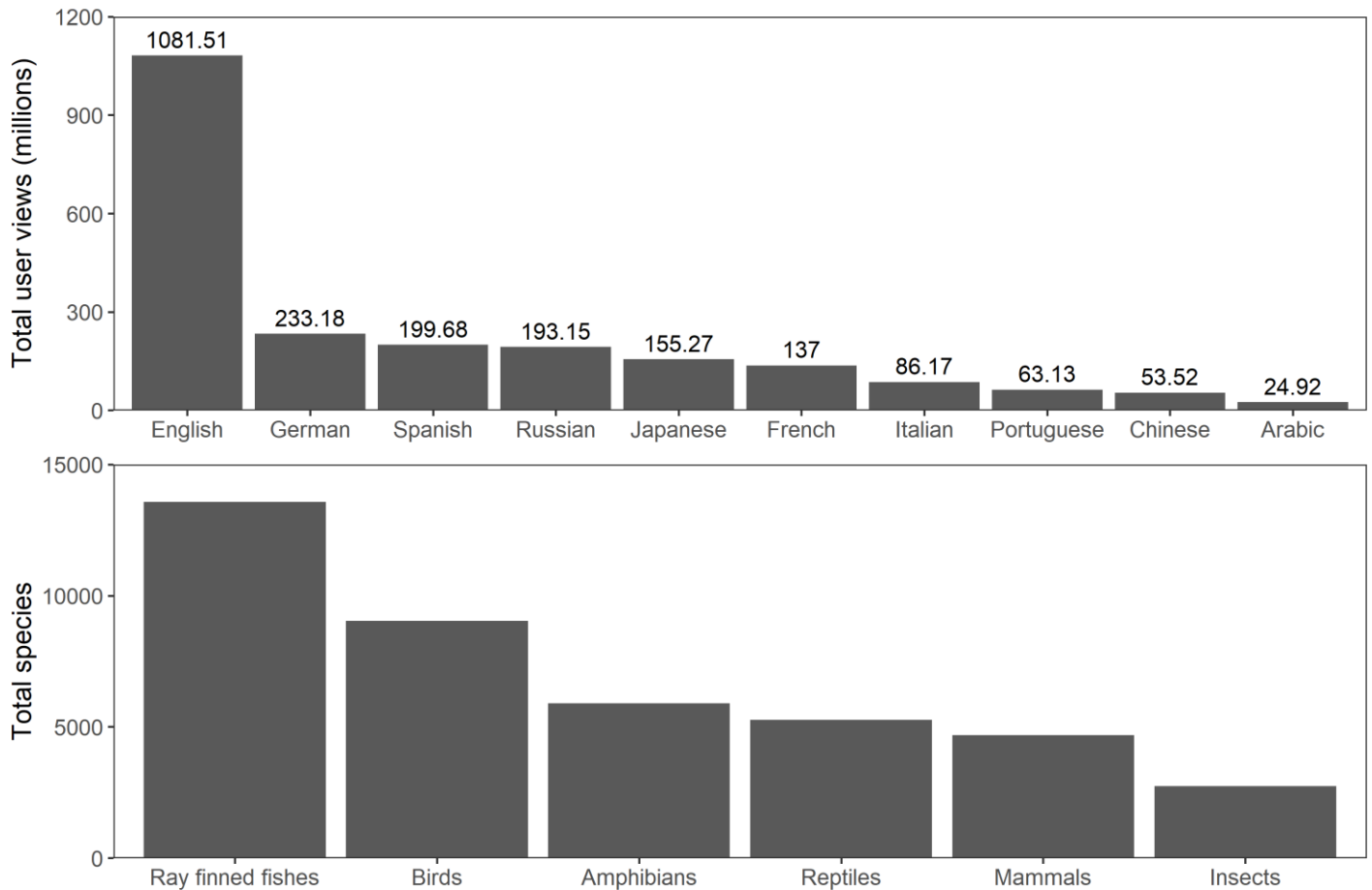


Figure 5.2 The total number of views and unique species for the initial Wikipedia view dataset, before removing incomplete series. Top: the total number of views for IUCN species in each Wikipedia language, in order of magnitude. Bottom: the total number of unique IUCN species ($n = 41,197$) with Wikipedia pages in each of 6 taxonomic classes, across all 10 languages of the top panel.

Absolute awareness of biodiversity

Among taxonomic classes, reptiles have consistently higher absolute awareness, appearing in the top 2 classes for 7/10 languages (Appendix 4, Figure S4.11). Amphibians on the other hand have consistently lower awareness, appearing in the bottom 2 classes for 8/10 languages. Some languages appear to have uniquely high absolute awareness for specific classes. For example, the ray-finned fish have the highest absolute awareness in the Japanese Wikipedia (Appendix 4, Figure S4.11). Across all languages, absolute awareness (total views) is significantly higher in traded species (Figure 5.2; $F = 15206.44$, $p < 0.001$, Appendix 4, Table S4.19), but not significantly different in pollinating species ($F = 0.3869$, $p = 0.5339$, Appendix 4, Table S4.18).

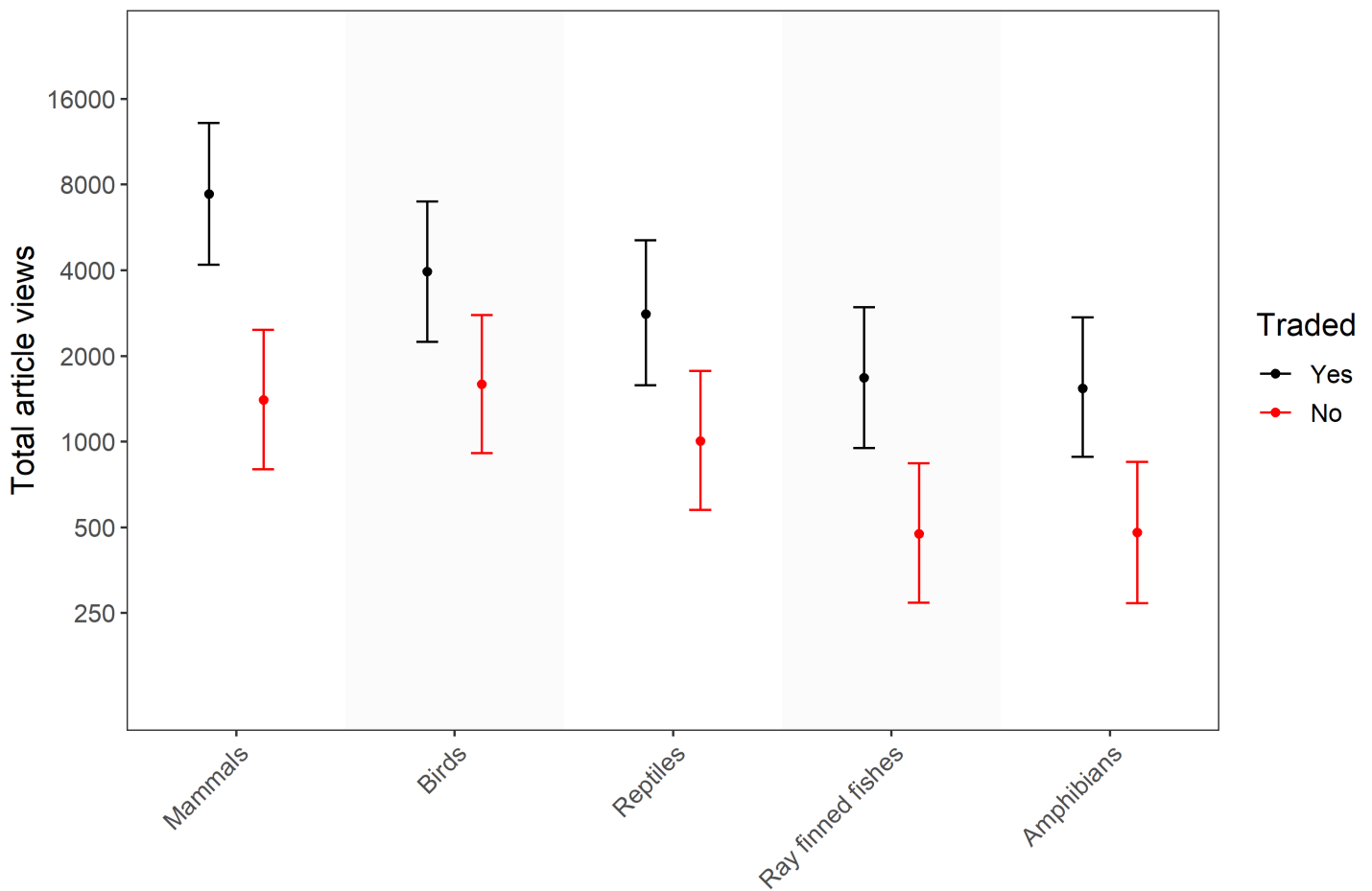


Figure 5.3 Total Wikipedia article views for traded and non-traded species across 6 taxonomic classes and 10 Wikipedia languages (Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian, Spanish). Predicted values were generated using a generalised linear mixed-effects model, modelling total article views as a function of the fixed effects taxonomic class, the presence of trade (Y/N), and their interaction, and the random effect language. Effect sizes were calculated by drawing fixed effects 1,000 times based on the variance-covariance matrix, and then calculating the median value (shown as points), and 2.5th and 97.5th percentiles (shown as error bars). Black error bars represent species that *are* known to be traded, and red error bars represent species that *are not* known to be traded or harvested. Taxonomic classes are ordered by the magnitude of total article views, from highest on the left to lowest on the right.

The SAI

The overall SAI for all taxa and languages was markedly affected by the inclusion of the French Wikipedia (see Appendix 4, Figure S4.6, and ‘Discussion’), meaning it was excluded from analyses presenting aggregated change at the overall level. With the exclusion of the French Wikipedia, the overall SAI increased from July 2015-March 2020, with marked declines in mid 2016 and 2018 (Figure 5.4). This overall increase in the SAI is largely robust to variable baselines (See Appendix 4, Figure S4.17, and ‘Discussion’), although average rate of change is marginally negative from mid 2017 and highly negative from mid 2019.

At the level of taxonomic class, jack-knifing trends by language again showed that the French Wikipedia was markedly affecting the overall trend (see Appendix 4, Figure S4.8, and ‘Discussion’). With the exclusion of the French Wikipedia, over the period July 2015-March 2020 all of the reptiles, ray-finned fish, mammals, and birds appear to have increased in awareness, whilst the amphibians and insects appear to have decreased (Figure 5.5). Birds experienced a peak in early 2017 (Figure 5), driven by an increase across multiple languages (Figure 5.6). The mammals appear to be experiencing a consistent and steady increase in awareness, particularly in the Japanese Wikipedia (Figure 5.6). The amphibians and insects both experienced a pronounced drop in awareness from the start of the series to mid 2016 before increasing, the cause of which is unclear. The trend for both the reptiles and insects is highly seasonal for multiple languages, peaking in July-August of each year, with the notable exception of the English language for insects (Figure 5.6).

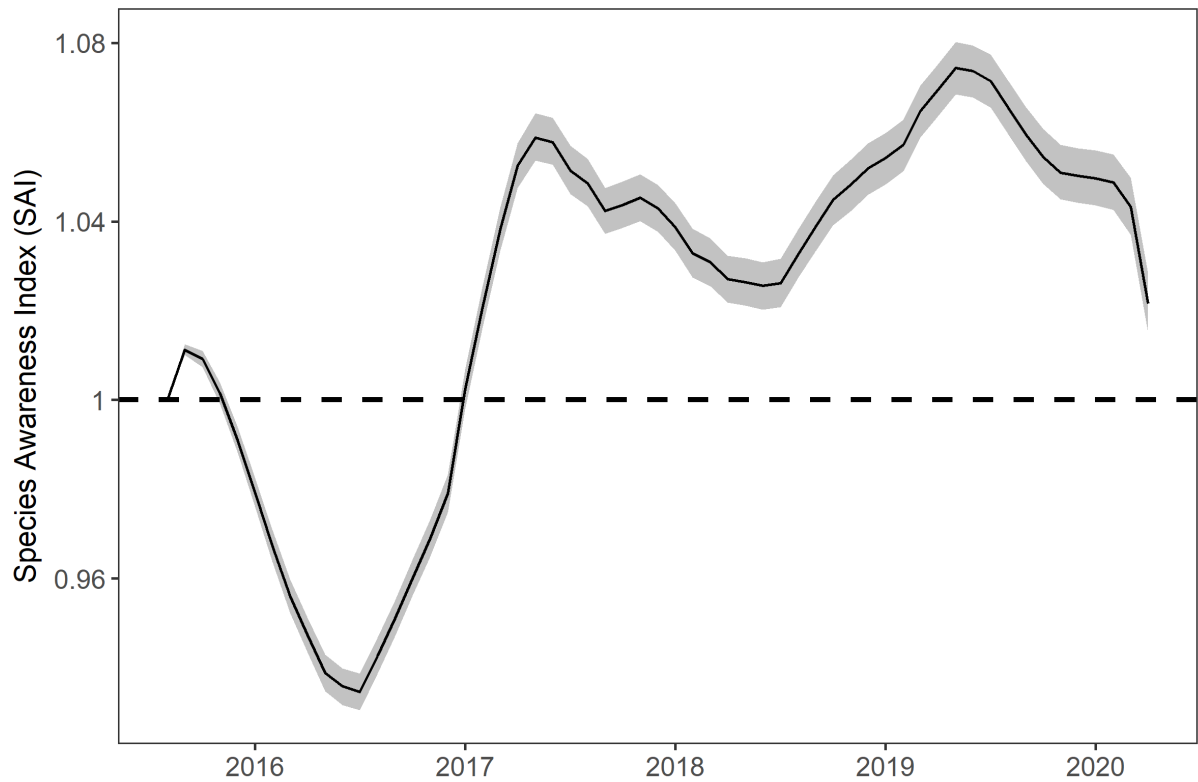


Figure 5.4. The SAI for 6 taxonomic classes (reptiles, ray-finned fish, mammals, birds, insects, and amphibians) and 9 Wikipedia languages (Arabic, Chinese, English, German, Italian, Japanese, Portuguese, Russian, and Spanish) for the period July 2015-March 2020. The French Wikipedia was removed here given its marked influence on the aggregated SAI (see Appendix 4, Figure S4.6). Here the black line represents the mean of the bootstrapped indices at each monthly timestep, and the grey band the 2.5th and 97.5th percentiles.

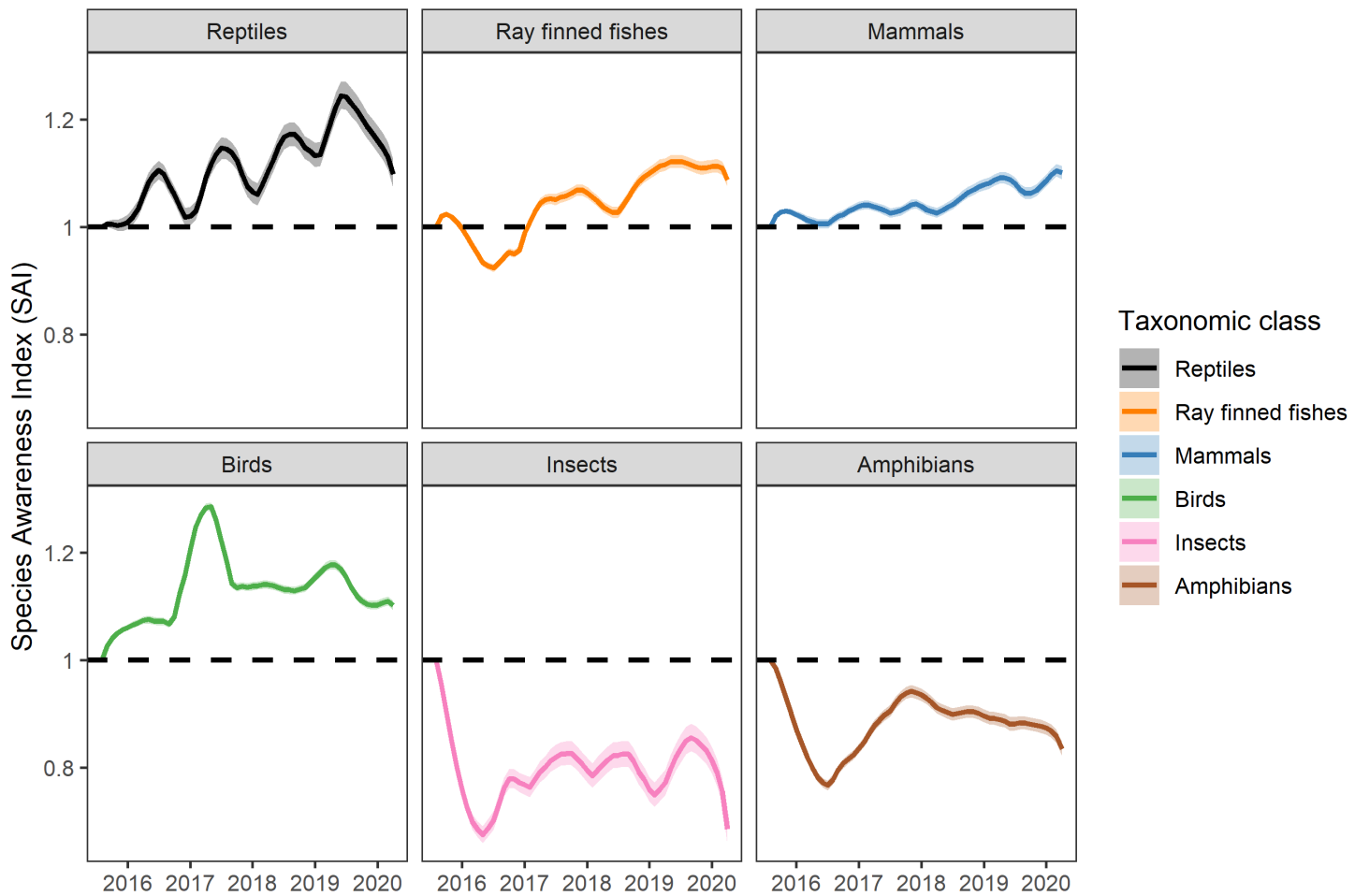


Figure 5.5 The SAI for 6 taxonomic classes (reptiles, ray-finned fishes, mammals, birds, insects, and amphibians) and 9 Wikipedia languages (Arabic, Chinese, English, German, Italian, Japanese, Portuguese, Russian, and Spanish), for the period July 2015-March 2020. The French Wikipedia was removed here given its marked influence on the aggregated SAI (see Appendix 4, Figure S4.8). Coloured lines represent the mean of the bootstrapped indices at each monthly time step, and coloured bands the 2.5th and 97.5th percentiles: reptiles (black), ray-finned fishes (orange), mammals (blue), birds (green), insects (pink), and amphibians (brown). Taxonomic class panels are ordered by the magnitude of overall increase in each taxonomic class.

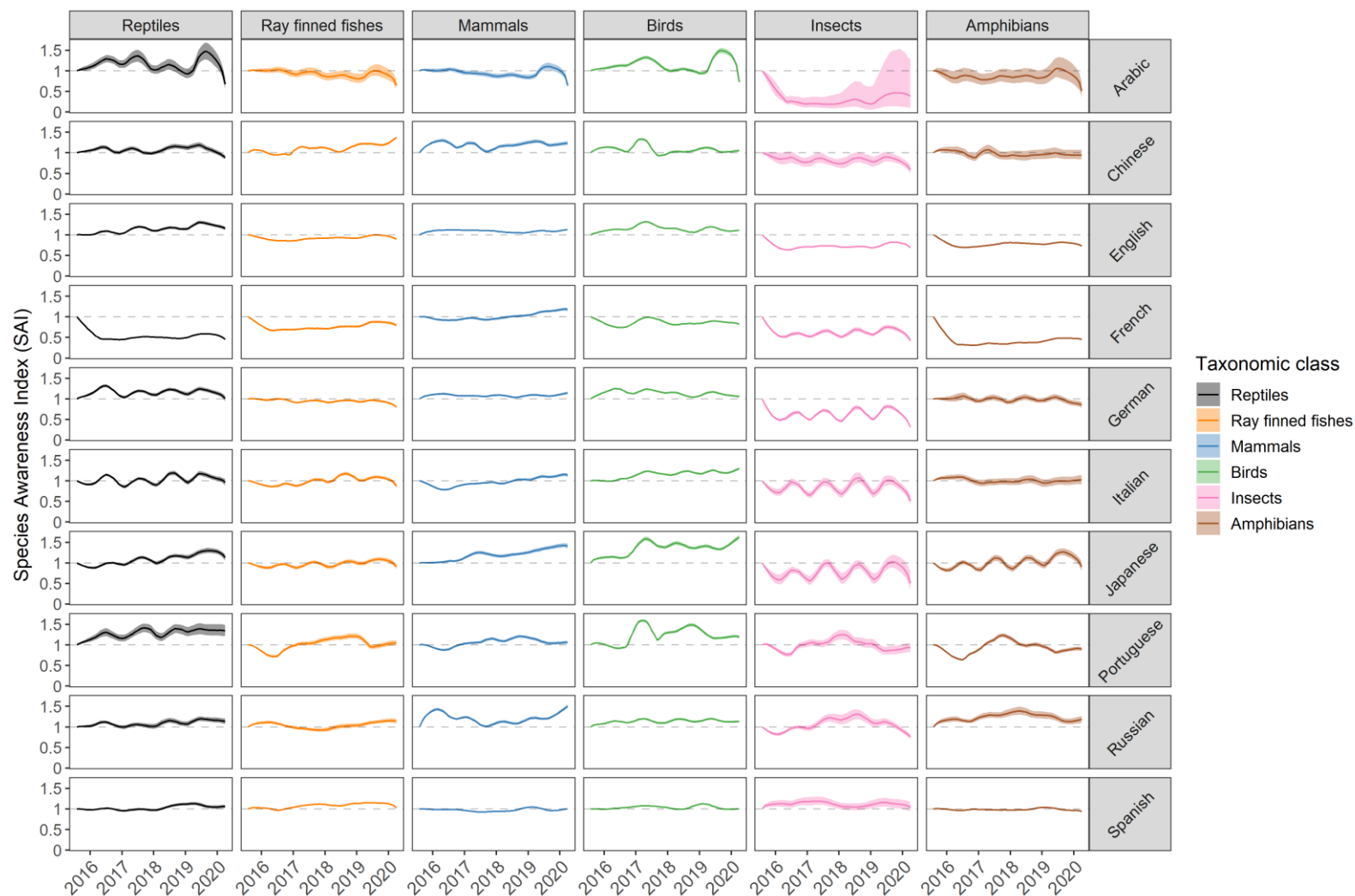


Figure 5.6 The SAI for 6 taxonomic classes across 10 Wikipedia languages, for the period July 2015-March 2020. Coloured lines represent the mean of the bootstrapped indices at each monthly each time step, and coloured bands the 2.5th and 97.5th percentiles: reptiles (black), ray-finned fishes (orange), mammals (blue), birds (green), insects (pink), and amphibians (brown). Taxonomic class panels are ordered by the magnitude of overall increase in each taxonomic class, and for language alphabetically.

Modelling average monthly rate of change in species page SAI

Average monthly rate of change in species page SAI for the period January 2016-January 2020 differed significantly for taxonomic class, language, and their interaction (Figure 5.7, Appendix 4, Table S4.22). At the level of taxonomic class, the reptiles and ray-finned fishes are increasing in awareness the fastest, and the insects and amphibians are either increasing slowly or declining (with the exception of the Japanese Wikipedia). Among languages, rate of change in species page SAI is highest in the Japanese and Portuguese Wikipedias, and lowest in the German and Chinese Wikipedias. Although absolute interest is significantly greater in traded species (See Figure 5.3, Appendix 4, Table S4.19), over the period January 2016-January 2020 average monthly rate of change in the species page SAI appears not to be related to either trade contribution (Appendix 4, Figure S4.14, Table S4.21) or pollination contribution (Appendix 4, Figure S4.14, Table S4.20).

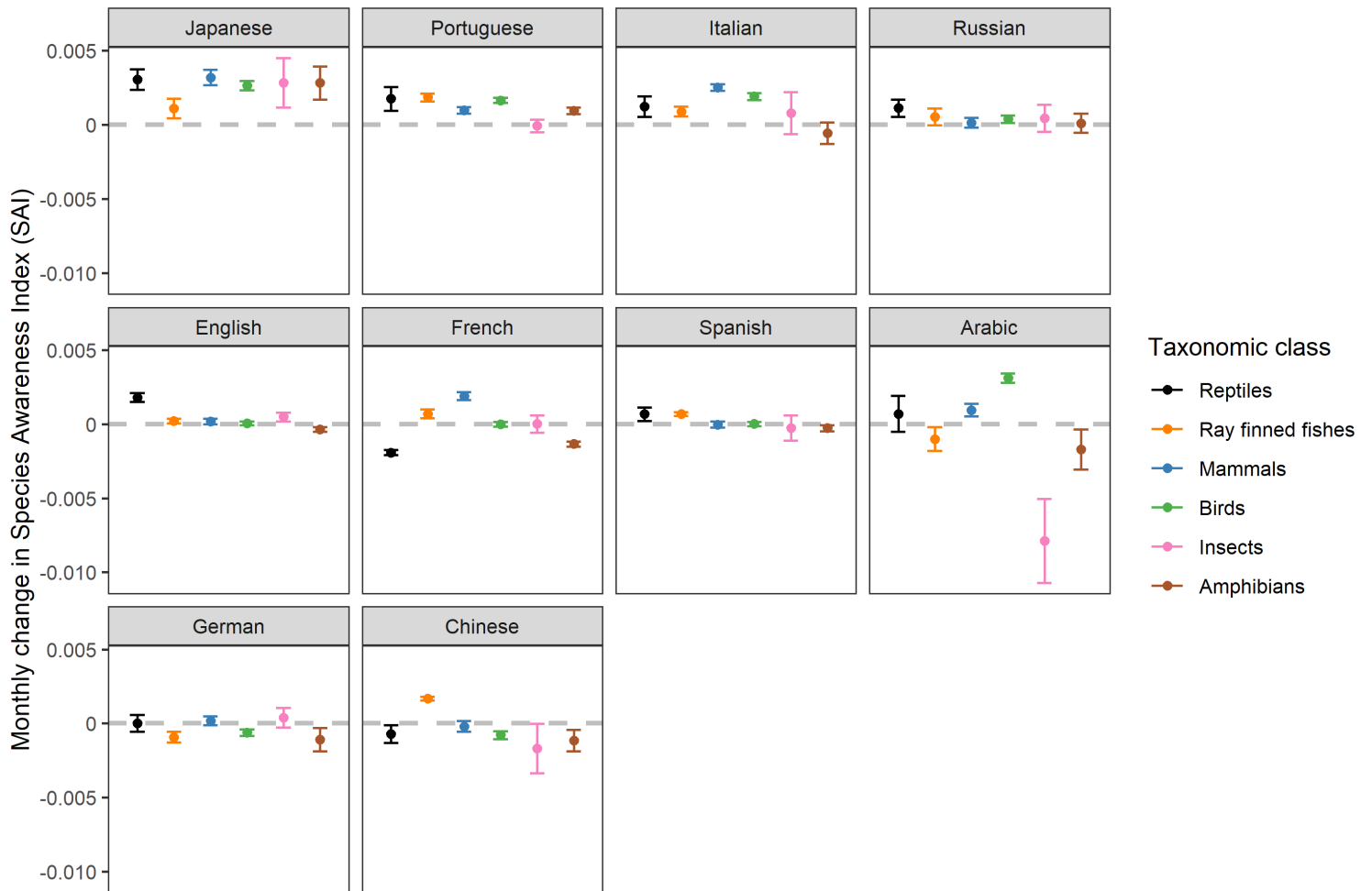


Figure 5.7 Average monthly rate of change for the species page SAI for 6 taxonomic classes across 10 Wikipedia languages. Errors bars represent the predicted values of a linear model, fitting average monthly change in the species page SAI as a function of taxonomic class, Wikipedia language, and their interaction. Fitted values were generated from the linear model with the R function `predict` (represented as points), and 95% confidence intervals from the fitted values $\pm 1.96 \times \text{standard error}$. The colour of error bars refers to taxonomic class: reptiles (black), ray-finned fishes (orange), mammals (blue), birds (green), insects (pink), and amphibians (brown).

Discussion

In this study I introduced the SAI, an index of change in public awareness of biodiversity. The SAI is derived from views of individual Wikipedia species pages, enabling investigation according to a variety of variables, such as taxonomy, language, geographic distribution, and ecosystem service provision. I show that awareness of biodiversity overall has marginally increased, with increases highest in the reptiles and ray-finned fish. Although biodiversity awareness is increasing overall, we also show that some groups (i.e. the amphibians and insects) are decreasing or only marginally increasing in awareness. Among languages, increases are highest in the Japanese and Portuguese Wikipedias, and lowest in the Chinese and German Wikipedias. Although my results suggest that awareness of biodiversity has increased since July 2015, increases do not appear to be related to a species' trade contribution or pollination contribution. As an indicator for biodiversity awareness, a Wikipedia derived metric such as the SAI represents a useful additional data source, given its explicit and unambiguous link to biodiversity itself at multiple scales (i.e. species, family, class).

The link between culture and perceived biodiversity value or awareness is widely recognised (Daniel et al. 2012, Cooper et al. 2019, Roll et al. 2016, Ladle et al. 2019), but for a culturomics metric such as the SAI the drivers of change are complex. Overall trends capture many different drivers of awareness, making it difficult to isolate the causes for a given increase or decrease. The Chinese Wikipedia, for example, shows a consistent decrease in awareness for 5 taxonomic classes, but a consistent increase for ray-finned fish. I hypothesised that this increase for ray-finned fish may be driven by increasing fish consumption, since seafood demand in China has increased significantly in recent years (FAO 2020). However, in a brief additional analysis, I found no significant difference between the rate of change for traded and non-traded ray-finned fish in the Chinese Wikipedia (Appendix 4, Figure S4.13). This would indicate that the greater rate of change for the

Chinese ray-finned fish may not be driven by consumption alone. The Japanese Wikipedia is also of note, given its consistent increase in awareness across all 6 taxonomic classes. This awareness increase is concordant with the results of conventional surveys, in which Japan has amongst the largest percentage point increases for familiarity with the term 'biodiversity' (UEBT 2019). However, it is unclear what may be driving this change in awareness. Counterfactual scenario modelling could help to better understand such relationships, previously demonstrated in a number of recent conservation culturomics studies (Acerbi et al. 2020, Veríssimo et al. 2020, Fernandez-Bellon & Kane 2019).

Although drivers of overall change in the SAI are complex, it is conspicuous that absolute awareness and change in awareness do not differ between pollinating and non-pollinating animals. Since traded species have high absolute awareness relative to non-traded species, biodiversity awareness likely does relate to its value. But for pollination contribution this relationship appears to be weak or non-existent. For three main reasons this may be the case: first, the impact pollination contribution has on biodiversity awareness will be highly taxa dependent, with groups more strongly associated with pollination experiencing a bigger increase in awareness; second, unlike a species traded for direct consumption (i.e. food), pollinators make an indirect contribution to people, making their benefit less intuitive to understand; and third, the nature of a pollinator is that it is often not deliberately sought by those that benefit from its contribution. Given its central role in the value of biodiversity, more work is required to understand the contribution ecosystem service provision makes to public biodiversity awareness. Since pollinators have been so well-publicised (Smith and Saunders 2016), one would expect that if awareness is not relatively high for pollinating species, it likely will not be for other service providing species.

Change in biodiversity awareness reflected by the SAI is related to, but largely distinct from absolute awareness. In other words, for a given group of species, high rate of change in awareness does not necessarily also mean high absolute awareness. Instead, one might expect that groups of high absolute awareness would often be stable or increasing slowly,

since the margin within which they can continue to increase is smaller. In the Japanese Wikipedia, for example, ray-finned fish have the highest absolute awareness, but the lowest average monthly rate of change in species page SAI. Future work should further explore the relationship between absolute and change in awareness, particularly for variables previously applied in absolute awareness research, such as threat status and venomosity (Roll et al. 2016), and geographic range and phenotypic distinctiveness (Ladle et al. 2019).

Strong seasonal patterns in awareness are apparent in the bootstrapped trends for some taxonomic groups (e.g. insects and reptiles) and languages (e.g. Japanese). Seasonality in species interest on Wikipedia has been demonstrated previously, with high levels of seasonality in organisms that tend to have a strong phenological component to their life-history (e.g. insects and flowering plants) (Mittermeier et al. 2019). In my analysis, languages such as English and Spanish have a conspicuous lack of a seasonal trend for all bootstrapped taxonomic classes. Most likely this results from the nature of English and Spanish as languages that are used widely in both the northern and southern hemispheres. Opposing peaks of interest in the northern and southern hemisphere summers will tend to cancel each other out, acting to smooth the overall bootstrapped trends.

The significant effect of the French Wikipedia on the overall SAI is surprising, given the low species page and random page views for the French Wikipedia. Inspection of the French Wikipedia shows that reptile species pages are primarily the cause of this problem. The French Wikipedia contains a large number of reptiles (Figure S4.1), which decrease rapidly at the start of the series (Figure 5.6). This decrease is compounded by the large number of reptiles in the French Wikipedia appearing in only that Wikipedia (Figure S4.15). This means that when an average is taken across languages, the large negative trends of the French reptiles are not mitigated against other less negative trends, meaning they come to dominate the index.

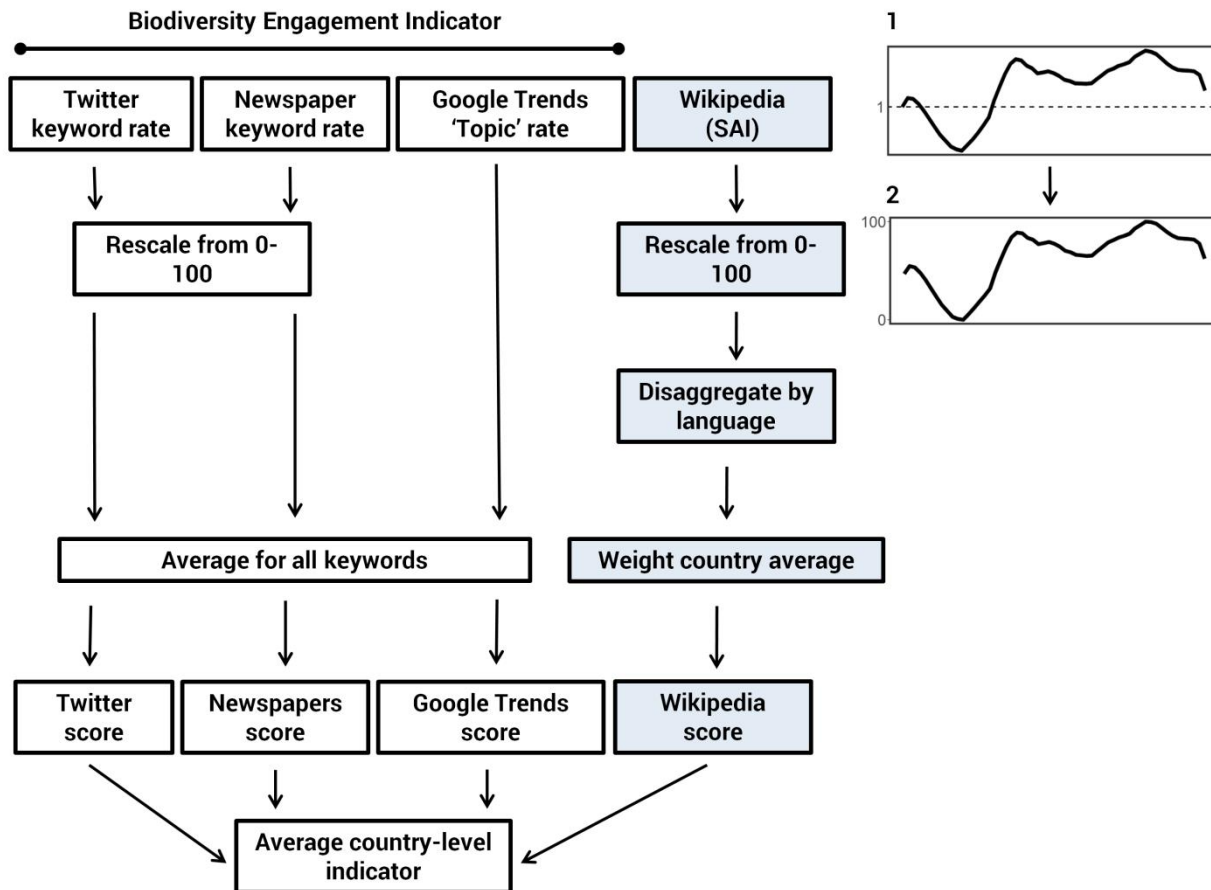


Figure 5.8 Schematic adapted from Cooper et al (2019), demonstrating one method through which the SAI might be incorporated into the Biodiversity Engagement Indicator for a single indicator of country-level biodiversity awareness. All text squares in white represent the methodological pathway for the original Biodiversity Engagement Indicator, and all squares in pale blue represent a potential additional pathway for combining the SAI. Box 1 (top right) represents the overall SAI scaled starting at 1. Box 2 represents the overall SAI rescaled between 0-100, consistent with the Biodiversity Engagement Indicator.

The SAI and Biodiversity Engagement have value as independent metrics, but as in Cooper et al. (2019) I emphasise the importance of combining multiple online platforms for inferring public biodiversity awareness. Particularly since the SAI provides an explicit link to biodiversity itself, its inclusion could provide a more holistic understanding as to how biodiversity awareness is changing. Combining the SAI with the Biodiversity Engagement Indicator presents two core challenges. First, both the Biodiversity Engagement Indicator and SAI are measured on different units, inherent to their underlying methodology. The Biodiversity Engagement Indicator is scaled on a 0-100 scale in a manner analogous to Google Trends. The SAI on the other hand is scaled relative to a benchmark index of 1, in an approach inspired by the LPI. Second, there are problems of geographic scale in combining the Biodiversity Engagement Indicator and SAI. Namely, the Biodiversity Engagement Indicator is aggregated at the country-level, whereas the SAI cuts across countries at the language-level.

Given differences in units and geographic scale, combining a Wikipedia metric and the Biodiversity Engagement Indicator is not simple. One potential solution could be to rescale the SAI on a 0-100 scale, disaggregate by language, and then calculate a weighted average among languages to reflect the proportion of users for a given country (Figure 5.8). The data for calculating such a weighting language level trend is provided by Wikipedia, in a format amenable to web-scraping (Wikimedia Traffic Analysis Report 2018). Such an approach would solve both the problem of differing units and geographic scale, transforming the SAI into a national metric amenable to averaging alongside a Twitter, Newspaper, and Google Trends score (Figure 5.8).

Despite providing a novel approach, online-derived metrics for biodiversity awareness are subject to limitations. Primarily, online metrics are proxies rather than direct measures of awareness. For the SAI, we can show whether views increased or decreased on average for a given grouping, from which we can infer exposure to species-related information. However, we cannot tell why a given page was visited, or whether information related to that page was

retained. Previous research has shown how on-site Google Analytics can be used to return a suite of metrics intimating as to the reasons for a given visit (Soriano-Redondo et al. 2016), such as the site used to reach a page, but for Wikipedia this data is not publicly available. Text-mining could help to quantify the type of information users are exposed to on Wikipedia. For example, calculating a text similarity for each species page to a reference text on pollination would provide an indication as to the pollination salience of a given species page. Similar approaches have been applied in the context of climate change and invasive species, using rate of threat-related terms as an indicator of threat salience (Jarić et al. 2020).

There are other limitations associated with the SAI. First, the short length of my Wikipedia time series presents a problem for the interpretation of our index. Future iterations of the SAI should aim to include views archived from before 2015, which presents additional problems in reconciling views aggregated across multiple formats. Second, an index based on internet activity will only be representative of those that have access to the internet, and use the internet to access Wikipedia. Given ~50% of the global population has access to the internet (The World Bank 2020), and ~15% of those regularly access Wikipedia (Graham et al. 2014), the SAI cannot be globally representative. In mainland China for example, the dominant online encyclopaedia is Baidu Baike, with Wikipedia views for the Chinese (zh) Wikipedia coming predominantly from Taiwan and Hong Kong (Wikimedia Traffic Analysis Report 2018). Future iterations of the SAI could incorporate page views for species on Baidu Baike, although this presents problems in requiring a distinct computational approach. Third, conclusions drawn from the SAI are susceptible to variations in the chosen baseline month. I therefore carried out an additional analysis to explore the robustness of my overall Species Awareness Index to variable baselines. For all potential baselines of the SAI (i.e. months), I recalculated the average rate of change from that point to the end of the series, and then superimposed these rates of change onto the overall SAI (Figure S4.17). Here I show that the average rate of change is stable or greater than 0 for all baselines between July 2015

and February 2017. The average rate of change then goes negative from the baselines of mid 2017, and highly negative from mid 2019 to the end of the series. Although my additional analysis shows that the SAI for all 6 taxonomic classes is largely robust to variable baselines, further research is required to fully understand the effect among all taxa and languages.

As global internet penetration increases and biodiversity continues to experience decline, digital metrics for public biodiversity awareness will become both more informative and more important. Here I presented the SAI, a metric derived from Wikipedia views depicting change in awareness for biodiversity online. I used this metric to show that overall awareness of biodiversity is increasing marginally, although this increase is inconsistent among taxonomic groups and languages. I also showed that such increases appear not to be related to either the pollination contribution or trade contribution of a species. I concluded by suggesting one approach through which the SAI could be combined with the Biodiversity Engagement Indicator, providing a more holistic understanding of public biodiversity awareness in the digital realm.

Chapter 6: Discussion and synthesis

Pollinator biodiversity has undergone significant change in the Anthropocene (Biesmeijer et al. 2006, Burkle et al. 2013, Potts et al. 2010). Such changes have often been associated with climate change and anthropogenic land use (Potts et al. 2010, Vanbergen, 2013, Winfree et al. 2011), both of which will continue to increase and expand over the coming decades (IPCC 2018; Laurance et al. 2014). Given the importance of animal pollinators to humanity, both to the structure of terrestrial ecosystems and crop pollination, understanding the extent to which pollinator biodiversity will change in the future is highly important (Brown et al. 2016).

In this thesis I applied a number of computational methods to investigate the causes and consequences of pollinator biodiversity change in the Anthropocene. First, I carried out a text-analysis review of the pollination literature, showing that although there is a marked bias towards the honey bees and bumblebees of North America and Europe, there is still pollination information for other geographic regions and taxonomic groups (Chapter 2). Second, I then used the output of this text-analysis (i.e. the set of animal genera extracted from a set of pollination related abstracts) to identify a set of potentially pollinating species in the PREDICTS database, which I then confirmed as likely pollinators using a manual check by myself and a group of pollination ecology experts (Chapter 3). I then used this pollinating subset of the PREDICTS database to investigate the global response of local pollinator biodiversity and land-use type and intensity (Chapter 3). Specifically, I showed that low levels of intensity can increase pollinator biodiversity, but that within a land use type increasing intensity tends to lead to decreases in pollinator biodiversity (Chapter 3). I also showed that pollinator responses to intensity differ among taxonomic groups, geographic region, and between analogues of intensity (i.e. total fertiliser application rate and a coarse categorical factor of intensity) (Chapter 3). Third, using the same set of likely pollinating species in PREDICTS, I then investigated the interactive effects of climate change and land use on local pollinator abundance (Chapter 4). I showed that land use and climate change have a significant interactive effect on insect pollinator abundance, such that climate change is associated with

a decrease in abundance on cropland, and an increase on primary vegetation (Chapter 4). I then used my interactive models, estimations of global crop production that is dependent on animal pollination, and projections of future climate change, to predict potential temporal and spatial regions of crop pollination risk (Chapter 4). Here I showed that under RCP 8.5—a high emissions future climate scenario—my total crop production risk index is predicted to increase by ~70% between 2016 and 2048. Again under RCP 8.5, I showed that the rate of change in total crop production risk could be up to ~25% higher than that of RCP 2.6 and 6.0, with abrupt periods of increased exposure in the 2030s. I further showed that crop pollination risk is highest, and predicted to increase most rapidly, in regions of sub-Saharan Africa and northern South America, primarily for cocoa, mango, pumpkin, melon, watermelon, and coffee production (Chapter 4). Fourth, I finished by introducing a new metric of public biodiversity awareness, derived from page views of animal species on Wikipedia (Chapter 5). I use this metric to show that, although awareness of biodiversity appears to be marginally increasing, these increases are likely not related to the contribution animals make to pollination (Chapter 5). In this last chapter, I highlight the core contributions I make to four areas of work, as well as the main challenges and future directions for each of these research areas.

Quantifying the geographic and taxonomic distribution of the animal pollination literature and improving tools for automated reviews

Contribution

Many studies make claims that most pollinator biodiversity change research concerns only the honey bees and bumble bees of North America and Europe (Ghazoul 2005, 2015; Goulson et al. 2015). However, attempts to quantify this extent are few, especially for the distribution of research among pollinating groups (although see Saunders et al. 2020). My work in Chapter 2 aimed to address this problem. I showed that at least 25% of pollination-related abstracts mention a species of honey bee and at least 20% a species of bumble bee, and approximately 46% of abstracts are focussed on either North America or Europe. This research builds on the work of Archer et al. (2014) and Ollerton (2017) through investigating taxonomic change over

of time, and investigating the interaction between taxonomy and geography in animal pollinators. Importantly, this research showed that there is a large quantity of animal pollinator related research based outside of North America and Europe, which could be used to supplement global models of pollinator biodiversity change. This work also used two named-entity recognition tools which are rarely applied in ecology and conservation: CLIFF-CLAVIN, a geoparser which attempts to resolve the main geographic focus of a piece of text (D'Ignazio et al. 2014); and the function `scrapenames` from the package `taxize`, a piece of software that extracts a taxonomic names from text (Chamberlain & Szöcs 2013). In Chapter 2 I discuss that named-entity recognition algorithms could be used to estimate a lower-bound for the number of pollinating species, and to predict likely plant-pollinator interactions.

As the quantity of literature in the field of pollination ecology continues to grow, incorporating all relevant literature in systematic maps and synthetic analyses will become increasingly difficult. In Chapter 2 I made some suggestions as to how this problem might be approached, using a novel synthesis path and text-analysis algorithms to shorten the synthesis gap, or in other words the lag between the practice of primary research and the synthesis of evidence (Westgate et al. 2018). A number of other papers have explored ways of shortening this synthesis gap, such as automating the identification of relevant texts (Cornford et al. 2021), and building software tools for article screening (Westgate 2019), but widespread uptake of these tools is still low.

Challenges and future directions

For the quantification of biases in the ecological literature, there are some key challenges that need to be overcome before such text-analysis of abstracts can be relied upon to make robust estimates. First, biases caused by the text-analysis algorithms themselves need to be more thoroughly checked, and where biases are found, these need to be accounted for in estimates of geographic and taxonomic distribution. In the context of the work I presented in Chapter 2, these biases will likely be strongest for the geoparser. CLIFF-CLAVIN was built for geoparsing online news articles, meaning it will likely perform less well when applied to ecological texts,

and may favour locations that tend to appear in the news articles from which it was trained (Imani et al. 2017). As mentioned in Chapter 2, one would expect that such biases likely favour the United States, meaning the actual disparity between the United States and the rest of the world may be smaller than suggested by my analysis. One approach to account for this potential bias could be to run CLIFF-CLAVIN on a set of texts with a known geographic distribution, and then use this to adjust any subsequent distributions outputted by CLIFF-CLAVIN. Work is currently underway exploring this approach with the Living Planet Database, preliminarily indicating that CLIFF-CLAVIN does indeed identify locations in the United States more often than expected by chance (Cornford et al. 2021). Second, biases caused by the extent to which abstracts represent full text need to be tested and accounted for in estimates of taxonomic and geographic distribution. For example, it could be that researchers have a tendency to over-emphasise certain taxa in abstracts, perhaps to increase the impact of their study. Understanding this bias is particularly important given the number of studies that assume the content of an abstract is representative of the taxa studied in that paper (Millard, et al. 2020a, da Silva et al. 2020, Santos et al. 2020). If taxonomic and geographic information in abstracts is not consistently representative of the research content of the full paper, then the distribution implied by these papers would be biased. Work is also underway exploring this, again with the Living Planet Index, although results at this stage are inconclusive (Cornford et al. 2021).

The causes of pollinator biodiversity change

Contribution

A number of prior studies have examined the effect of land use on pollinator biodiversity change (Biesmeijer et al. 2006, Goulson et al. 2015, Winfree et al. 2011), described by some as likely the greatest driver of change (Goulson et al. 2015). Climate change and its interaction with land use has also been described as a significant driver of pollinator biodiversity change (Settele et al. 2016, Vanbergen, 2013), often investigated in the context of localised studies

(e.g. (Zaragoza-Trello et al. 2021)). However, for both land use and the interaction of climate change with land use, as far as I know there are no cross-taxonomic global studies.

In Chapter 3 I used a set of likely pollinating species in the PREDICTS database to model the response of pollinator biodiversity to land-use type and intensity, using a space-for-time approach. Counter to many studies which suggest consistently negative effects of anthropogenic activity on pollinator biodiversity, I show that pollinator biodiversity change is likely more heterogeneous. First, I demonstrate that pollinator biodiversity change follows a 'hump-shaped' response, with low disturbance associated with higher biodiversity (relative to primary vegetation), and high intensity associated with lower biodiversity (relative to primary vegetation). Such a relationship has been shown in a number of other localised studies (e.g. Lázaro et al. 2016), and likely relates to the increased heterogeneity of habitats at low intensity (the intermediate disturbance hypothesis) (Connell 1978). Second, I show that vertebrate pollinators are more resilient to anthropogenic land use (and intensity) than invertebrate pollinators. Vertebrate pollinator resilience has been found in a number of localised studies (e.g. Tscharrntke et al. 2008), but not in previous synthetic studies where the effect diminishes when accounting for differences in methodology among studies (Winfree et al. 2011). My work in Chapter 3 indicates that vertebrate pollinators are indeed less sensitive, likely resulting from their nature as bigger bodied, more mobile species (Henle et al. 2004). Third, I show that tropical pollinator biodiversity is more sensitive to the effects of land use intensity on cropland, with high anthropogenic intensity associated with significant declines on tropical cropland, and no change in non-tropical cropland. Relative to the non-tropics, tropical pollinator biodiversity has previously been found to be more sensitive (e.g. Newbold et al. 2020), but as far as I know my work here represents the first time this has been confirmed at large-scale. Such sensitivity is likely predicted by the combined effects of novel extreme climate and historical anthropogenic land use. The tropical zone is both being exposed to extreme novel temperatures more quickly, and has a shorter history of anthropogenic land use, which in the temperate zone has likely already filtered out the more sensitive species (Balmford 1996).

Fourth, I show that the effects of anthropogenic land use are not always negative, with some groups (e.g. flies) experiencing an increase in pollinator biodiversity in response to land use intensity. Such a positive response of flies to anthropogenic land use intensity has been found previously in local studies (Biesmeijer et al. 2006, Jauker et al. 2009, Weiner et al. 2011), here confirmed at the global scale.

In Chapter 4 I used the same set of pollinating species to investigate the interactive effect of climate change and land use on local pollinator abundance. For insects, this interactive effect is such that abundance declines in regions where cropland coincides with high historical temperature change, and increases in regions of primary vegetation. Such a response has been found previously (Outhwaite et al. in review), likely relating to the nature of tropical insects as highly sensitive to novel extreme temperatures that fall outside their thermal limits (Deutsch et al. 2008). I show that this response is likely stronger in pollinating insects than non-pollinating insects and also vertebrates. From my analysis the mechanisms of this sensitivity are unclear, but as I describe in Chapter 4, it's likely that some combination of traits and a reliance on floral resources predicts this high sensitivity.

Challenges and future directions

There is a lot we understand about the causes of pollinator biodiversity change, but still some key challenges and questions for future research. Here I briefly highlight three areas which require additional work.

The primary challenge for pollinator biodiversity change research I feel lies in the geographical distribution of research. In Chapters 3 and 4 I showed that tropical pollinator biodiversity is highly sensitive. In the tropics, high agricultural intensity is associated with much greater declines than in the non-tropical zone (Chapter 3). The tropics are also being exposed to novel extreme temperatures more quickly, with this change associated with significant declines in local pollinator abundance on croplands (Chapter 4). Given that many prior synthetic studies on pollinator biodiversity change focus on the temperate zone, it therefore seems likely that

these previous studies have underestimated the effects of anthropogenic activity on pollinator biodiversity. Although the two studies I carried out here help to reduce this bias, there is still a significant bias in the data towards non-tropical regions, and within the tropics towards a select few countries (e.g. South Africa, Kenya, Ghana, Madagascar). Moving forward we need to integrate more primary research from the tropics into global models of pollinator biodiversity change.

Second, the models I present in both Chapters 3 and 4 are correlative, meaning they do not reveal the mechanisms of biodiversity change, with the observed patterns potentially explained by a set of correlated variables that I did not consider. This is particularly the case for the interaction between climate change and land use tested in Chapter 4. Standardised temperature change could be correlated with recent land-use intensity, or some measure of current climate. Future work therefore needs to establish the extent to which change in climate, current climate, historical land use, and current land-use intensity predict variation in pollinator biodiversity between tropical and non-tropical regions, with all variables included in a single model. Estimations of historical and current climate are feasible for much of the world (e.g. Frieler et al. 2017, Harris et al. 2020), but data for historical land use is, as far as I know, not available for much of the world.

Third, for historical land use there is a related challenge associated with baselines for pollinator biodiversity change. The baseline for the models I present in this thesis, which used the PREDICTS database, is 'primary vegetation', or in other words vegetation that has not been disturbed by humans. However, as I discuss in Chapter 3, the nature of this baseline differs between the tropical and temperate zone, meaning it does not necessarily represent a consistent, perfectly natural baseline. Research increasingly shows that human activity goes back thousands of years (Mottl et al. 2021), meaning the land-use factor described here as 'primary vegetation' will often not be a truly undisturbed habitat. Conservation biology more broadly has real difficulties with baselines, notably in the Living Planet Index that is anchored to a baseline of 1970 (McRae et al. 2017). Often it is either unclear where baselines should

be set, or the most logical baseline is clear but the relevant data are unavailable (Mihoub et al. 2017). One potential way forward is to test the sensitivity of predictions to changes in the baseline, as I did in Chapters 3 and 5 for pollinator biodiversity and changes in biodiversity awareness (see Figure S2.6 and Figure S4.17), and then quantify some degree of uncertainty according to the variation among baselines. Although such an approach still means model uncertainty, it is at least transparent that baselines are a cause of variation, and it enables the researcher to test the influence of arbitrary baseline selection on their model predictions.

The consequences of pollinator biodiversity change

Contribution

One of the core questions for pollinator biodiversity research concerns the extent to which future changes in pollinator biodiversity might relate to the stability of crop production and the availability of animal pollination dependent crops (Potts et al. 2016). A number of studies have found that localised reductions in yield can occur through reductions in local pollinator abundance (Kevan 1977, Stephen 1955, Watanabe 1994), whilst others have suggested that we risk widespread shortfalls from continued pollinator losses (Aizen et al. 2008, Winfree, 2008). Although large shortfalls are not yet thought to have occurred on a broad scale researchers postulate that as pollinator biodiversity continues to change we risk them become increasingly common (Winfree 2008). My work in Chapter 4 uses a model of the interactive effects of climate change and land-use on local pollinator abundance to make the first systematic predictions of the temporal and spatial distribution of crop pollination risk. Although there are a number of studies and reviews examining the interactive effects of climate change and land-use on pollinators (e.g. Settele et al. 2016, Vanbergen, 2013), as far I know there are none that combine models with estimates of animal pollination dependent crop production to predict risk to global crop pollination.

Challenges and future directions

The models I present here make coarse predictions on the temporal and spatial distribution of crop pollination risk. To increase the robustness of crop pollination risk scenarios there are a number of challenges that need to be addressed. Here I highlight three challenges, and suggest ways in which each may be overcome. Each of these points are in addition to the limitations I discuss in Chapter 4 itself.

First, we need to understand better how the interactive effect of climate change and land use differs between wild and managed pollinators, since the resilience of managed pollinators will likely predict a locality's ability to adapt to wild pollinator losses. In the future, as localities lose wild pollinator abundance through the interactive effects of climate change and land use, there will likely be an increased demand and incentive for shipping in pollinators (vanEngelsdorp & Meixner 2010). If the response for managed pollinators (e.g. *Apis*, *Bombus*, *Osmia*, *Megachile*) is similar to all pollinators (i.e. farmers are unable to maintain managed bee abundance in croplands that are experiencing extreme high novel temperatures), shipping in these bees will likely experience a diminishing return, such that production of the crop becomes more expensive, and at worst economically unviable. But if the response of domesticated pollinators differs (i.e. farmers are able to maintain managed bee abundance), it may be the case that crop pollination can be buffered against the effects of wild pollinator losses, albeit with some increase in cost associated with managed bee shipping. Here I did not distinguish between wild and managed pollinators in the development of my models. Future work needs to tease these two groups apart for a more refined understanding of our ability to adapt to changing crop pollination risk.

Second, we need higher quality and more recent estimates of global crop production. Monfreda et al. (2008) is a valuable resource, but the data represents crop production from over 20 years ago, and relies on country-level interpolations for many highly pollination-dependent crops (e.g. cocoa). To understand better how these crops will be affected, we need better estimates of where they tend to grow and at what relative quantity. To be of use to crop pollination risk models, these estimates need to be at the level of spatial cells. Climate change

and land use do not occur at the spatial scale of countries, meaning FAO national reports alone are of minimal use, especially for large countries (e.g. Brazil, China, India).

Third, to better understand how localised shortfalls might affect globalised supply of crops dependent on animal pollination, we need to link models of localised crop pollination risk to estimates of global trade networks. Virtual pollination trade flows (see Silva et al. (2021)) represent a means through which this could be achieved. Silva et al. (2021) released data for the flow of animal-pollination-dependent crop production between countries, with quantities for the flow between each exporter and its corresponding importers. Each metric tonne of production risk I describe in Chapter 4 (see Figure 3) has a country in which it was produced (i.e. export risk), which can be ascribed to an importer (i.e. import risk) according to the proportional flow reported by Silva et al. (2021). Such work would provide a coarse estimate as to how many countries could be indirectly affected by changes in crop pollination risk, and to what extent. One might expect that countries that have low export production risk (e.g. the United Kingdom) could have high import production risk. Given the relationship between supply diversity and food-shock incidence (Gomez et al. 2021), countries of high import production risk that import from only a few countries will likely be most vulnerable to the knock-on consequences of localised pollinator losses.

Developing online metrics for pollinator biodiversity awareness

Contribution

Pollinator biodiversity has undoubtedly undergone significant change through the effects of climate change and anthropogenic land use. In regions of high agricultural dependence on pollinators, such change in agricultural intensity and climate change could have consequences for crop pollination. Although the potential for these effects has been well publicised (Carrington 2013, Milman 2020), it is unclear the extent to which publicity has translated into changes in awareness of the importance of biodiversity and pollinators. In Chapter 5, I developed a new approach for quantifying changes in awareness of biodiversity, which I use

to show that awareness of biodiversity has likely increased, but not awareness of pollinating species. I then suggest an approach through which this metric could be combined with the Biodiversity Engagement Indicator (Cooper et al. 2019).

Challenges and future directions

My work provides a novel means through which to quantify changes in biodiversity awareness, and therefore pollinator biodiversity. Regardless, the field still has a long way to go before it can claim to make broad, robust conclusions that are of genuine use to policy-makers. Here I highlight some of the problems associated with quantifying biodiversity awareness more broadly, before discussing some potential areas of research and concern in the context of pollinator biodiversity specifically.

To understand how attempts to quantify changes in global biodiversity awareness might move forward, it is first helpful to understand their current state. At present there are effectively two main philosophies relating to the way in which culturomics—the quantitative study of human culture and behaviour using digitised data—can be used to infer biodiversity awareness: in the first, the frequency of biodiversity and conservation related key words online are used to try and infer directly awareness of issues (e.g. Cooper et al. 2019); in the second, awareness is quantified at the level of species through either online page views (e.g. the Species Awareness Index in Chapter 5) or a frequency of species names in text (e.g. Ladle et al. 2019) from which awareness of the importance of biodiversity is indirectly inferred. For example, a greater rate of change in page views for ecosystem service providing species, relative to non-service providers, might suggest that awareness of the importance of biodiversity is increasing. The first of these approaches (i.e. keyword level inference) works well across platforms at scale, but is disconnected from biodiversity itself, and tracks a set of words that are not in the lexicon of most people. Or in other words, it tracks changes in awareness for individuals that tend to use these words already, which will more often be those working in the biodiversity sciences. The second of these approaches (i.e. species level inference) overcomes this problem by focussing on biodiversity itself, but for page views is limited due to the difficulty in

understanding why a particular page was visited. As I discuss in Chapter 5, owners of websites can retrieve visit statistics for their own site through services such as Google Analytics, which intimate as to the reason for a given visit, but such data are unavailable for Wikipedia. For example, Google Analytics can show which other sites tend to be used to reach a site, and for how long a site is visited (Soriano-Redondo et al. 2017).

Although understanding the reason for a given species page view on Wikipedia is challenging, recent research indicates that one of the core drivers is physical encounters with species (Mittermeier et al. 2021). For example, Mittermeier et al. (2021) showed that for a given Wikipedia language, species found within countries of that language tend to get more views, and that the frequency of sightings reported online explained a significant quantity of the variation in pageviews. An interesting complement to my work in Chapter 5, therefore, could be an index of change in awareness derived from the frequency of animal species photos uploaded online. A photo uploaded online, assuming it is of an organism which can be accurately identified, represents an unambiguous physical encounter with that species. Each photo has a timestamp indicating the approximate period of encounter, and typically 25% are geotagged such that the location of the species can be resolved to an unambiguous location (August et al. 2020). In combination with the Species Awareness Index, an online photo index would help to establish the extent to which change in species page views are predicted by physical encounters with biodiversity. Such an index could be built using an approach similar to August et al. (2020), in which photos are pulled from Flickr, and then identified using a species identifying algorithm such as PlantNet. Developing a species identifying algorithm for insects will be challenging, however with advances in technology (e.g. Buschbacher et al. 2020) it may be that such an approach is feasible in the future. In the manner of the Species Awareness Index (Chapter 5), the frequency of identified species could then be converted to a rate of change, adjusted for the background rate of change in uploaded photos, and then bootstrapped for groups of species.

In tackling pollinator biodiversity change, and biodiversity change more generally, there is also the question of how changes in online awareness relate to changes in behaviour in the physical world. Given that the link between awareness and behaviour can be weak (Asvatourian et al. 2018), it may be that aggregate changes in online awareness are poorly correlated with aggregate changes in behaviour. Assuming this is the case, conservation culturomics researchers need to develop means through which behaviours can be tracked online as well as awareness. This represents a much tougher challenge, since behaviours relevant to pollinator biodiversity change more often leave only a minimal trace online. The localised application of pesticides or fertiliser (e.g. in urban green spaces), for example, or the destruction of wild flower habitats, do not as far as I know leave a digital footprint. One proxy for these sorts of behaviour, however, could be online purchases. An online purchase represents an unambiguous action, which depending on the nature of the purchase, could be defined as either a negative or positive action on behalf of pollinator biodiversity. For example, the purchase of weed-killers or insecticides might indicate negative actions on behalf of pollinator biodiversity, whereas purchases of bee hotels or wild flower seed mixes positive actions. Purchase data is not made available by companies such as Amazon, but it could potentially be inferred from the frequency of reviews. Such an approach would require a robust methodology for filtering out fake reviews, but it could be a fruitful approach for quantifying large-scale changes in pollinator beneficial behaviours, and how these changes relate to awareness and conservation campaigns.

More broadly for awareness of pollinator biodiversity change, I think the field of pollinator and insect conservation needs to challenge the language it uses, and think carefully about the way in which this language is interpreted by the online and print media. I feel that the language used, particularly in press releases and in some papers, will likely have caused some confusion for members of the public with respect to insect pollinator conservation. Such confusion is most recently exemplified by Sánchez-Bayo & Wyckhuys (2019) and the subsequent media coverage following this paper. Sánchez-Bayo & Wyckhuys (2019) carried

out a review of the evidence of insect declines, drawing some quite apocalyptic conclusions, such as that “insects as a whole will go down the path of extinction in a few decades”. This paper was picked up by the Guardian, who ran with the headline “Plummeting insect numbers threaten collapse of nature”, and the tagline “insects could vanish within a century at current rate of decline” (Carrington 2019). Sánchez-Bayo & Wyckhuys (2019) has since been subject to multiple critiques regarding its methodology and language (Komonen et al. 2019, Simmons et al. 2019, Thomas et al. 2019). In short, the conclusions it draws, and much of the subsequent media coverage, are not supported by the breadth of evidence. Pollinator biodiversity is clearly highly important, but if apocalyptic language is used when it is not warranted, the public will not pay attention when it is. Perhaps more importantly, if and when apocalyptic language is warranted, it needs to acknowledge better the heterogeneity of risks to both pollinator biodiversity change and ecosystem services. In reality some geographic regions and some taxonomic groups will be more susceptible to environmental changes, and only some crops at higher risk in consequence.

Conclusion

The Earth’s biosphere is changing rapidly. Animal pollinators have not escaped this change. Climate change and anthropogenic land use are largely responsible, driving significant shifts in pollinator biodiversity. The work I present in this thesis provides a significant advance in our understanding of the causes of pollinator biodiversity change, and the potential consequences for crop pollination risk. I also contribute towards the developing field of conservation culturomics, introducing a new metric of public biodiversity awareness derived from Wikipedia page views. Such metrics will likely become increasingly useful in the future as internet penetration increases and pollinator biodiversity continues to experience rapid change. Over the coming decades the consequences of climate change and land use for pollinator biodiversity will grow, and so too will risk for the ecosystem service these animals provide. These risks are big, but they’re not insurmountable. If carbon emissions and intensive agriculture can be kept in check, and if agriculture shifts such that it avoids the more severe

effects of climate change, then humanity's ability to adapt will be greater. Life on Earth will undoubtedly look quite different by 2050. But I take solace in the hope that, however fast biodiversity changes, the ingenuity and generosity of people will change faster.

References

- Acerbi, A. et al. (2020). The impact of the “World's 25 Most Endangered Primates” list on scientific publications and media. *Journal for Nature Conservation*, 54(125794).
- Ackerman, J.D. (2000). Abiotic pollen and pollination: ecological, functional, and evolutionary perspectives. *Pollen and Pollination*. 222: 167–185.
- Aguirre-Gutiérrez, J. et al. (2016). Functional traits help to explain half-century long shifts in pollinator distributions. *Scientific Reports* 6:1–13.
- Aizen, M.A. and Harder, L.D. (2009). The Global Stock of Domesticated Honey Bees Is Growing Slower Than Agricultural Demand for Pollination. *Current Biology* 19:915–918.
- Aizen, M.A. et al. (2008). Long-Term Global Trends in Crop Yield and Production Reveal No Current Pollination Shortage but Increasing Pollinator Dependency. *Current Biology* 18:1572–1575.
- Aizen, M.A. et al. (2009). How much does agriculture depend on pollinators? Lessons from long-term trends in crop production. *Annals of Botany* 103:1579–1588.
- Aizen, M.A. et al. (2019). Global agricultural productivity is threatened by increasing pollinator dependence without a parallel increase in crop diversification. *Global Change Biology* 25:3516–3527.
- Akella, L. et al. (2012). NetiNeti: discovery of scientific names from text using machine learning methods. *BMC Bioinformatics* 13:211.
- Amano, T. et al. (2016). Languages Are Still a Major Barrier to Global Science. *PLOS Biology* 14(12): e2000933.
- Archer, C.R. et al. (2014). Economic and ecological implications of geographic bias in pollinator ecology in the light of pollinator declines. *Oikos* 123:401–407.
- Asvatourian, V. et al. (2018). Relationship between pro-environmental attitudes and behaviour and dietary intake patterns. *Sustainable Production and Consumption* 16:216–226.
- August, T.A. et al. (2020). AI Naturalists Might Hold the Key to Unlocking Biodiversity Data in Social Media Imagery. *Patterns* 1:100116.
- Baldock, K.C.R. et al. (2015). Where is the UK's pollinator biodiversity? The importance of urban areas for flower-visiting insects. *Proceedings of the Royal Society B: Biological Sciences* 282(1803).
- Baldock, K.C.R. et al. (2019). A systems approach reveals urban pollinator hotspots and conservation opportunities. *Nature Ecology and Evolution* 3:363–373.
- Balmford, A. (1996). Extinction filters and current resilience: the significance of past selection pressures for conservation biology. *Trends Ecol. Evol.* 1:193–196.

-
- Barlow, J. et al. (2018). The future of hyperdiverse tropical ecosystems. *Nature* 559:517–526.
- Barrett, S.C.H. (2010). Understanding plant reproductive diversity. *Philosophical Transactions of the Royal Society* 365:99–109.
- Bartomeus, I. et al. (2017). On the inconsistency of pollinator species traits for predicting either response to land-use change or functional contribution. *Oikos* 127:306–315.
- Bartomeus, I. et al. (2018). Historical collections as a tool for assessing the global pollination crisis. *Phil. Trans. R. Soc. B* 374: 1763
- Beaman, R.S. and Conn, B.J. (2003). Automated geoparsing and georeferencing of Malesian collection locality data. *Telopea* 10:43–52.
- Beckmann, M. and von Wehrden, H. (2012). Where you search is what you get: literature mining - Google Scholar versus Web of Science using a data set from a literature search in vegetation science. *Journal of Vegetation Science* 23:1197–1199.
- Bellard, C. et al. (2012). Impacts of climate change on the future of biodiversity. *Ecology Letters* 15:365–377.
- Benton, T.G. et al. (2002). Linking agricultural practice to insect and bird populations: a historical study over three decades. *Journal of Applied Ecology* 39:673–687.
- Benton, T.G. et al. (2003). Farmland biodiversity: is habitat heterogeneity the key? *Trends in Ecology & Evolution* 18:182–188.
- Bessa-Gomes, C. et al. (2004). Allee effects, mating systems and the extinction risk in populations with two sexes. *Ecology Letters* 7:802–812.
- Biesmeijer, J.C. et al. (2006). Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science* 313:351–354.
- Blaauw, B.R. and Isaacs, R. (2014). Flower plantings increase wild bee abundance and the pollination services provided to a pollination-dependent crop. *Journal of Applied Ecology* 51:890–898.
- Blitzer, E.J. et al. (2016). Pollination services for apple are dependent on diverse wild bee communities. *Agriculture, Ecosystems and Environment* 221:1–7.
- Blowes, S.A. et al. (2019). The geography of biodiversity change in marine and terrestrial assemblages. *Science* 366:339–345.
- Boggs, C.L. (2016). The fingerprints of global climate change on insect populations. *Current Opinion in Insect Science* 17:69–73.
- Bommarco, R. et al. (2014). Extinction debt for plants and flower-visiting insects in landscapes with contrasting land use history *Diversity and Distributions* 20:591–599.
- Bommarco, R. et al. (2010). Dispersal capacity and diet breadth modify the response of wild bees to habitat loss. *Proceedings of the Royal Society B: Biological Sciences*, 277(1690): 2075–2082.

-
- Brittain, C.A. et al. (2010). Impacts of a pesticide on pollinator species richness at different spatial scales. *Basic and Applied Ecology* 11:106–115.
- Brown, M.J.F. et al. (2016). A horizon scan of future threats and opportunities for pollinators and pollination. *PeerJ* 4:e2249.
- Burivalova, Z. et al. (2015). Avian responses to selective logging shaped by species traits and logging practices. *Proc. R. Soc. B* 282:20150164.
- Burkle, L.A. et al. (2013). Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. *Science* 339:1611–5.
- Buschbacher, K. et al. (2020). Image-based species identification of wild bees using convolutional neural networks. *Ecological Informatics* 55:101017.
- Butchart, S.H.M. et al. (2010). Global biodiversity: indicators of recent declines. *Science* 328:1164–8.
- Cameron, S.A. et al. (2011). Patterns of widespread decline in North American bumble bees. *Proceedings of the National Academy of Sciences* 108:662–667.
- Carrington, D. (2013). Loss of wild pollinators serious threat to crop yields, study finds. *The Guardian* [Online]. Available at: <https://www.theguardian.com/environment/2013/feb/28/wild-bees-pollinators-crop-yields> [Accessed: 12 August 2021].
- Carrington, D. (2019). Plummeting insect numbers ‘threaten collapse of nature’. *The Guardian* [Online]. Available at: <https://www.theguardian.com/environment/2019/feb/10/plummeting-insect-numbers-threaten-collapse-of-nature> [Accessed: 12 August 2021].
- Ceballos, G. et al. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances* 1:e1400253.
- Chamberlain, S.A. and Szöcs, E. (2013). taxize: taxonomic search and retrieval in R. *F1000Research* 2:191.
- Chao, A. et al. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology letters*, 8(2): 148–159.
- Chen, I.-C. et al. (2011). Rapid range shifts of species associated with high levels of climate warming. *Science* 333:1024–6.
- Chiang, Y.-Y. (2017). Unlocking Textual Content from Historical Maps - Potentials and Applications, Trends, and Outlooks. In: International Conference on Recent Trends in Image Processing and Pattern Recognition. *Springer* 111–124.
- Claus, G. et al. (2018). Challenges in Cocoa Pollination: The Case of Côte d’Ivoire. In: Pollination in Plants. InTech.
- Clough, Y. et al. (2014). Density of insect-pollinated grassland plants decreases with increasing surrounding land-use intensity. *Ecology Letters* 17:1168–1177.
- Cohen, K.B. and Hunter, L. (2008). Getting Started in Text Mining. *PLoS Computational Biology* 4(1): e20.

-
- Collen, B. (2009). Monitoring change in vertebrate abundance: the Living Planet Index. *Conservation Biology*, 23(2), pp 317–327.
- Colwell, R.K. et al. (2012). Coextinction and Persistence of Dependent Species in a Changing World. *Annual Review of Ecology, Evolution, and Systematics* 43:183–203.
- Connell, J.H. (1978). Diversity in Tropical Rain Forests and Coral Reefs. *Science* 199:1302–1310.
- Cooper, M.W. et al. (2019). Developing a global indicator for Aichi Target 1 by merging online data sources to measure biodiversity awareness and engagement. *Biological Conservation* 230:29–36.
- Cornford, R. et al. (2021). Fast, scalable, and automated identification of articles for biodiversity and macroecological datasets. *Global Ecology and Biogeography* 30:339–347.
- Cornford, R. et al. (in prep). Automated Synthesis of Biodiversity Knowledge Requires Better Tools and Standardised Research Output.
- Correia, R.A. et al. (2017). Internet scientific name frequency as an indicator of cultural salience of biodiversity. *Ecological Indicators* 78:549–555.
- Correia, R.A. et al. (2018). Nomenclature instability in species culturomic assessments: Why synonyms matter. *Ecological Indicators* 90:74–78.
- Crall, J.D. et al. (2018). Neonicotinoid exposure disrupts bumblebee nest behavior, social networks, and thermoregulation. *Science* 362:683–686.
- Cusser, S., Neff, J.L. and Jha, S. (2015). Land use change and pollinator extinction debt in exurban landscapes. *Insect Conservation and Diversity* 8:562–572.
- Cusser, S., Neff, J.L. and Jha, S. (2018). Land-use history drives contemporary pollinator community similarity. *Landscape Ecology* 33:1335–1351.
- D'Ignazio, C. et al. (2014). CLIFF-CLAVIN: Determining geographic focus for news articles. NewsKDD: Data Science for News Publishing, at KDD, 2014.
- da Silva, A.F. et al. (2020). Taxonomic bias in amphibian research: Are researchers responding to conservation need? *Journal for Nature Conservation* 56:125829.
- Dainese, M. et al. (2019). A global synthesis reveals biodiversity-mediated benefits for crop production. *Science Advances* 5:eaax0121.
- Dalsgaard, B. (2020). Land-Use and Climate Impacts on Plant–Pollinator Interactions and Pollination Services. *Diversity* 12:168.
- Daniel, T.C. et al. (2012). Contributions of cultural services to the ecosystem services agenda. *Proceedings of the National Academy of Sciences*, 109(23):8812–8819.
- De Palma, A. et al. (2015). Ecological traits affect the sensitivity of bees to land-use pressures in European agricultural landscapes. *Journal of Applied Ecology* 52:1567–1577.

-
- De Palma, A. et al. (2016). Predicting bee community responses to land-use changes: Effects of geographic and taxonomic biases. *Scientific Reports* 6:31153.
- De Palma, A. et al. (2018). Challenges with inferring how land-use affects terrestrial biodiversity: Study design, time, space and synthesis. *Advances in ecological research*. 58: 163-199
- Deans, A.M. et al. (2007). Hoverfly (Syrphidae) Communities Respond to Varying Structural Retention After Harvesting in Canadian Peatland Black Spruce Forests. *Environmental Entomology* 36:308–318.
- DeFries, R. and Rosenzweig, C., (2010). Toward a whole-landscape approach for sustainable land use in the tropics. *Proceedings of the National Academy of Sciences*. 107(46):19627-19632.
- Degrandi-Hoffman, G. et al. (2019). The Economics of Honey Bee (Hymenoptera: Apidae) Management and Overwintering Strategies for Colonies Used to Pollinate Almonds. *Journal of Economic Entomology* 112:2524–2533.
- Deutsch, C. A. et al. (2008). Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences* 105.18: 6668-6672.
- Dicks, L.V. et al. (2021). A global-scale expert assessment of drivers and risks associated with pollinator decline. *Nature Ecology and Evolution* 5:1453–1461
- Didham, R.K. et al. (2007). Interactive effects of habitat modification and species invasion on native species decline. *Trends in Ecology & Evolution* 22:489–496.
- Didham, R.K. et al. (2020). Interpreting insect declines: seven challenges and a way forward. *Insect Conservation and Diversity* 13:103–114.
- Dirzo, R. et al. (2014). Defaunation in the Anthropocene. *Science* 345:401–406.
- Donald, P.F. et al. (2001). Agricultural intensification and the collapse of Europe's farmland bird populations. *Proc. R. Soc. Lond. B*. 268:25–29.
- Dornelas, M. et al. (2014). Assemblage time series reveal biodiversity change but not systematic loss. *Science* 344:296–299.
- Dukas, R. and Morse, D.H. (2003). Crab spiders affect flower visitation by bees. *Oikos* 101(1): 157–163.
- Dutton, E.M. and Frederickson, M.E. (2012). Why ant pollination is rare: New evidence and implications of the antibiotic hypothesis. *Arthropod-Plant Interactions* 6:561–569.
- Easton, A.H. and Goulson, D. (2013). The Neonicotinoid Insecticide Imidacloprid Repels Pollinating Flies and Beetles at Field-Realistic Concentrations. *PLoS ONE* 8:e54819.
- Eilers, E.J. et al. (2011). Contribution of Pollinator-Mediated Crops to Nutrients in the Human Food Supply. *PLOS ONE* 6:e21363.
- Elbasiouny, H. et al. (2014). Spatial variation of soil carbon and nitrogen pools by using ordinary Kriging method in an area of north Nile Delta, Egypt. *Catena* 113:70–78.

-
- Ellis, E.C. and Ramankutty, N. (2008). Putting people in the map: anthropogenic biomes of the world. *Frontiers in Ecology and the Environment* 6:439–447.
- Embury-Dennis, T. (2017) Scientists warn of 'ecological Armageddon' after study shows flying insect numbers plummeting 75%. *The Independent*.
<http://www.independent.co.uk/news/science/flying-insects-numbers-drop-ecological-armageddon-75-per-cent-plummet-a8008406.html> [Accessed 03 December 2020].
- Endress, P.K. (2001). The Flowers in Extant Basal Angiosperms and Inferences on Ancestral Flowers. *International Journal of Plant Sciences* 162:1111–1140.
- European Commission (2018). EU Pollinators Initiative.
- FAO (2018). The International Pollinator Initiative plan of action 2018-2030.
- Fernández-Bellón, D., & Kane, A. (2020). Natural history films raise species awareness—A big data approach. *Conservation Letters*. 13(1):e12678.
- Ferreira, C. et al. (2015). The evolution of peer review as a basis for scientific publication: directional selection towards a robust discipline? *Biol. Rev* 91:3
- Fink, C., Hausmann, A. and Di Minin, E., (2020). Online sentiment towards iconic species. *Biological Conservation*, 241:108289.
- Fitter, A.H. and Fitter, R.S.R. (2002). Rapid changes in flowering time in British plants. *Science* 296:1689–1691.
- Fleming, T.H. and Muchhala, N. (2008). Nectar-feeding bird and bat niches in two worlds: pantropical comparisons of vertebrate pollination systems. *Journal of Biogeography* 35:764–780.
- Fleming, T.H. et al. (2009). The evolution of bat pollination: a phylogenetic perspective. *Annals of Botany* 104:1017–1043.
- Food and Agriculture Organization (FAO), (2020). ASFIS list of species for fishery statistics purposes.
- Frieler, K. et al. (2017). Assessing the impacts of 1.5 • C global warming-simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b). *Geosci. Model Dev* 10:4321–4345.
- Frishkoff, L.O. et al. (2016). Climate change and habitat conversion favour the same species. *Ecology Letters* 19:1081–1090.
- Frishkoff, L.O. et al. (2018). Do correlated responses to multiple environmental changes exacerbate or mitigate species loss? *Oikos* 127:1724–1734.
- Fukano, Y. et al. (2020). Zoos and animated animals increase public interest in and support for threatened animals. *Science of The Total Environment*, 704:135352.
- Gallai, N. et al. (2009). Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics* 68:810–821.

-
- Garibaldi, L.A. et al. (2009). Pollinator shortage and global crop yield. <http://www.tandfonline.com/action/authorSubmission?journalCode=kcib20&page=instructions> 2:37–39.
- Garibaldi, L.A. et al. (2011). Global growth and stability of agricultural yield decrease with pollinator dependence. *Proceedings of the National Academy of Sciences* 108:5909–5914.
- Garibaldi, L.A. et al. (2013). Wild pollinators enhance fruit set of crops regardless of honey bee abundance. *Science* 340:1608–1611.
- Garibaldi, L.A. et al. (2016). Mutually beneficial pollinator diversity and crop yield outcomes in small and large farms. *Science* 351:388–391.
- Garnett, T. (2013). Food sustainability: problems, perspectives and solutions. *Proceedings of the Nutrition Society* 72:29–39.
- Garratt, M.P.D. et al. (2014). Avoiding a bad apple: Insect pollination enhances fruit quality and economic value. *Agriculture, Ecosystems and Environment* 184:34–40.
- Gathmann, A. and Tscharntke, T. (2002). Foraging ranges of solitary bees. *Journal of Animal Ecology* 71:757–764.
- Genersch, E. et al. (2010). The German bee monitoring project: a long term study to understand periodically high winter losses of honey bee colonies. *Apidologie* 41:332–352.
- Ghazoul, J. (2005). Buzziness as usual? Questioning the global pollination crisis. *Trends in Ecology & Evolution* 20:367–373.
- Ghazoul, J. (2015). Qualifying pollinator decline evidence. *Science* 348:981–2.
- Gomez, M. et al. (2021). Supply chain diversity buffers cities against food shocks. *Nature* 595(7866):250–254.
- Gonzalez, A. et al. (2016). Estimating local biodiversity change: a critique of papers claiming no net loss of local diversity. *Ecology* 97:1949–1960.
- Goodrich, B.K. (2019). Do more bees imply higher fees? Honey bee colony strength as a determinant of almond pollination fees. *Food Policy* 83:150–160.
- Goulson, D. (2003). Conserving wild bees for crop pollination. *Journal of Food Agriculture and Environment* 1:142–144.
- Goulson, D. (2013). An overview of the environmental risks posed by neonicotinoid insecticides. *Journal of Applied Ecology* 50:977–987.
- Goulson, D. et al. (2005). Causes of rarity in bumblebees. *Biological Conservation* 122:1–8.
- Goulson, D. et al. (2015). Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* 347:6229.
- Goulson, D. et al. (2008). Decline and Conservation of Bumble Bees. *Annual Review of Entomology* 53:191–208.

-
- Graham, M. et al. (2014). Uneven geographies of user-generated information: patterns of increasing informational poverty. *Annals of the Association of American Geographers*, 104(4):746–764.
- Griffiths, T.L. et al. (2004). Finding scientific topics. *PNAS* 101:5228–5235.
- Grimmer, J. and Stewart, B.M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* 21:267–297.
- Gritta, M. et al. (2018). What's missing in geographical parsing? *Language Resources and Evaluation* 52:603–623.
- Guralnick, R. and Hill, A. (2009). Biodiversity informatics: automated approaches for documenting global biodiversity patterns and processes. *Bioinformatics* 25:421–428.
- Gurevitch, J. et al. (2018). Meta-analysis and the science of research synthesis. *Nature* 555:175–182.
- Haddaway, N.R. and Westgate, M.J. (2018). Predicting the time needed for environmental systematic reviews and systematic maps. *Conservation Biology* 33:434–443.
- Haenke, S. et al. (2009). Increasing syrphid fly diversity and density in sown flower strips within simple vs. complex landscapes. *Journal of Applied Ecology* 46:1106–1114.
- Hagen, M. and Kraemer, M. (2010). Agricultural surroundings support flower–visitor networks in an Afrotropical rain forest. *Biological Conservation* 143:1654–1663.
- Hahn, M. and Brühl, C.A. (2016). The secret pollinators: an overview of moth pollination with a focus on Europe and North America. *Arthropod-Plant Interactions* 10:21–28.
- Hall, D.M. et al. (2017) The city as a refuge for insect pollinators. *Conservation Biology* 31: 24–29
- Hallmann, C. A. et al. (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLOS ONE*, 12(10):e0185809.
- Halsch, C.A. et al. (2021). Insects and recent climate change. *Proceedings of the National Academy of Sciences of the United States of America* 118:2002543117.
- Hamel, R.E. (2008). The dominance of English in the international scientific periodical literature and the future of language use in science. *AILA Review* 20:53–71.
- Han, F. et al. (2012). From where did the Western honeybee (*Apis mellifera*) originate? *Ecology and Evolution* 2:1949–1957.
- Hansen, M.C. et al. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160):850-853.
- Harlan, J. (1992). Crops and Man. *American Society of Agronomy*.
- Harris, I. et al. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data* 7:1–18.
- Hayter, K.E. and Cresswell, J.E. (2006). The influence of pollinator abundance on the dynamics and efficiency of pollination in agricultural *Brassica napus* : implications for landscape-scale gene dispersal. *Journal of Applied Ecology* 43:1196–1202.

-
- Hendershot, J.N. et al. (2020). Intensive farming drives long-term shifts in avian community composition. *Nature* 579:393–396.
- Henle, K. et al. (2004). Predictors of species sensitivity to fragmentation. *Biodiversity & Conservation*, 13(1): 207-251.
- Holzschuh, A. et al. (2012). Landscapes with wild bee habitats enhance pollination, fruit set and yield of sweet cherry. *Biological Conservation* 153:101–107.
- Høye, T. et al. (2013) Shorter flowering seasons and declining abundance of flower visitors in a warmer Arctic. *Nature Climate Change* 3: 759–763.
- Hudson, L.N. et al. (2014). The PREDICTS database: A global database of how local terrestrial biodiversity responds to human impacts. *Ecology and Evolution* 4:4701–4735.
- Hudson, L.N. et al. (2017). The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. *Ecology and Evolution* 7:145–188.
- Hurt, G. C. et al. (2011). Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Clim. Change* 109:117–161
- ICCO (2014). ICCO (International Cocoa Organization) Annual Report. Abidjan.
- Imani, M.B. et al. (2017). Focus location extraction from political news reports with bias correction. In: 2017 IEEE International Conference on Big Data (BIGDATA). 1956–1964.
- IPBES (2016). The assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination and food production. S.G. Potts, V. L. Imperatriz-Fonseca, and H. T. Ngo (eds). Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Bonn, Germany.
- IPBES (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo (eds). IPBES secretariat, Bonn, Germany.
- IPCC (2018). Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5° above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.
- IUCN (2021). The IUCN Red List of Threatened Species [Online]. Available at: www.iucnredlist.org.
- Jarić, I. et al. (2020). Societal attention toward extinction threats: a comparison between climate change and biological invasions. *Scientific Reports* 10(1):1-9.

- Jauker, F. et al. (2009). Pollinator dispersal in an agricultural matrix: Opposing responses of wild bees and hoverflies to landscape structure and distance from main habitat. *Landscape Ecology* 24:547–555.
- Johansson, F. et al. (2020). Temperate insects with narrow seasonal activity periods can be as vulnerable to climate change as tropical insect species. *Scientific Reports* 10:1–8.
- Jones, R.O. and Chapman, S.K. (2011). The roles of biotic resistance and nitrogen deposition in regulating non-native understory plant diversity. *Plant and Soil* 345:257–269.
- Joppa, L.N. et al. (2016). Filling in biodiversity threat gaps: Only 5% of global threat data sets meet a 'gold standard'. *Science* 352:416–418.
- Joppa, L.N. et al. (2011). How many species of flowering plants are there? *Proc. R. Soc. B.* 278:554–9.
- Kennedy, C.M. et al. (2013). A global quantitative synthesis of local and landscape effects on wild bee pollinators in agroecosystems. *Ecology Letters*. 16:584–599.
- Kerr, J.T. (2001). Butterfly Species Richness Patterns in Canada: Energy, Heterogeneity, and the Potential Consequences of Climate Change. *Conservation Ecology* 5(1)
- Kerr, J.T. et al. (2015). Climate change impacts on bumblebees converge across continents. *Science* 349:177–80.
- Kessler, A. and Baldwin, I.T. (2002). *Manduca quinquemaculata's* optimization of intra-plant oviposition to predation, food quality, and thermal constraints. *Ecology* 83:2346–2354.
- Kevan, P.G. (1977). Blueberry crops in Nova Scotia and New Brunswick—pesticides and crop reductions. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 25:61–64.
- Khanal, U. et al. (2018). Farmers' Adaptation to Climate Change, Its Determinants and Impacts on Rice Yield in Nepal. *Ecological Economics* 144:139–147.
- Kleijn, D. et al. (2009). On the relationship between farmland biodiversity and land-use intensity in Europe. *Proceedings of the Royal Society B: Biological Sciences* 276:903–909.
- Kleijn, D. et al. (2015). Delivery of crop pollination services is an insufficient argument for wild pollinator conservation. *Nature Communications* 6:36.
- Klein A.-M. et al. (2007) Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society of London series B* 274:303-313.
- Klein, A.-M. et al. (2003). Fruit set of highland coffee increases with the diversity of pollinating bees. *Proceedings of the Royal Society of London: Biological Sciences* 270(1518):955–961
- Klink, R. van et al. (2020). Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science* 368:417–420.

- Komonen, A. et al. (2019). Alarmist by bad design: Strongly popularized unsubstantiated claims undermine credibility of conservation science. *Rethinking Ecology* 4:17.
- Kosior, A. et al. (2007). The decline of the bumble bees and cuckoo bees (Hymenoptera: Apidae: Bombini) of Western and Central Europe. *Oryx* 41:79.
- Kremen, C. et al. (2002). Crop pollination from native bees at risk from agricultural intensification. *Proceedings of the National Academy of Sciences of the United States of America* 99:16812–16816.
- Kühnel, S. and Blüthgen, N. (2015). High diversity stabilizes the thermal resilience of pollinator communities in intensively managed grasslands. *Nature Communications* 6:1–10.
- Kunin, W.E. (2019). Robust evidence of declines in insect abundance and biodiversity. *Nature* 574:641–642.
- Kuussaari, M. et al. (2007). Determinants of local species richness of diurnal Lepidoptera in boreal agricultural landscapes. *Agriculture, Ecosystems & Environment* 122:366–376.
- Labandeira, C.C. et al. (2007). Pollination drops, pollen, and insect pollination of Mesozoic gymnosperms. *International Association for Plant Taxonomy* 56:663–695.
- Ladle, R.J. et al. (2016). Conservation culturomics. *Frontiers in Ecology and the Environment* 14:269–275.
- Ladle, R.J. et al. (2019). A culturomics approach to quantifying the salience of species on the global internet. *People and Nature* 1:524–532.
- Lajeunesse, M.J. (2016). Facilitating systematic reviews, data extraction and meta-analysis with the METAGEAR package for R. *Methods in Ecology and Evolution* 7:323–330.
- Laurance, W.F. et al. (2014). Agricultural expansion and its impacts on tropical nature. *Trends in Ecology and Evolution* 29:107–116.
- Lautenbach, S. et al. (2012). Spatial and temporal trends of global pollination benefit. *PLoS ONE*, 7(4):e35954.
- Lavoie, M.-C. et al. (2014). Devices for preventing percutaneous exposure injuries caused by needles in healthcare personnel. *Cochrane Database of Systematic Reviews* 3:1–76.
- Lázaro, A. et al. (2016). Effects of grazing intensity on pollinator abundance and diversity, and on pollination services. *Ecological Entomology* 41:400–412.
- Lázaro, A. et al. (2016). Effects of grazing intensity on pollinator abundance and diversity, and on pollination services. *Ecological entomology* 41(4):400–412.
- Le B. Hooke, R. et al. (2012). Land transformation by humans: A review. *GSA today* 22(12):4–10
- Leadley, P.W. et al. (2013). Progress towards the Aichi Biodiversity Targets: An Assessment of Biodiversity Trends, Policy Scenarios and Key Actions. Montreal.

-
- Leary, P.R. et al. (2007). uBioRSS: Tracking taxonomic literature using RSS. *Bioinformatics* 23:1434–1436.
- Lee, H. et al. (2019). Pollination Markets and the Coupled Futures of Almonds and Honey Bees: Simulating Impacts of Shifts in Demands and Costs. *American Journal of Agricultural Economics* 101:230–249.
- Leidner, J.L. and Lieberman, M.D. (2011). Detecting geographical references in the form of place names and associated spatial natural language. *Sigspatial Special* 3:5–11.
- Lenda, M. et al. (2020). Effects of amusing memes on concern for unappealing species. *Conservation Biology* 34(5):1200–1209.
- Leung, B. et al. (2020). Clustered versus catastrophic global vertebrate declines. *Nature* 588:267–271.
- Lever, J.J. et al. (2014). The sudden collapse of pollinator communities. *Ecology Letters* 17:350–359.
- Lewis, S.L. and Maslin, M.A. (2015). Defining the Anthropocene. *Nature* 519:171–180.
- Lister, B.C. and Garcia, A. (2018). Climate-driven declines in arthropod abundance restructure a rainforest food web. *Proceedings of the National Academy of Sciences of the United States of America* 115:E10397–E10406.
- Lobell, D.B. et al. (2008). Prioritizing Climate Change Adaptation Needs for Food Security in 2030. *Science* 319:607–610.
- Loh, J. et al. (2005). The Living Planet Index: using species population time series to track trends in biodiversity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 360(1454):289–295.
- Lortie, C.J. (2014). Formalized synthesis opportunities for ecology: systematic reviews and meta-analyses. *Oikos* 123:897–902.
- Mace, G. M. et al. (2018). Aiming higher to bend the curve of biodiversity loss. *Nature Sustainability* 1(9):448–451.
- Macgregor, C.J. et al. (2019). Moth biomass increases and decreases over 50 years in Britain. *Nature Ecology and Evolution* 3:1645–1649.
- Machovina, B. et al. (2015). Biodiversity conservation: The key is reducing meat consumption. *Science of The Total Environment* 536:419–431.
- Maggi, F. et al. (2019). PEST-CHEMGRIDS, global gridded maps of the top 20 crop-specific pesticide application rates from 2015 to 2025. *Scientific Data* 6(1):1–20.
- Mantyka-Pringle, C.S. et al. (2011). Interactions between climate and habitat loss effects on biodiversity: a systematic review and meta-analysis. *Global Change Biology* 18:1239–1252.
- Marini, L. et al. (2014). Contrasting effects of habitat area and connectivity on evenness of pollinator communities. *Ecography* 37:544–551.

-
- Marshall, L. et al. (2017). The interplay of climate and land use change affects the distribution of EU bumblebees. *Global Change Biology* 24:101–116.
- Martin, E.A. et al. (2019). The interplay of landscape composition and configuration: new pathways to manage functional biodiversity and agroecosystem services across Europe. *Ecology Letters* 22:1083–1094.
- Matsumura, C. et al. (2004). Invasion status and potential ecological impacts of an invasive alien bumblebee, *Bombus terrestris* L. (Hymenoptera: Apidae) naturalized in Southern Hokkaido, Japan. *Global Environmental Research* 8:51–66.
- Mayer, C. et al. (2011). Pollination ecology in the 21st century: key questions for future research. *Journal of Pollination Ecology* 3:8–23.
- Mccallum, M.L. and Bury, G.W., (2013). Google search patterns suggest declining interest in the environment. *Biodiversity and Conservation* 22(6-7):1355-1367.
- McCrea, R. et al. (2015). Climate Change Skepticism and Voting Behavior: What Causes What? *Environment and Behavior*. 48:1309–1334.
- McKinney, M.L. (2006). Urbanization as a major cause of biotic homogenization. *Biological Conservation* 127:247–260.
- Mcowen, C.J. et al. (2016). Sufficiency and Suitability of Global Biodiversity Indicators for Monitoring Progress to 2020 Targets. *Conservation Letters* 9:489–494.
- McQuillan, P.B. and Hingston, A.B. (1999). Displacement of Tasmanian native megachilid bees by the recently introduced bumblebee *Bombus terrestris* (Linnaeus, 1758) (Hymenoptera : Apidae). *Australian Journal of Zoology* 47:59–65.
- McRae, L. et al. (2017). The Diversity-weighted Living Planet Index: controlling for taxonomic bias in a global biodiversity indicator. *PLOS ONE* 12(1):e0169156.
- Melidonis, C.A. and Peter, C.I., (2015). Diurnal pollination, primarily by a single species of rodent, documented in *Protea foliosa* using modified camera traps. *South African Journal of Botany* 97:9-15.
- Michel, J.-B. et al. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* 331:176–182.
- Michener, C. D. (2007). *The Bees of the World*. Johns Hopkins University Press.
- Mihoub, J.-B. et al. (2017). Setting temporal baselines for biodiversity: the limits of available monitoring data for capturing the full impact of anthropogenic pressures. *Scientific Reports* 7(1):1–13.
- Millard, J. et al. (2021a). Global effects of land-use intensity on local pollinator biodiversity. *Nature Communications* 12:2902.
- Millard, J. et al. (2021b). The species awareness index as a conservation culturomics metric for public biodiversity awareness. *Conservation Biology* 35:472–482.
- Millard, J. et al. (2020a). Text-analysis reveals taxonomic and geographic disparities in animal pollination literature. *Ecography* 43:44–59.

-
- Millard, J. et al. (2020b). Frequency of animal genera in English primary research abstracts returned by Scopus with the term "pollinat*". *Figshare* dataset. <https://doi.org/10.6084/m9.figshare.8326973.v1>
- Milman, O. (2020). Loss of bees causes shortage of key food crops, study *finds* *The Guardian* [Online]. Available at: <https://www.theguardian.com/environment/2020/jul/29/bees-food-crops-shortage-study> [Accessed: 12 August 2021].
- Mittermeier, J. C. et al (2019). A season for all things: Phenological imprints in Wikipedia usage and their relevance to conservation. *PLOS Biology* 17(3):e3000146.
- Mittermeier, J.C. et al. (2021). Birds that are more commonly encountered in the wild attract higher public interest online. *Conservation Science and Practice* 3:e340.
- Monfreda, C. et al. (2008). Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles* 22(1):1–19
- Mongeon, P. and Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics* 106:213–228.
- Montero-Castaño, A. and Vilà, M. (2012). Impact of landscape alteration and invasions on pollinators: a meta-analysis. *Journal of Ecology* 100:884–893.
- Mora, C. et al. (2011). How Many Species Are There on Earth and in the Ocean? *PLoS Biology* 9(8):e1001127.
- Morales, C.L. et al. (2013). Rapid ecological replacement of a native bumble bee by invasive species. *Frontiers in Ecology and the Environment* 11:529–534.
- Morandin, L.A. and Winston, M.L. (2005). Wild Bee Abundance and Seed Production in Conventional, Organic, and Genetically Modified Canola. *Ecological applications* 15(3):871-881
- Mottl, O. et al. (2021). Global acceleration in rates of vegetation change over the past 18,000 years. *Science* 372:860–864.
- Mueller, N.D. et al. (2012). Closing yield gaps through nutrient and water management. *Nature* 490:254–257.
- Myers, S.A. et al. (2012). Rainfall can explain adaptive phenotypic variation with high gene flow in the New Holland Honeyeater (*Phylidonyris novaehollandiae*). *Ecology and Evolution* 2:2397–2412.
- National Research Council (2007). Status of Pollinators in North America. Washington, D.C. National Academies Press.
- New, T.R. (2004). Moths (Insecta: Lepidoptera) and conservation: background and perspective. *Journal Of Insect Conservation* 8:79–94.
- Newbold, T. et al. (2013). Ecological traits affect the response of tropical forest bird species to land-use intensity. *Proceedings of the Royal Society of London B: Biological Sciences* 280(1750):1–8

-
- Newbold, T. et al. (2014a). A global model of the response of tropical and sub-tropical forest biodiversity to anthropogenic pressures. *Proceedings of the Royal Society B: Biological Sciences* 281(1792):1–10
- Newbold, T. et al. (2014b). Functional traits, land-use change and the structure of present and future bird communities in tropical forests. *Global Ecology and Biogeography* 23:1073–1084.
- Newbold, T. et al. (2015a). Global effects of land use on local terrestrial biodiversity. *Nature* 520:45–50.
- Newbold, T. (2015b). StatisticalModels: Functions for Generating, Analysing, Checking and Plotting Statistical Models. <https://github.com/timnewbold/StatisticalModels>
- Newbold, T. et al. (2018). Widespread winners and narrow-ranged losers: Land use homogenizes biodiversity in local assemblages worldwide. *PLOS Biology* 16:e2006841.
- Newbold, T. et al. (2020). Tropical and Mediterranean biodiversity is disproportionately sensitive to land-use and climate change. *Nature Ecology & Evolution* 4(12):1630–1638.
- Noriega, J.A. et al. (2018). Research trends in ecosystem services provided by insects. *Basic and Applied Ecology* 26:8–23.
- Nunez-Mir, G.C. et al. (2016). Automated content analysis: addressing the big literature challenge in ecology and evolution. *Methods in Ecology and Evolution* 7:1262–1272.
- O'Mara-Eves, A. et al. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews* 4(5):1–22
- Öckinger, E. et al. (2010). Life-history traits predict species responses to habitat area and isolation: a cross-continental synthesis. *Ecology Letters* 13:969–979.
- Oldroyd, B.P. (1999). Coevolution while you wait: *Varroa jacobsoni*, a new parasite of western honeybees. *Trends in Ecology & Evolution* 14:312–315.
- Oliver, I. et al. (2016). Additive and synergistic effects of land cover, land use and climate on insect biodiversity. *Landscape Ecology* 31:2415–2431.
- Oliver, T.H. and Morecroft, M.D. (2014). Interactions between climate change and land use change on biodiversity: attribution problems, risks, and opportunities. *Wiley Interdisciplinary Reviews: Climate Change* 5:317–335.
- Oliver, T.H. et al. (2012). Population resilience to an extreme drought is influenced by habitat area and fragmentation in the local landscape. *Ecography* 36:579–586.
- Ollerton J. (2021). Pollinators & Pollination: Nature and Society. Pelagic Publishing, Exeter.
- Ollerton, J. (2017). Pollinator Diversity: Distribution, Ecological Function, and Conservation. *Annual Review of Ecology, Evolution, and Systematics* 48:353–376.
- Ollerton, J. and Liede, S. (1997). Pollination systems in the Asclepiadaceae: a survey and preliminary analysis. *Biological Journal of the Linnean Society* 62:593–610.

-
- Ollerton, J. et al. (2014). Extinctions of aculeate pollinators in Britain and the role of large-scale agricultural changes. *Science* 346:1360–1362.
- Ollerton, J. et al. (2011). How many flowering plants are pollinated by animals? *Oikos* 120:321–326.
- Outhwaite, C.L. et al. (2020). Complex long-term biodiversity change among invertebrates, bryophytes and lichens. *Nature Ecology and Evolution* 4:384–392.
- Outhwaite, C.L. et al. (in review). Agriculture and Climate Change Interact to Drive Insect Biodiversity Declines Worldwide.
- Papworth, S.K. et al. (2015). Quantifying the role of online news in linking conservation research to Facebook and Twitter. *Conservation Biology* 29(3):825–833.
- Parker, W.E. and Howard, J.J., (2001). The biology and management of wireworms (*Agriotes* spp.) on potato with particular reference to the UK. *Agricultural and Forest Entomology* 3(2):85–98.
- Parmesan, C. (2007). Influences of species, latitudes and methodologies on estimates of phenological response to global warming. *Global Change Biology* 13:1860–1872.
- Parr, C.S. et al. (2012). Evolutionary informatics: unifying knowledge about the diversity of life. *Trends in Ecology & Evolution* 27:94–103.
- Pauw, A. (2007). Collapse of a pollination web in small conservation areas. *Ecology* 88:1759–1769.
- Pereira, H.M. et al. (2010). Scenarios for global biodiversity in the 21st century. *Science* 330:1496–501.
- Pilling, E.D. and Jepson, P.C. (1993). Synergism between EBI fungicides and a pyrethroid insecticide in the honeybee (*Apis mellifera*). *Pesticide Science* 39:293–297.
- Pimm, S.L. et al. (1995). The Future of Biodiversity. *Science* 269:347–350.
- Pisa, L.W. et al. (2015). Effects of neonicotinoids and fipronil on non-target invertebrates. *Environmental Science and Pollution Research* 22:68–102.
- Polis, G.A. et al. (2000). When is a trophic cascade a trophic cascade? *Trends in Ecology & Evolution* 15:473–475.
- Potts, S.G. et al. (2010a). Declines of managed honey bees and beekeepers in Europe. *Journal of Apicultural Research* 49:15–22.
- Potts, S.G. et al. (2010b). Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution* 25:345–353.
- Potts, S.G. et al. (2016). Safeguarding pollinators and their values to human well-being. *Nature* 540:220–229.
- Powney, G.D. et al. (2019). Widespread losses of pollinating insects in Britain. *Nature Communications* 10(1):1–6.
- Prescott-Allen, R. and Prescott-Allen, C. (1990). How many plants feed the world? *Conservation Biology* 4:365–374.

-
- Proctor, M. (1996). *The Natural History of Pollination*. Harper Collins.
- Pullin, A.S. and Stewart, G.B. (2005). Guidelines for Systematic Review in Conservation and Environmental Management. *Conservation Biology* 20:1647–1656.
- Purvis, A. et al. (2018). Modelling and Projecting the Response of Local Terrestrial Biodiversity Worldwide to Land Use and Related Pressures: The PREDICTS Project. In: *Advances in Ecological Research*. 58: 201–241.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (<https://www.R-project.org/>)
- Rader, R. et al. (2014). The winners and losers of land use intensification: Pollinator community disassembly is non-random and alters functional diversity. *Diversity and Distributions* 20:908–917.
- Raguso, R.A. and Willis, M.A. (2002). Synergy between visual and olfactory cues in nectar feeding by naïve hawkmoths, *Manduca sexta*. *Animal Behaviour* 64:685–695.
- Rathbone, J. et al. (2015). Better duplicate detection for systematic reviewers: evaluation of Systematic Review Assistant-Deduplication Module. *Systematic Reviews* 4:6.
- Rech, A.R. et al. (2016). The macroecology of animal versus wind pollination: ecological factors are more important than historical climate stability. *Plant Ecology and Diversity* 9:253–262.
- Regan, E. C. et al. (2015). Global Trends in the Status of Bird and Mammal Pollinators. *Conservation Letters* 8(6):397–403.
- Richards, A.J. (2001). Does Low Biodiversity Resulting from Modern Agricultural Practice Affect Crop Pollination and Yield? *Annals of Botany* 88:165–172.
- Ricketts, T.H. et al. (2008). Landscape effects on crop pollination services: are there general patterns? *Ecology Letters* 11:499–515.
- Riddiford, L.M. et al. (2003). Insights into the molecular basis of the hormonal control of molting and metamorphosis from *Manduca sexta* and *Drosophila melanogaster*. *Insect Biochemistry and Molecular Biology* 33:1327–1338.
- Ridding, L.E. et al. (2020). Long-term change in calcareous grassland vegetation and drivers over three time periods between 1970 and 2016. *Plant Ecol.* 221:377–394.
- Rigby, R. A. et al. (2008). A framework for modelling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution. *Comput. Stat. Data Anal.* 53:381–393
- Roberge, J.M., (2014). Using data from online social networks in conservation science: which species engage people the most on Twitter? *Biodiversity and Conservation* 23(3):715–726.
- Roll, U. et al. (2017). Using machine learning to disentangle homonyms in large text corpora. *Conservation Biology* 32:716–724.
- Roll, U. et al. (2016). Using Wikipedia page views to explore the cultural importance of global reptiles. *Biological Conservation* 204:42–50.

-
- Roquer-Beni, L. et al. (2020). A novel method to measure hairiness in bees and other insect pollinators. *Ecology and Evolution* 10:2979–2990.
- Rosindell, J., & Wong, Y. (2020). OneZoom. (<https://www.onezoom.org/>)
- Roskov Y. et al. (2017). Species 2000 & ITIS Catalogue of Life, 2017 Annual Checklist.
- Roubik, D. (2001). Ups and downs in pollinator populations: when is there a decline? *Conservation Ecology* 5(1).
- Rucker, R.R. et al. (2012). Honey Bee Pollination Markets and the Internalization of Reciprocal Benefits. *American Journal of Agricultural Economics* 94:956–977.
- Sánchez-Bayo, F. and Goka, K. (2014). Pesticide Residues and Bees – A Risk Assessment *PLoS ONE* 9:e94482.
- Sánchez-Bayo, F. and Wyckhuys, K.A.G. (2019). Worldwide decline of the entomofauna: A review of its drivers. *Biological Conservation* 232:8–27.
- Santeramo, F.G. et al. (2018). Drivers of grain price volatility: A cursory critical review. *Agricultural Economics* 64:347–356.
- Santos, J.W. dos et al. (2020). Drivers of taxonomic bias in conservation research: a global analysis of terrestrial mammals. *Animal Conservation* 23:679–688.
- Sarkar, I.N. (2007). Biodiversity informatics: organizing and linking information across the spectrum of life. *Briefings in Bioinformatics* 8:347–357.
- Saunders, D.A. et al. (1991). Biological Consequences of Ecosystem Fragmentation: A Review. *Conservation Biology* 5:18–32.
- Saunders, M.E. et al. (2020). Moving On from the Insect Apocalypse Narrative: Engaging with Evidence-Based Insect Conservation. *BioScience* 70(1):80-89
- Scheffers, B.R. et al. (2019). Global wildlife trade across the tree of life. *Science*, 366(6461):71-76.
- Schmidt, N.M. et al. (2016). An ecological function in crisis? The temporal overlap between plant flowering and pollinator function shrinks as the Arctic warms. *Ecography* 39:1250–1252.
- Schmuck, R. et al. (2003). Field relevance of a synergistic effect observed in the laboratory between an EBI fungicide and a chloronicotinyl insecticide in the honeybee (*Apis mellifera* L, Hymenoptera). *Pest Management Science* 59:279–286.
- Schoen, D.J. et al. (1996). How does self-pollination evolve? Inferences from floral ecology and molecular genetic variation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 351:1281–1290.
- Schroth, G. et al. (2016). Vulnerability to climate change of cocoa in West Africa: Patterns, opportunities and limits to adaptation. *Science of The Total Environment* 556:231–241.
- Scoble, M.J. (1992). *The Lepidoptera: Form, Function and Diversity*. Oxford University Press.

- Scopus (2018). Scopus Content Policy and Selection
<https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection>
- Secretariat of the Convention on Biological Diversity (2020). Zero Draft of the Post-2020 Global Biodiversity Framework.
- Seibold, S. et al. (2019). Arthropod decline in grasslands and forests is associated with landscape-level drivers. *Nature* 574:671–674.
- Semmens, T.D. et al. (1993). *Bombus terrestris* (L.) (Hymenoptera: Apidae) now established in Tasmania. *Australian Journal of Entomology* 32:346–346.
- Settele, J. et al. (2016). Climate change impacts on pollination. *Nature Plants* 2(7):1–3.
- Shemilt, I. et al. (2014). Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews. *Research Synthesis Methods* 5:31–49.
- Sherren, K. et al. (2017). Conservation culturomics should include images and a wider range of scholars. *Frontiers in Ecology and the Environment* 15(6):289–290.
- Shuler, R.E. et al. (2005) Farming practices influence wild pollinator populations on squash and pumpkin. *Journal of Economic Entomology* 98:790–795.
- Silva, F.D.S. et al. (2021). Virtual pollination trade uncovers global dependence on biodiversity of developing countries. *Science Advances* 7:6636–6646.
- Simmons, B.I. et al. (2019). Worldwide insect declines: An important message, but interpret with caution. *Ecology and Evolution* 9:3678–3680.
- Smith, T. J., & Saunders, M. E. (2016). Honey bees: the queens of mass media, despite minority rule among insect pollinators. *Insect Conservation and Diversity* 9(5):384–390.
- Soriano-Redondo, A. et al. (2017). Internet-based monitoring of public perception of conservation. *Biological Conservation* 206:304–309.
- Soroye, P. et al. (2020). Climate change contributes to widespread declines among bumble bees across continents. *Science* 367:685–688.
- Speight, M.C.D. (2017). Species Accounts of European Syrphidae 2017 Syrph the Net, the Database of European Syrphidae (Diptera).
- Spooner, F.E.B. et al. (2018). Rapid warming is associated with population decline among terrestrial birds and mammals globally. *Global Change Biology* 24:4521–4531.
- Ssymank, A. et al. (2008). Pollinating Flies (Diptera): A major contribution to plant diversity and agricultural production. *Biodiversity* 9:86–89.
- Stavert, J.R. et al. (2016). Hairiness: The missing link between pollinators and pollination. *PeerJ* 4:e2779.
- Steffan-Dewenter, I. and Westphal, C. (2007). The interplay of pollinator diversity, pollination services and landscape change. *Journal of Applied Ecology* 45:737–741.

-
- Steffen, W. et al. (2015). The trajectory of the Anthropocene: The Great Acceleration. *The Anthropocene Review* 2:81–98.
- Stephen, W.P. (1955). Alfalfa Pollination in Manitoba. *Journal of Economic Entomology* 48:543–548.
- Stokstad, E. (2007). The case of the empty hives. *Science* 316(5827):970–972.
- Tamaddoni-Nezhad, A. et al. (2013). Construction and Validation of Food Webs Using Logic-Based Machine Learning and Text Mining. *Advances In Ecological Research* 49:225–289.
- Tardy, C. (2004). The role of English in scientific communication: lingua franca or Tyrannosaurus rex? *Journal of English for Academic Purposes* 3:247–269.
- Teichroew, J.L. et al. (2017). Is China's unparalleled and understudied bee diversity at risk? *Biological Conservation* 210:19–28.
- Tenkanen, H. et al. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports* 7(1):1–11.
- Tews, J. et al. (2004). Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. *Journal of Biogeography* 31:79–92.
- Thessen, A.E. et al. (2012). Applications of Natural Language Processing in Biodiversity Science. *Advances in Bioinformatics* 2012:1–17.
- The World Bank. (2020). Individuals using the internet (% of population). Available from <https://data.worldbank.org/indicator/IT.NET.USER.ZS>
- Thomas, C.D. et al. (2004). Extinction risk from climate change. *Nature* 427:145–148.
- Thomas, C.D. et al. (2019). "Insectageddon": A call for more robust data and rigorous analyses. *Global Change Biology* 25:1891–1892.
- Thomas, J.A. et al. (2004). Comparative Losses of British Butterflies, Birds, and Plants and the Global Extinction Crisis. *Science* 303:1879–1881.
- Thuiller, W. et al. (2005). Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences* 102:8245–8250.
- Tittensor, D.P. et al. (2014). A mid-term analysis of progress toward international biodiversity targets. *Science* 346:241–244.
- Toledo-Hernández, M. et al. (2020). Hand pollination, not pesticides or fertilizers, increases cocoa yields and farmer income. *Agriculture, Ecosystems & Environment* 304:107160.
- Trisos, C.H. et al. (2020). The projected timing of abrupt ecological disruption from climate change. *Nature* 580:496–501.
- Tscharntke, T et al. (2005). Landscape perspectives on agricultural intensification and biodiversity–ecosystem service management. *Ecology letters* 8(8):857–874.
- Tscharntke, T. et al. (2002). Contribution of small habitat fragments to conservation of insect communities of grassland–cropland landscapes. *Ecological Applications* 12:354–363.

-
- Tscharntke, T. et al. (2008). Landscape constraints on functional diversity of birds and insects in tropical agroecosystems. *Ecology* 89:944–951.
- UNEP CBD. (2010). Strategic Plan for Biodiversity 2011-2020.
- Union for Ethical BioTrade (UEBT), (2019). Biodiversity Barometer. Amsterdam. (2019)
- van der Kooi, C.J. and Ollerton, J. (2020). The origins of flowering plants and pollinators. *Science* 368:1306–1308.
- van Engelsdorp, D. and Meixner, M.D. (2010). A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *Journal of Invertebrate Pathology* 103:80–95.
- van Engelsdorp, D. et al. (2008). A Survey of Honey Bee Colony Losses in the U.S., Fall 2007 to Spring 2008. *PLoS ONE* 3(12):e4071.
- van Engelsdorp, D. et al. (2009). Colony Collapse Disorder: A Descriptive Study. *PLoS ONE* 4(8):e6481.
- van Vuuren, D.P. et al. (2011). The representative concentration pathways: an overview. *Climatic Change* 109(1):5–31.
- Vanbergen, A.J (2013). Threats to an ecosystem service: pressures on pollinators. *Frontiers in Ecology and the Environment* 11:251–259.
- Vanbergen, A.J. (2014). Landscape alteration and habitat modification: impacts on plant–pollinator systems. *Current Opinion in Insect Science* 5:44–49.
- Velthuis, H.H.W. and van Doorn, A. (2006). A century of advances in bumblebee domestication and the economic and environmental aspects of its commercialization for pollination. *Apidologie* 37:421–451.
- Veríssimo, D. et al. (2020). Did the movie Finding Dory increase demand for blue tang fish? *Ambio* 49(4):903-911.
- Vit, P. et al. (2004). Quality standards for medicinal uses of Meliponinae honey in Guatemala, Mexico and Venezuela. *Bee World* 85:2–5.
- Wardhaugh, C.W. (2015). How many species of arthropods visit flowers? *Arthropod-Plant Interactions* 9:547–565.
- Watanabe, M.E. (1994). Pollination worries rise as honey bees decline. *Science* 265(5176): 1170
- Weiner, C.N. et al. (2011). Land use intensity in grasslands: Changes in biodiversity, species composition and specialisation in flower visitor networks. *Basic and Applied Ecology* 12:292–299.
- Weiner, C.N. et al. (2014). Land-use impacts on plant–pollinator networks: interaction strength and specialization predict pollinator declines. *Ecology* 95:466–474.
- West, P.C. et al. (2014). Leverage points for improving global food security and the environment. *Science* 345:325–328.

-
- Westergaard, D. et al. (2018). A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. *PLoS Comp. Biol.* 14:2.
- Westgate, M.J. (2019). revtools: An R package to support article screening for evidence synthesis. *Research Synthesis Methods* 10:606–614.
- Westgate, M.J. and Lindenmayer, D.B. (2016). The difficulties of systematic reviews. *Conservation Biology* 31:1002–1007.
- Westgate, M.J. et al. (2015). Text analysis tools for identification of emerging topics and research gaps in conservation science. *Conservation Biology* 29:1606–1614.
- Westgate, M.J. et al. (2018). Software support for environmental evidence synthesis. *Nature Ecology & Evolution* 2:588–590.
- Whitfield, C.W. et al. (2006). Thrice Out of Africa: Ancient and Recent Expansions of the Honey Bee, *Apis mellifera*. *Science* 314:642–645.
- Wikimedia Traffic Analysis Report. (2018). Wikistats. Available from <https://stats.wikimedia.org/wikimedia/squids/SquidReportPageViewsPerCountryOverview.htm>
- Williams, G.R. et al. (2010). Colony Collapse Disorder in context. *Bioessays* 32:845.
- Williams, J.J. et al. (2020a). Human-dominated land uses favour species affiliated with more extreme climates, especially in the tropics. *Ecography* 43(3):391–405.
- Williams, J.J. & Newbold, T. (2020b). Local climatic changes affect biodiversity responses to land use: A review. *Diversity & Distributions* 26: 76–92.
- Williams, P.H. and Osborne, J.L. (2009). Bumblebee vulnerability and conservation world-wide. *Apidologie* 40:367–387.
- Winfree, R. (2008). Pollinator-Dependent Crops: An Increasingly Risky Business. *Current Biology* 18(20):968–969.
- Winfree, R. et al. (2018). Species turnover promotes the importance of bee diversity for crop pollination at regional scales. *Science* 359:791–793.
- Winfree, R. et al. (2011). Native Pollinators in Anthropogenic Habitats. *Annu. Rev. Ecol. Evol. Syst* 42:1–22.
- Wood, T.J. and Goulson, D. (2017). The environmental risks of neonicotinoid pesticides: a review of the evidence post 2013. *Environmental Science and Pollution Research* 24:17285–17325.
- Woodcock, B.A. et al. (2016). Impacts of neonicotinoid use on long-term population changes in wild bees in England. *Nature Communications* 7:12459.
- Woodcock, B.A. et al. (2019). Meta-analysis reveals that pollinator functional diversity and abundance enhance crop pollination and yield. *Nature Communications* 10:1–10.
- Xiao, Y. et al. (2016). The diverse effects of habitat fragmentation on plant–pollinator interactions. *Plant Ecology* 217:857–868.

Zaragoza-Trello, C., Vilà, M. and Bartomeus, I. (2021). Interaction between warming and landscape foraging resource availability on solitary bee reproduction. *Journal of Animal Ecology* 90:2536–2546

Appendix 1: Chapter 2 Supplementary Information

Contents

Figure S1.1	187
Figure S1.2	188
Figure S1.3	189
Table S1.4	190
Figure S1.5	193
Figure S1.6	194
Figure S1.7	196
Figure S1.8	198
Figure S1.9	200

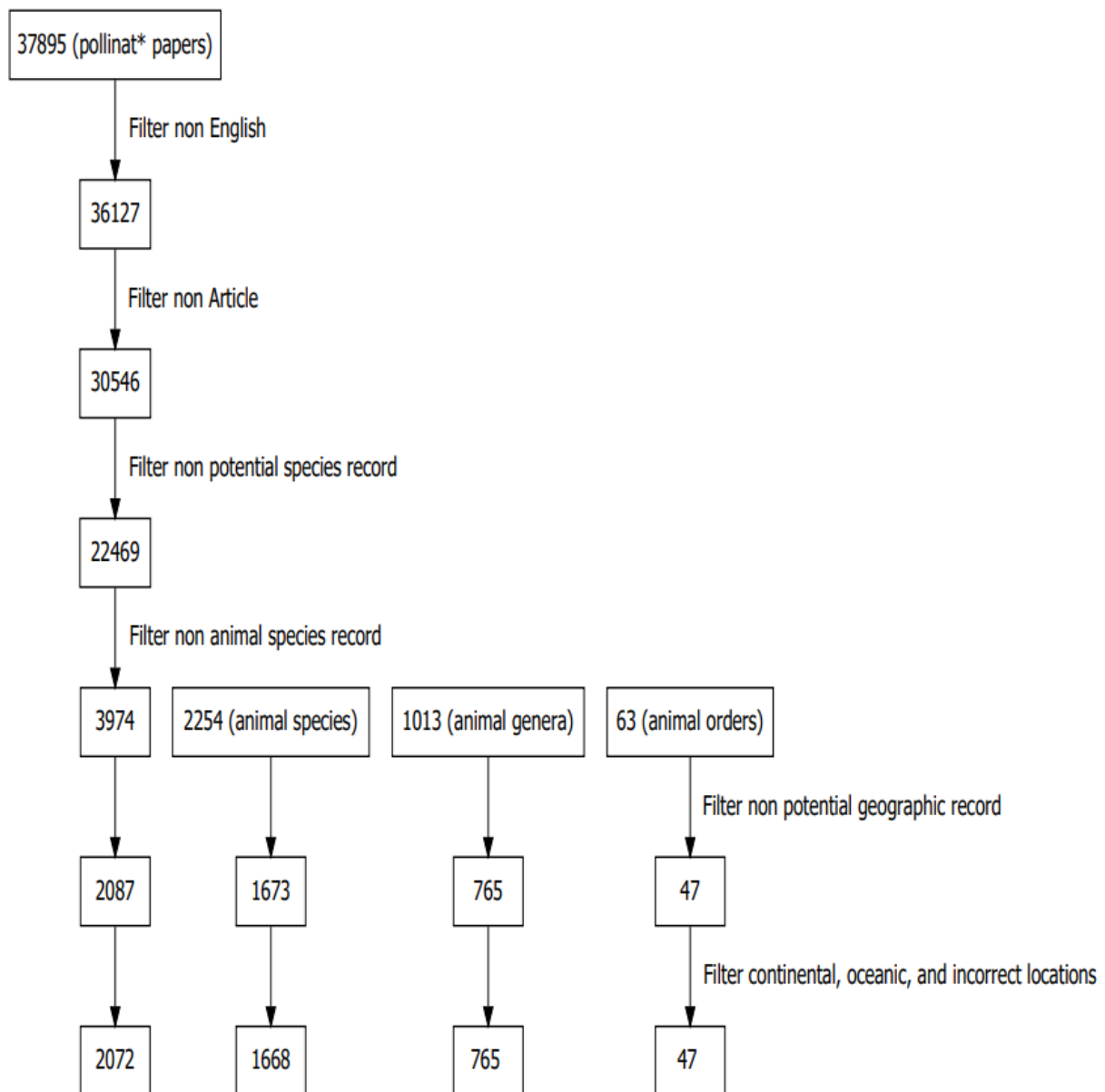


Figure S1.1 PRISMA diagram for pollination paper selection. 37895 abstracts were returned through entering the search term pollinat* in Scopus. These abstracts were filtered for English language (“Filter non English”), primary research articles (“Filter non Article”), any potential species records (“Filter non potential species record”), confirmed animal species (“Filter non animal species record”), geographic locations (“Filter non potential geographic record”), and those that do not contain only a continental, oceanic, and incorrect locations.

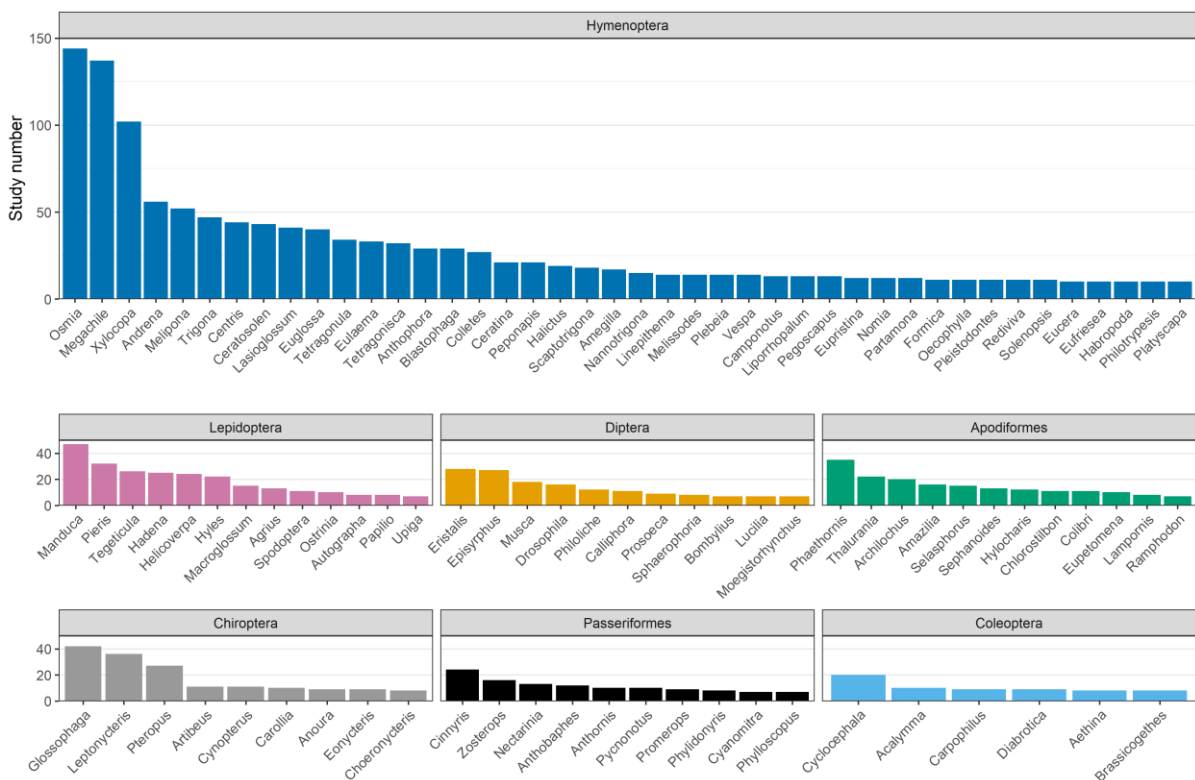


Figure S1.2 Distribution of all animal genera, with the exception of *Apis* and *Bombus*, occurring in 10 or more studies related to pollination. *Apis* and *Bombus* have been excluded here to better represent distribution for less well studied genera (see Figure 2.3 for comparable values for *Apis* and *Bombus*).

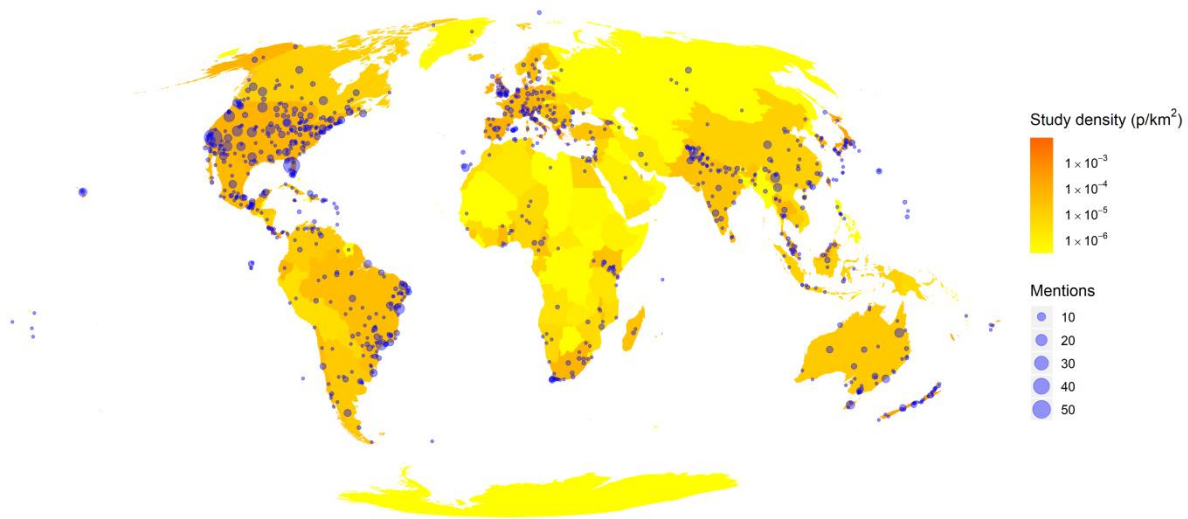


Figure S1.3 Global study density of animal pollinator related studies, aggregated at country level and adjusted for country area. Study densities were calculated by counting the number of abstracts with their “major” focus in each country, and then dividing this value by country area. All oceanic and otherwise obviously incorrect “minor” mentions, as well as “minor” mentions that could only be resolved to a unit larger than a country, were removed. Densities were $\log_{10}$ -transformed. Partially transparent blue points (“minor” mentions) represent the number of abstracts in which CLIFF-CLAVIN resolved that location.

Table S1.4 Summary output for a poisson generalised linear model, predicting annual study count against year (1961-2017) and taxonomic genera (Apis, Bombus, Other).

	Estimate	Std. Error	P value
Intercept	-1.069e+02	2.561e+00	<2e-16
Year	5.357e-02	1.274e-03	<2e-16
Apis	3.178e+00	3.166e-02	<2e-16
Bombus	3.130e+00	3.224e-02	<2e-16

Validation

I carried out a series of checks of the validity of the outputs of our text analysis methods. I validated my outputs at three levels: first, the accuracy of the extraction of animal species names; second, at the level of abstract subject area, to determine whether we had selected abstracts that are typically related to animal pollination ecology; and third, the accuracy and potential bias of the geographic locations as determined by CLIFF-CLAVIN.

I estimated the completeness and accuracy of the animal species extraction by sampling and manually searching approximately 1% of the original full set of abstracts (300 in total). For any random samples, I used the R function `set.seed()` to seed the random number generator, and then sampled abstracts at random. Given that the taxonomic extraction algorithm attempts to resolve each animal as its accepted name, in order to fairly judge its effectiveness, any manual searching would have to attempt to resolve accepted names in a similar way. For each potential animal species record identified manually, I searched for the species in the COL hoping to confirm as an accepted name. If the species was not accepted but recognised by the COL as a synonym, I changed the species record for that abstract to the accepted name. If the COL did not recognise the potential animal species as either an accepted name or a synonym, I then searched the website Discover Life (<http://www.discoverlife.org/>) for the accepted name and changed the record if appropriate. I removed any potential species records that I could not confirm as either an accepted name or a synonym verified by either the COL or Discover Life.

After manually confirming accepted animal species, I then compared these outputs to the performance of the algorithm. 79.5% of the animal species records manually extracted were found by the automated algorithm (i.e. a 20.5% omission error). Precision on the other hand was high at 100%, meaning that the algorithm found no animals which were not in that given abstract (i.e. a 0% commission error).

I also conducted a validation to investigate whether considering only Latin binomial names influenced the taxonomic bias in the pollination literature (Figure S1.5). This analysis shows that the abstract count for honey bees and bumble bees is underestimated by searching for Latin binomials, potentially by a factor of ~2. This is because honey bees and bumblebees are referred to by their common name more often than other species. I counted the number of “pollinat*” abstracts mentioning two typical common name spellings for *Apis* and *Bombus*, and then to control for string number, analogous strings for both *Osmia* and *Megachile*: *Apis* - “honey bee” and “honeybee”; *Bombus* - “bumble bee” and “bumblebee”; *Osmia* - “mason bee” and “mason-bee”; *Megachile* - “leafcutter bee” and “leaf-cutter bee”. I included *Osmia* and *Megachile* as a control, to investigate common name frequency for less well-known species. For *Apis* and *Bombus*, including abstracts mentioning a common name doubles their respective abstract count. For *Osmia* and *Megachile*, the inclusion of common names increases abstract count by 4.6% and 7.2% abstracts respectively. These results would indicate that whilst *Apis* and *Bombus* study count is underestimated, for other less well-known taxa the Latin binomial will be an effective indicator of study effort. I also reasoned that, whilst my analysis might underestimate for *Apis* and *Bombus*, including only the Latin binomial would help to reduce false positive rate.

I conducted an additional validation to investigate whether excluding taxonomic names above the level of species influenced the taxonomic bias in the pollination literature. This analysis indicates that although representation for some families (hummingbirds, fig wasps, and hoverflies) may be underestimated, the overall trend is likely similar (Figure S1.6). I counted the number of “pollinat*” abstracts mentioning family names for each of 4 well-known pollination-related families (hummingbirds, fig wasps, hoverflies, hawk-moths, and leaf-nosed bats), selected from each of the top 5 orders (Hymenoptera, Lepidoptera, Diptera, Apodiformes, and Chiroptera). For each family, I searched for four common, Latin, and pluralised family names: hummingbirds (“humming-bird”, “hummingbird”, “Hummingbird”, “Trochilidae”); fig wasps (“fig wasp”, “Fig wasp”, “fig-wasp”, “Agaonidae”); hoverflies

("Hoverfly", "hoverflies", "hoverfly", "Syrphidae"); hawk-moths ("Hawk-moth", "hawk moth", "hawk-moth", "Sphingidae") and leaf-nosed bats ("Phyllostomidae", "leaf-nosed bat", "leaf nosed bat", "Leaf-nosed bat"). The number of abstracts for hummingbirds, hoverflies, and fig wasps all increased by more than a factor of ~2 with the inclusion of family names, with the leaf-nosed bats making only a marginal increase in total abstract number (Figure S1.6).

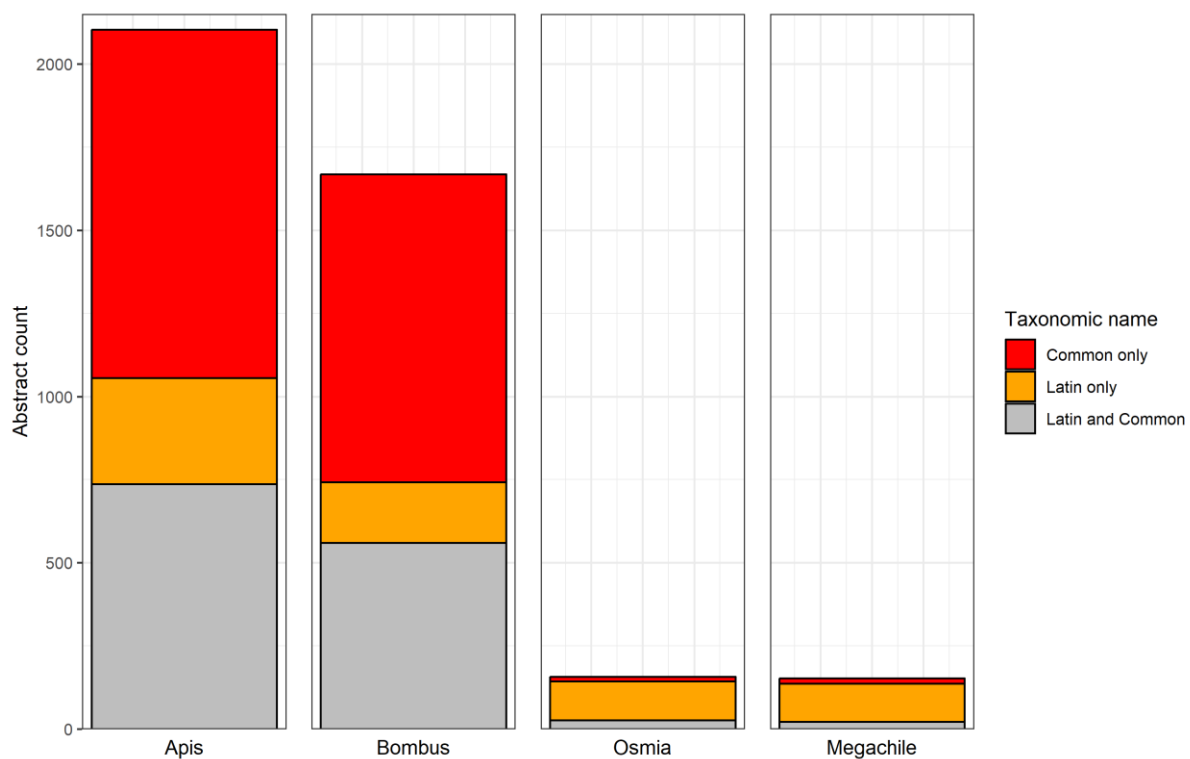


Figure S1.5 Frequency of “pollinat” Scopus abstracts containing a common name and Latin binomial for each of *Apis* (honey bee), *Bombus* (bumblebee), *Osmia* (mason bee), and *Megachile* (leafcutter bee). Grey bars represent abstracts containing both a Latin binomial and common name for that genus. Orange bars represent abstracts containing only a Latin binomial for a species in that genus. Red bars represent abstracts containing only a common name for that genus.

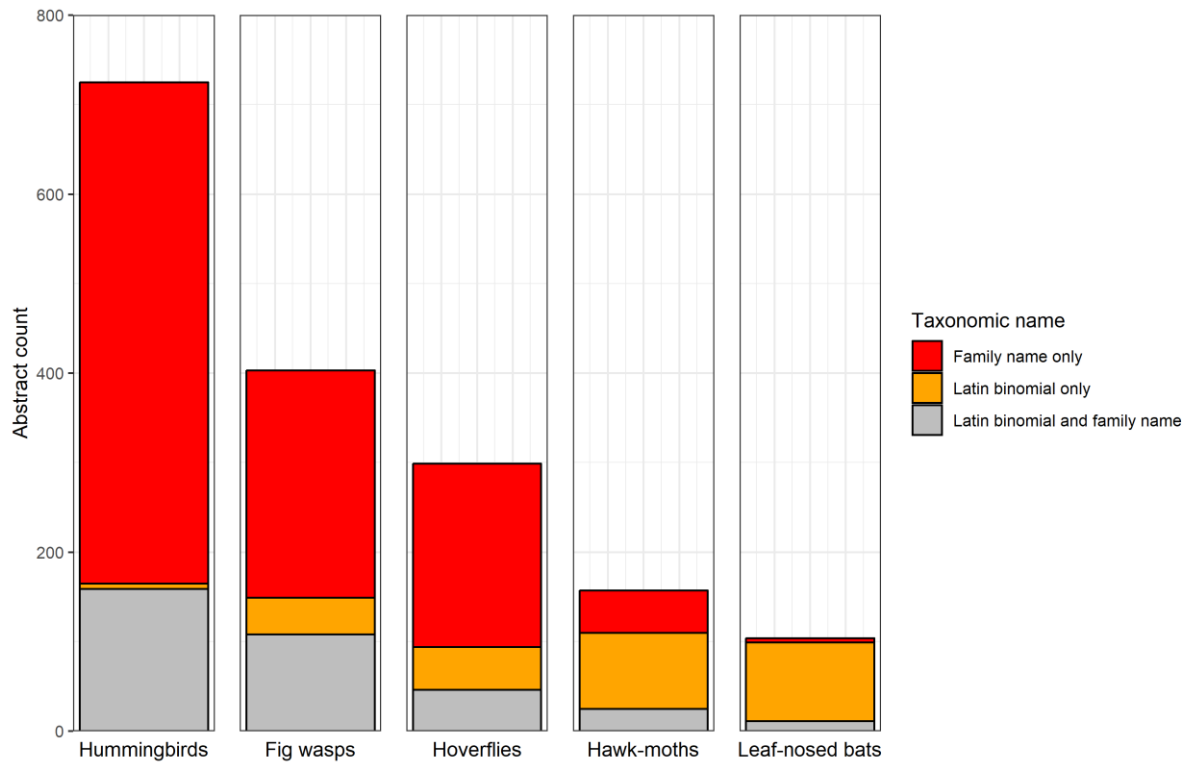


Figure S1.6 Frequency of “pollinat” Scopus abstracts containing a family name (either common or Latin) and Latin binomial for each of the hummingbirds (Trochilidae), fig wasps (Agaonidae), hoverflies (Syrphidae), hawk-moths (Sphingidae), and leaf-nosed bats (Phyllostomidae). Grey bars represent abstracts containing both a Latin binomial and the family name for that family. Orange bars represent abstracts containing only a Latin binomial for a species in that family. Red bars represent abstracts containing only the family name for that family.

To validate the subject areas of the identified abstracts, I randomly sampled 100 abstracts from the subset of original abstracts that also mentioned an animal species (approximately 2.5% of the total). I then read each abstract and title, assigning the subject area as any of three categories: general pollination ecology, pollinator status or habitat disturbance, and other pollinator related literature (Figure S1.7). No abstracts were totally unrelated to pollination. Abstracts on general pollination ecology included any studies on visitation, efficiency, pollinator movement ecology, pollinator foraging behaviour, pollination syndromes, plant-pollinator networks, and pollination dependent crop yields. Abstracts on pollinator status included any studies on pollinator population trends, diversity, abundance, ecological impacts, and habitat disturbance. All “other” abstracts concerned analyses of population genetics, pest/disease management, pollinator predation, invasive species management, animal floral mimicry, pollinator mating behaviour, pollinator awareness, pollinator learning behaviour, and pollinator nesting behaviour.

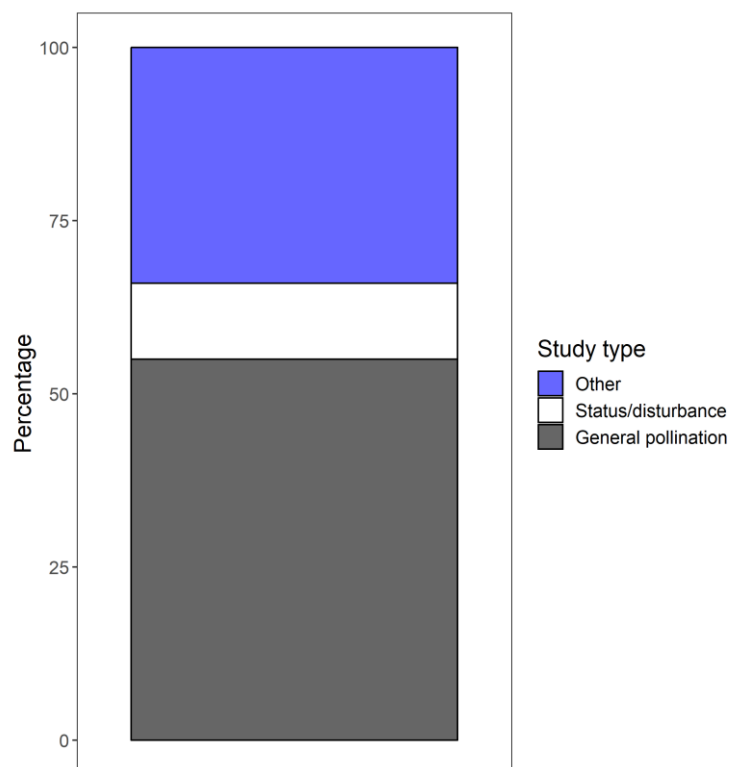


Figure S1.7 Distribution of study types for a random sample of 100 of the subset of abstracts that mentioned an animal species. Abstracts on general pollination ecology included any studies related to pollinator flower visitation. Abstracts on pollinator status included any studies on population status and disturbance. All “other” abstracts concerned analyses of genetics, species management, and behaviours/evolution not directly related to pollinator flower visitation.

To assesses whether CLIFF-CLAVIN introduced geographic biases, I manually geoparsed the 100 abstracts previously sampled for study topic. Initially I identified and recorded all text locations in each abstract. I then geocoded each of these locations using Google Maps, recording the respective coordinates. Both the automatically geoparsed (after removal of obvious mistakes, continental, and oceanic “minor” mentions) and the manually geoparsed map show very similar distributions (Figure S1.8), with high densities in North America, South America, Europe, south and South East Asia, and Australia. CLIFF-CLAVIN appears to have slightly over-estimated the number of locations in North America and Europe. The geographic distribution for CLIFF-CLAVIN appears particularly similar in Africa, with clusters in South Africa and Kenya for both automatic and manual geoparsing. The similarity between the automatic and manually geoparsed data is likely genuine and not a sampling effect. The manually geoparsed abstracts are a randomly sampled subset of those that have already been automatically geoparsed, meaning the manual and automatic geoparse compares like-with-like.

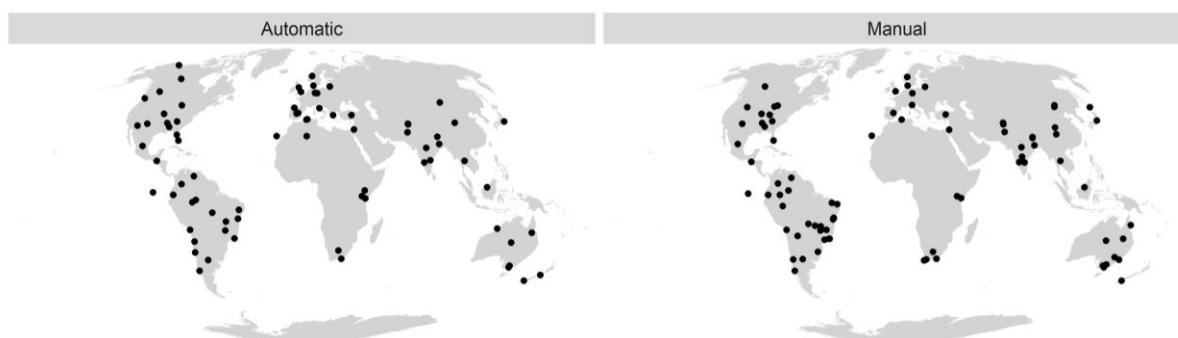


Figure S1.8 The geographic distribution of a random sample of 100 abstracts related to pollination and containing an animal species, geoparsed automatically and manually. Continental, oceanic, and obviously incorrect “minor” mentions were removed from the automatically generated locations.

Given Germany, a country previously reported as a key contributor of pollinator disturbance data (Archer *et al.* 2014), didn't appear in the top 15 countries for pollination-related study in our analysis, I performed a separate check to ensure this was not a mistake in the geoparsing (Figure S1.9). Specifically, I performed a character string match between each abstract and a set of strings representing the exact names of all countries. I assumed that the number of abstracts mentioning a country string would provide a coarse indication of relative study frequency, albeit through an approach less sophisticated than CLIFF-CLAVIN. In other words, I assumed that if the term "Germany" appeared frequently across abstracts, this would indicate that CLIFF-CLAVIN likely underestimated its importance. I found that CLIFF-CLAVIN ranked Germany as the 22nd most important country for pollination related studies, whilst an exact character string match ranked it at 17th. This suggests that CLIFF-CLAVIN hasn't greatly underestimated the frequency of studies occurring in Germany. The United States has a much lower representation through character string matches, likely reflecting a tendency in US abstracts to mention the specific locality without the country string.

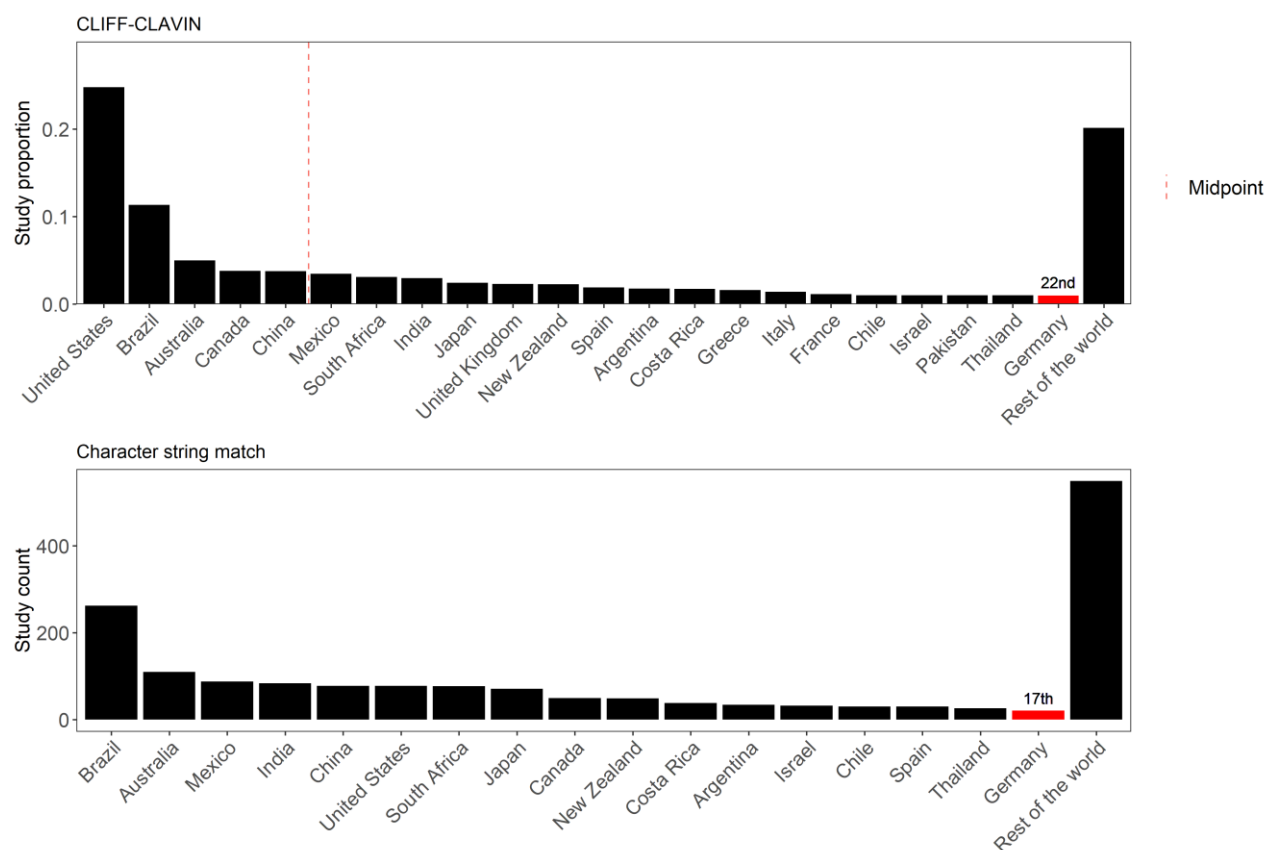


Figure S1.9 Geographic distribution of the pollination literature, resolved through exact character string matches and CLIFF-CLAVIN. I deemed exact character string matches a coarse and imperfect check on CLIFF-CLAVIN. The red dotted line represents the study proportion midpoint. Consistent with CLIFF-CLAVIN, exact character string matches also return Germany outside of the top 15 countries, at 22nd for CLIFF-CLAVIN and 17th for exact character string matches.

Appendix 2: Chapter 3 Supplementary Information

Contents

Supplementary Figures	203
Figure S2.1	203
Figure S2.2	204
Figure S2.3	205
Figure S2.4	206
Figure S2.5	207
Figure S2.6	208
Figure S2.7	209
Figure S2.8	210
Figure S2.9	211
Figure S2.10	212
Figure S2.11	213
Figure S2.12	214
Figure S2.13	215
Supplementary Tables.....	216
Table S2.15	216
Table S2.16	216
Table S2.17	217
Table S2.18	218
Table S2.19	220
Table S2.20	221
Table S2.21	222
Table S2.22	222
Table S2.23	223
Table S2.24	223
Table S2.25	224
Table S2.26	225
Table S2.27	226
Table S2.28	227
Table S2.29	228
Table S2.30	228
Table S2.31	229

Table S2.32	230
Table S2.33	231
Table S2.34	232

Supplementary Figures

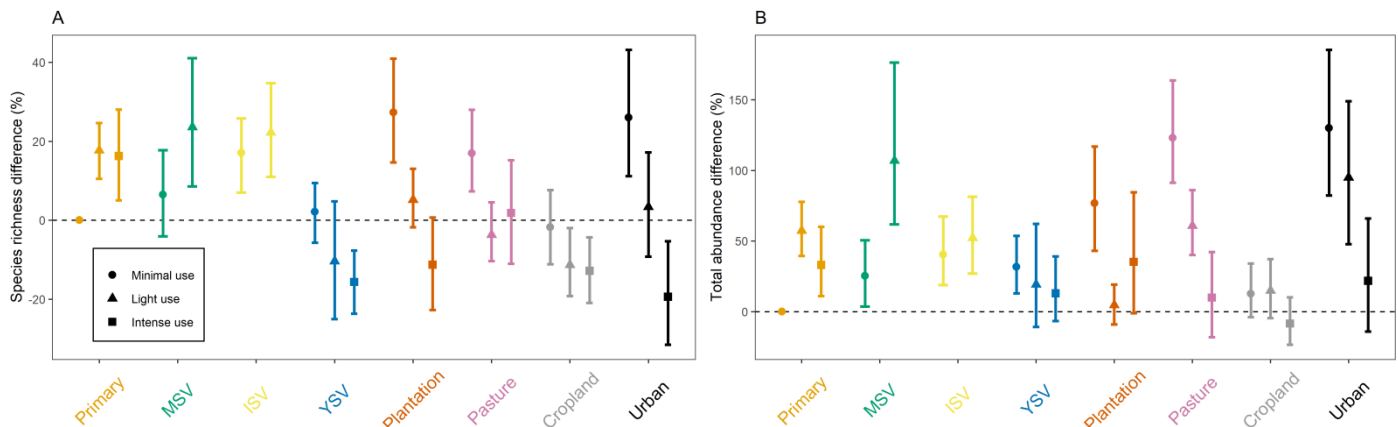


Figure S2.1 Response of pollinator biodiversity to land-use type and land-use intensity for species richness and total abundance, both fit using generalised linear mixed-effects models. A) species richness, fit using a poisson error generalised linear mixed-effects model; and B) total abundance, fit using a zero-inflated negative binomial error generalised linear mixed-effects model. Colours represent land-use type: orange (primary vegetation, Primary), green (mature secondary vegetation, MSV), yellow (intermediate secondary vegetation, ISV), blue (young secondary vegetation, YSV), dark orange (plantation forest, Plantation), pink (pasture), grey (cropland), and black (urban), and point shape represents land-use intensity: circle (minimal uses), triangles (light use), and squares (intense use). Effect sizes were adjusted to a percentage by drawing fixed effects 1,000 times based on the variance-covariance matrix, expressing each fixed effect as a percentage of the baseline (primary vegetation minimal use), and then calculating the median value (shown as points), and 2.5th and 97.5th percentiles (shown as error bars).

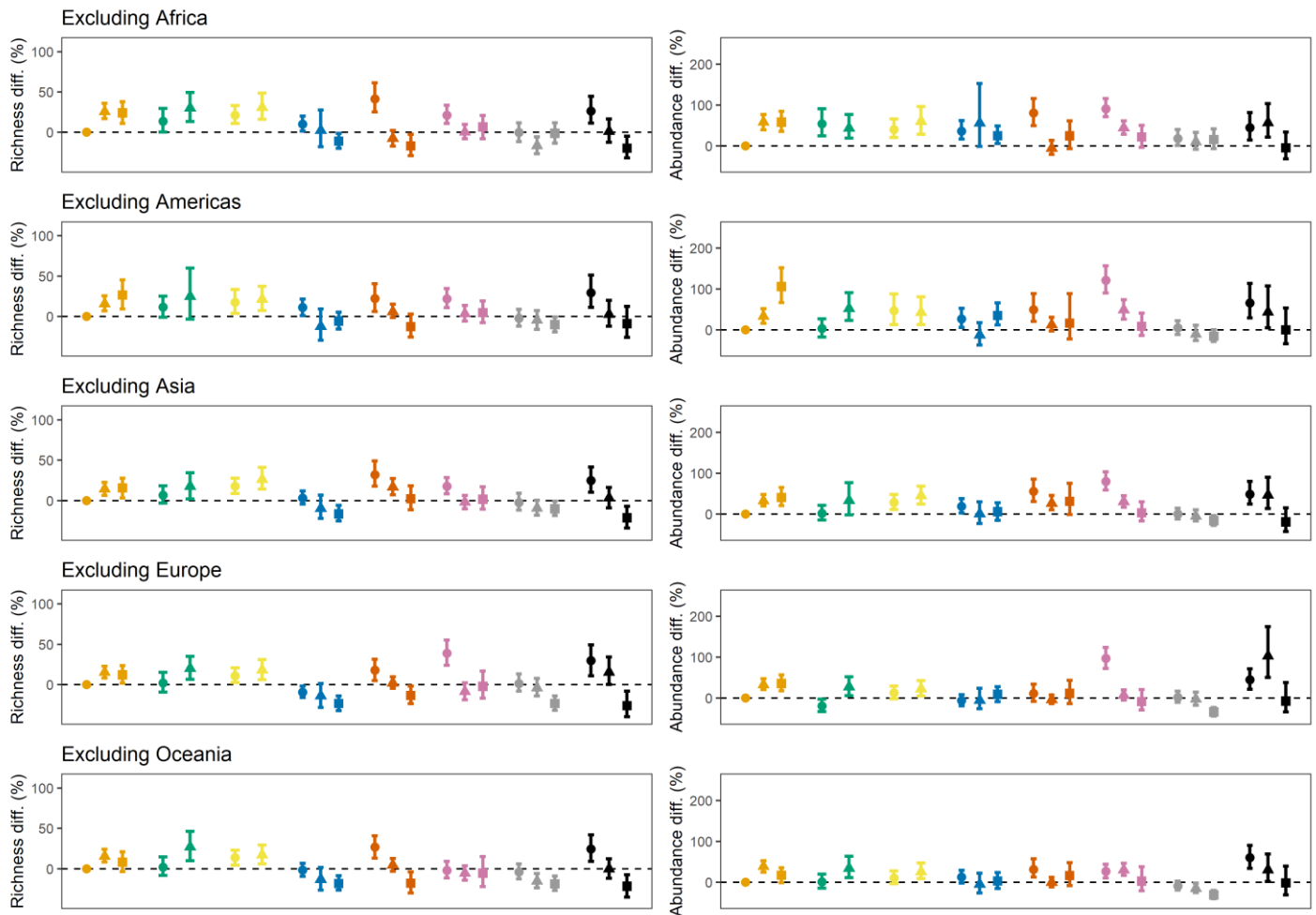


Figure S2.2 Response of pollinator biodiversity to land-use type and land-use intensity for species richness and total abundance (fit using the approach in the main text), jack-knifed by UN regions (Africa, Americas, Asia, Europe, and Oceania). Colours represent land-use type: orange (primary vegetation), green (mature secondary vegetation), yellow (intermediate secondary vegetation), blue (young secondary vegetation), dark orange (plantation forest), pink (pasture), grey (cropland), and black (urban), and point shape represents land-use intensity: circle (minimal uses), triangles (light use), and squares (intense use). Effect sizes were adjusted to a percentage by drawing fixed effects 1,000 times based on the variance-covariance matrix, expressing each fixed effect as a percentage of the baseline (primary vegetation minimal use), and then calculating the median value (shown as points), and 2.5th and 97.5th percentiles (shown as error bars).

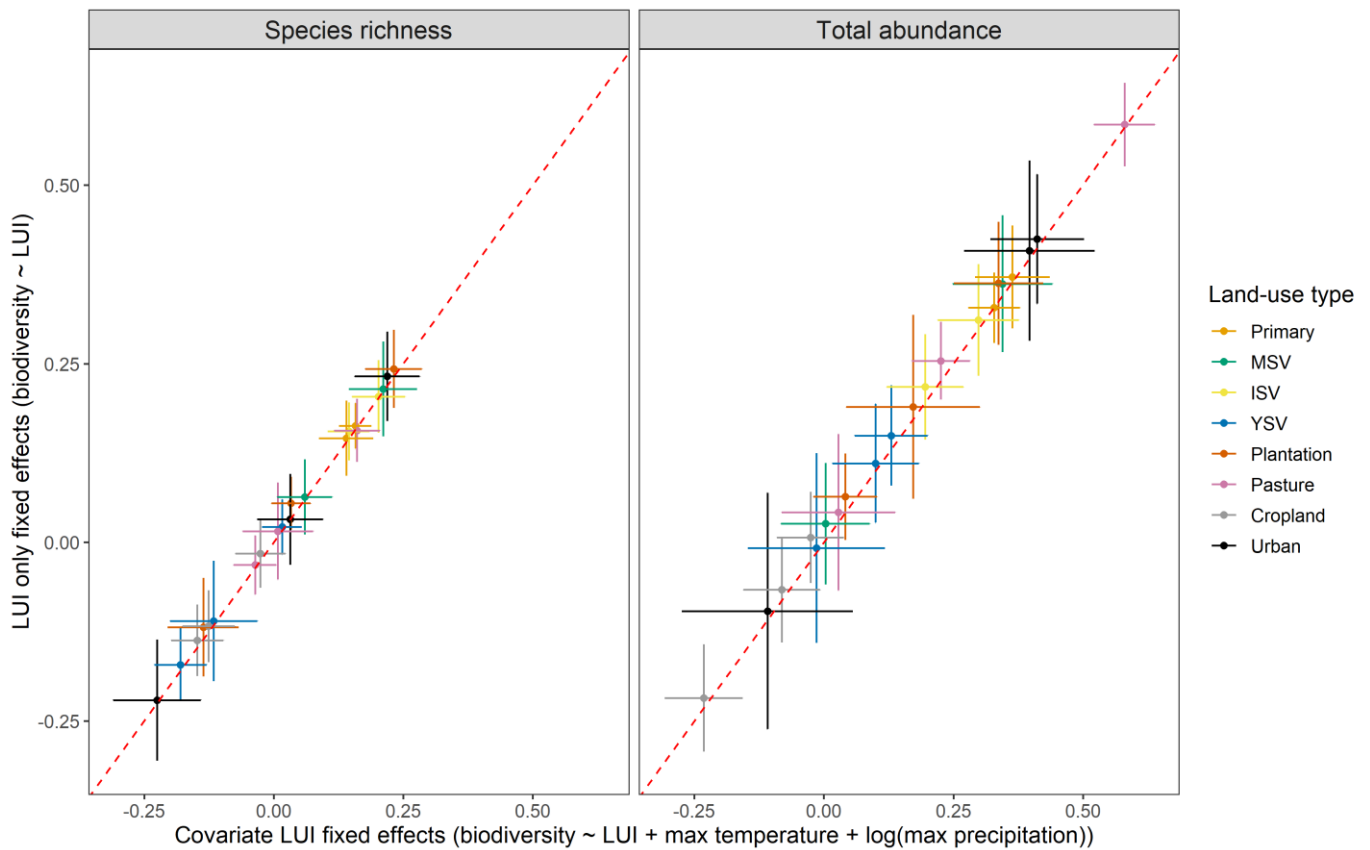


Figure S2.3 Fixed effects for a model including climate covariates (max temperature of the hottest month and the total precipitation of the wettest month, both over the 12 months previous to the end data of each sample), predicting either species richness (left panel) or abundance (right panel), plotted against the same fixed effects for the LUI-only model fitted in the main text. Crosses for each point represent the standard error for both the LUI + climate and LUI-only models. Here the diagonal dotted red line represents a line gradient of 1 (i.e. $y=x$), showing that the inclusion of climate covariates does not markedly change the predictions.

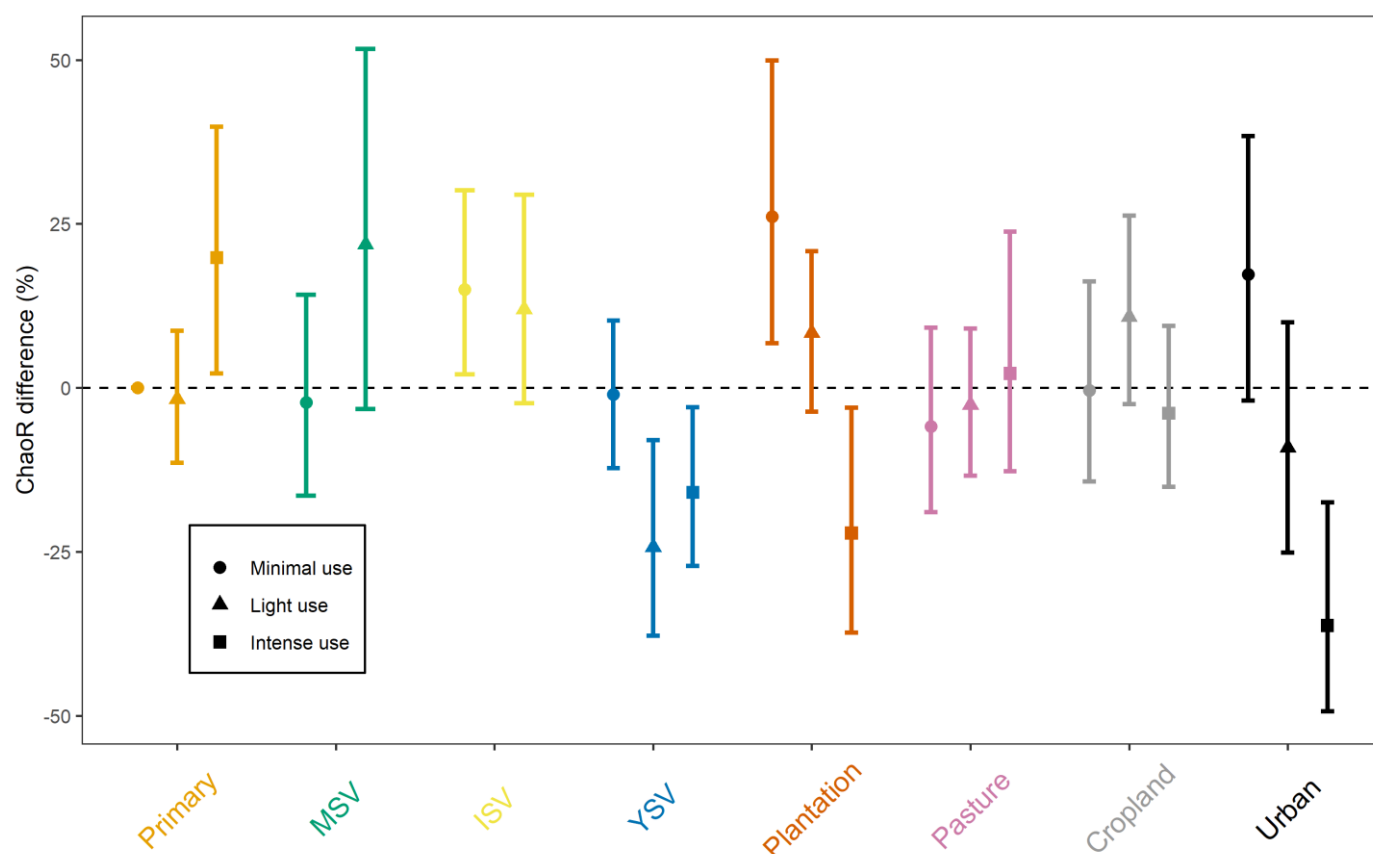


Figure S2.4 Response of pollinator Chao's species richness to land-use type and land-use intensity. Colours represent land-use type: orange (primary vegetation, Primary), green (mature secondary vegetation, MSV), yellow (intermediate secondary vegetation, ISV), blue (young secondary vegetation, YSV), dark orange (plantation forest, Plantation), pink (Pasture), grey (Cropland), and black (Urban), and point shape represents land-use intensity: circle (minimal uses), triangles (light use), and squares (intense use). Effect sizes were adjusted to a percentage by drawing fixed effects 1,000 times based on the variance-covariance matrix, expressing each fixed effect as a percentage of the baseline (primary vegetation minimal use), and then calculating the median value (shown as points), and 2.5th and 97.5th percentiles (shown as error bars).

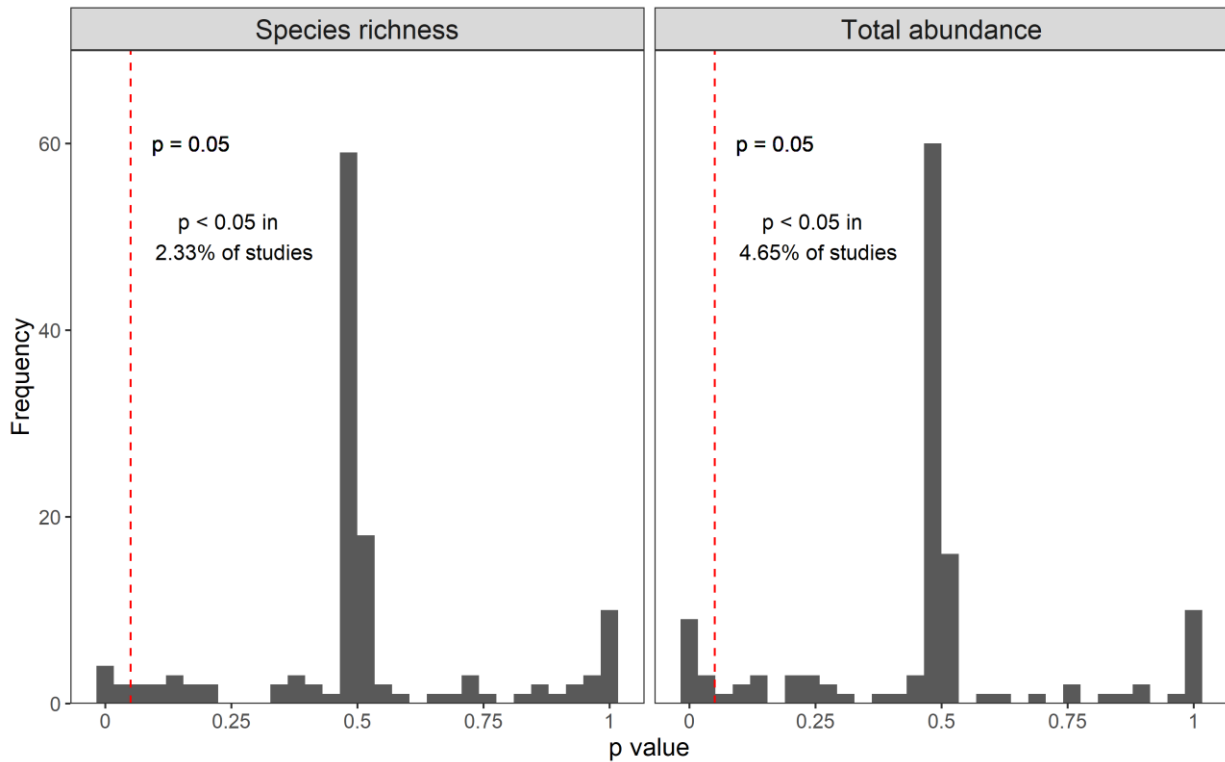


Figure S2.5 The distribution of P values for a set of one-sided Moran's I tests for spatial autocorrelation, calculated for each study for species richness (left panel) and total abundance (right panel). The red dotted line in both panels represents a P value of 0.05. For species richness $p < 0.05$ in 2.33% of studies, and for total abundance $p < 0.05$ in 4.65% of studies. By chance, in the absence of spatial autocorrelation, we would expect 5% of studies to return a P value < 0.05 .

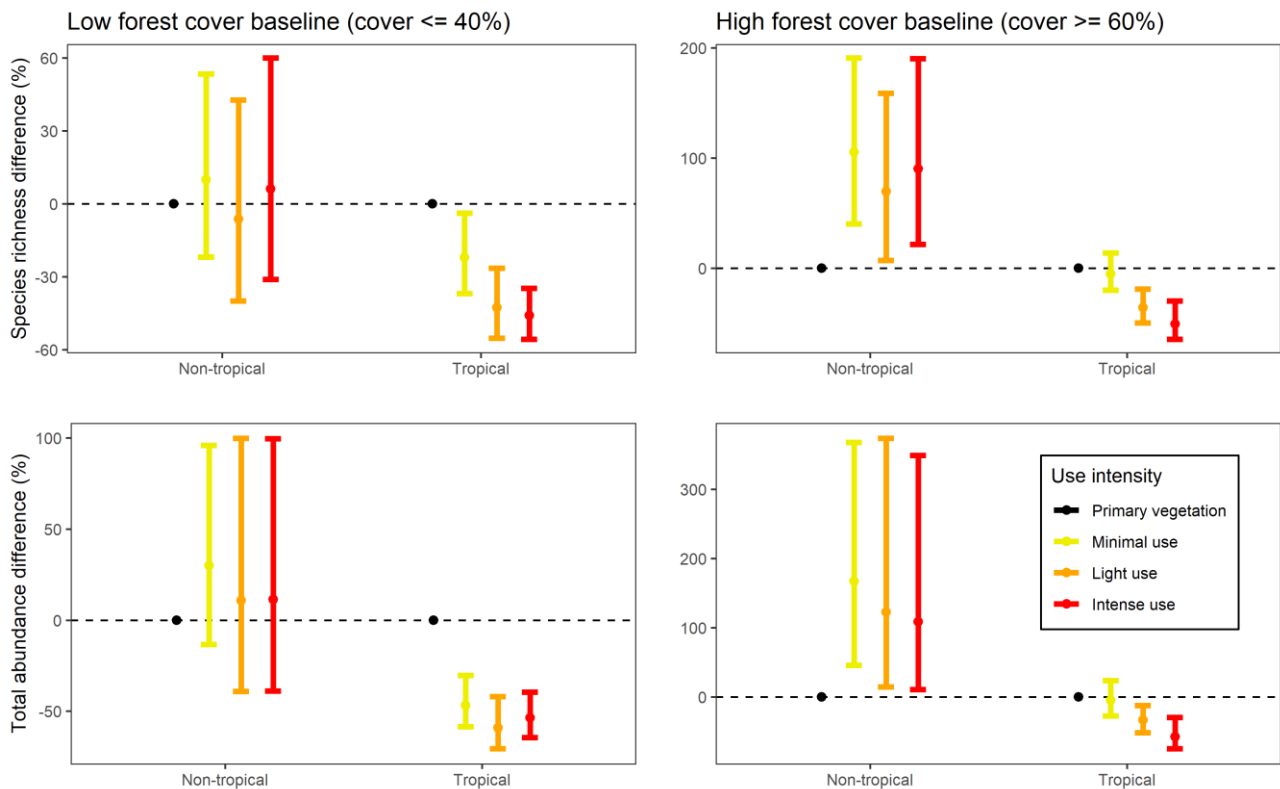


Figure S2.6 Response of pollinators to land-use intensity on cropland, for non-tropical and tropical sites, when the primary vegetation minimal use baseline is shifted between high and low forest cover. Each panel represents a linear or generalised linear mixed-effects model for a given biodiversity metric. Left-hand panels (low forest cover baseline) represent species richness and total abundance predictions for cropland between tropical and non-tropical regions when the baseline is subset for only low forest cover sites ($\leq 40\%$ forest cover). Right-hand panels (high forest cover baseline) represent species richness and total abundance predictions for cropland when the baseline is subset for only high forest cover sites ($\geq 60\%$ forest cover). Effect sizes were adjusted to a percentage by sampling fixed effects 1,000 times based on the variance-covariance matrix, expressing each fixed effect as a percentage of the value in primary vegetation for that geographical zone, and then calculating the median value (shown as points), and 2.5th and 97.5th percentiles (shown as error bars).

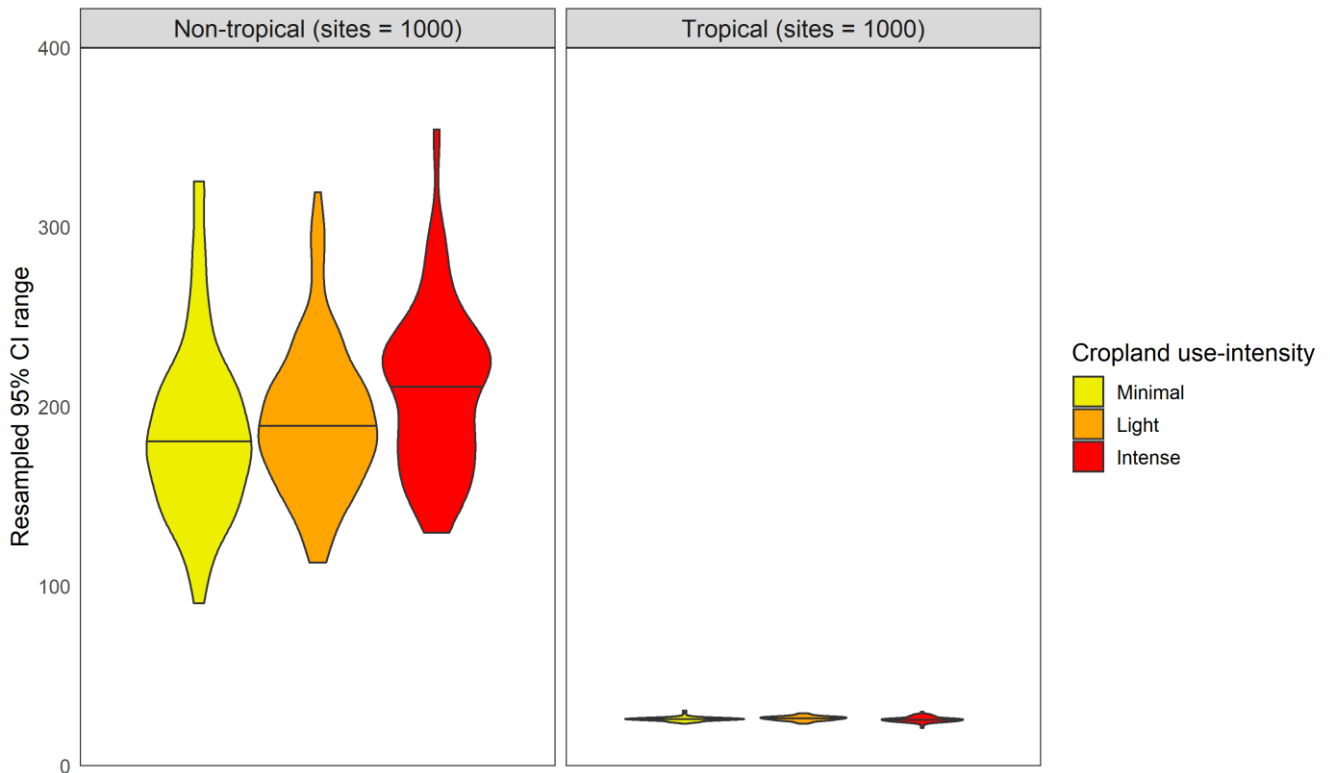


Figure S2.7 Resampled 95% confidence interval ranges for total abundance for tropical and non-tropical sites. 1000 sites were resampled from each of the tropical and non-tropical sites a total of 100 times, and then for each group of 2000 (tropical and non-tropical), total abundance was fitted as a function of land-use intensity, geographical zone, and their interaction. Each violin represents the distribution of the 95% confidence interval size for all samples in each land-use intensity—geographical zone combination (the black line represents the median for all samples), indicating that there is greater variation in non-tropical responses even when sample size is controlled. For tropical sites, the distribution of the violin is close to the median since total site number (from which the sample is taken) is 1052, meaning each re-sample for the tropics is effectively all sites.

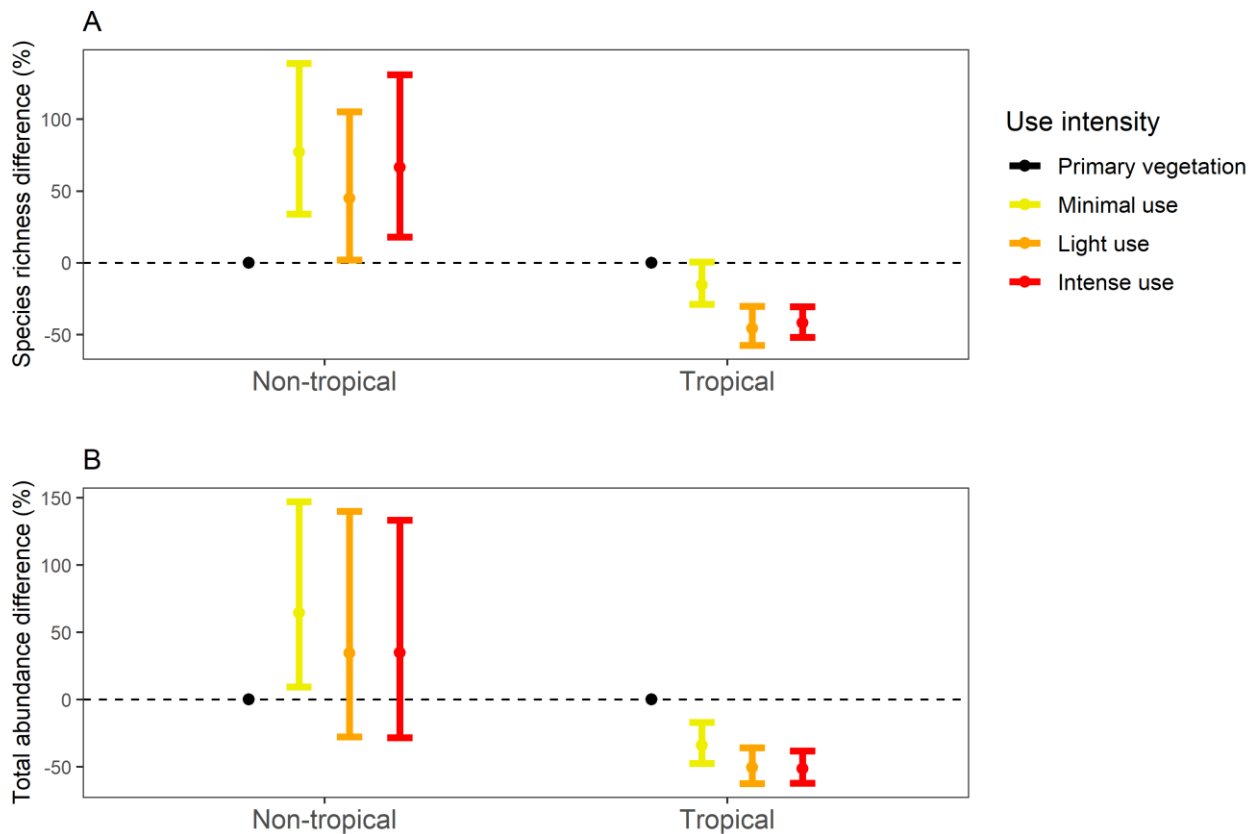


Figure S2.8 Response of the main crop pollinators (bees, wasps, beetles, thrips, flies, birds, and bats) to land-use intensity on cropland, for non-tropical and tropical sites. Each panel represents a linear or generalised linear mixed-effects model for a given biodiversity metric: A, species richness; and B, total abundance. Colours represent the land-use intensity level, with primary vegetation (minimal use) as the reference factor: black (primary vegetation, minimal use); yellow (cropland, minimal use), orange (cropland, light use), and red (cropland, intense use). Effect sizes were adjusted to a percentage by sampling fixed effects 1,000 times based on the variance-covariance matrix, expressing each fixed effect as a percentage of the value in primary vegetation for that geographical zone, and then calculating the median value (shown as points), and 2.5th and 97.5th percentiles (shown as error bars).

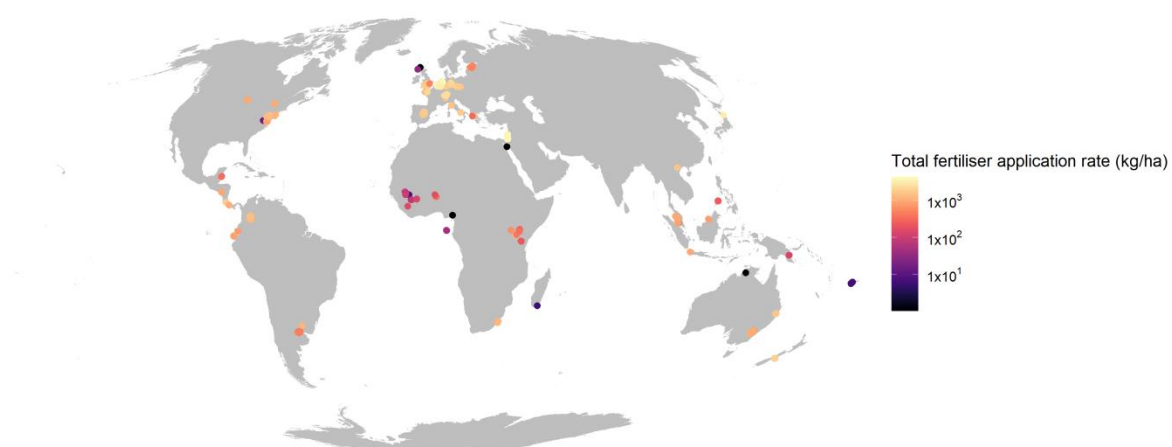


Figure S2.9 Site level total fertiliser application rate (+1) for cropland sites in PREDICTS that contain pollinating species, a pair of geographical coordinates, and a fertiliser application rate for that coordinate (n=1560).

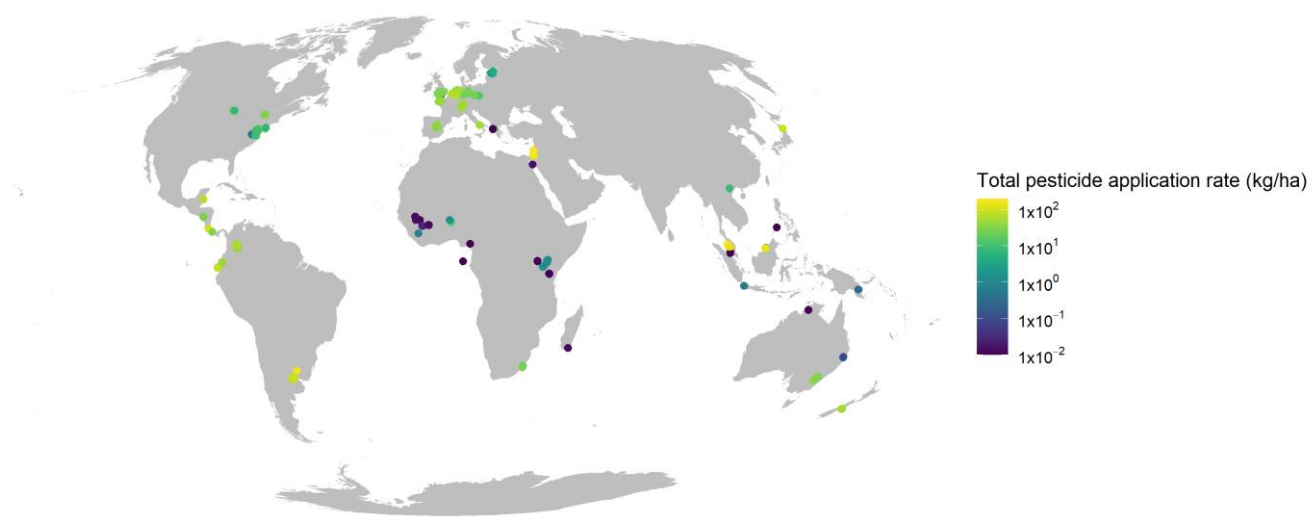


Figure S2.10 Site level total pesticide application rate (+0.01) for cropland sites in PREDICTS that contain pollinating species, a pair of geographical coordinates, and a pesticide application rate for that coordinate (n=1560).

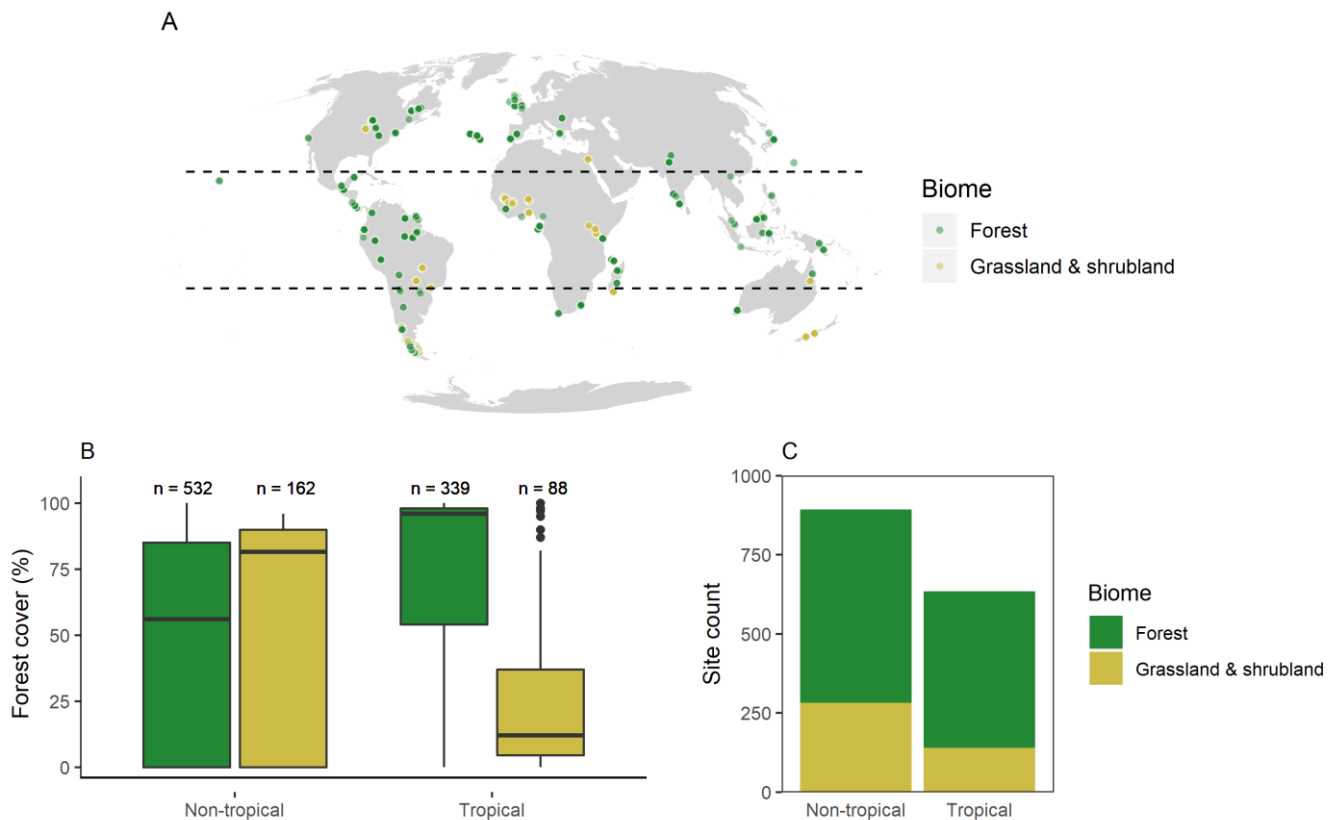


Figure S2.11 The structure of primary vegetation minimal use intensity sites in PREDICTS that contain pollinating species (i.e. the baseline for Figures 2, 3, and 4). A) The global geographic distribution of 'forest' (green) and 'grassland & shrubland' (yellow) sites according to the terrestrial ecoregions of the world (Olson et al 2001). The dashed lines represent latitudes of 23.5 and -23.5, which divide tropical from non-tropical regions. B) Boxplots for the distribution of forest cover (Hansen et al 2013) between 'grassland' and 'forest & shrubland' sites in the tropical and non-tropical zones. Here the box extends from the 25th to the 75th percentiles, the dark black inner line corresponds to the median, the upper and lower whiskers to 1.5 x IQR, and black dots to any sites beyond 1.5 x IQR. Sample size number for each grouping is represented above each boxplot. C) The frequency of 'forest' and 'grassland & shrubland' sites between the tropical and non-tropical zones.

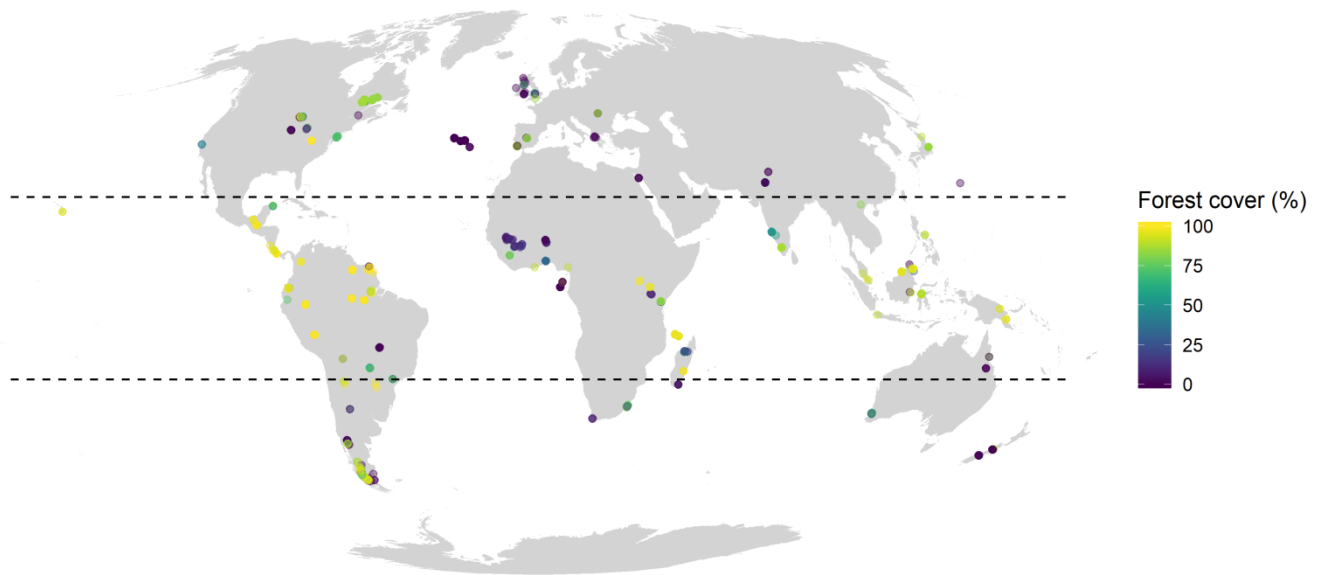


Figure S2.12 The global geographic distribution of forest cover (Hansen et al 2013) for primary vegetation minimal use intensity sites in PREDICTS that contain pollinating species. Here the dashed lines represent latitudes of 23.5 and -23.5, which divide tropical from non-tropical regions.

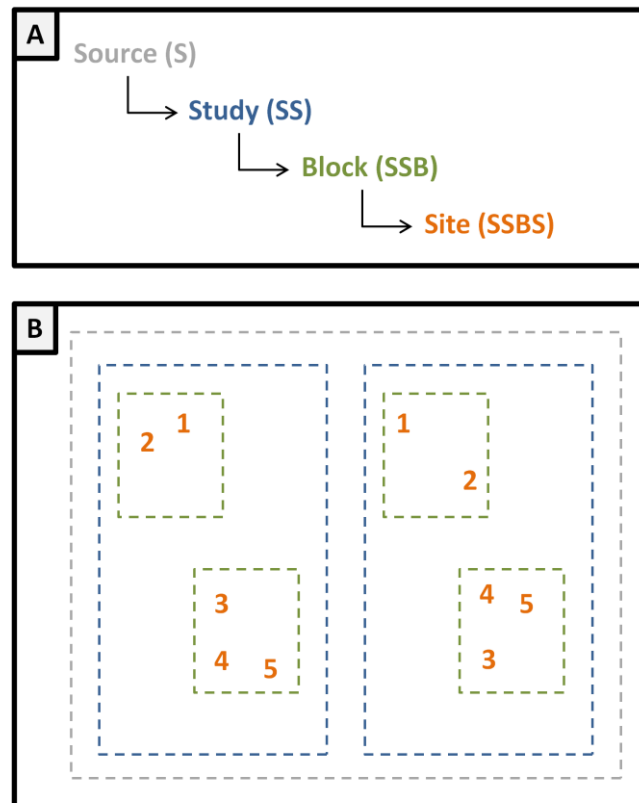


Figure S2.13 Schematic of the nested structure of the PREDICTS database. A) The PREDICTS database is nested at four levels: 'source', 'study', 'block', and 'site', represented in the database as the columns S, SS, SSB, and SSBS. A source represents a unique paper, a study an experiment within a source that has a consistent sampling methodology, a block a collection of sites in a distinct spatial cluster, and site a geographic location at which biodiversity was sampled. B) The structure of the PREDICTS database is such that there can be multiple studies, blocks, and sites nested within each other.

Supplementary Tables

Table S2.15 List of the 7 pollination ecologists consulted as a check-and-balance on our approach for identifying pollinating species in the PREDICTS database.

Pollination ecology expert	Taxonomic speciality	Continental speciality
Opeyemi Adedaja	Insects	Africa
Sabrina Gavini	Broad	South America
Esther Kioko	Insects	Africa
Michael Kuhlmann	Bees	Europe
Jeff Ollerton	Broad	Europe
Zong-Xin Ren	Insects	Asia
Manu Saunders	Insects	Australasia

Table S2.16 The total number of animal pollinating species in the PREDICTS database, aggregated by taxonomic order and class (see Figure 3.1).

Order	Class	N (species)
Lepidoptera	Insecta	2398
Hymenoptera	Insecta	988
Passeriformes	Aves	412
Coleoptera	Insecta	247
Diptera	Insecta	228
Apodiformes	Aves	99
Chiroptera	Mammalia	51
Psittaciformes	Aves	23
Soricomorpha	Mammalia	13
Primates	Mammalia	11
Dasyuromorphia	Mammalia	7
Rodentia	Mammalia	6
Squamata	Reptilia	5
Columbiformes	Aves	4
Didelphimorphia	Mammalia	3
Diprotodontia	Mammalia	3
Macroscelidae	Mammalia	2
Scandentia	Mammalia	1
Thysanoptera	Insecta	1

Table S2.17 ANOVA tables for species richness, $\log(\text{total abundance} + 1)$, and $\log(\text{Simpson diversity} + 1)$ predicted as a function of land use-type, land-use intensity, and their interaction, and the random intercepts study (SS), block (SSB), and site (SSBS). Statistics are for either a two-sided mixed-effects generalised linear model (species richness) or a two-sided mixed-effects linear model (total abundance and species richness). *P* values are not included here for species richness since deriving these for mixed effects generalised linear models is problematic.

Response variable	Explanatory variable	df	Sum Sq	Mean Sq	F value	P value
Species richness	Land-use type	7	62.608	8.9440	8.9440	-
	Land-use intensity	2	18.877	9.4384	9.4384	-
	Type * Intensity	12	99.648	8.3040	8.3040	-
Total abundance	Land-use type	7	32.173	4.5962	8.0346	9.196e-10
	Land-use intensity	2	5.500	2.7502	4.8075	0.008192
	Type * Intensity	12	80.646	6.7205	11.7481	< 2.2e-16
Simpson diversity	Land-use type	7	2.6035	0.37193	4.4150	6.648e-05
	Land-use intensity	2	1.9661	0.98303	11.6691	8.749e-06
	Type * Intensity	12	4.6201	0.38501	4.5703	2.122e-07

Table S2.18 AIC values for all the models presented in Figures 3.2, 3.3, 3.4, and 3.5, as well as the insignificant model fitting fertiliser and geographical zone as an interaction. Models in which AIC is lowest are indicated by †. Significant full models are highlighted in bold.

Figure	Fixed effect structure	AIC
LUI only models (Figure 3.2)	Richness ~ LUI	33734.62 †
	Richness ~ 1	33871.57
	Abundance ~ LUI	20110.66 †
	Abundance ~ 1	20237.72
	Diversity ~ LUI	3698.13
	Diversity ~ 1	3641.20 †
LUI * zone models (Figure 3.3)	Richness ~ LUI * zone	13318.63 †
	Richness ~ LUI	13333.56
	Richness ~ 1	13375.75
	Abundance ~ LUI * zone	8093.51 †
	Abundance ~ LUI	8101.58
	Abundance ~ 1	8126.23
	Diversity ~ LUI * zone	1841.00
	Diversity ~ LUI	1830.72 †
	Diversity ~ 1	1843.73
Fertiliser * zone models (no main text figure)	Richness ~ fertiliser * zone	12757.00
	Richness ~ fertiliser	12756.18 †
	Richness ~ 1	12833.61
	Abundance ~ fertiliser * zone	9903.95
	Abundance ~ fertiliser	9902.60 †
	Abundance ~ 1	9967.06
	Diversity ~ fertiliser * zone	1293.94
	Diversity ~ fertiliser	1285.62 †
	Diversity ~ 1	1289.99

LUI * order models (Figure 3.4)	Richness ~ LUI * order	14838.27 †
	Richness ~ LUI	15957.05
	Richness ~ 1	16002.50
	Abundance ~ LUI * order	10604.53 †
	Abundance ~ LUI	11132.19
	Abundance ~ 1	11156.47
	Diversity ~ LUI * order	1808.68 †
	Diversity ~ LUI	1817.79
	Diversity ~ 1	1829.32
Fertiliser * order models (Figure 3.5)	Richness ~ fertiliser * order	10520.86 †
	Richness ~ fertiliser	11599.57
	Richness ~ 1	11598.16
	Abundance ~ fertiliser * order	8034.78 †
	Abundance ~ fertiliser	8508.81
	Abundance ~ 1	8505.94
	Diversity ~ fertiliser * order	1325.33 †
	Diversity ~ fertiliser	1345.22
	Diversity ~ 1	1339.64

Table S2.19 Overall site representation for Figure 3.2, the overall response of pollinator biodiversity to LUI (land-use intensity and land-use type combined). The land-use types categories are as follows: PV (Primary vegetation), MSV (mature secondary vegetation), ISV (intermediate secondary vegetation), YSV (young secondary vegetation), PF (plantation forest) P (pasture), C (cropland), and U (urban). The intensity categories are as follows: MU (minimal use), LU (light use), and IU (intense use).

LUI	N (sites)
PVMU	1560
PVLU	996
PVIU	381
MSVMU	212
MSVLU	179
ISVMU	276
ISVLU	155
YSVMU	312
YSVLU	97
YSVIU	260
PFMU	214
PFLU	674
PFIU	110
PMU	446
PLU	532
PIU	122
CMU	426
CLU	649
CIU	658
UMU	173
ULU	136
UIU	43

Table S2.20 Model summary for species richness predicted as a function of a combined factor for land-use intensity and type (LUI) and the random intercepts study (SS), block (SSB), and site (SSBS) with a two-sided mixed-effects generalised linear model (see Figure 3.2 for predicted values).

Term	estimate	std.error	statistic	p.value
(Intercept)	1.32896397	0.09640307	13.7854939	3.116264e-43
LUI-PVLU	0.16226839	0.03122186	5.1972684	2.022382e-07
LUI-PVIU	0.14887082	0.05095798	2.9214428	3.484143e-03
LUI-MSVMU	0.06379620	0.05242243	1.2169638	2.236180e-01
LUI-MSVLU	0.21482279	0.06615476	3.2472766	1.165151e-03
LUI-ISVMU	0.15557023	0.04056605	3.8349856	1.255717e-04
LUI-ISVLU	0.20433055	0.05091357	4.0132829	5.988009e-05
LUI-YSVMU	0.02190840	0.03809628	0.5750797	5.652374e-01
LUI-YSVLU	-0.11002557	0.08427108	-1.3056149	1.916835e-01
LUI-YSVIU	-0.17073604	0.05050497	-3.3805792	7.233322e-04
LUI-PFMU	0.24269714	0.05451398	4.4520166	8.506757e-06
LUI-PFLU	0.05463885	0.03714311	1.4710359	1.412814e-01
LUI-PFIU	-0.11853355	0.06879615	-1.7229678	8.489436e-02
LUI-PMU	0.15678740	0.04408967	3.5561029	3.763967e-04
LUI-PLU	-0.03122227	0.04119039	-0.7579990	4.484516e-01
LUI-PIU	0.01590149	0.06782871	0.2344359	8.146466e-01
LUI-CMU	-0.01540370	0.04786049	-0.3218458	7.475695e-01
LUI-CLU	-0.11717122	0.05037602	-2.3259323	2.002216e-02
LUI-CIU	-0.13700839	0.04998200	-2.7411547	6.122367e-03
LUI-UMU	0.23228075	0.06259899	3.7106150	2.067564e-04
LUI-ULU	0.03215269	0.06336714	0.5074032	6.118719e-01
LUI-UIU	-0.22080354	0.08463465	-2.6089024	9.083313e-03

Table S2.21 Model summary for $\log(\text{total abundance} + 1)$ predicted as a function of a combined factor for land-use intensity and type (LUI) and the random intercepts study (SS) and block (SSB) with a two-sided mixed-effects linear model (see Figure 3.2 for predicted values).

term	estimate	std.error	statistic	p.value
(Intercept)	3.066663564	0.13047422	23.50398104	1.441767e-66
LUI-PVLU	0.328309492	0.04936891	6.65012621	3.125781e-11
LUI-PVIU	0.371272193	0.07205361	5.15272138	2.631861e-07
LUI-MSVMU	0.026126195	0.08501411	0.30731599	7.586111e-01
LUI-MSVLU	0.361740682	0.09563360	3.78256885	1.563882e-04
LUI-ISVMU	0.217422866	0.07366012	2.95170376	3.169719e-03
LUI-ISVLU	0.310991569	0.07822650	3.97552697	7.089066e-05
LUI-YSVMU	0.149413495	0.07011363	2.13101940	3.311893e-02
LUI-YSVLU	-0.008061529	0.13255730	-0.06081543	9.515078e-01
LUI-YSVIU	0.110668437	0.08303010	1.33287130	1.826132e-01
LUI-PFMU	0.362561497	0.08599708	4.21597432	2.516897e-05
LUI-PFLU	0.063777193	0.06055574	1.05319819	2.922826e-01
LUI-PFIU	0.189445958	0.12859996	1.47314167	1.407558e-01
LUI-PMU	0.584565062	0.05851292	9.99035868	2.329859e-23
LUI-PLU	0.254135112	0.05423478	4.68583282	2.837077e-06
LUI-PIU	0.041868650	0.10917168	0.38351202	7.013506e-01
LUI-CMU	0.006691476	0.06373908	0.10498232	9.163925e-01
LUI-CLU	-0.066291107	0.07378357	-0.89845354	3.689713e-01
LUI-CIU	-0.217808927	0.07491060	-2.90758476	3.653874e-03
LUI-UMU	0.424397518	0.09029529	4.70010712	2.644965e-06
LUI-ULU	0.408046211	0.12605275	3.23710689	1.212531e-03
LUI-UIU	-0.096260227	0.16519365	-0.58271142	5.601051e-01

Table S2.22 Cropland site representation for Figure 3.3, the overall response of pollinator biodiversity to land-use intensity between the non-tropical and tropical zones

Geographical zone	LUI	N (sites)
Non-tropical	Primary vegetation	893
Non-tropical	Minimal use cropland	245
Non-tropical	Light use cropland	492
Non-tropical	Intense use cropland	578
Tropical	Primary vegetation	634
Tropical	Minimal use cropland	181
Tropical	Light use cropland	157
Tropical	Intense use cropland	80

Table S2.23 Model output for species richness predicted as a function of a combined factor for land-use intensity and type (LUI), geographical zone (tropics/non-tropics), and the random intercepts study (SS), block (SSB), and site (SSBS) with a two-sided mixed-effects generalised linear model (see Figure 3.3 for predicted values).

Term	estimate	std.error	statistic	p.value
(Intercept)	1.2092244	0.1822545	6.634814	3.249122e-11
LUI-Minimal use	0.3735829	0.1303490	2.866020	4.156686e-03
LUI-Light use	0.1965426	0.1779420	1.104532	2.693624e-01
LUI-Intense use	0.3271289	0.1769412	1.848800	6.448668e-02
zone-Tropics	0.2388790	0.2471688	0.966461	3.338136e-01
LUI-Minimal use:zone-Tropics	-0.5307787	0.1477803	-3.591674	3.285612e-04
LUI-Light use:zone-Tropics	-0.6940822	0.2054538	-3.378289	7.293833e-04
LUI-Intense use:zone-Tropics	-0.9160952	0.1980057	-4.626611	3.716984e-06

Table S2.24 Model output for log(total abundance + 1) predicted as a function of a combined factor for land-use intensity and type (LUI), geographical zone (tropics/non-tropics), and the random intercepts study (SS) and block (SSB) with a two-sided mixed-effects linear model (see Figure 3.3 for predicted values).

Term	estimate	std.error	statistic	p.value
(Intercept)	2.7469928	0.2460986	11.1621626	8.118052e-23
LUI-Minimal use	0.4692762	0.1790959	2.6202502	8.831766e-03
LUI-Light use	0.3010843	0.2895618	1.0397928	2.990043e-01
LUI-Intense use	0.2816584	0.2848026	0.9889603	3.232719e-01
zone-Tropics	0.5182296	0.3380493	1.5330002	1.270054e-01
LUI-Minimal use:zone-Tropics	-0.8576550	0.2028786	-4.2274302	2.434102e-05
LUI-Light use:zone-Tropics	-0.9745301	0.3161887	-3.0821156	2.146717e-03
LUI-Intense use:zone-Tropics	-0.9607852	0.3106910	-3.0924140	2.083223e-03

Table S2.25 Cropland site representation for Figure 4, the response of 6 taxonomic orders (Hymenoptera, Diptera, Lepidoptera, Coleoptera, Passeriformes, and Apodiformes) to land-use intensity on cropland. Note that the total number of sites for each order-intensity group will not sum to the total of 3006, since some sites are represented across multiple groups.

Taxonomic order	LUI	N (sites)
Hymenoptera	Primary vegetation	186
Hymenoptera	Minimal use cropland	180
Hymenoptera	Light use cropland	564
Hymenoptera	Intense use cropland	578
Diptera	Primary vegetation	25
Diptera	Minimal use cropland	50
Diptera	Light use cropland	91
Diptera	Intense use cropland	72
Lepidoptera	Primary vegetation	317
Lepidoptera	Minimal use cropland	76
Lepidoptera	Light use cropland	112
Lepidoptera	Intense use cropland	44
Coleoptera	Primary vegetation	342
Coleoptera	Minimal use cropland	90
Coleoptera	Light use cropland	112
Coleoptera	Intense use cropland	44
Passeriformes	Primary vegetation	637
Passeriformes	Minimal use cropland	120
Passeriformes	Light use cropland	53
Passeriformes	Intense use cropland	61
Apodiformes	Primary vegetation	108
Apodiformes	Minimal use cropland	33
Apodiformes	Light use cropland	20

Table S2.26 Model output for species richness predicted as a function of taxonomic order and a combined factor for land-use intensity and type (LUI), and the random intercepts study (SS), block (SSB), and site (SSBS) with a two-sided mixed-effects generalised linear model (see Figure 3.4 for predicted values).

Term	estimate	std.error	statistic	p.value
(Intercept)	2.20840153	0.1991447	11.0894288	1.411863e-28
Order-Diptera	-0.83341036	0.1530930	-5.4438158	5.215114e-08
Order-Lepidoptera	-1.13968603	0.1506288	-7.5661918	3.843246e-14
Order-Coleoptera	-1.55508820	0.1592342	-9.7660458	1.574787e-22
Order-Passeriformes	-1.14836510	0.2985974	-3.8458648	1.201280e-04
Order-Apodiformes	-0.93317182	0.3072722	-3.0369547	2.389814e-03
LUI-Minimal use	0.37282903	0.2126349	1.7533761	7.953747e-02
LUI-Light use	-0.17614884	0.1930156	-0.9126147	3.614452e-01
LUI-Intense use	-0.04344631	0.1967355	-0.2208362	8.252200e-01
Order-Diptera:LUI-Minimal use	-3.18780581	0.4359546	-7.3122431	2.627192e-13
Order-Lepidoptera:LUI-Minimal use	-0.69525855	0.2226143	-3.1231526	1.789249e-03
Order-Coleoptera:LUI-Minimal use	-0.59864325	0.2378589	-2.5168001	1.184260e-02
Order-Passeriformes:LUI-Minimal use	-0.15672200	0.2441114	-0.6420102	5.208665e-01
Order-Apodiformes:LUI-Minimal use	-0.54035561	0.2681862	-2.0148525	4.392011e-02
Order-Diptera:LUI-Light use	-0.18953673	0.1934202	-0.9799221	3.271246e-01
Order-Lepidoptera:LUI-Light use	-0.54235377	0.1986485	-2.7302187	6.329232e-03
Order-Coleoptera:LUI-Light use	-0.48525618	0.2173994	-2.2320957	2.560864e-02
Order-Passeriformes:LUI-Light use	0.07107473	0.2844426	0.2498737	8.026850e-01
Order-Apodiformes:LUI-Light use	-0.77090778	0.3367013	-2.2895892	2.204514e-02
Order-Diptera:LUI-Intense use	-0.91597731	0.1926870	-4.7537046	1.997228e-06
Order-Lepidoptera:LUI-Intense use	-0.91324565	0.2345930	-3.8928932	9.905577e-05
Order-Coleoptera:LUI-Intense use	-0.27001790	0.2304695	-1.1715992	2.413580e-01
Order-Passeriformes:LUI-Intense use	-0.40540007	0.2270651	-1.7853916	7.419781e-02

Table S2.27 Model output for $\log(\text{total abundance} + 1)$ predicted as a function of taxonomic order and a combined factor for land-use intensity and type (LUI), and the random intercepts study (SS) and block (SSB) with a two-sided mixed-effects linear model (see Figure 3.4 for predicted values).

term	estimate	std.error	statistic	p.value
(Intercept)	4.73922131	0.2829571	16.7489059	6.296397e-47
Order-Diptera	-1.44912912	0.2550315	-5.6821574	1.438105e-08
Order-Lepidoptera	-1.73354917	0.2270221	-7.6360361	2.999194e-14
Order-Coleoptera	-1.80793239	0.2292210	-7.8872900	4.291811e-15
Order-Passeriformes	-2.30268897	0.4118455	-5.5911479	7.372749e-08
Order-Apodiformes	-2.11948413	0.4227288	-5.0138155	1.094958e-06
LUI-Minimal use	-1.10153763	0.2873542	-3.8333794	1.291541e-04
LUI-Light use	-1.69955568	0.2656710	-6.3972204	1.875913e-10
LUI-Intense use	-1.40296862	0.2701440	-5.1934097	2.223549e-07
Order-Diptera:LUI-Minimal use	-0.99823141	0.3047054	-3.2760539	1.062431e-03
Order-Lepidoptera:LUI-Minimal use	0.12833239	0.2894333	0.4433920	6.575106e-01
Order-Coleoptera:LUI-Minimal use	0.43580445	0.2891418	1.5072346	1.318455e-01
Order-Passeriformes:LUI-Minimal use	1.32596000	0.3318642	3.9954900	6.598938e-05
Order-Apodiformes:LUI-Minimal use	0.92948978	0.3435239	2.7057499	6.848714e-03
Order-Diptera:LUI-Light use	1.03211340	0.2820608	3.6591877	2.566325e-04
Order-Lepidoptera:LUI-Light use	0.39551699	0.2543970	1.5547232	1.201104e-01
Order-Coleoptera:LUI-Light use	0.44574567	0.2562287	1.7396399	8.201790e-02
Order-Passeriformes:LUI-Light use	1.69308440	0.3458009	4.8961252	1.025282e-06
Order-Apodiformes:LUI-Light use	0.98890956	0.3481864	2.8401732	4.536227e-03
Order-Diptera:LUI-Intense use	-0.18244690	0.2881395	-0.6331894	5.266497e-01
Order-Lepidoptera:LUI-Intense use	-0.16831841	0.2791143	-0.6030447	5.465186e-01
Order-Coleoptera:LUI-Intense use	0.05306171	0.2808995	0.1888993	8.501829e-01
Order-Passeriformes:LUI-Intense use	1.05068214	0.3225579	3.2573444	1.136272e-03

Table S2.28 Model output for $\log(\text{Simpson diversity} + 1)$ predicted as a function of taxonomic order and a combined factor for land-use intensity and type (LUI), and the random intercepts study (SS) and block (SSB) with a two-sided mixed-effects linear model (see Figure 3.4 for predicted values).

Term	estimate	std.error	statistic	p.value
(Intercept)	1.58383270	0.09627049	16.4519030	2.162291e-45
Order-Diptera	-0.06544099	0.09414709	-0.6950930	4.870580e-01
Order-Lepidoptera	-0.16830797	0.08304914	-2.0266070	4.282549e-02
Order-Coleoptera	-0.27585315	0.08486779	-3.2503870	1.170032e-03
Order-Passeriformes	-0.34473774	0.13620897	-2.5309473	1.217500e-02
Order-Apodiformes	-0.25772117	0.14095948	-1.8283351	6.885026e-02
LUI-Minimal use	0.12659585	0.11562792	1.0948553	2.738108e-01
LUI-Light use	-0.08470655	0.10185990	-0.8315986	4.057828e-01
LUI-Intense use	-0.03002987	0.10306170	-0.2913776	7.708067e-01
Order-Diptera:LUI-Minimal use	-0.50068423	0.16753455	-2.9885431	2.828102e-03
Order-Lepidoptera:LUI-Minimal use	-0.09384230	0.13338019	-0.7035700	4.817930e-01
Order-Coleoptera:LUI-Minimal use	-0.20716970	0.12646894	-1.6381074	1.015956e-01
Order-Passeriformes:LUI-Minimal use	-0.02003844	0.12992647	-0.1542290	8.774498e-01
Order-Apodiformes:LUI-Minimal use	-0.10811138	0.13493663	-0.8012011	4.231267e-01
Order-Diptera:LUI-Light use	-0.15765579	0.10769809	-1.4638680	1.433469e-01
Order-Lepidoptera:LUI-Light use	-0.10832984	0.10223874	-1.0595772	2.894413e-01
Order-Coleoptera:LUI-Light use	-0.10885346	0.10298404	-1.0569935	2.906226e-01
Order-Passeriformes:LUI-Light use	0.05550724	0.13072949	0.4245962	6.711757e-01
Order-Apodiformes:LUI-Light use	-0.13518246	0.14040177	-0.9628259	3.357357e-01
Order-Diptera:LUI-Intense use	-0.11398265	0.11574556	-0.9847691	3.248249e-01
Order-Lepidoptera:LUI-Intense use	-0.26533172	0.11391675	-2.3291720	1.992766e-02
Order-Coleoptera:LUI-Intense use	0.06450120	0.11161580	0.5778859	5.633927e-01
Order-Passeriformes:LUI-Intense use	-0.26787298	0.11890101	-2.2529076	2.438168e-02

Table S2.29 Cropland site representation for Figure 5, the response of 6 taxonomic orders (Hymenoptera, Diptera, Lepidoptera, Coleoptera, Passeriformes, and Apodiformes) to total fertiliser application rate (kg/ha) on cropland. Note that the total number of sites for each order-intensity group will not sum to the total of 2190, since some sites are represented across multiple groups.

Taxonomic order	N (sites)
Hymenoptera	1355
Passeriformes	683
Lepidoptera	258
Coleoptera	246
Apodiformes	244
Diptera	213

Table S2.30 Model output for species richness predicted as a function of $\log_{10}(\text{fertiliser application rate})$, taxonomic order, and the random intercepts study (SS), block (SSB), and site (SSBS) with a two-sided mixed-effects generalised linear model (see Figure 3.5 for predicted values.

term	estimate	std.error	statistic	p.value
(Intercept)	-4.952880	1.9133878	-2.588539	9.638395e-03
$\log_{10}(\text{fert})$	1.763928	0.6462196	2.729610	6.340926e-03
Order-Coleoptera	4.066850	2.1788416	1.866519	6.196874e-02
Order-Diptera	-6.581990	2.2395758	-2.938945	3.293316e-03
Order-Hymenoptera	7.782467	1.9674357	3.955640	7.633008e-05
Order-Lepidoptera	6.816010	2.1208452	3.213818	1.309829e-03
Order-Passeriformes	5.717434	1.7523151	3.262789	1.103217e-03
$\log_{10}(\text{fert})$:Order-Coleoptera	-1.397608	0.7389379	-1.891374	5.857445e-02
$\log_{10}(\text{fert})$:Order-Diptera	2.184480	0.7458250	2.928945	3.401149e-03
$\log_{10}(\text{fert})$:Order-Hymenoptera	-2.044201	0.6595253	-3.099504	1.938449e-03
$\log_{10}(\text{fert})$:Order-Lepidoptera	-2.372238	0.7163021	-3.311784	9.270311e-04
$\log_{10}(\text{fert})$:Order-Passeriformes	-1.753834	0.5951118	-2.947067	3.208041e-03

Table S2.31 Model output for $\log(\text{total abundance} + 1)$ predicted as a function of $\log_{10}(\text{fertiliser application rate})$, taxonomic order, and the random intercepts study (SS) and block (SSB) with a two-sided mixed-effects linear model (see Figure 3.5 for predicted values.

Term	estimate	std.error	statistic	p.value
(Intercept)	1.54088771	1.6289045	0.94596564	3.443103e-01
$\log_{10}(\text{fert})$	-0.05127043	0.5798805	-0.08841551	9.295537e-01
Order-Coleoptera	1.73923685	2.0015144	0.86896046	3.852573e-01
Order-Diptera	-8.43201071	1.9876442	-4.24221343	2.316702e-05
Order-Hymenoptera	4.91168470	1.8050664	2.72105487	6.581797e-03
Order-Lepidoptera	3.02481460	1.9954106	1.51588579	1.300592e-01
Order-Passeriformes	0.37425442	1.3446391	0.27833076	7.807799e-01
$\log_{10}(\text{fert})$:Order-Coleoptera	-0.41448339	0.6991881	-0.59280667	5.534798e-01
$\log_{10}(\text{fert})$:Order-Diptera	3.07974048	0.6879237	4.47686346	7.887758e-06
$\log_{10}(\text{fert})$:Order-Hymenoptera	-0.99775501	0.6265387	-1.59248750	1.114068e-01
$\log_{10}(\text{fert})$:Order-Lepidoptera	-0.88228527	0.6962327	-1.26722749	2.053937e-01
$\log_{10}(\text{fert})$:Order-Passeriformes	-0.03415814	0.4917438	-0.06946330	9.446260e-01

Table S2.32 Model output for $\log(\text{Simpson diversity} + 1)$ predicted as a function of $\log_{10}(\text{fertiliser application rate})$, taxonomic order, and the random intercepts study (SS) and block (SSB) with a two-sided mixed-effects linear model (see Figure 3.5 for predicted values.

term	estimate	std.error	statistic	p.value
(Intercept)	0.31314652	0.9728151	0.32189726	0.7475667
$\log_{10}(\text{fert})$	0.31316180	0.3390191	0.92372904	0.3557465
Order-Coleoptera	0.06297832	1.0489090	0.06004174	0.9521328
Order-Diptera	-0.75504668	1.5538186	-0.48592974	0.6270741
Order-Hymenoptera	1.39980252	1.0140713	1.38037886	0.1676404
Order-Lepidoptera	1.25212031	1.0516298	1.19064739	0.2339989
Order-Passeriformes	0.78924098	0.9157879	0.86181636	0.3889024
$\log_{10}(\text{fert})$:Order-Coleoptera	-0.05337955	0.3650140	-0.14623974	0.8837564
$\log_{10}(\text{fert})$:Order-Diptera	0.23220352	0.5204045	0.44619817	0.6555056
$\log_{10}(\text{fert})$:Order-Hymenoptera	-0.39618381	0.3504534	-1.13048915	0.2584168
$\log_{10}(\text{fert})$:Order-Lepidoptera	-0.46442622	0.3650117	-1.27235982	0.2034519
$\log_{10}(\text{fert})$:Order-Passeriformes	-0.26699620	0.3178015	-0.84013513	0.4009433

Table S2.33 Pseudo R squared for all main models included in the main text.

	Model	Conditional pseudo R squared	Marginal pseudo R squared
Figure 3.2	Richness ~ LUI	0.706	0.004
	Abundance ~ LUI	0.876	0.009
Figure 3.3	Richness ~ LUI * zone	0.714	0.009
	Abundance ~ LUI * zone	0.887	0.010
Figure 3.4	Richness ~ LUI * order	0.753	0.149
	Abundance ~ LUI * order	0.850	0.094
	Diversity ~ LUI * order	0.828	0.055
Figure 3.5	Richness ~ fertiliser * order	0.775	0.169
	Abundance ~ fertiliser * order	0.841	0.114
	Diversity ~ fertiliser * order	0.700	0.058

Table S2.34 Crops for which fertiliser application rate estimates are available in the Earthstat data, each with estimates of nitrogen, phosphorus, and potassium.

Crop
Barley
Cassava
Cotton
Groundnut
Maize
Millet
Oilpalm
Potato
Rapeseed
Rice
Rye
Sorghum
Soybean
Sugarbeet
Sugarcane
Sunflower
Wheat

Appendix 3: Chapter 4 Supplementary Information

Contents

Figure S3.1	234
Figure S3.2	235
Figure S3.3	236
Figure S3.4	237
Figure S3.5	238

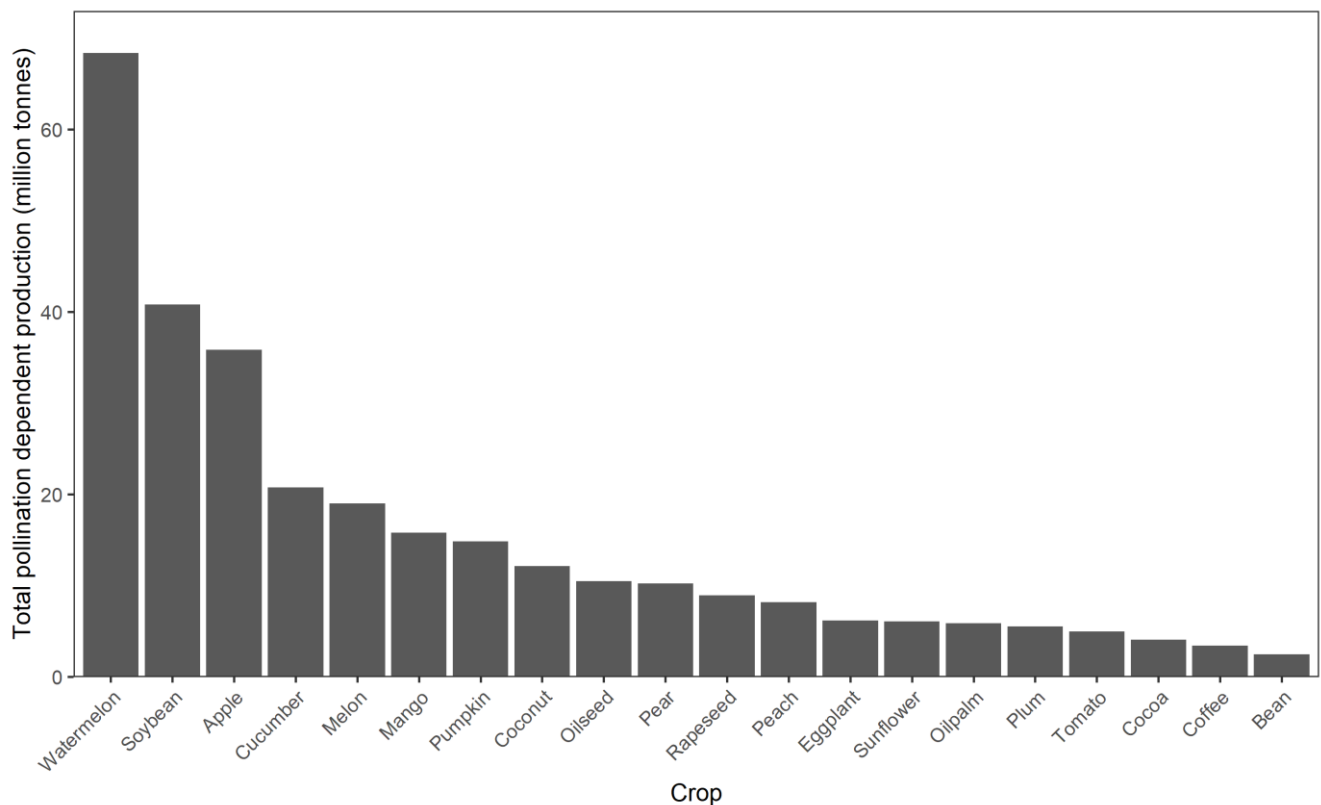


Figure S3.1 Total pollination dependent production for the 20 crops with the highest pollination dependent production. Total production values are for the year 2000, taken from Monfreda et al (2008). Pollination dependent production is calculated by multiplying total crop production for each crop by the pollination dependence ratios for that crop, as reported in Klein et al (2007). For any Monfreda crop represented by multiple dependence ratios, I took the pollination dependence to be the mean of the ratios for that crop.

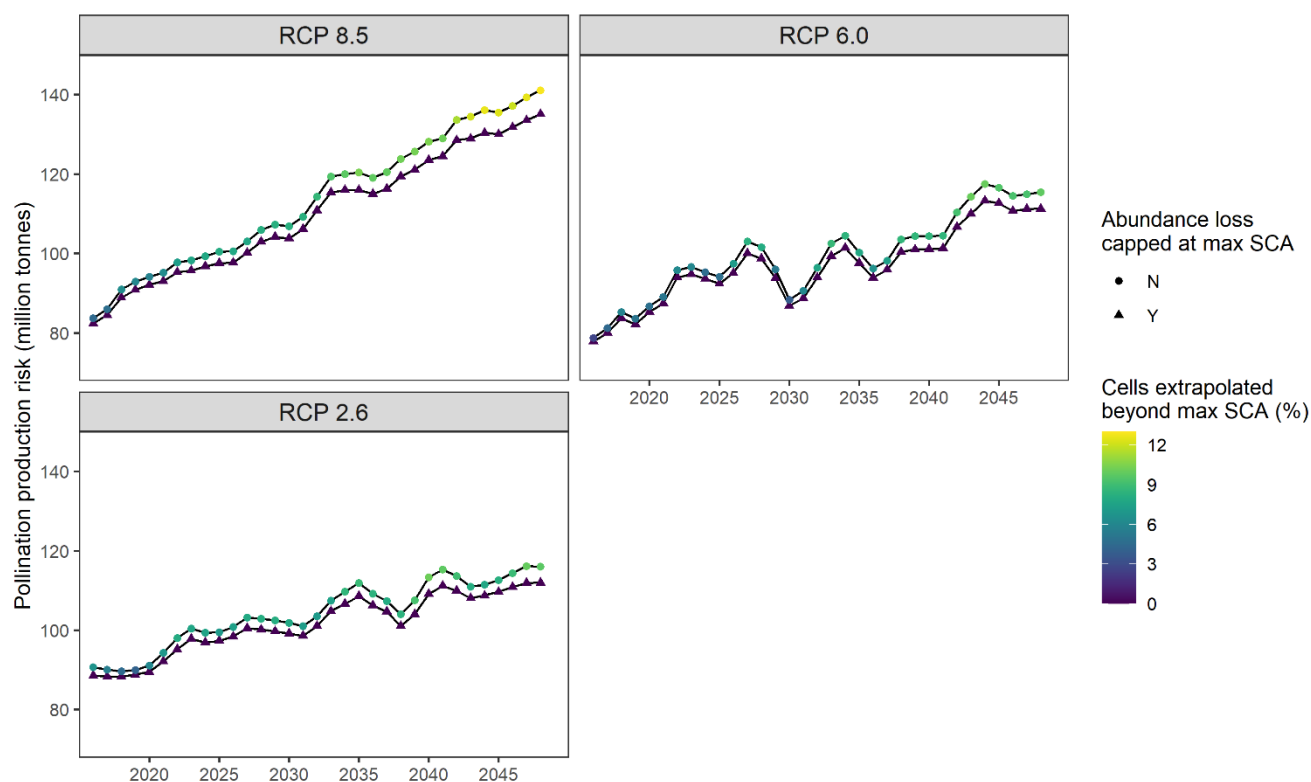


Figure S3.2 Projected change in total production risk under three RCP scenarios (8.5, 6.0, and 2.6), for the average of four climate models (GFDL, HadHEM2, IPSL, and MIROC5). Here circles represent a projection in which abundance loss is extrapolated beyond the maximum standardised temperature anomaly (STA) (as in Figure 4.3), whereas triangles represent a projection in which abundance loss is capped at an STA of 1.58 (2dp). Colour refers to the percentage of cells at each time step that have been extrapolated beyond the maximum STA in Figure 4.3, which in the projection in which abundance loss is capped (triangles) will always be 0.

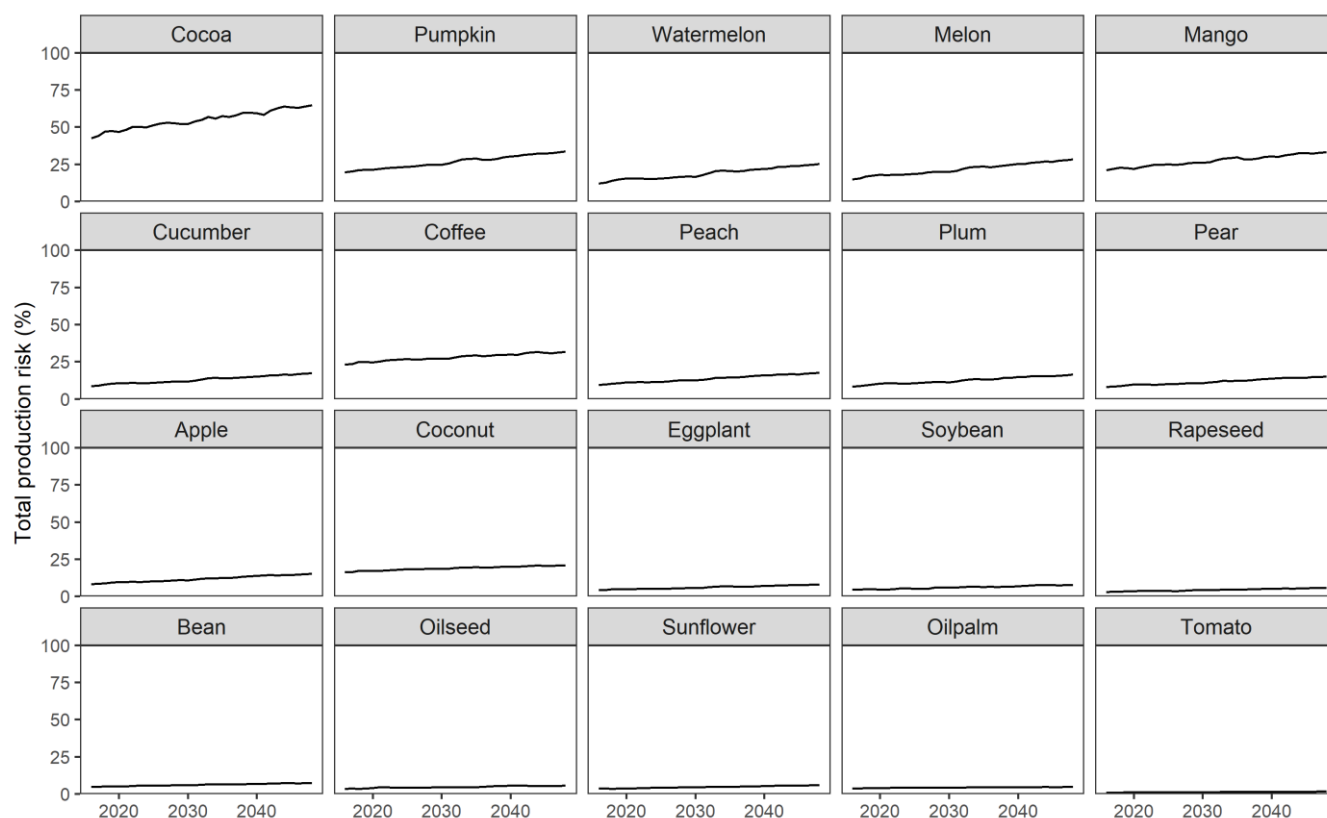


Figure S3.3 Change in proportional production risk for each crop, projected under RCP scenario 8.5 from the average of four climate models (GFDL, HadGEM2, IPSL, and MIROC5), for the top 20 crops by total pollination dependent production (ordered by rate of change). For each crop global standardised temperature anomaly was projected for all areas of pollination-dependent cropland to 2050, using the 3-year average approach as described in Figure 4.3. For each value of standardised temperature anomaly, insect pollinator abundance was predicted according to the model in Figure 1. Insect pollinator abundance at each cell at each time step was then expressed as proportional loss compared to cropland that has experienced no warming (i.e. standardised temperature anomaly of 0). Animal pollination-dependent production at each cell for each crop was then adjusted for the predicted proportional loss of insect pollinator abundance, and then summed for all cells at each time step. The sum at each time step was then divided by the total production for that crop and multiplied by 100, giving a percentage of total production at risk.

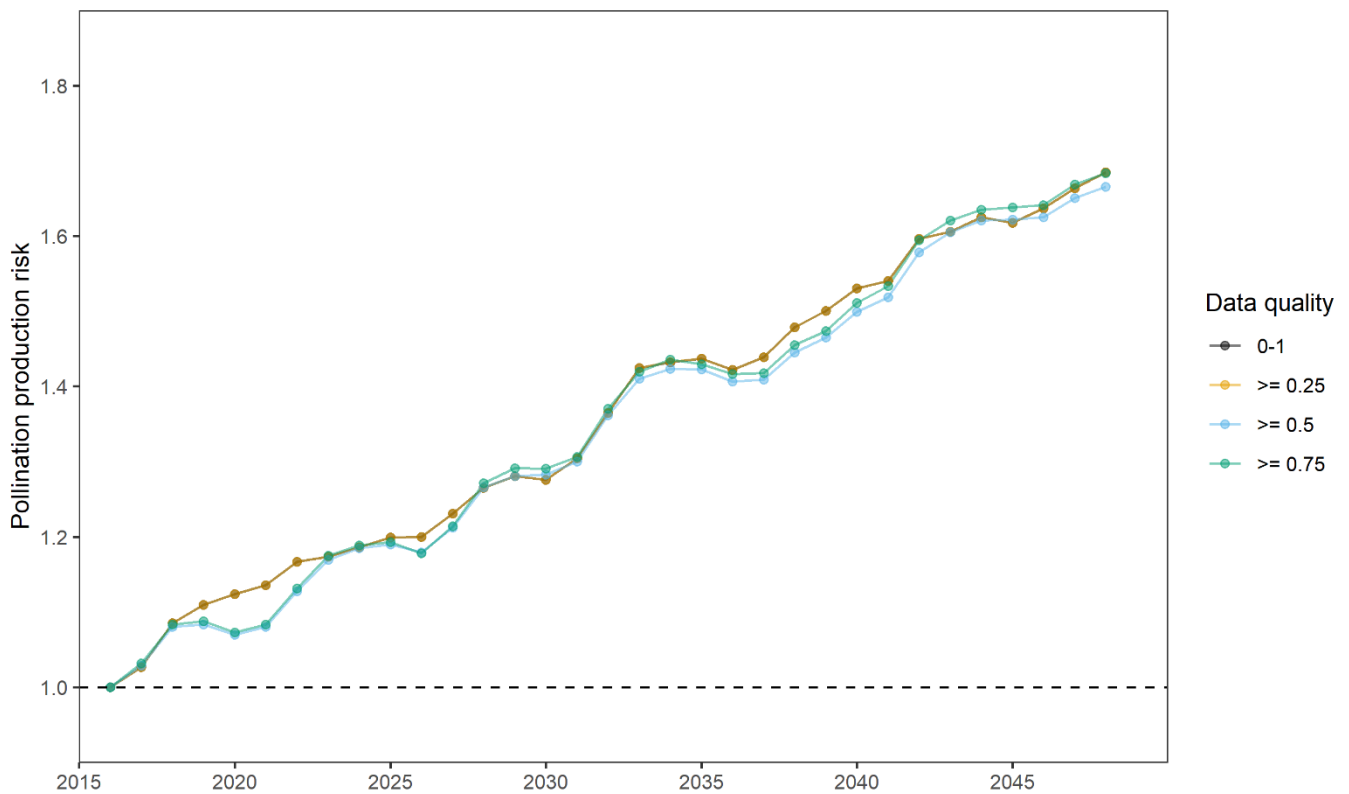


Figure S3.4 Change in production risk for the average of four climate models (GFDL, HadHEM2, IPSL, and MIROC5) under RCP 8.5, here adjusted to an index for each of 4 data quality subsets. For each coloured line crop production was subsetting according to four levels of data quality reported in Monfreda et al (2008), before the projection was then rerun: 1, county level census data; 0.75, state level census data; 0.5, regional interpolation from census data; 0.25, country level census data; 0, no census data. For each simulation here I converted total production risk to an index by calculating the percentage change between each time point, and then calculating the cumulative product of these percentage changes (starting at an initial index of 1).

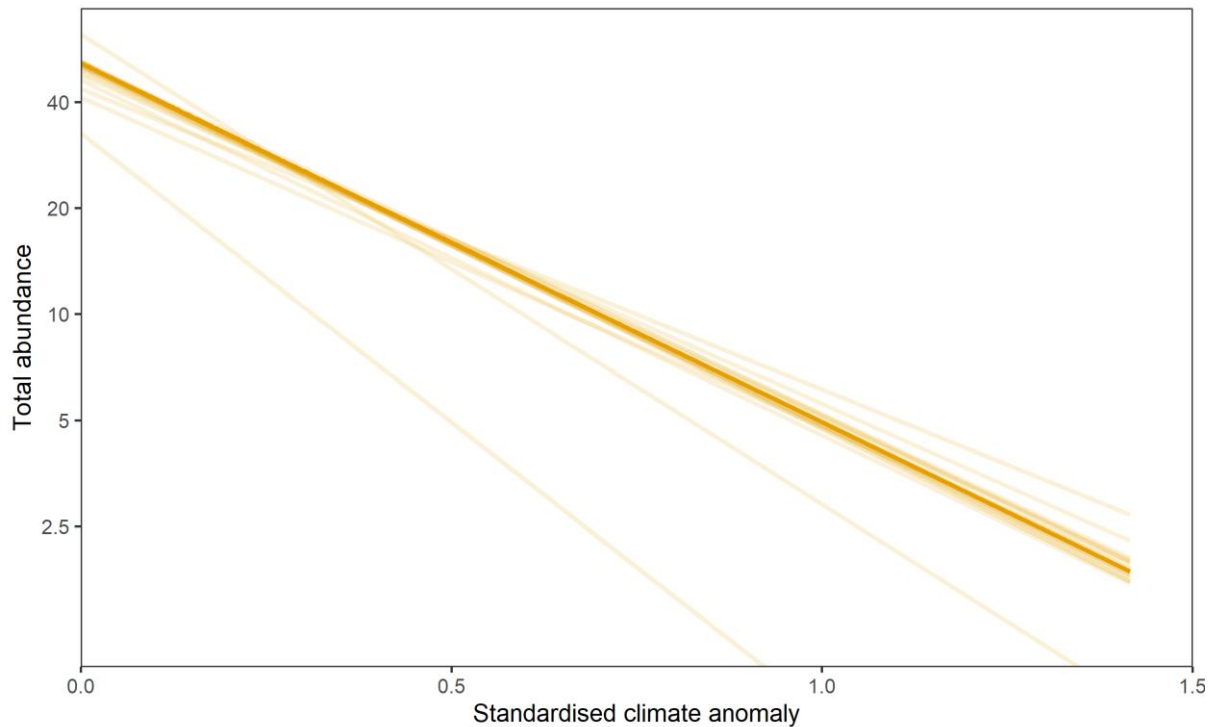


Figure S3.5 Response of pollinating dipteran and hymenopteran total abundance to standardised temperature anomaly on cropland (note that abundance is plotted on a $\log_e$ scale), predicted using a linear mixed-effects model. Each pale orange line represents the predicted change in total abundance with one family removed from the overall dataset (i.e. jack-knife removal with replacement). The darker orange line represents the overall predicted change in total abundance with all families included ($n = 61$).

Appendix 4: Chapter 5 Supplementary Information

Contents

Extended methodology for pollinator dataset construction	241
Supplementary figures	245
Figure S4.1.....	245
Figure S4.2.....	246
Figure S4.2.....	247
Figure S3.4.....	248
Figure S4.5.....	249
Figure S4.4.....	250
Figure S4.5.....	251
Figure S4.8.....	252
Figure S4.9.....	253
Figure S4.10.....	254
Figure S4.11.....	255
Figure S4.12.....	256
Figure S4.13.....	257
Figure S4.14.....	258
Figure S4.15.....	259
Figure S4.16.....	260
Figure S4.17.....	261
Supplementary tables	262
Table S4.18.....	262
Table S4.19.....	262
Table S4.20.....	263
Table S4.21.....	263
Table S4.22.....	264
Table S4.23.....	264
Table S4.24.....	265
Table S4.25.....	265
Table S4.26.....	266
Table S4.27.....	266
Table S4.28.....	267

Table S4.29	268
-------------------	-----

Extended methodology for pollinator dataset construction

I first created a list of possible pollinating animal genera through automatic text analysis of the pollination literature. I considered the pollination literature to be any primary research article published in English returned through a search for the term “pollinat*” in Scopus, and which mentioned an animal species in the abstract. I considered a possible pollinating genus to be any animal genus appearing as part of a Latin binomial in a pollination-related abstract returned from Scopus. Genera scraping was accomplished using the *Taxonfinder* and *Neti Neti* algorithms implemented in the ‘taxize’ R package (Chamberlain 2013), with animal species confirmed through a series of character string matches to the Catalogue of Life (see Millard et al 2020a for a detailed methodology).

For each possible pollinating genus, I then read the abstracts in which these animals appeared, searching for evidence confirming that genus as pollinating. For any situation in which the abstract was inconclusive, I also searched the full text of the paper for more definitive evidence. For each confirmed pollinating genus, I then assigned a level of experimental confidence between 1 and 4 based on the type of evidence (following (Ollerton and Liede 1997)): 1) experimental evidence confirming pollination; 2) evidence of pollen carrying; 3) evidence of nectar/pollen feeding; 4) evidence of non-destructive/non-predatory flower visitation. My process of assigning evidence was one of maximisation: I read abstracts for each genera searching for the highest level of evidence, either until I could be sure that the confidence value should be 1, or I ran out of abstracts for that genera. Non-destructive flower visitor refers to any animal which visits a flower without causing damage to the plant. This meant the exclusion of most ants, which are typically referred to as poor pollen vectors, given that they damage pollen through secretions from the meta-pleural gland (Dutton and Frederickson 2012). Non-predatory flower visitor refers to any animal which visits for some purpose other than predation. This meant the exclusion of animals such as crab spiders, which predate on pollinators during visitation, and therefore contribute

minimally or negatively to pollination (Dukas and Morse 2003). I did not classify broad statements as evidence for pollination—for example, one study stated that *Phylidonyris novaehollandiae* is a “key pollinator” (Myers et al 2012)—unless it was associated with specific evidence reinforcing that statement, or some claim that pollination in that genus is “well-known” or “widely acknowledged”.

Given that we only had direct evidence for a sample of all pollinating genera, I then searched for higher-level clades of pollinators. From the confirmed pollinators in the original list of genera, I identified all unique families with at least one pollinator. For each family, I assessed the breadth of evidence for pollination through consulting the abstracts and taxonomic reference books. For any family with evidence of pollination across multiple branches of that family, and no evidence of any species definitely not pollinating, I assumed that the whole family is pollinating. If unable to extrapolate across the whole family, I then searched progressively lower taxonomic clades (i.e. subfamily, tribe, subtribe), searching for the point at which I could be relatively confident that the entire clade contributes to pollination. If unable to extrapolate for a given clade, I kept only the genera with direct evidence. For example, within the family Formicidae (ants), I found no clade across which I could confidently extrapolate, and so kept only those genera for which we had direct evidence. I assigned a level of confidence for these extrapolated clade designations, which varied between 5.1 and 5.4, reflecting the quality of evidence available for most species in that clade. The ‘5’ indicates an extrapolated clade: 5.1, experimental evidence across multiple groups within that clade; 5.2, evidence of pollen carrying across multiple groups; 5.3, evidence of nectar/pollen feeding across multiple groups; 5.4, evidence of non-destructive/non-predatory flower visitation across multiple groups. After checking all families represented by the original list of genera, I then inspected additional resources, searching for any pollinating families I may have missed. Each of these resources included evidence of only nectar/pollen feeding or flower visitation, meaning all families extrapolated through additional resources were assigned a confidence level of either 5.3 or 5.4.

Weighting the Species Awareness Index (SAI)

In addition to the un-weighted index presented in the main text, I also explored 2 types of weighting approach (a sample weighting and a user weighting) for an overall SAI including all 10 Wikipedia languages and 41,197 IUCN species. My sample weighting accounts for species occurring in very few or many languages, whilst my user weighting adjusts either for the number of people using each language on the internet, or the number of views for each language on Wikipedia. Given the limitations inherent in each of these approaches, I present the results of weighting the SAI here. Future work should further explore the benefits of weighting the SAI, aiming to derive a metric both robust to the inclusion of languages and representative of the global population.

I explored a sample weighting to account for the uneven distribution of unique species in each language (see Figure S4.16). After averaging each species across languages, I then multiplied each averaged species by a weighting according to the number of languages in which that species occurred. A species occurring in one language was assigned a weighting of 0.1, a species in 2 languages a weighting of 0.2, up to a weighting of 1 for a species occurring in 10 languages. I then bootstrapped these sample weighted species using the approach described in the main text.

I explored a user weighting to derive a metric representative of the global population, aiming to increase the relevance of the SAI to global policy. For my internet user weighting, I converted internet users by language from the Internet World Stats (<https://www.internetworldstats.com/stats7.htm>) to a proportion, and then used that proportion to weight the mean for each species across languages. The Internet World Stats internet user data is calculated from the total number of speakers for each language (counting each speaker once), multiplied by the internet penetrance of the countries in which those people reside (Table S11). I supplemented the Internet World Stats data with data for the Italian language, which falls outside the top 10 languages on the internet. I took internet

penetrance data for Italy from Statista (www.statista.com), and population data for Italy from Istat (<https://www.istat.it/>). My values for the Italian language will therefore be a slight underestimate, since I do not include first language speakers outside of Italy. For my Wikipedia user weighting, I converted the total number of species views for each language to a proportion (Table S12), and then used those proportions to weight the mean for each species across languages.

Given my sample weighting adjusts *after* a species average is taken across languages, and our user weighting adjusts *during* the average, I was able to combine the approaches to explore their effect in combination and independently (Figure S4.16). Here the sample weighting approaches (left panel) drop markedly less at the start of the series, indicating that the high magnitude change caused by the French reptiles (see Figure S4.15, Figure 5.7, and 'Extended Discussion') has been mitigated. Similarly, both user weightings drop markedly less at the start of the series, with the internet user rating running marginally higher. Preliminarily, it looks like a sample weighted approach may be a feasible alternative to the exclusion of the French Wikipedia. More work is required to investigate the robustness of a sample weighted approach to language exclusion.

There are limitations associated with user weighting the SAI, which will need to be addressed before such an approach becomes workable. The primary problem for a user weight is the difficulty in extrapolating awareness for a given language to all speakers of that language. For example, weighting for the number of internet users of the Chinese (Mandarin) language makes the assumption that views predominantly coming from Taiwan and Hong Kong are representative of all of mainland China, which may be unlikely. Further work is required to determine the validity of extrapolating views for a given language to all speakers of that language.

Supplementary figures

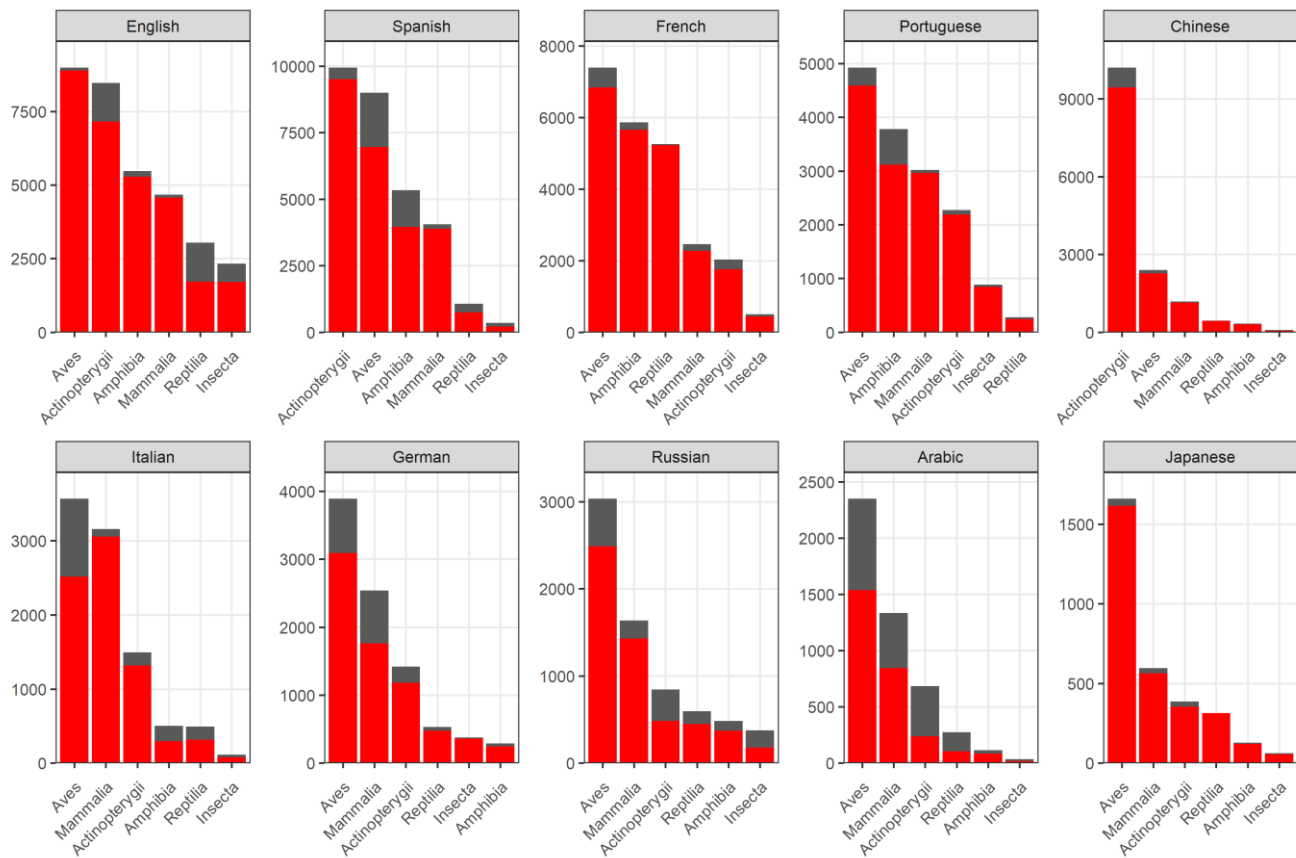


Figure S4.1 The number of complete Wikipedia view series for each taxonomic class in each language. Red bars represent species with complete series (i.e. data for each month over the period July 2015-March 2020), from which the SAI was calculated. Black bars represent the total species for each taxonomic class in each language, including species for which the series of views is incomplete.

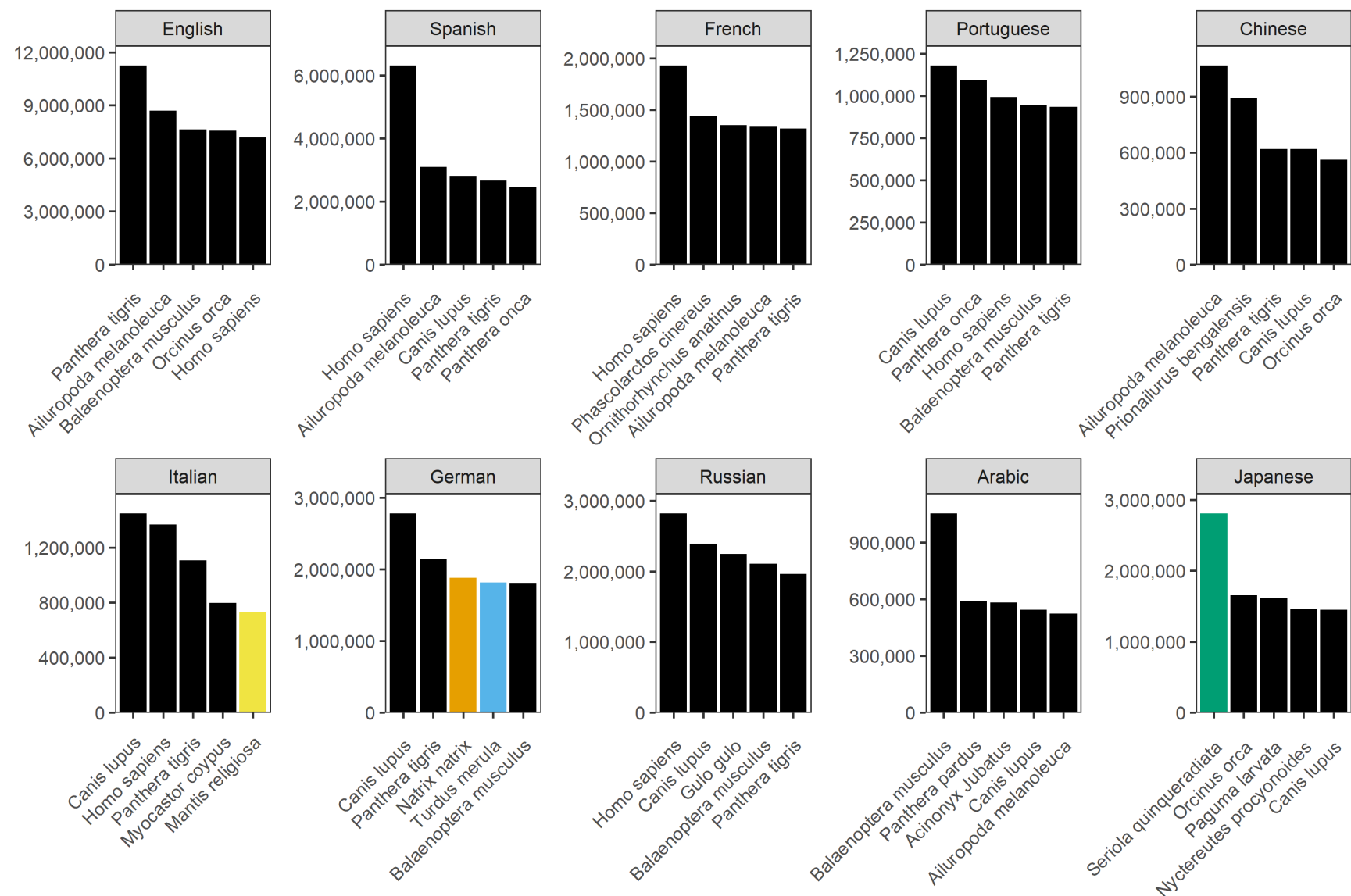


Figure S4.2 The top 5 viewed species in each Wikipedia language. Each colour fill refers to a taxonomic class, consistent across all 10 panels: black (mammals), yellow (insects), green (ray-finned fish), orange (reptiles), and blue (birds).

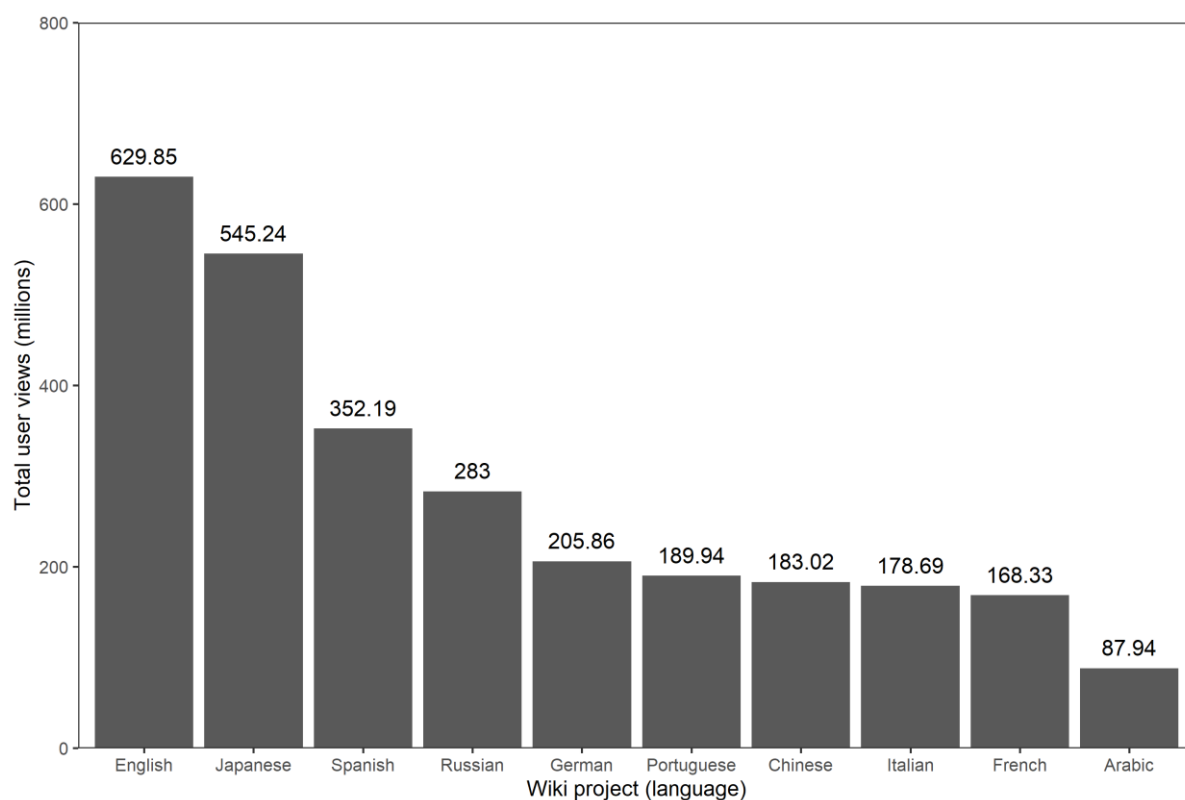


Figure S4.2 The total number of Wikipedia views for the set of random pages in each language, before removing incomplete series.

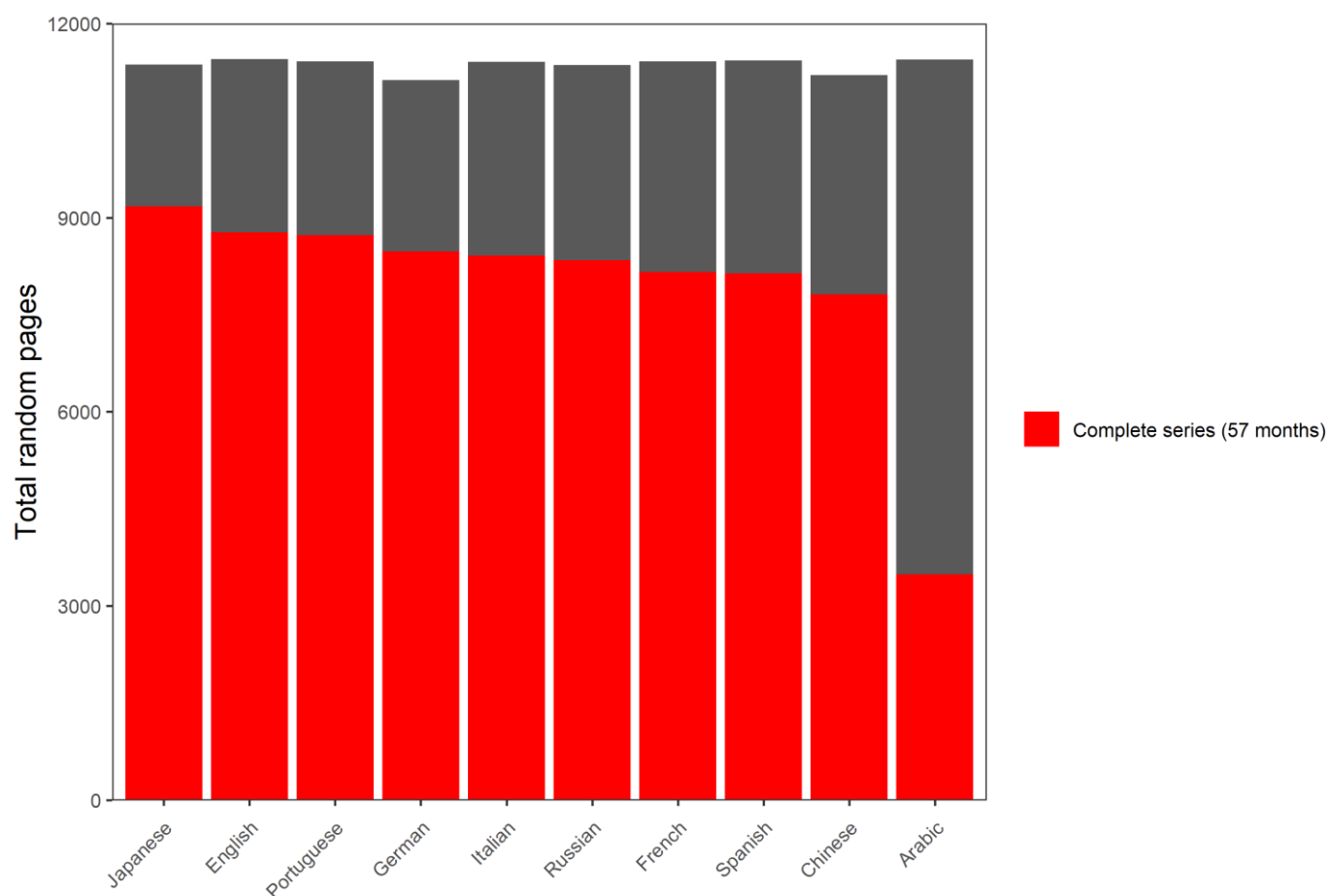


Figure S3.4 The total number of complete series pages for each set of random pages in each language. Red bars represent pages with complete series (i.e. data for each month over the period July 2015-March 2020), from which the random index was calculated. Black bars represent the total number of random pages in each language, including pages for which the series of views is incomplete.

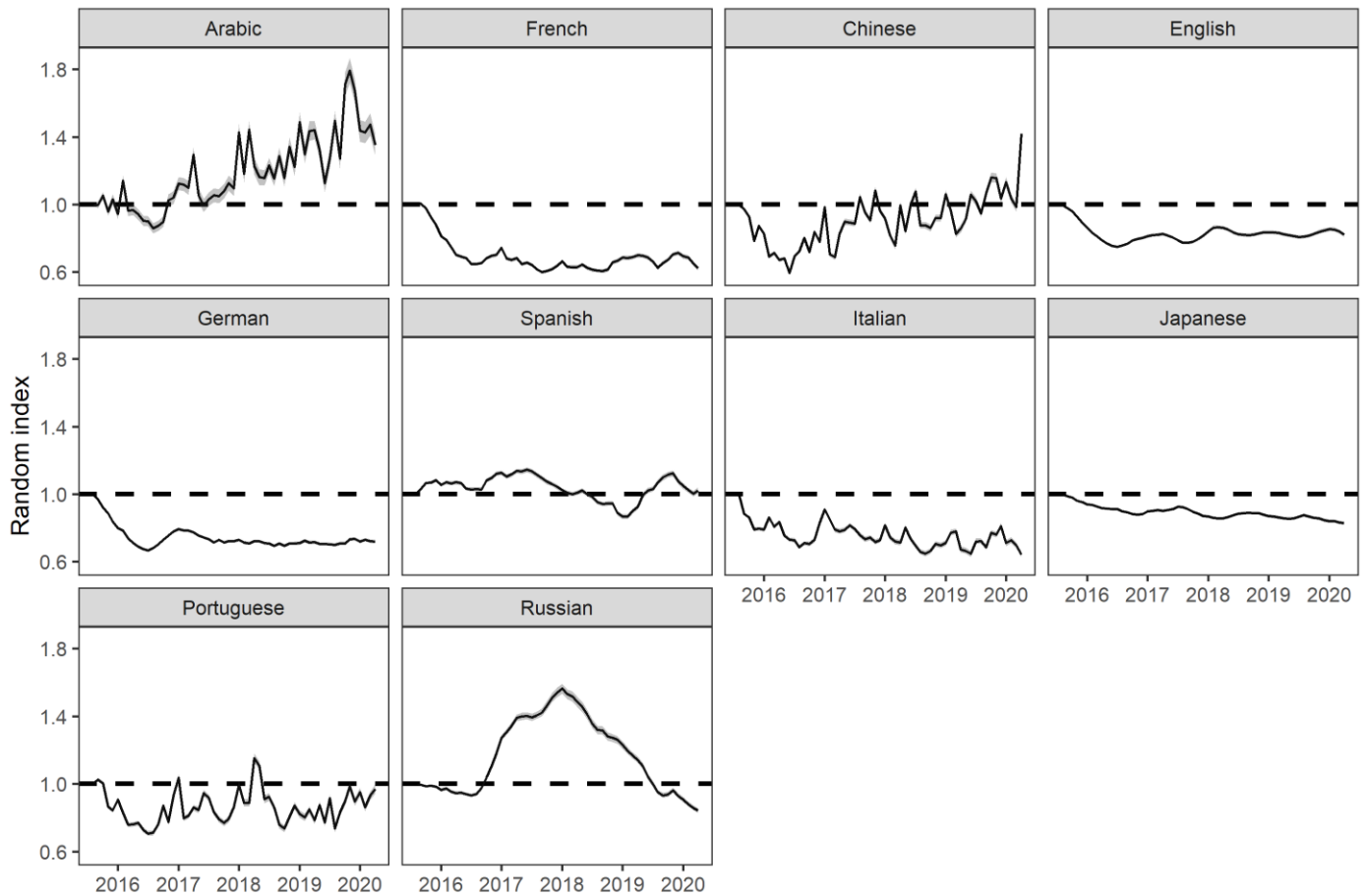


Figure S4.5 The random index for the set of complete series random pages in each language. Black lines represent the mean of the bootstrapped indices at each timestep, and grey bars the 2.5th and 97.5th percentiles. The random index provides an estimate of the overall change in popularity for each Wikipedia language (i.e. a decrease in the random index for a given language would imply that the popularity of Wikipedia in that language has decreased). This will differ from the trend in the total number of views for each language, which is heavily influenced by particularly popular pages, and confounded by increasing page number. To calculate the species page SAI, the rate of change in the mean random index is subtracted from the rate of change in each species page.

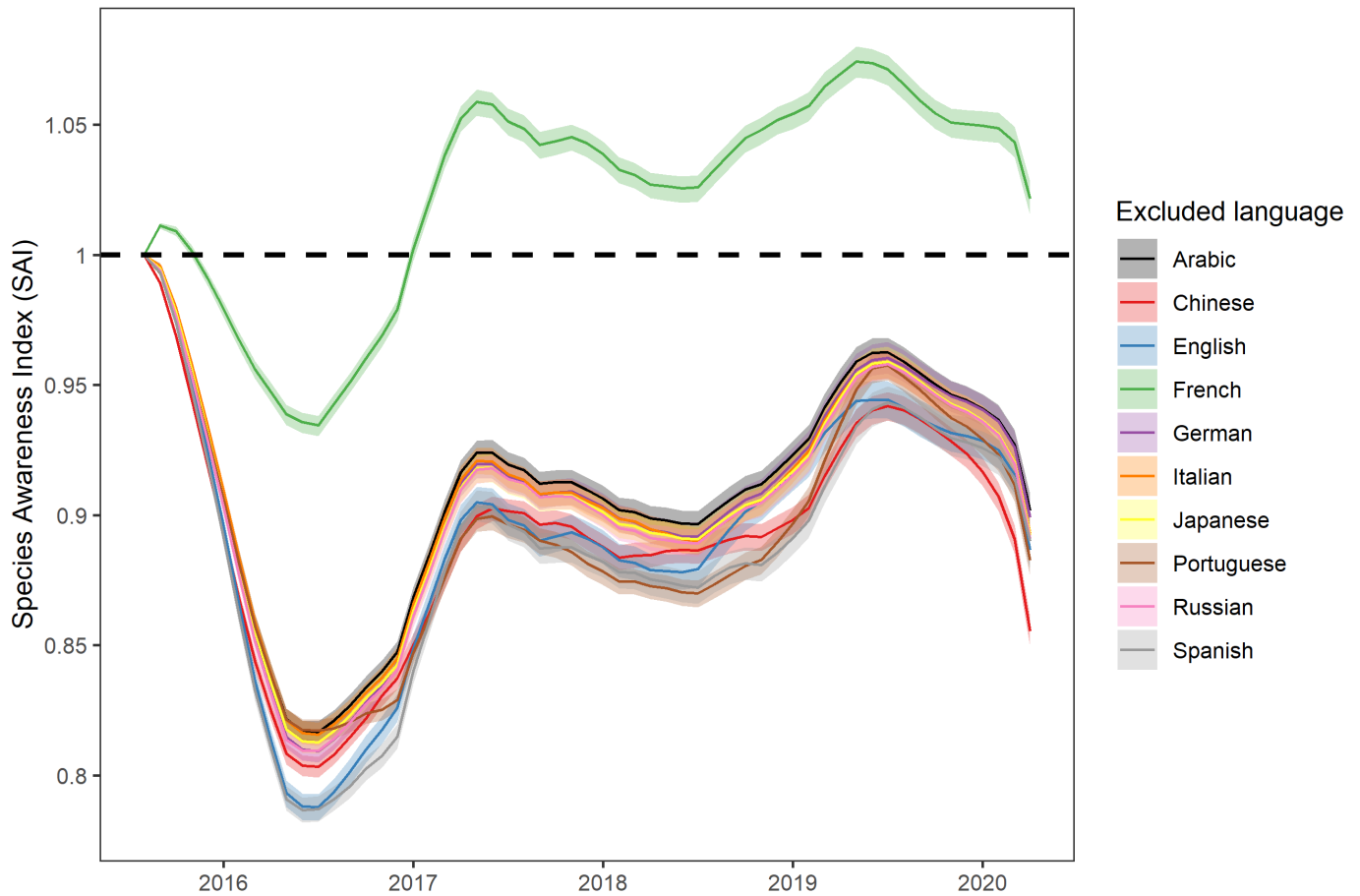


Figure S4.4 The overall SAI for all species pooled, jack-knifed for language. Each line represents the mean of the bootstrapped indices, and each band the 2.5th and 97.5th percentiles. The colour of each line represents the overall SAI if that language is excluded, providing an indication as to how single languages influence the overall SAI: black (Arabic), red (Chinese), blue (English), green (French), purple (German), orange (Italian), yellow (Japanese), brown (Portuguese), pink (Russian), and grey (Spanish). Here the French language has a significant effect on the overall trend, meaning it was removed in the main text. This marked effect is driven by the large number of reptiles in the French Wikipedia (Figure S4.1), which decrease rapidly at the start of the series (Figure 5.6). This decrease is compounded by the large number of reptiles in the French Wikipedia appearing in only that Wikipedia (Figure S4.15).

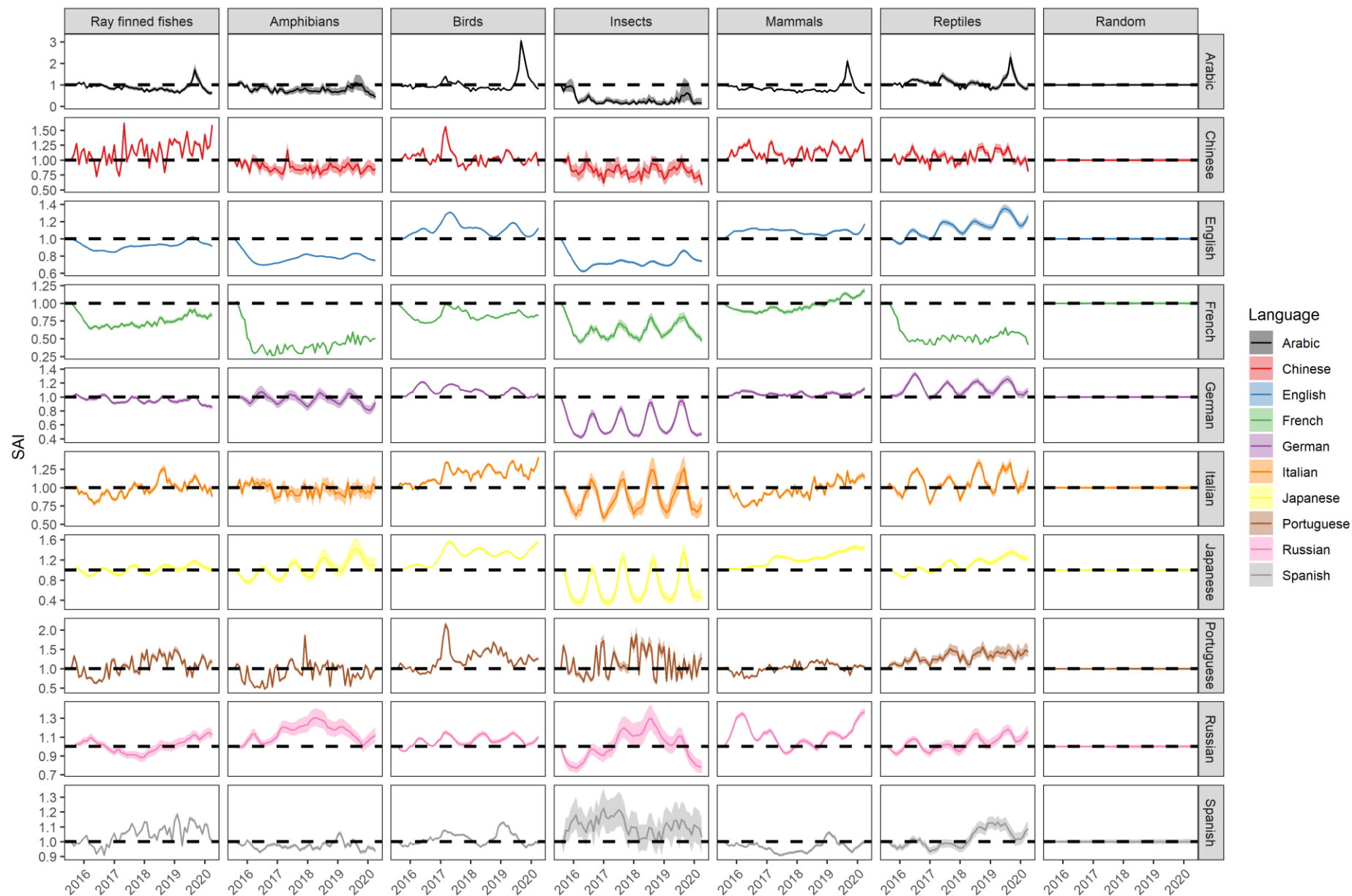


Figure S4.5 The overall SAI for each taxonomic class in each language before a final loess smooth (span = 0.3) is applied to each species page SAI. More tortuous trends are those for which the page views in that grouping tend to be lower. The “Random” column represents the random index adjusted for itself, indicating that the adjustment is functioning correctly.

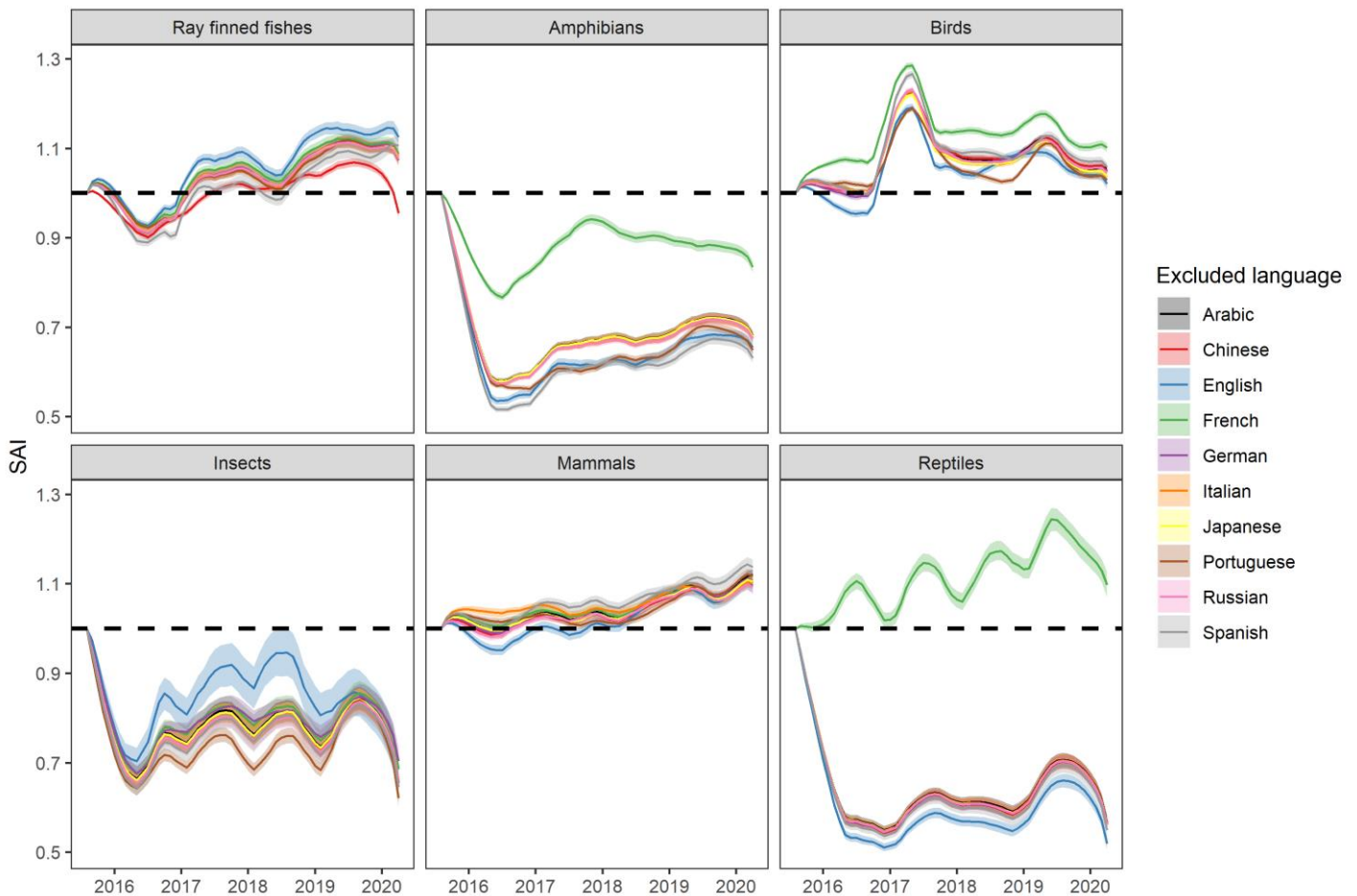


Figure S4.8 The overall SAI for each taxonomic class, jack-knifed for language. Each line represents the mean of the bootstrapped indices, and each band the 2.5th and 97.5th percentiles. The colour of each line represents the overall SAI if that language is excluded, providing an indication as to how single languages influence the overall SAI: black (Arabic), red (Chinese), blue (English), green (French), purple (German), orange (Italian), yellow (Japanese), brown (Portuguese), pink (Russian), and grey (Spanish). Here the French language had a significant effect on the trend for amphibians and reptiles, meaning it was removed in the main text.

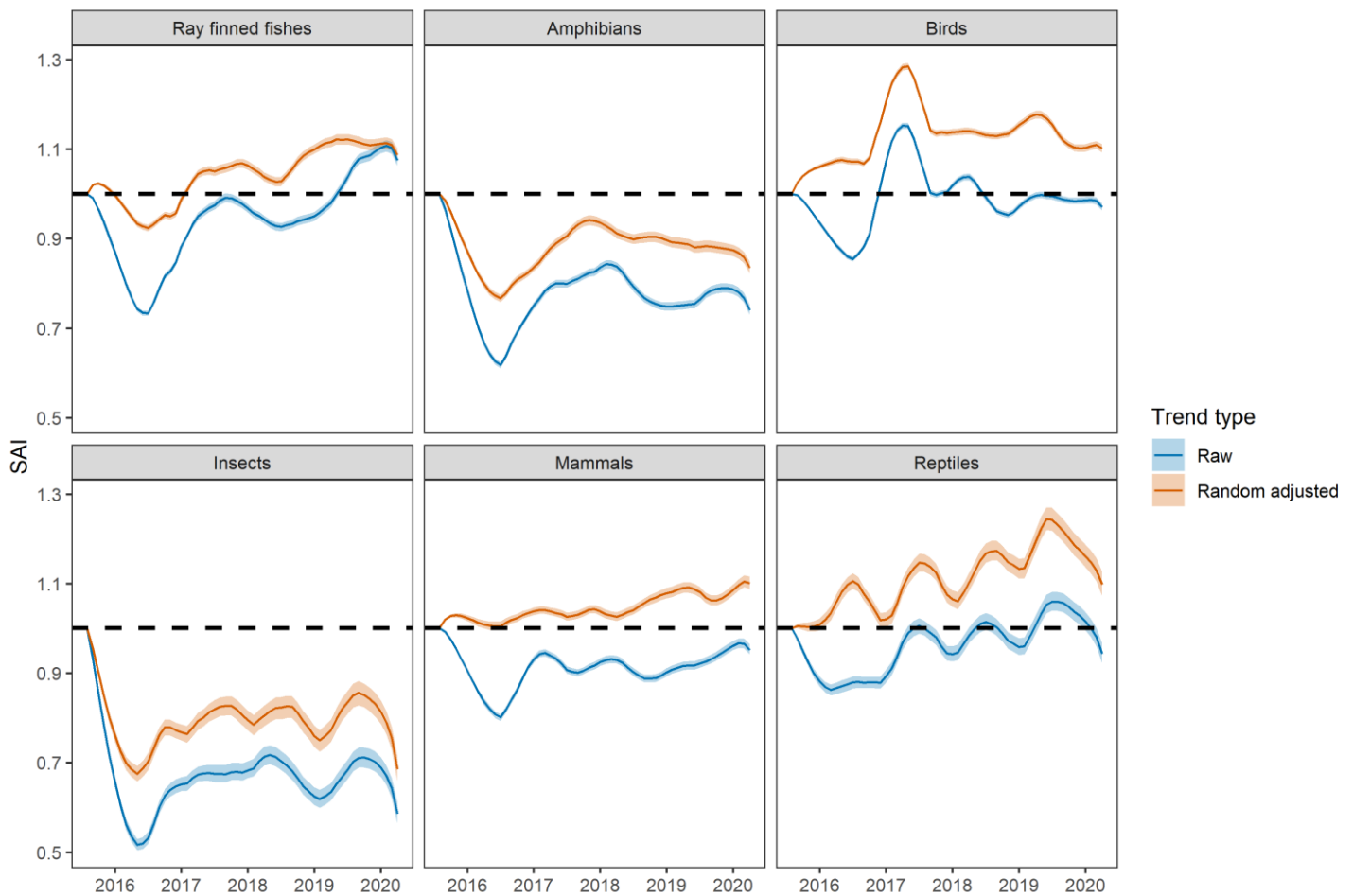


Figure S4.9 The overall SAI and SAI without a random adjustment for each taxonomic class. Each line represents the mean of the bootstrapped indices, and each band the 2.5th and 97.5th percentiles. Blue lines represent the trend without subtracting the random index, and orange lines represent the trend with the random index subtracted. In other words, the orange line represents the overall trend in awareness without the underlying residual trend of Wikipedia. Importantly, the random adjustment here cuts across multiple languages, which differ in their underlying trend of random pages. Each species page trend is adjusted for the language of that species page, which means when adjusted species are averaged across languages, that species then captures random adjustment moving in multiple directions.

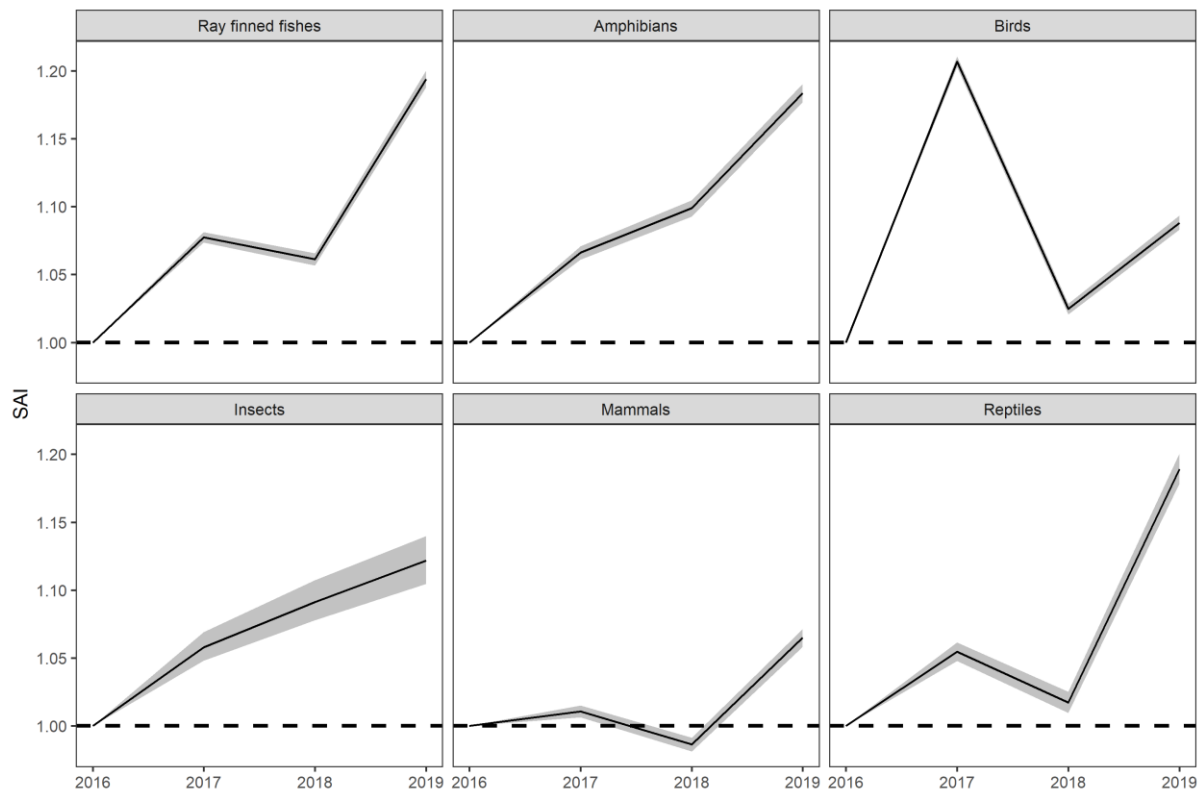


Figure S4.10 The overall SAI for each taxonomic class, based on the average monthly views per year. Here all languages (Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian, and Spanish) are included, with no additional loess smoothing of the species page SAI. Although annual trends may be more useful in the future, given this current time series is represented by only four points I opted not to include the annual trends in the main text.

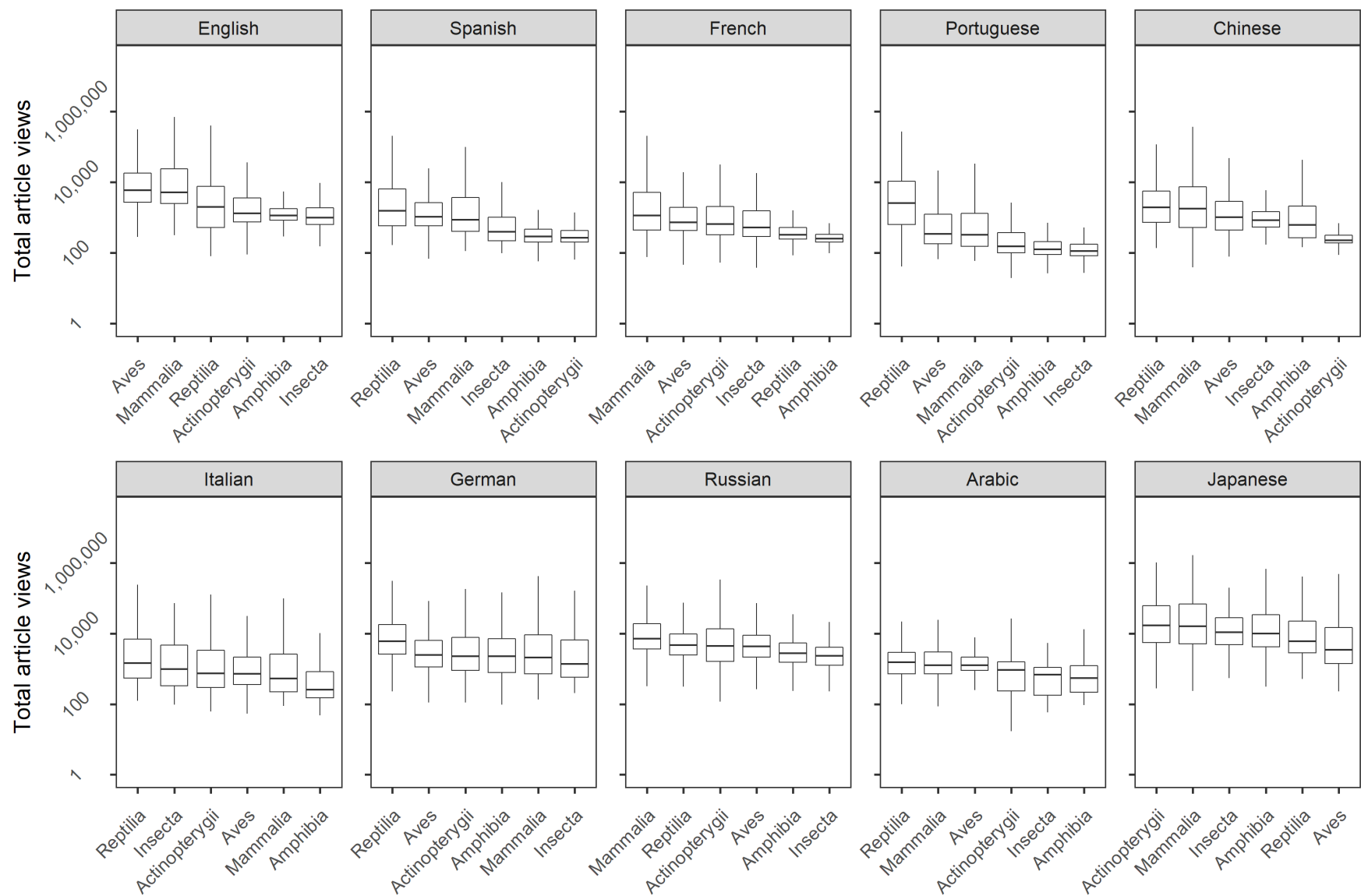


Figure S4.11 The distribution of page views for each taxonomic class in each language over the period 1st July 2015- 31st March 2020, as an indication of absolute awareness.

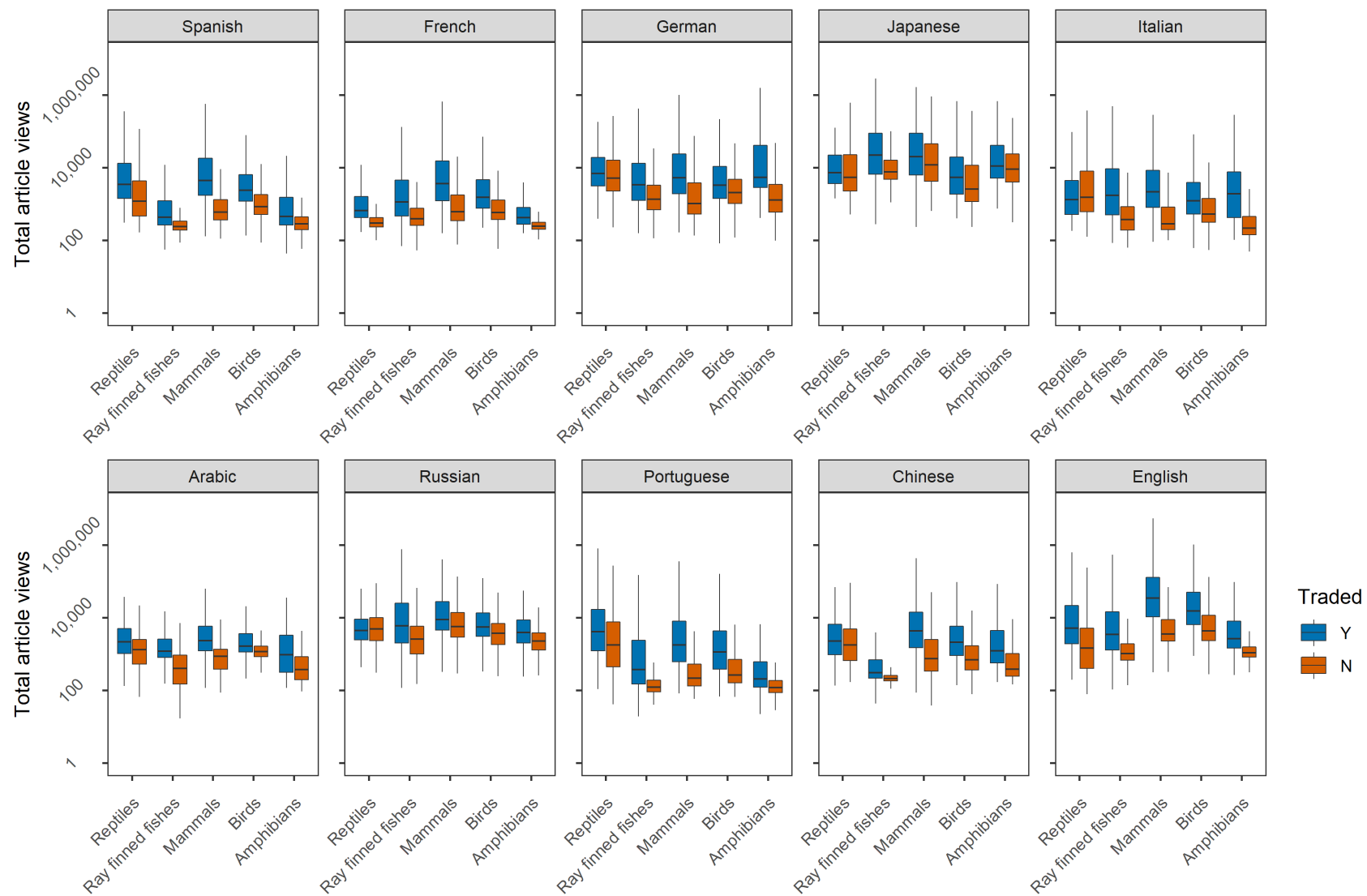


Figure S4.12 The distribution of views for traded and non-traded species in each taxonomic class in each language, for the period 1st July 2015- 31st March 2020. Blue boxplots represent species that are known to be traded, and red boxplots represent species that are not known to be traded.

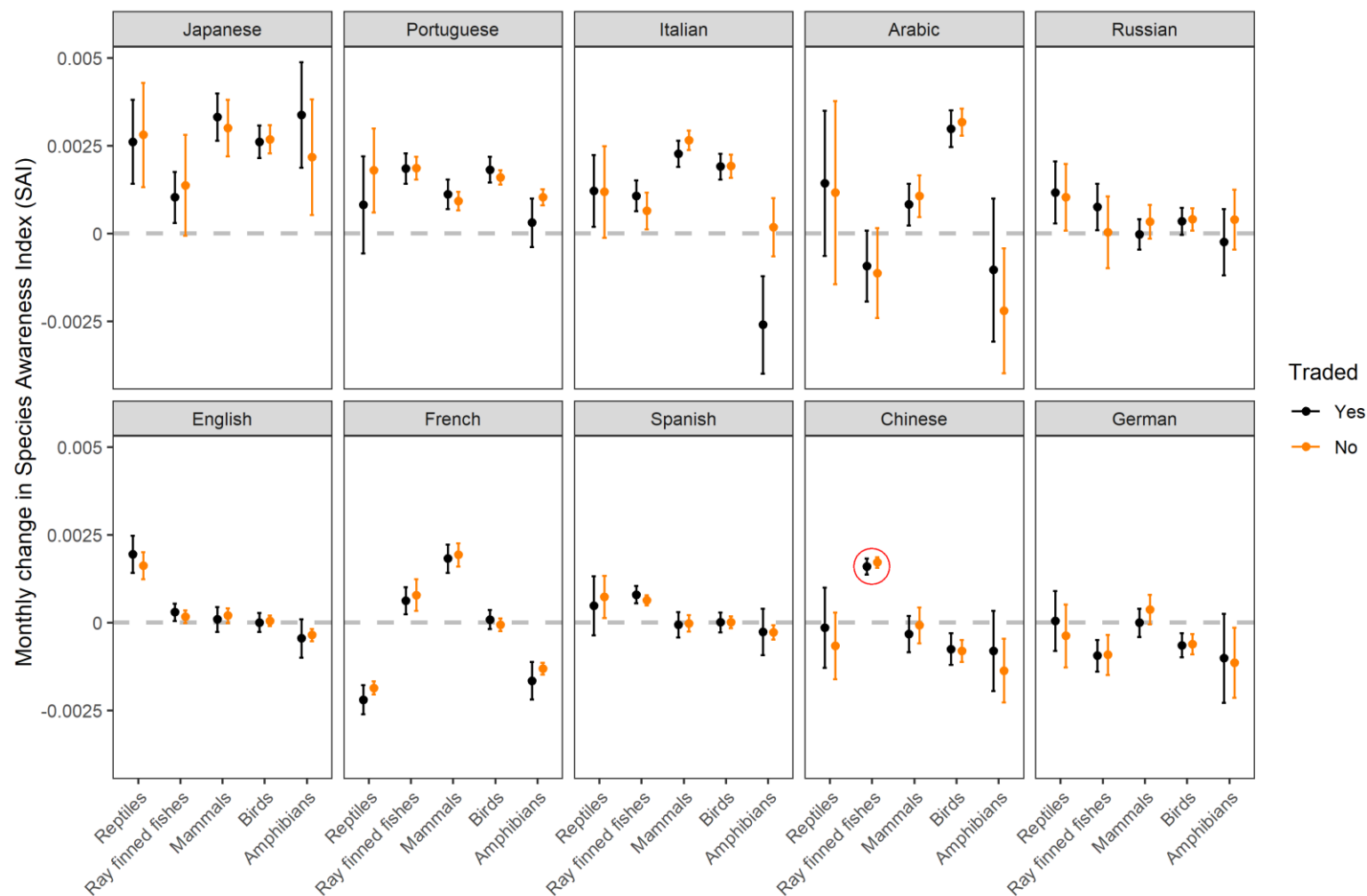


Figure S4.13 Average monthly rate of change for the species page SAI for 6 taxonomic classes across 10 Wikipedia languages. Errors bars represent the predicted values of a linear model, fitting average monthly change in the species page SAI as a function of taxonomic class, Wikipedia language, trade contribution (Y/N), and their interaction. Black bars refer to species that are traded, whereas orange bars refer to species which are not traded. The red circle highlights the Chinese ray-finned fish, in which both traded and non-traded species have a similar rate of change (see Discussion).

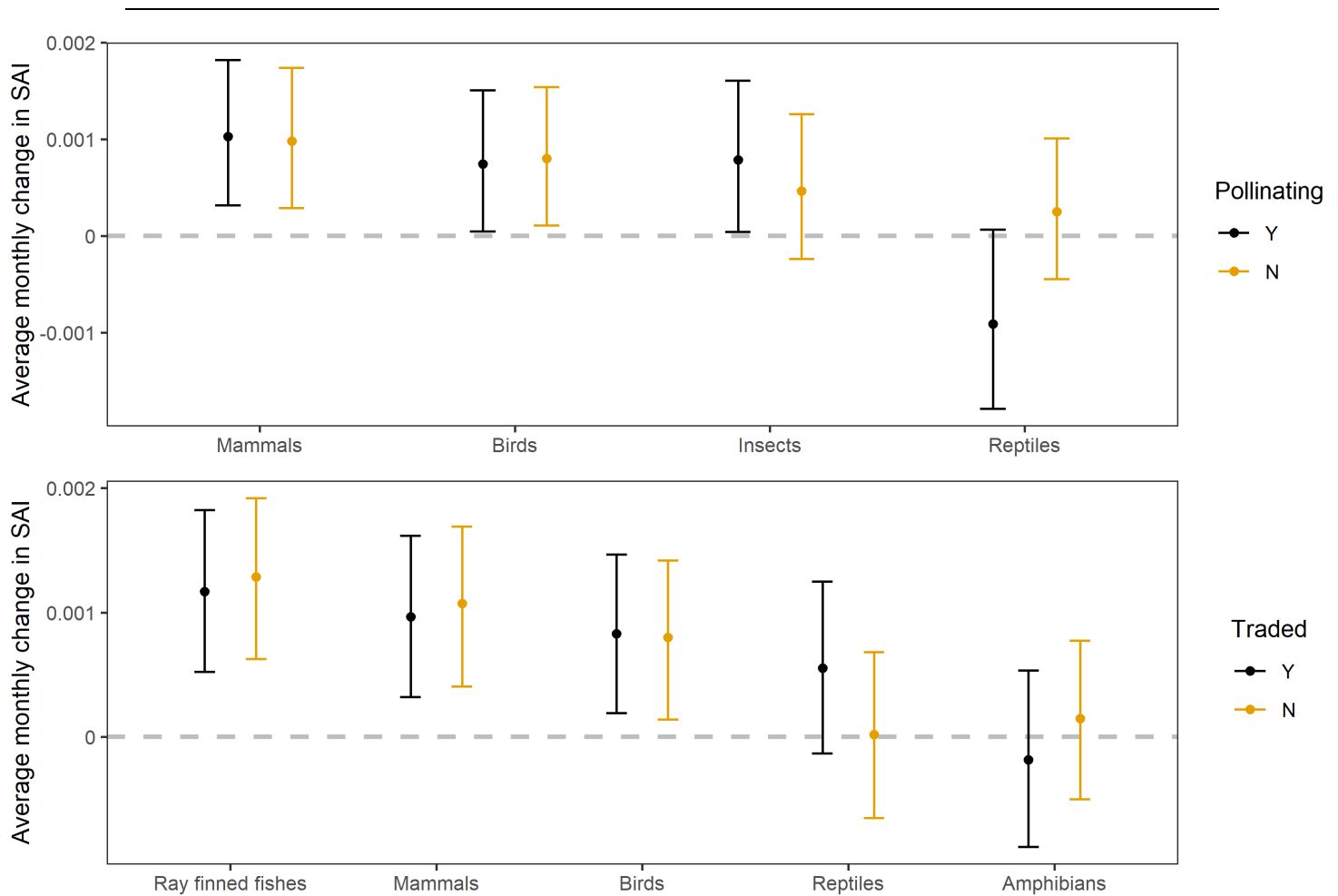


Figure S4.14 Average monthly rate of change for the species page SAI as a function of pollination contribution and trade contribution. Top: a mixed effects model for average rate of change in species page SAI as a function of taxonomic class, pollination contribution (Y/N), their interaction, and language (random effect). Bottom: a mixed effects model for average rate of change in species page SAI as a function of taxonomic class, traded species, and language (random effect). Effect sizes were calculated for both panels by drawing fixed effects 1,000 times based on the variance-covariance matrix, and then calculating the median value (shown as points), and 2.5th and 97.5th percentiles (shown as error bars). Note that monthly rate of change for reptiles is much lower here given the inclusion of the French Wikipedia (see jack-knife figures S6 and S8).

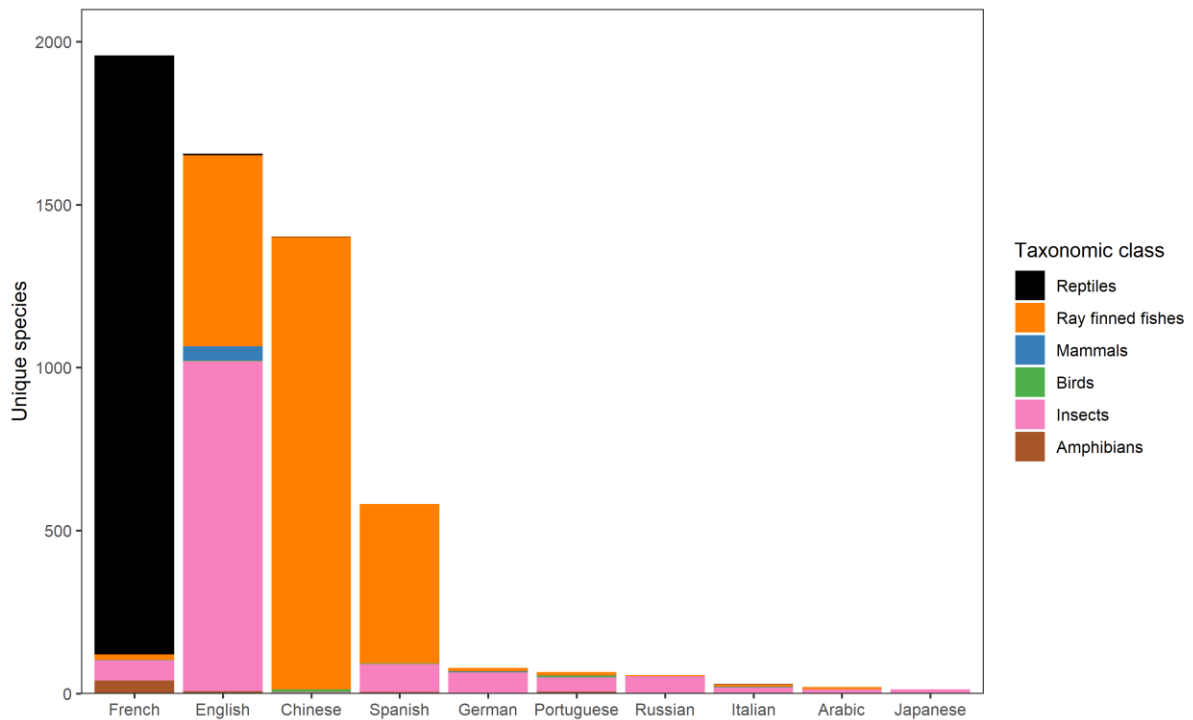


Figure S4.15 The total number of species appearing in only one language, for each of the top 10 languages (by active user) on Wikipedia. Fill colour represents the taxonomic class for species in only one language: reptiles (black), ray-finned fishes (orange), mammals (blue), birds (green), insects (pink), and amphibians (brown). The French Wikipedia contains the largest number of unique species, a high proportion of which are reptiles.

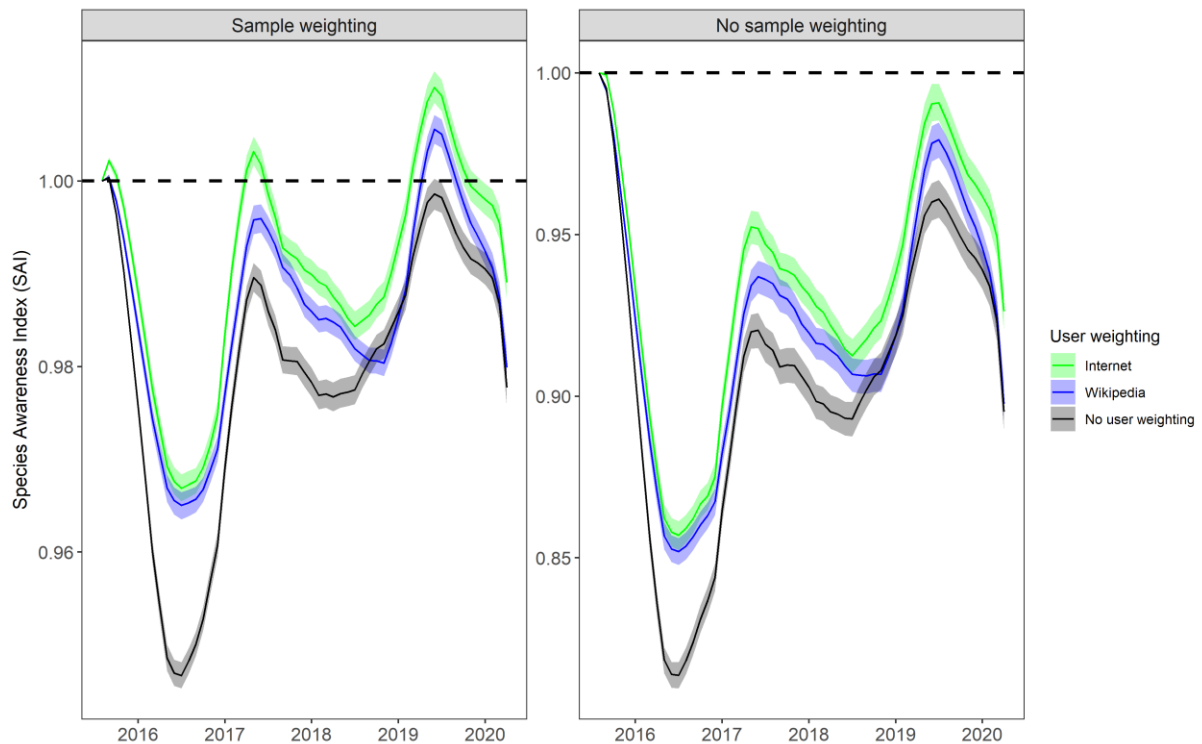


Figure S4.16 The Species Awareness Index (SAI) for 6 taxonomic classes (reptiles, ray-finned fish, mammals, birds, insects, and amphibians) and 10 Wikipedia languages (Arabic, Chinese, English, German, Italian, Japanese, Portuguese, Russian, and Spanish) for the period July 2015-March 2020, weighted according to a series of approaches. Each line represents the mean of the bootstrapped indices, and each band the 2.5th and 97.5th percentiles. Left: a set of sample weighted approaches, in which the average change for a single species is adjusted for the number of languages in which that species occurs. Right: the non-sample weighted approach explored in the main text. Colours in both the left and right panel refer to an additional user weighting, adjusting each species page for either the number of people using that language on the internet (green), or searching for animal species in that language on Wikipedia (blue). The black line represents the non-user weighted approach explored in the main text.

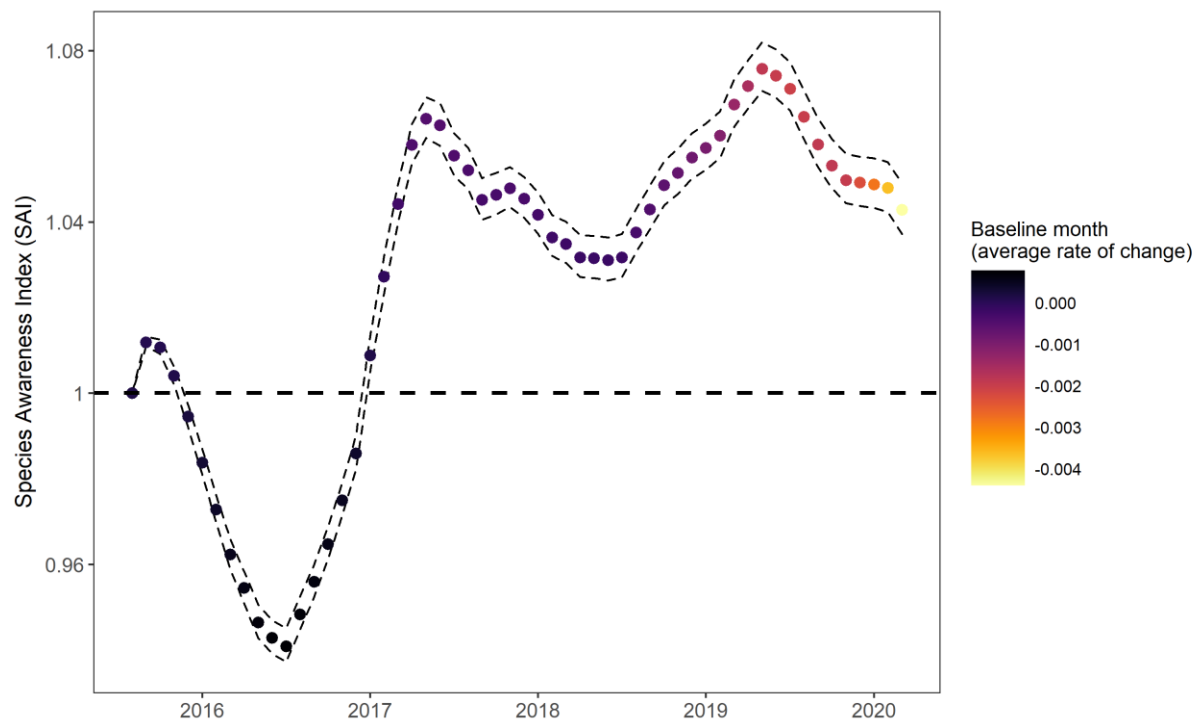


Figure S4.17 The Species Awareness Index (SAI) for 6 taxonomic classes (reptiles, ray-finned fish, mammals, birds, insects, and amphibians) and 9 Wikipedia languages (Arabic, Chinese, English, German, Italian, Japanese, Portuguese, Russian, and Spanish) for all potential baselines in the period July 2015-March 2020. Here the colour of each point represents the average rate of change from that point to the end of the series, providing an indication of robustness to variable baselines. The dashed lines represent the 2.5th and 97.5th percentiles for the overall SAI.

Supplementary tables

Table S4.18 ANOVA table for a mixed effects linear model predicting the log₁₀(total views) for each species page as a function of taxonomic class, pollination contribution, their interaction, and a random effect for language.

Fixed effect	Sum of squares	Mean square	F value	P value
Pollination contribution	0.18	0.178	0.3869	0.5339
Taxonomic class	889.74	296.579	645.8999	<0.001
Pollination contribution * Taxonomic class	78.57	26.192	57.0409	<0.001

Table S4.19 ANOVA table for a mixed effects linear model predicting the log₁₀(total views) for each species page as a function of taxonomic class, trade contribution, their interaction, and a random effect for language.

Fixed effect	Sum of squares	Mean square	F value	P value
Trade contribution	5212.6	5212.6	15206.44	<0.001
Taxonomic class	5813.2	1453.3	4239.59	<0.001
Trade contribution * Taxonomic class	401.5	100.4	292.78	<0.001

Table S4.20 ANOVA table for a mixed effects linear model predicting the average rate of change in species page SAI as a function of taxonomic class, pollination contribution, their interaction, and a random effect for language.

Fixed effect	Sum of squares	Mean square	F value	P value
Pollination contribution	0.00013237	0.00013237	4.3662	0.036662
Taxonomic class	0.00198949	0.00066316	21.8750	<0.001
Pollination contribution * Taxonomic class	0.00047564	0.00015855	5.2299	0.001314

Table S4.21 ANOVA table for a mixed effects linear model predicting the average rate of change in species page SAI as a function of taxonomic class, trade contribution, their interaction, and a random effect for language.

Fixed effect	Sum of squares	Mean square	F value	P value
Trade contribution	0.0000017	0.00000165	0.0422	0.8373260
Taxonomic class	0.0118755	0.00296888	75.6620	<0.001
Trade contribution * Taxonomic class	0.0008680	0.00021699	5.5301	<0.001

Table S4.22 ANOVA table for a linear model predicting the average rate of change in species page SAI as a function of taxonomic class, language, and their interaction.

Fixed effect	Degrees of freedom	Sum of squares	Mean square	F value	P value
Taxonomic class	5	0.0321	0.0064104	162.044	< 0.001
Language	9	0.0631	0.0070092	177.181	< 0.001
Taxonomic class * Language	45	0.0599	0.0013317	33.662	< 0.001

Table S4.23 AIC and Δ AIC values for a set of mixed effects generalised linear models: one fitting $\log_{10}(\text{total views})$ as a function of an interaction between pollination contribution and taxonomic class, and three a series of candidate null models. All models were fit with one random effect (language).

Model fixed effects	AIC	Δ AIC
Pollination contribution * Taxonomic class	183919.7	0
Pollination contribution	188112.0	4192.25
Taxonomic class	184245.5	325.76
Intercept (1)	188306.3	4386.6

Table S4.24 AIC and Δ AIC values for a set of mixed effects generalised linear models: one fitting $\log_{10}(\text{total views})$ as a function of an interaction between trade contribution and taxonomic class, and three a series of candidate null models. All models were fit with one random effect (language).

Model fixed effects	AIC	ΔAIC
Trade contribution * Taxonomic class	254981.1	0
Trade contribution	276230.9	2124.98
Taxonomic class	276276.0	2129.48
Intercept (1)	295182.9	4020.18

Table S4.25 AIC and Δ AIC values for a set of mixed effects generalised linear models: one fitting the average monthly rate of change in species page SAI as a function of an interaction between pollination contribution and taxonomic class, and three a series of candidate null models. All models were fitted with one random effect (language).

Model fixed effects	AIC	ΔAIC
Pollination contribution * Taxonomic class	-583810.4	0
Pollination contribution	-583776.7	33.7
Taxonomic class	-583865.6	-55.2
Intercept (1)	-583796.0	14.4

Table S4.26 AIC and Δ AIC values for a set of mixed effects generalised linear models: one fitting the average monthly rate of change in species page SAI as a function of an interaction between trade contribution and taxonomic class, and one candidate null model. Only one candidate null model was fit given the models for taxonomic class only and intercept only were fit to a different number of observation (trade data was not found for Squamate reptiles). All models were fit with one random effect (language).

Model fixed effects	AIC	ΔAIC
Trade contribution * Taxonomic class	-914369.5	0
Trade contribution	-914065.2	304.2876

Table S4.27 AIC and Δ AIC values for a set of linear models: one fitting the average monthly rate of change in species page SAI as a function of an interaction between taxonomic class and language, and three a series of candidate null models.

Model fixed effects	AIC	ΔAIC
Taxonomic class * Language	-950543.4	0
Taxonomic class	-947577.1	2966.3
Language	-948712.2	1831.2
Intercept (1)	-946797.9	3745.5

Table S4.28 The total number of internet users in each language, both in absolute terms and as a proportion.

Language	Total internet users	Proportion
English	1,186,451,052	0.35202835
Chinese (Mandarin)	888,453,068	0.26361026
Spanish	363,684,593	0.10790777
Arabic	237,418,349	0.07044369
Portuguese	171,750,818	0.05095967
French	151,733,611	0.04502043
Japanese	118,626,672	0.03519737
Russian	116,353,942	0.03452303
German	92,525,427	0.02745294
Italian	43,330,632	0.01285650

Table S4.29 The total number of species page views for each Wikipedia, both in absolute terms and as a proportion.

Language	Total internet users (millions)	Proportion
English (en)	1081.51	0.48551984
German (de)	233.18	0.10468097
Spanish (es)	199.68	0.08964189
Russian (ru)	193.15	0.08671039
Japanese (ja)	155.27	0.06970501
French (fr)	137	0.06150310
Italian (it)	86.17	0.03868410
Portuguese (pt)	63.13	0.02834081
Chinese (zh)	53.52	0.02402661
Arabic (ar)	24.92	0.01118728