THE WORLD'S LARGEST CARNIVORES

UNDERSTANDING DECLINES AND RECOVERIES

THOMAS FREDERICK JOHNSON

MANUELA GONZÁLEZ-SUÁREZ
NICK ISAAC
AGUSTIN PAVIOLO

The fate of the world's largest carnivores: understanding declines and recoveries

A thesis submitted for the degree of

Doctor of Philosophy

School of Biological Sciences

University of Reading

Thomas Frederick Johnson

April 2021

Primary supervisor: Dr Manuela González-Suárez

Additional supervisors: Dr Nick J. B. Isaac & Dr Agustin Paviolo

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Thomas Frederick Johnson

Acknowledgments

Over the last three and half-years, there have been moments of loneliness, self-doubt, ill-health, and frustration, but these have largely been fleeting because I have been surrounded by such wonderful people.

Manuela, working with you has been a blast. When I struggled, you noticed and you reached out, always guiding me to a solution. When I flourished, you let me run with ideas and put my stamp on the work. These qualities made you a wonderful academic advisor, but also a dear friend, and long may this continue. I will always remember the first time you reviewed my work and the blood red of comments and tracked-changes that awaited me; you have certainly made more resilient to critique and I now see these changes for what they are, a desire to see your students improve and succeed. On a personal level, watching you indulge in hysterics when I questioned your age will always tickle me, and our trip to have caipirinha’s on the Copacabana will always be a highlight. I look forward to continuing our work together in the future!

Nick, whilst our chats were brief and infrequent, your contributions were vital. At times, it felt like an hour with you could push the work forward weeks. In hindsight, I just wish I had engaged with you more.

Agustin, thank you for all of your advice and support over the years. A particular highlight was visiting you in Iguazu, where you brought some on-the-ground carnivore experience to the work. I especially enjoyed seeing a tapir and rosewood trees!

Amy, where do I begin? You have had the unenviable task of being carer, advisor and confidant over the years, keeping me healthy and reminding me of the big picture – the things that really should matter to me. But we both know your contribution far exceeds any thank you I can offer, as I can’t imagine the last few years without you by my side. That said, I should certainly endeavour to give you the best thanks I can. As Oscar Wilde famously said, and I paraphrase - all good stories begin at the end. So here we stand, in 2021, preparing to move to a new city and home, our fourth and eighth, respectively. But this feels different, despite being together for nearly ten years, this feels like a big step, or a new phase in our lives, where we will finally be going to place we would happily live and work forever. To get to this place, I think back over the compromises we have made in the last ten years, and I am drawn to the sheer disparity between us. Yes, I may have shifted to less field-oriented work, but you picked up your life and friendships and moved across the country with me three times! I think this is the perfect example of how selfless you

really are. I once said you were the kindest person I have ever met, and whilst your teeth and talons may have sharpened, I stand by that statement, as these testing but wonderful years would not have happened without your willingness to join me on this journey. Thinking back over these years, I am sure you will agree they been wonderful, ranging from the people we have met, to days pottering in our beautiful garden, and more recently extending our family with the arrival of Maggie and Benji. Although, as I write this, Benji has just chewed a hole in the carpet… I could not be prouder of having you as my partner, and I am so grateful you chose me too. Your contribution towards this thesis has been enormous, and that is why I was so pleased you agreed to include your art throughout it, which acts as a mere symbol of your unrivalled emotional contribution.

Mum and dad, I could have written a thesis just describing all of the amazing things you have done for me, but I want to just focus on two. Firstly, I want to thank you for the freedom and privilege you have provided that have led me to this place. Whilst you have offered opinions on what I should do for a career, where we have discussed options such as becoming an automotive engineer – I hope you agree I made the right choice dad - you always presented these as options. You allowed me to find the right career for me, and helped along the way, supporting me financially as a student, driving me to interviews, offering your thoughts and sensible concerns before fieldwork, and fundamentally, encouraging me to do what I thought was best. Knowing I have always had you as a safety net has provided this freedom, and I can't stress how privileged I am to have you as my parents. Secondly, I want to thank you for simply being you. Everyday, I can see so much of your personalities within me. From mum: compassion for others, perseverance and a desire to succeed in work, and a work-ethic that some may consider obsessive. From dad: a healthy appetite for conflict that has allowed me to challenge views, and a steadfast stance to make sure I don't lose sight of what's important to me. I see these traits within me every day, and I attribute so much of my successes to the values you have instilled in me. Thank you!

Mark, having you as an older brother has been a blessing, and I suspect I have arrived at this point – with a finished thesis - because I have followed so many of your steps. You were the first person I knew that embarked on exotic 'travelling' holidays, making my life-changing trips to South Africa and the Seychelles feel a little less scary. Your academic success was an inspiration and showed me that hard-work can lead to great things, and whilst my academic success never matched yours, sub-consciously competing with you has certainly provided me with opportunities that I would not have dreamt of. You taught me to ride a bike, introduced me to good food and showed me how to cook. You bought me my first ever copy of 'Origin of Species', now my favourite book. You showed me the Attenborough 'Planet Earth' series, the documentary that I think accelerated my interest in nature and led to the career I have dreamt of, and landed in. I could provide many more examples like this, where small relatively passive actions have shaped so much of me. I am reluctant to continue praising you, as a good brother naturally would be, but you have been and will continue to be so important to me.

To the amazing friends I have made in Reading - you know who you are – you have cheered my successes, comforted my failures, helped me when I was stuck, provided the best advice, and fundamentally, made my PhD fun, even during the tough times. I believe Bob Dylan* said ‘I wish there was a way to know you are in the good ole days, before you have actually left them’. To be frank, the idea of moving to Reading hardly appealed to Amy and I, but we had no idea we would meet people like you. Now I am preparing to leave, I am left wondering ‘were these my good ole days’.

**Correction, it was actually Andy Bernard from the American docuseries ‘The Office’*

There are so many others that warrant a thank you, ranging from past supervisors to close friends, but I am simply too tired to carry on.

Contents

Impact statement

Of the four research chapters in this thesis, two are published:

Johnson, T.F., Isaac, N.J., Paviolo, A. and González-Suárez, M., 2021. Handling missing values in trait data. *Global Ecology and Biogeography*, 30(1), pp.51-62.

Johnson, T.F., Kent, H., Dunn, G., Dommett, L., Penwill, N., Francis, T., and González-Suárez, M., 2021. classecol: classifiers to understand public opinions of nature. *Methods in Ecology and Evolution*

Additional contributions

Johnson T.F., et.al, (I2021) Effect of park use and landscape structure on COVID-19 transmission rates. Science of the Total Environment

Shackelford G, et.al, [**13th author**] (2020) Accumulating evidence using crowdsourcing and machine learning: a living bibliography about existential risk and global catastrophic risk. Futures

Johnson T.F., (2019) Exceptional nest attendance and solo breeding attempt by an African White-backed Vulture. Vulture News

Johnson M., et. al, [**6th author**] (2019) Which patients with ischaemic priapism require further investigation for malignancy? International journal of impotence research

Johnson T.F., Murn C. (2019) Interactions between Pied Crows and nesting African White-backed Vultures. Ethology, Ecology & Evolution

Johnson M.J., et al. [**7th author**] (2018) An Analysis of the Frequency of Y-Chromosome Microdeletions and the Determination of a Threshold Sperm Concentration for Genetic Testing in Infertile Men. British Journal of Urology

Johnson T.F., Brown T. J., Richardson D.S., Dugdale H.L. (2018) The importance of post-translocation assessment of Seychelles warbler habitat and resource-use: directing island restoration and management. Journal of Ornithology

With additional publications in review or revision:

Johnson T.F., Greenwell M.P. (In revision) Is sustainability advertising just a public relations stunt. Sustainability science

Greenwell M.P., **Johnson T.F.** (In review) Vote, Tweet, Campaign, Repeat: Using Twitter to predict environmental voting intentions. Scientific Reports

Johnson T.F. Murn C (In revision) Lack of adaptive nest-site selection in a social and group-living vulture. Ibis

Johnson T.F. Murn C (In revision) Density of Pied Crows Corvus albus within two protected areas of South Africa. Bothalia

Abstract

Humans are reliant on biodiversity and the ecosystem services that biodiversity provides. However, the status of biodiversity is unclear, with reports of both declines and recoveries, suggesting a new complex biodiversity change narrative. Large carnivores are a perfect example of this complex narrative. On one side, there are extensive reports of declines, but more recently, some populations appear to be recovering and expanding their range. In this thesis, I use interdisciplinary approaches covering evolutionary biology, biodiversity science, data science, and social science to explore influences of biodiversity change, specifically understanding declines and recoveries in these large carnivore species. I start by introducing the current literature on biodiversity change and the status of large carnivores, highlighting weaknesses in the available methods and data. In Chapter 2, I review one of the methodological weaknesses – approaches for handling missing trait values – providing some recommendations and warnings. In Chapter 3, I addressed known data biases in large carnivore population trends and build a new dataset obtained from an extensive and systematic search of the literature for population data. In Chapter 4, I develop a trait-based model exploring environmental and anthropogenic factors influencing large carnivore population trends and describe the status of the carnivore guild. Findings show diverse factors influence population change, but that this guild has and will remain relatively stable between 1970 and 2050. These results provide optimism for the status of large carnivores, and biodiversity more generally. However, I suspect our model failed to capture important characteristics on human perceptions and tolerances of carnivores, which could impact their population status. As a result, in Chapter 5, I develop a machine learning text classifier to measure public opinions of nature from social media. This, or tools like it, could be effective at capturing these previously undetected tolerance features at a global scale. Finally, chapter 6 summarises the collective thesis findings and offers my thoughts on the status of large carnivores and biodiversity change research.

Chapter 1: Introduction

Ecosystems offer a wealth of features that sustain human life (Rockström et al., 2009; Pecl et al., 2017). Biodiversity is a core component of this, underpinning support services like photosynthesis and pollination (Hanley et al., 2015), provision of services like food and medicine (Nielsen et al., 2018), and cultural services like outdoor recreation (Belaire et al., 2015). These biodiversity and nature-related services make humans healthier and happier (Díaz et al., 2018). However, with the continued expansion of humans into natural spaces and the resulting decline in biodiversity, there is a risk that ecosystem services could become degraded (Rockström et al., 2009; Newbold et al., 2016; Leclère et al., 2020).

Whilst the importance of biodiversity for ecosystem services is clear, the current status of the biodiversity underpinning these services is less clear. In recent decades, a wealth of evidence has been released suggesting biodiversity is experiencing declines (Newbold et al., 2015, 2016), with some suggesting declines are reminiscent of a mass extinction event (Ceballos et al., 2015). For example, in one analysis, wildlife population abundances are reported to have declined by 68% since 1970 (WWF, 2020b). Whilst this is compelling and alarming, a new biodiversity change narrative has developed in recent years casting doubt on this storyline instead suggesting that declines are somewhat offset or balanced by recoveries (Dornelas et al., 2019). This is perfectly emphasised by a reanalysis of the data showing a 68% decline, where removing just 3% of the extremely declining populations switches the overall trend into an increase (Leung et al., 2020). In essence, biodiversity change is complex, and the status of biodiversity is unclear.

Large terrestrial carnivores, the focus of this thesis, are a clear representation of the complexity of the new biodiversity change narrative. For example, a number of studies have reported dramatic population declines in these charismatic megafauna; in Africa (Bauer et al., 2015; Riggio et al., 2016), Asia (Harrison et al., 2016), Europe (Wolf & Ripple, 2017), North America (Woodroffe, 2000), and South America (Altrichter et al., 2006). Furthermore, some species have become extinct in the last few hundred years, and others are surviving in less than 10% of their historic range (Wolf & Ripple, 2017) – emphasising the precarity of these species' extinction status. Yet, there is also evidence that populations are beginning to recover (Chapron et al., 2014), and restore their historic distribution (Cimatti et al., 2021). All of this considered, the status of these large carnivores and the features influencing their population change remains unclear. This lack of clarity is particularly problematic, as large carnivores are essential for maintaining community composition through top-down trophic forcing (Atkins et al., 2019) and the landscape of fear (Suraci et al., 2016) which regulate prey abundance and movement, respectively. As a result, when the status of large carnivores is uncertain, the status of the entire ecological community becomes uncertain. Given their importance in the wider community, large carnivores are regularly described as indicator species, as a healthy carnivore population implies the lower trophic levels are also healthy, although this is very context specific (Sergio et al., 2008a). Moving beyond the ecological value, large carnivores are also culturally important to humans, and so there is often a public desire to protect these species, which explains why carnivores are the recipients of large proportions of the available conservation funding. All things considered, this presents a compelling case to improve our understanding of the status of large carnivores.

Understanding biodiversity change

There are a plethora of reasons for which carnivores should be conserved simply for the benefits they provide to humans, over and above their intrinsic value (i.e. they deserve not to go extinct). For example, large carnivores are important for maintaining ecosystem function as they regulate community composition (Ripple et al., 2014), but also provide really tangible cultural benefits e.g. acting as flagship species (Clucas et al., 2008). To conserve these species, and the benefits they provide, we must understand their population status.

One of the core challenges in understanding biodiversity change (or the status of a population) is overcoming extreme data biases. These biases occur across multiple features of the biodiversity change data: temporally e.g. there are fewer records in the past; spatially e.g. some countries or regions are underrepresented; taxonomically e.g. there are fewer records in some species; or on the influences of biodiversity change itself e.g. wildlife population monitoring is less common in areas more likely to experience land-use change. All of these biases are present (to some degree) in the large publicly available biodiversity change datasets, including in large carnivores. For example, population trend and occurrence data are abundant for some species, yet completely absent or extremely limited for others (Figure 1). Even well-studied carnivore species rarely have trend data for more than one population and location, and these are found primarily in wealthy countries (2), where there are fewer carnivore species to protect.

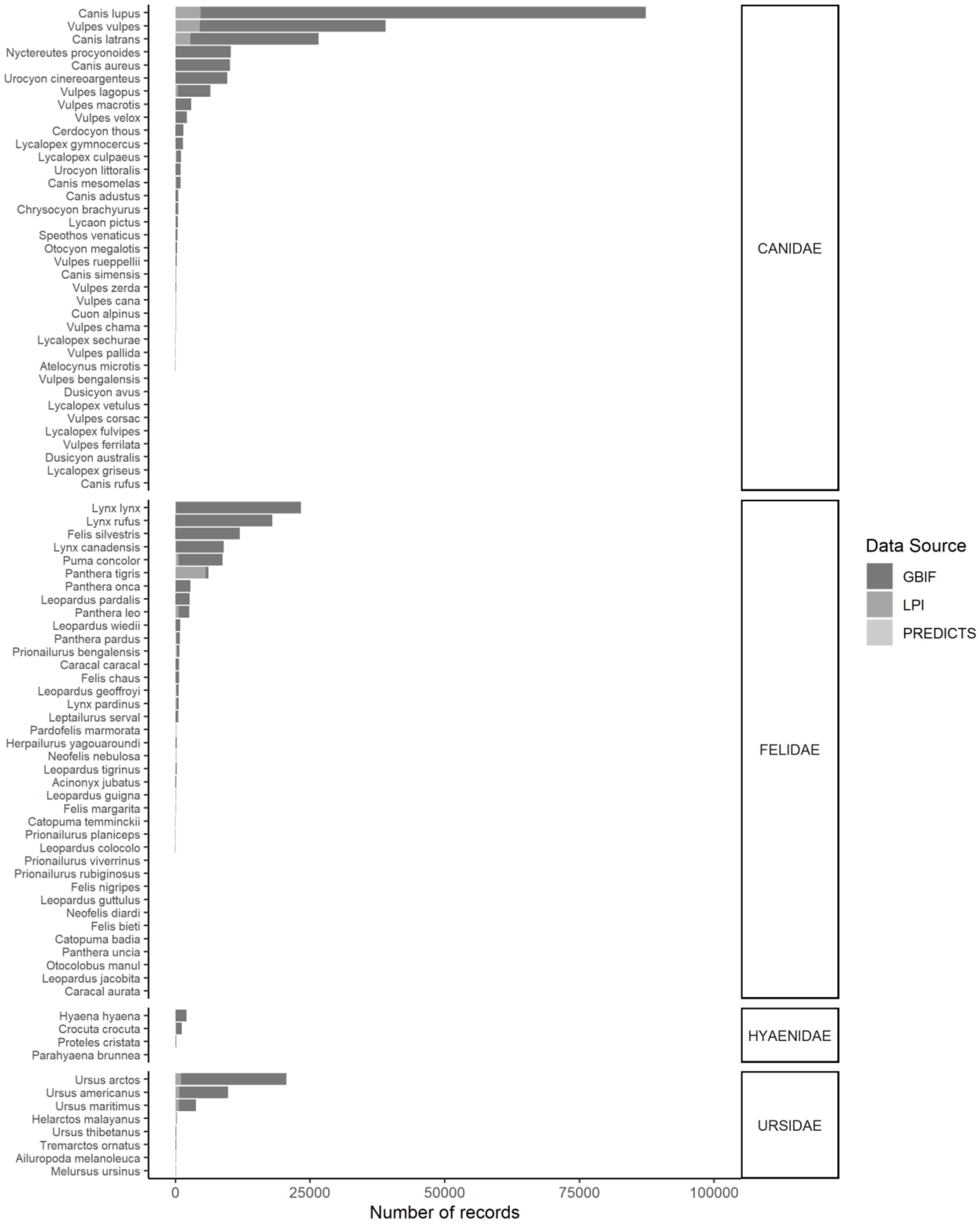


Figure 1. Frequency of occurrence and population trend records for the 87 species in the families *Canidae*, *Felidae*, *Hyaenidae* and *Ursidae* of the order *Carnivora*. Occurrence records were pulled from GBIF (GBIF.org, 2018), and trend records from the Living Planet Index (WWF, 2020a), and PREDICTS (Hudson et al., 2017).

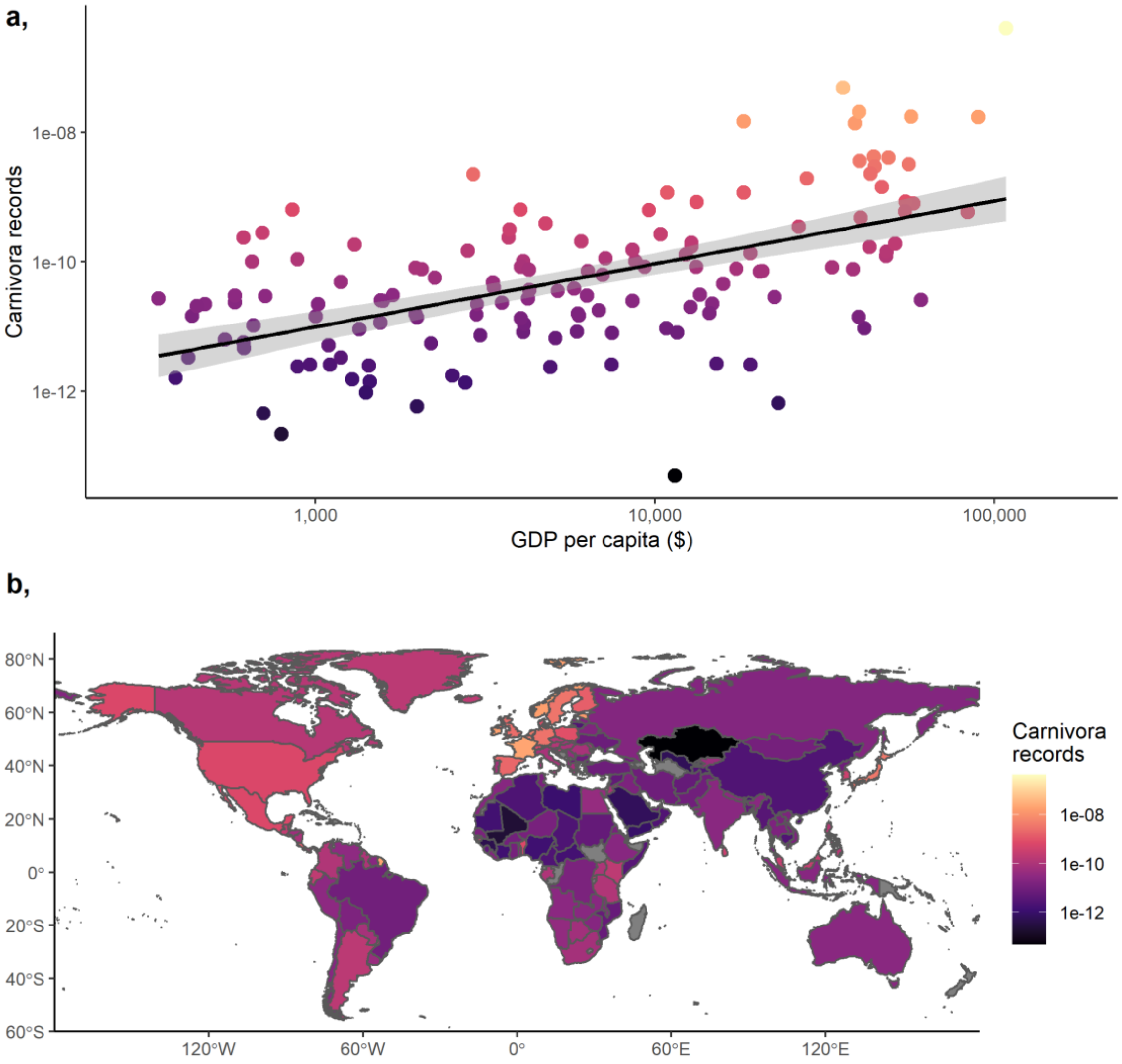


Figure 2. National *Carnivora* records plotted against gross-domestic product (GDP) per capita on the log10 scale (a), and spatially (b). *Carnivora* records are defined as the frequency of occurrence and trend records from GBIF (GBIF.org, 2018), Living Planet Index (WWF, 2020a), and PREDICTS (Hudson et al., 2017) within each country, which are then scaled by dividing the frequency by the countries size (in square kilometres) and the number of extant *Carnivora* species within the country. These *Carnivora* species include the 87 species in the families *Canidae*, *Felidae*, *Hyaenidae* and *Ursidae* of the order *Carnivora*. This '*Carnivora* records' variable is plotted on the log10 scale.

Even if all biodiversity change data were bias-free, our ability to understand biodiversity change is hampered by the availability and representativeness of data about drivers of change in biodiversity, especially traits. Traits, which describe the life-history characteristics of species (e.g. body mass and gestation length), have become pivotal in developing a more mechanistic understanding of ecology (Webb et al., 2010). However, traits are plagued with missing data biases (Etard et al., 2020), as collecting these life-history trait information is challenging, and much of the available data represents the common and more charismatic species (González-Suárez et al., 2012). This is problematic for inference, as conclusions can only represent this biased subset instead of the full taxonomic array. I address this feature in Chapter 2: Handling missing values in trait data.

Biases can also occur in other types of data (e.g. land-use and climate data). Remote sensing has relatively mitigated these biases, albeit only for certain types of data. Cryptic data sources, like poaching, persecution, and hunting pressures, can have a substantial impact on biodiversity change (Benítez-López et al., 2017) but are hard to measure at the global scale e.g. unlike remote sensing. This could be problematic, as features like land-use change may not necessarily be driving biodiversity change, but instead the cryptic features that come with land-use change could be the ones causing impacts. For example, wildlife-vehicle collisions can result in wildlife mortality events, but the accompanying threat of poaching access and land encroachment can also be important (Laurance et al., 2009). Capturing these cryptic features could be essential for understanding biodiversity change, and increasingly, ecologists are turning to novel data science technique to gather information about threats to biodiversity (Di Minin et al., 2018). I utilise these data science techniques in Chapter 5 ‘classecol: a text classifier to understand public opinions of nature’ to explore one of these cryptic features – a tolerance for nature.

Addressing data biases (e.g. in traits) and capturing cryptic influences of biodiversity change is important, but equally, biodiversity change needs to considered over very long timescales. This is important for two core reasons. Firstly, biodiversity change estimation is more accurate over longer time periods, capturing the natural fluctuation in population and community dynamics (Fournier et al., 2019). Secondly, and perhaps more importantly, shorter-term monitoring is more vulnerable to shifting baseline syndrome, whereby biodiversity is assessed relative to an already degraded baseline, instead of an intact system (Soga & Gaston, 2018). On this point, we are constrained by severe limitations in the data, and many previous biodiversity change assessments only stretch back to a baseline of 1970 (e.g. Living Planet Index). As this data, and the status of biodiversity, is lost beyond the past 50 years, it is important to consider proxies that can represent this status. Previous work on large carnivores has used range maps to estimate the species current extant relative to their historic ones (Wolf & Ripple, 2017). Using these proxies of historic loss alongside data for recent losses (e.g. population trends) views the biodiversity change narrative more holistically – as explored in Chapter 4: A global assessment of large carnivores.

A further problem with understanding biodiversity change is accounting for the sheer complexity and frequency of the features influencing change. These could fall into three groups: stressors (i.e. features that stress a population like poaching), supporters (i.e. features that support the population like conservation effort), and interactions (i.e. how the impact of features could change depending on the presence of other features like conservation effort having greatest value when poaching pressure is present). Much of the recent high-profile global biodiversity change research has focussed primarily on the stressors, especially climate and land-use change (Newbold et al., 2015; Spooner et al., 2018; Daskalova et al., 2020). Whilst this work has been important, ignoring the supporter features and the interactions could limit the robustness of the inference and any projection. For example, perhaps a shift from natural to agricultural land would have less of an impact in places with good governance (Amano et al., 2018). Furthermore, it is unlikely that all species will respond equally to this land-use change, where some species may decline whilst others prosper. Capturing the complexity of the system could be essential for understanding biodiversity change. However, capturing this complexity is challenging. In large carnivores, there are more than 75 IUCN threats (or stressors) expected to influence population change (Figure 3). Accounting for this massive array of stressors is simply unfeasible, and even any extensive analyses would ignore the presence of the supporters and interactions. This likely explains why much of the biodiversity change research has focussed on smaller themes (e.g. climate and land-use change). However, to develop accurate projections of biodiversity change, it is important to take this more holistic approach, which I explore in Chapter 4: A global assessment of the status of large predators.

One problem that could not be addressed within the thesis was capturing interactions between carnivores and the wider community. Working at the population level, as in Chapter 4, can only provide limited insight as its unclear whether any observed population change is driven by external influences (e.g. like climate change), or is a consequence of natural fluctuations in prey abundance, or even competition within a given trophic level. To better understand these population dynamics in the face of external influences, alternate approaches like Lotka-Volterra or network-based modelling would be required. However, these approaches require high temporal resolution datasets, that are only available for a few sites across the planet, and so are unfit for asking large global biodiversity change questions – the focus of this thesis. As a result, throughout the thesis, I have compromised a lack of understanding about the rest of the ecological community to answer broader macroecological questions about the status of large carnivores.

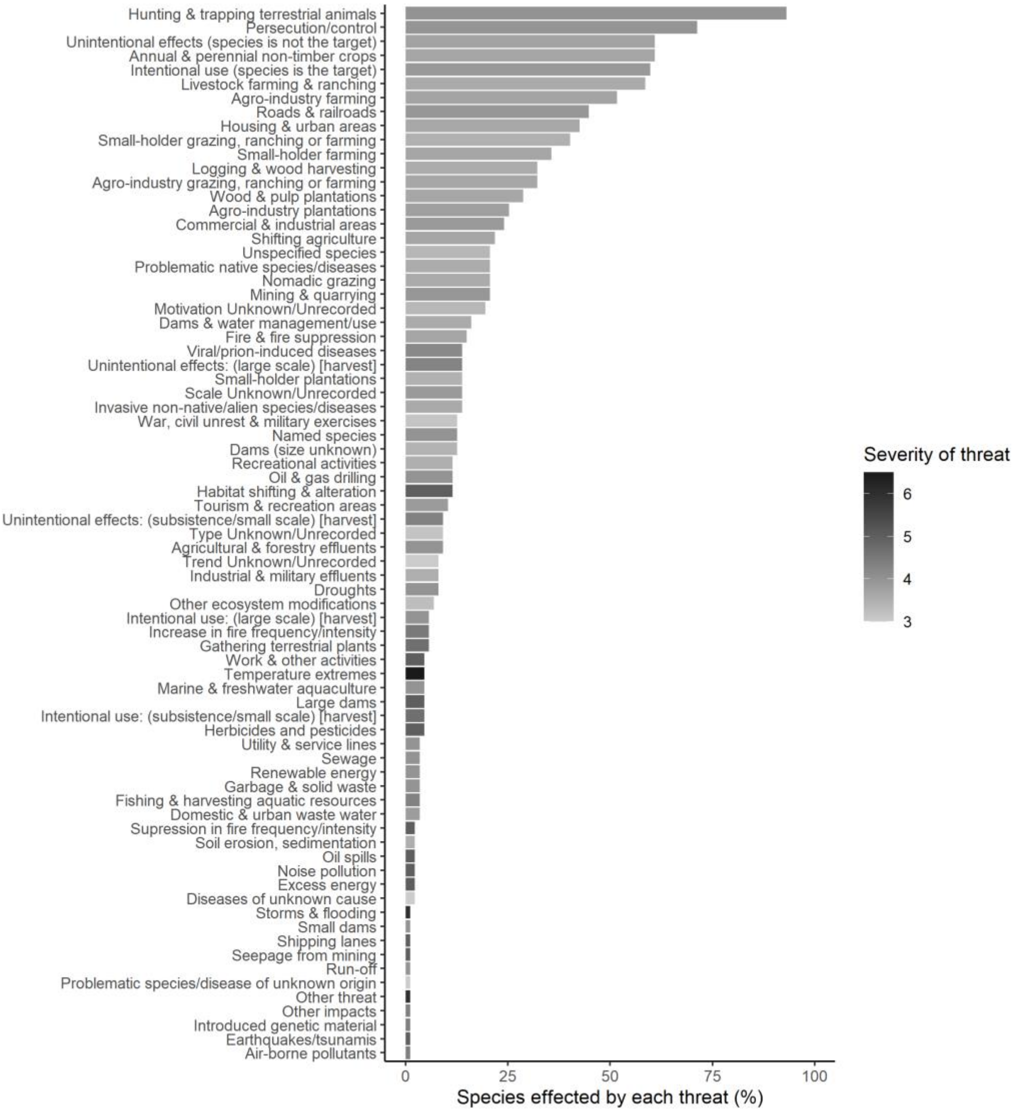


Figure 3. Percentage of the 87 species in the families *Canidae*, *Felidae*, *Hyaenidae* and *Ursidae* of the order *Carnivora*, effected by each of the IUCN threat classifications (IUCN, 2018), and the mean severity of the threat across species (high impact = 10, low impact = 1, range in data: 3 - 6). Threats specific to only one species were removed.

Thesis overview

Understanding biodiversity change in large carnivores is important, as these species are key in maintaining community composition through top-down trophic forcing - regulating the populations of prey and smaller carnivores, as well as the ecosystem services humanity relies on (Ripple et al., 2014). As indicator species', understanding the status of carnivores could also offer insight into the general status of biodiversity (Dalerum et al., 2008). In this work, I take a methodological and research synthesis approach to understanding decline and recoveries in the world's largest carnivores, specifically species in the families *Canidae*, *Felidae*, *Hyaenidae*, and *Ursidae* of the order *Carnivora*.

Chapter 2: Handling missing values in trait data - As our understanding of biodiversity change is hindered by incomplete trait datasets, I explore different approaches for handling missing trait data, and determine how missing values can limit inference. The aim of this work was to identify recommendations and warnings for handling trait values, which could then be used in Chapter 4.

Chapter 3: CaPTrends: A global database of population trends in large terrestrial *Carnivorans* – To address the data biases identified in Figure 1 & 2, I conducted a systematic search of the literature to bolster the available population trend data for our target carnivore species. This produced a dataset with three times more large carnivore records than in the Living Planet Database (WWF, 2020a).

Chapter 4: A global assessment of large carnivores – Utilising the data from Chapter 3, I conducted a global assessment of the status of large carnivores, identifying how intrinsic, extrinsic, and interactive features influence population change. This population change model is novel, in that it utilises both quantitative and qualitative record of population change to boost the sample size, and the spatio-taxonomic coverage. I then use this model to describe status of large carnivores across space, time and the phylogeny.

Chapter 5: `classecol` - classifiers to understand public opinions of nature – After identifying that cryptic features were likely important influences of carnivore population trends in Chapter 4, I developed a machine learning tool which can be used to assess people's perceptions of nature (i.e. tolerance of nature and wildlife) at a global scale. I anticipate development of tools that can capture cryptic variables like 'tolerance' will be important for improving the socio-ecological scale and predictive accuracy of biodiversity change models.

We close the thesis by discussing the importance of our findings, highlighting limitations in the work, and proposing next steps.


A. Cardwell

Chapter 2: Handling missing values in trait data

Authors: Thomas F. Johnson[1], Nick J. B. Isaac[2], Agustin Paviolo[3,4], Manuela González-Suárez[1]

Corresponding author: Thomas Frederick Johnson (Thomas.frederick.johnson@outlook.com)

1: Ecology and Evolutionary Biology, School of Biological Sciences, University of Reading, Reading, UK

2: Biodiversity Science Area, Centre for Ecology and Hydrology, Wallingford, OX10 8BB, UK,

3: Instituto de Biología Subtropical, CONICET-Universidad Nacional de Misiones, Bertoni 85, (N3370AIA) Puerto Iguazú, Misiones, Argentina

4: Asociación Civil Centro de Investigaciones del Bosque Atlántico, Bertoni 85, (N3370AIA) Puerto Iguazú, Misiones, Argentina

Author contribution statement: TFJ and MGS conceived the idea; TFJ, NI and MGS designed the methodology; TFJ analysed the data and led the writing of the manuscript with substantial contributions by MGS and NI. All authors interpreted results and gave final approval for publication.

Citation: **Johnson, T.F.,** Isaac, N.J., Paviolo, A. and González-Suárez, M., 2021. Handling missing values in trait data. *Global Ecology and Biogeography,* 30(1), pp.51-62.

Abstract

Aim

Trait data are widely used in ecological and evolutionary phylogenetic comparative studies, but often values are not available for all species of interest. Researchers traditionally have excluded species without data from analyses, but estimation of missing values using imputation has been proposed as a better approach. However, imputation methods have largely been designed for randomly missing data, yet trait data are often not missing at random (e.g. more data for bigger species). Here we evaluate the performance of approaches for handling missing values considering biased datasets.

Location

Any

Time period

Any

Major taxa studied

Any

Methods

We simulated continuous traits and separate response variables to test performance of nine imputation methods and complete-case analysis (excluding missing values from the dataset) under biased missing data scenarios. We characterized performance by estimating error in imputed trait values (deviation from the true value) and inferred trait-response relationships (deviation from the true relationship between a trait and response).

Results

Generally, *Rphylopars* imputation produced the most accurate estimate of missing values and best preserved the response-trait slope. However, estimates of missing data were still inaccurate, even with only 5% of values missing. Under severe biases, errors were high with every approach. Imputation was not always the best option, with complete-case analysis frequently outperforming *Mice* imputation*,* and to a lesser degree *BHPMF* imputation. *Mice*, a popular approach, performed poorly when the response variable was excluded from the imputation model.

Main conclusions

Imputation can effectively handle missing data under some conditions, but is not always the best solution. None of the methods we tested could effectively deal with severe biases, which may be common in trait datasets. We recommend rigorous data checking for biases before and after imputation and propose variables that can assist researchers working with incomplete datasets to detect data biases and minimise errors.

Keywords: BHPMF, Functional trait, Imputation, Life-history trait, MAR, MCAR, MNAR, Missing data, Multiple imputation chained equations, Rphylopars.

Introduction

Trait data describe the characteristics of individuals of a population or species (Webb et al., 2010). Trait-based analyses have been essential for improving our understanding of ecological and evolutionary processes, for example, identifying: negative impacts of climate change on biodiversity (Lancaster et al., 2017; Pacifici et al., 2017), common life-history strategies among invasive species (González-Suárez et al., 2015; Allen et al., 2017), and evolutionary changes in reproductive traits (Baker et al., 2020). Large-scale modelling studies like these are increasing in popularity, and often require trait data for numerous species and across taxonomic groups (Ríos-Saldaña et al., 2018). However, trait datasets can contain many missing values and these values can be missing with a bias (Sandel et al., 2015; Roth et al., 2018). For example, in a widely used mammalian trait dataset (Jones et al*.* 2009), species with smaller body mass values are more likely to have missing data for other traits, and this bias in missing data can impact inferences in comparative analyses (González-Suárez, Lucas, & Revilla, 2012).

The literature recognizes three broad types of missing data mechanisms: missing completely at random (MCAR), were there is no bias and records represent a random sample; missing at random (MAR), were missing data can be explained by available variables e.g. we know about the bias and can account for it statistically; and missing not at random (MNAR), were missing data cannot be explained by available information e.g. we either do not know about the bias, or lack associated information that could account for it statistically (Little & Rubin, 2002).

Currently there are at least 160 packages for handling missing data available on the R-CRAN repository (Josse et al., 2020). A simple, common approach is 'complete-case analysis', i.e. to exclude all observations with any missing values. This approach is robust when there is no bias (MCAR missing data); bias in the missing values can lead to erroneous inferences. Imputation, estimating missing values, is an alternative approach to handle missing data that can bypass this disadvantage (Little & Rubin, 2002). Imputation methods range from simple approaches like filling missing values with an average, to more complex approaches like estimating missing values using statistical models e.g. regression and random forest. Models can also be made more complex e.g. adding hierarchical information, allowing censored observations and weighting observations. There are also approaches specifically designed for handling values with extreme bias (MNAR), as well as methods for imputing missing response (sometimes called outcome or dependent variable) values - see Molenberghs, Fitzmaurice, Kenward, Tsiatis, & Verbeke (2015) for a more comprehensive description of methods.

Imputation can be applied to any dataset but is particularly useful for trait data because traits are often correlated (e.g. body mass is correlated with body length) and shaped by evolutionary history. Therefore, correlations and phylogenetic information can be used to predict missing trait values more accurately (Penone et al., 2014; Swenson, 2014). Previous studies have suggested that imputation in ecological and evolutionary studies generally outperforms complete-case analysis (Little & Rubin, 2002; Penone et al., 2014; Kim et al., 2018). However, imputation can only be successful if it accounts for the mechanism by which data are missing. If the imputation model cannot account for this mechanism (e.g. under extreme biases like MNAR), it is plausible that imputation may in fact amplify error in inference.

In this manuscript, we evaluate the performance of different approaches for handling missing trait data, considering the following questions: How effective is imputation at estimating missing values and making inference? Which imputation method is best? Is imputation better than complete-case analysis? How does the amount of missing data and presence of bias affect results? Expanding on previous comparisons of imputation methods, we introduce two new bias types, compare six additional imputation methods, evaluate the implications of including the response variable within the imputation, and propose steps for detecting erroneous imputation. Our study is most relevant for phylogenetic comparative studies but still applies to wider missing data scenarios.

Materials and methods

Data simulation

We simulated 40 datasets, each with 500 species using the *simtraits* function (Goolsby et al., 2017). Each dataset included four trait-predictor variables (“traits” hereafter) and one response variable. The 40 datasets represent 10 replicates (seeds 1-10) of four dataset types reflecting the combination of two correlation levels among traits (weak Pearson R^2= 0.2 or strong R^2= 0.6) with two response-trait slopes (no relationship ~0 or positive ~0.7). Traits were simulated under a Brownian model of evolution, with a gaussian distribution of values ranging from 0 to 10 to mimic the distribution of real trait data on a logarithmic scale (a transformation often used in comparative studies). The impact of phylogenetic signal strength on imputation performance was already tested by Kim et al. (2018) and Molina-Venegas et al. (2018), so we standardised Pagels lambda between the phylogeny and traits at ~ 1. The response was simulated as a product of a trait, rather than through the phylogeny, and has a gaussian distribution ranging from 0 to 10. We aimed to represent response variables used in comparative analyses such as extinction risk or population trend, rather than allometric relationships.

From each of the 40 original datasets, we removed trait values to create scenarios reminiscent of real trait datasets. Values were removed from between 5% and 80% of the species (in 5% intervals), across 11 distinct *bias types* (or missing data mechanisms) – see Appendix A Table A1.1. As a control, one mechanism was to remove trait values completely at random simulating the MCAR category. Two mechanisms stratified deletion with trait values removed evenly over the phylogeny and trait range. The remaining mechanisms explored four *bias types* likely to occur in trait datasets: Trait – large trait values more likely to be missing; Response – trait values more likely to be missing in species with larger responses; Trait & response – trait values more likely to be missing in species with large trait and large response values; Phylogeny – trait values more likely to be missing in certain clades (Fig 1, Table A1.1).

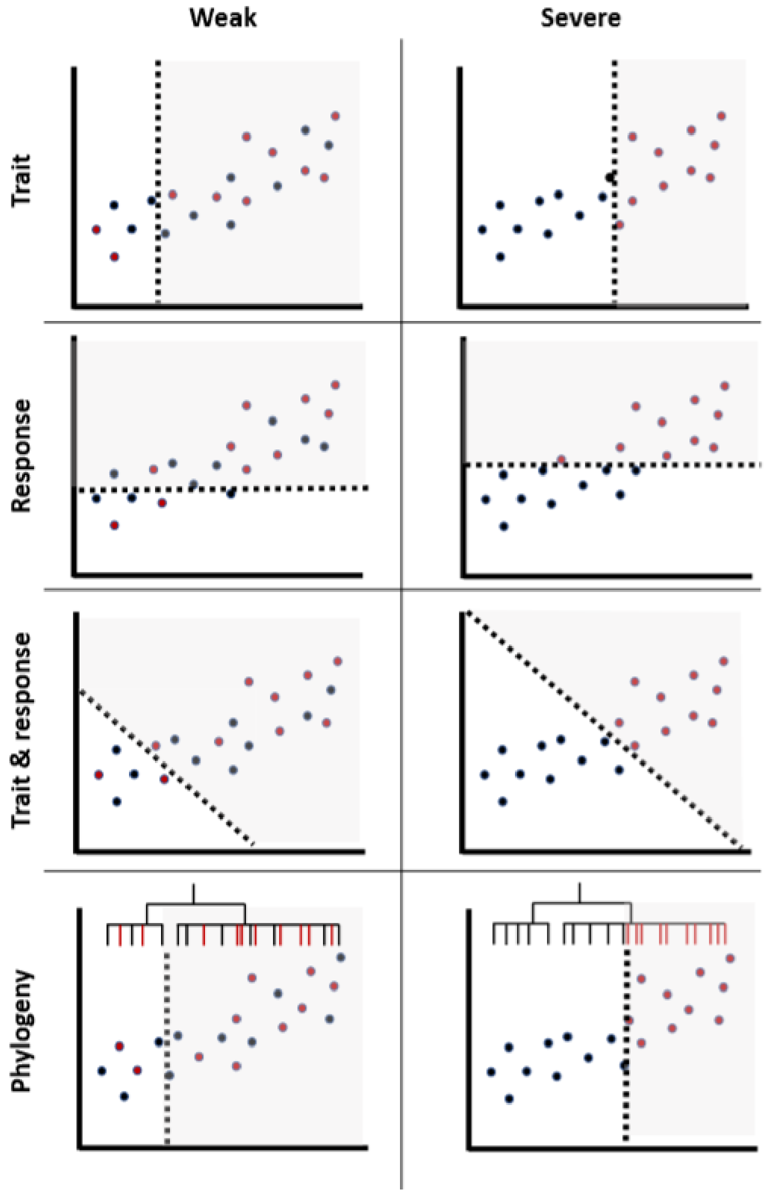


Figure 1. Schematic illustration of the effects of different biases. Panels contain an example scatterplot (x: trait, y: response) depicting a positive trend. In each panel there are 20 points each

representing a species, of which 50% are missing trait values (shown in red). Dotted lines illustrate a removal threshold based on the percentage of missing data (*missingness*) and *bias type*, which shows a different mechanism by which data are missing: Trait – large trait values more likely to be missing; Response – trait values more likely to be missing in species with larger responses; Trait & response – trait values more likely to be missing in species with large trait and large response values; Phylogeny – trait values more likely to be missing in certain clades. For each *bias type* we illustrate two severities: left panels show weak severity in which species are split into two groups, and species in the shaded area are 1.33 times more likely to be removed; right panels show severe severity in which values are systematically removed from large to small (all values removed from the shaded area). For further descriptions of these biases see Appendix A1.

Within each of these four *bias types* we tested two bias severity levels: weak – a conservative lower-end estimate for how much bias exists in trait data; and severe – an upper-end estimate aimed at testing how methods perform under the most extreme biases. Under a weak trait bias, the distribution of trait values becomes marginally skewed and the central point is shifted, but the range of values is largely preserved. Under a severe trait bias the distribution is truncated and the range reduced with extreme skew and shift in the central point. The weak and severe biases replicate the MAR and MNAR categories respectively. Appendix A1 provides a comprehensive description and justification of the bias severities. In total, across all dataset types, levels of missing data (*missingness*), and *bias types*, we generated 7040 datasets.

Data imputation

Testing all available imputation methods was not feasible; instead we expanded upon previous ecological and evolutionary imputation studies (Penone et al., 2014; Poyatos et al., 2018) to compare methods already identified as effective with new, promising methods. In total, we evaluated the performance of nine imputation methods available from three R packages (R 3.5.0, R Core Team, 2018): *BHPMF*: Bayesian hierarchal probabilistic matrix factorization (Schrodt et al., 2015), *Rphylopars* (Goolsby et al., 2017), and *Mice*: multiple imputation chained equations (Van Buuren & Groothuis-Oudshoorn, 2011). We summarize these approaches below providing a more detailed description in Appendix A2.

BHPMF is a machine learning technique that takes a sparse trait matrix and uses Bayesian probabilistic matrix factorisation to impute and estimate uncertainty in values, but it is not fully Bayesian in that imputation and analysis are not conducted simultaneously (Molenberghs et al., 2015). *BHPMF* provides a multi-level hierarchal framework which can control for spatial and taxonomic structures - see Schrodt et al (2015) for a comprehensive description of *BHPMF*. This hierarchical framework, coupled with the overall flexibility of probabilistic matrix factorisation (e.g. can handle non-linear relationships and binary categories), makes *BHPMF* versatile and potentially robust. However, unlike other approaches it is unable to make estimates for species for which all trait values are missing.

Rphylopars is a maximum likelihood frequentist method that uses a phylogeny and a sparse trait matrix to simultaneously estimate across-species (phylogenetic) and within-species (phenotypic) trait covariance (similar to a phylogenetic mixed model) to reconstruct the ancestral state and impute missing values (Goolsby et al 2017). This method is designed explicitly for phylogenetic imputation and requires a phylogenetic tree, which means the success of *Rphylopars* imputation depends on the phylogenetic signal in a trait - with low signal, the phylogeny may just add noise. An earlier version of *Rphylopars* was amongst the most accurate methods examined by Penone et al (2014).

Mice is the most general and flexible of the imputation packages used in this study, offering 24 different methods of imputation from which we explored three: 1) Predictive mean matching - imputes data by matching observed values between traits, then populates missing values in incomplete traits by adopting information from the matched species. This is the default *Mice* approach for continuous data and was considered the best overall method in Poyatos et al. (2018). 2) Bayesian linear regression – uses a linear model between traits to estimate missing values. This method is rarely tested and struggles with non-linear relationships, but is appealing to researchers familiar with linear regression. 3) Random forest - uses machine learning to produce and aggregate regression trees of the observed data and impute missing values. A similar imputation method ‘missForest’ was found to be effective by Penone et al (2014) with results comparable to *Rphylopars* and *Mice predictive mean matching*.

The three imputation approaches we explore fall into two groups: single imputation (*BHPMF* & *Rphylopars*) – where each missing value is populated by one estimate (but can have an associated variance); and multiple imputation (*Mice*) – where each missing value is assigned multiple estimates from a stochastic draw of the distribution (Little & Rubin, 2002). If the objective of the imputation is to produce estimates of missing values, e.g. to fill gaps on a dataset, single imputation is considered most effective, as the stochastic draws in multiple imputation add error (Van Buuren, 2012). However, if the objective is to model imputed values against another variable, the added error in the multiple imputation is advantageous, as when the trait data are modelled, the within- and among-dataset errors are pooled, inflating the standard error and reducing the type-1 error rate (Van Buuren, 2012). Whilst this makes multiple imputation more robust to type-1 errors, it does not necessarily mean multiple imputation can more accurately predict the slope within a model, as whilst this slope will have a greater standard error, it may still have the wrong direction.

Phylogenetic imputation

Imputation has been suggested to improve when phylogenetic information is incorporated (Penone et al., 2014; Kim et al., 2018). To test this, we imputed missing data with *BHPMF* and *Mice,* incorporating and ignoring phylogenetic information (for *Rphylopars* a phylogeny is required). *BHPMF* is unable to automatically process phylogenies, but its hierarchal nature can support taxonomies. We created a hierarchical node structure reminiscent of a taxonomy by splitting the phylogeny. For *Mice* we used phylogenetic eigenvectors that described the relationship between the phylogeny and traits (Diniz et al., 2015). Eigenvectors that were effective predictors of a trait were included as predictors within the imputation. We provide a comprehensive description of these approaches in Appendix A3.

Including a response variable in the imputation

The standard practice in comparative studies that use imputation is to impute values using only the traits and where relevant the phylogeny. However, the medical statistics literature recommends including every variable you plan to analyse, including the response, within the imputation model (Moons et al., 2006; Sterne et al., 2009). Including a response within the imputation of traits, which will then be modelled against the response in later analyses, appears circular and poor practice. However, in the event the trait has a response bias, including it within the imputation could control for this bias, and shift data from the MNAR to the MAR category, where imputation is more robust. We test this by performing each imputation with the response present and absent.

Error calculation

Imputation error: Is there a difference between the true and imputed values?

We compared true to imputed trait values under each of the nine imputation approaches (using the mean value across the repeated imputations for *Mice*) estimating the root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{Im} - y_{Tr})^2}$$

Where *N* is the number of imputed values ranging from 25 (5% of 500) to 400 (80% of 500), y_{Im} is the imputed value for a given observation and y_{Tr} is its true value.

RMSE's units are the same as the trait (range 0 – 10). We show alternative error metrics (mean absolute error, median absolute error, and R^2 between true and imputed values) in Appendices A5 and B4.

Mice guidelines stress that multiple imputation is not effective at solely predicting missing values and should be used for inference after model averaging instead. However, in the event inference is prone to error (where the imputed response-trait relationship deviates from the true relationship), it is important to consider how the imputation of missing values influenced this error. Conversely, it is also plausible that a method could produce inaccurate estimates of missing values, but still produce valid inference. Thus, assessing error in both the imputations and the inference (see below) provides a more holistic view of the imputation approach, which can help determine the point at which imputation becomes unreliable.

Slope error: Is there a difference between the true and imputed response-trait slope?

We fitted linear regressions with the imputed datasets to describe the response-trait relationship, recording the slope and associated standard error. We checked assumptions (e.g. normality) in a subset of these models, which were acceptable regardless of bias or amount of missing data. As *Mice* repeats the imputation process numerous times, we fitted multiple regressions using each of the imputed sets and then averaged the slope coefficients. To estimate 'slope error', we calculated the absolute difference between the imputed slope (or the complete-case slope) and the true slope. This 'slope error' metric illustrates how wrong the imputed slope could be, with the next step showing whether the estimated slope coefficient differed significantly from the true slope. Previous studies have considered how imputation can alter inference, focusing on allometric relationships between traits (Penone et al., 2014) and the impact on functional diversity metrics (Kim et al., 2018). Here we explored how imputation affects the relationship between traits and a separate response variable.

Slope significance: Is the difference between the true and imputed slope significant?

We tested if imputed slopes differed significantly from the true slope and the complete-case slope using the *t* statistic (Cohen et al., 2003):

$$t = \frac{Slope1 - Slope2}{\sqrt{{SE_{Slope1}}^2 + {SE_{Slope2}}^2}}$$

Where *Slope1* is the true slope coefficient and *Slope 2* is the imputed or complete-case slope coefficient. SE_{Slope1} is the standard error of the true slope and SE_{Slope2} is the standard error of the imputed or complete-case slope. We calculated degrees of freedom as the total sample size from the true relationship dataset plus the imputed or complete-case dataset, minus 4. We estimated significance at the 95% level. The *Mice* model slopes were averaged across each of the repeats, and the standard errors were pooled by calculating the within- and among-dataset variation, following Little and Rubin (2002). Incorporating the within and among-dataset variation inflates the standard errors around the slope. This is a key advantage to the *Mice* approaches, as whilst slope error could be high, the inflated standard errors around the slope reduce the probability of the imputed slope differing significantly from the true slope, and the likelihood of type-1 errors.

Data analysis

To understand the factors influencing the different error estimates we fitted regression models with various predictors (details below and in Appendix A6) and dataset seed as a random intercept effect. We used linear mixed models for numeric-continuous responses, with a $\log_{10}$ transformation on imputation error and a square-root transformation on slope error, and logistic mixed models for binary responses (e.g. significant or non-significant difference between the imputed relationship and the true relationship). In all cases we ensured model assumptions were met. Summary statistics display the mean ± standard deviation.

Modelling imputation error

We modelled imputation error as a function of six predictors: *imputation approach, bias type, missingness* (percentage of missing values in a dataset)*, response in imputation, initial slope direction* (positive or none) and *between-trait correlation* (Table A6.1). We included interaction terms between *imputation approach* and *bias type,* as well as *imputation approach* and *missingness*. We also tested whether including the response in the imputation improved accuracy by testing an interaction between *response in imputation, imputation approach* and *initial slope direction.* In some cases, the imputation resulted in implausible values; we removed any records with an RMSE greater than 10 to reduce the effect of these outliers.

Modelling slope error

We modelled slope errors separately for dataset types with initial positive relationships (response-trait slope ~ 0.7) and with no initial relationship (response-trait slope ~ 0). We tested as predictors: *imputation approach, bias type, missingness,* and *between-trait correlation*, as well as interactions of *imputation approach* with *bias type,* and *missingness*. We ran this model first including complete-case as a category within the *imputation approach* factor to identify scenarios where imputation is worse than complete-case analysis. This required excluding *response in imputation* as a predictor because this variable was not applicable for complete-case records. Second, we excluded the complete-case records and tested *response in imputation* as a factor, including an interaction with *imputation approach*.

Predicting imputation and slope error

We predicted imputation error using the variables: *missingness, phylogenetic clustering,* and *change in mean* (difference in mean before and after imputation). To predict slope error and significance we used the variables: *missingness, phylogenetic clustering, change in mean,* and *change in slope* (between imputation and complete-case). For all models we grouped the datasets with positive and no relationship slopes because in a real scenario a user would not know the true relationship.

Results

Including phylogenetic information generally improved imputation performance in every method (Appendix B1); thus, we focused on phylogenetic imputation methods, showing results for non-phylogenetic approaches in Appendix B3.

Which method performs best?

Imputed values were most accurate with *Rphylopars* (Table B2.1), which had consistently lower imputation errors in every *bias type*. However, *BHPMF* was the best approach when missing data exceeded 60% with a severe bias on the trait, and *Mice random forest, BHPMF* and *Rphylopars* were comparable when missing data exceeded 40% with a severe bias on the phylogeny (Fig 2; Fig B4.2). Imputation error results are similar regardless of whether the true response-trait slope was positive or had no relationship (Fig B4.1), and results are similar across all imputation error metrics (Fig B4.2 – B4.5).

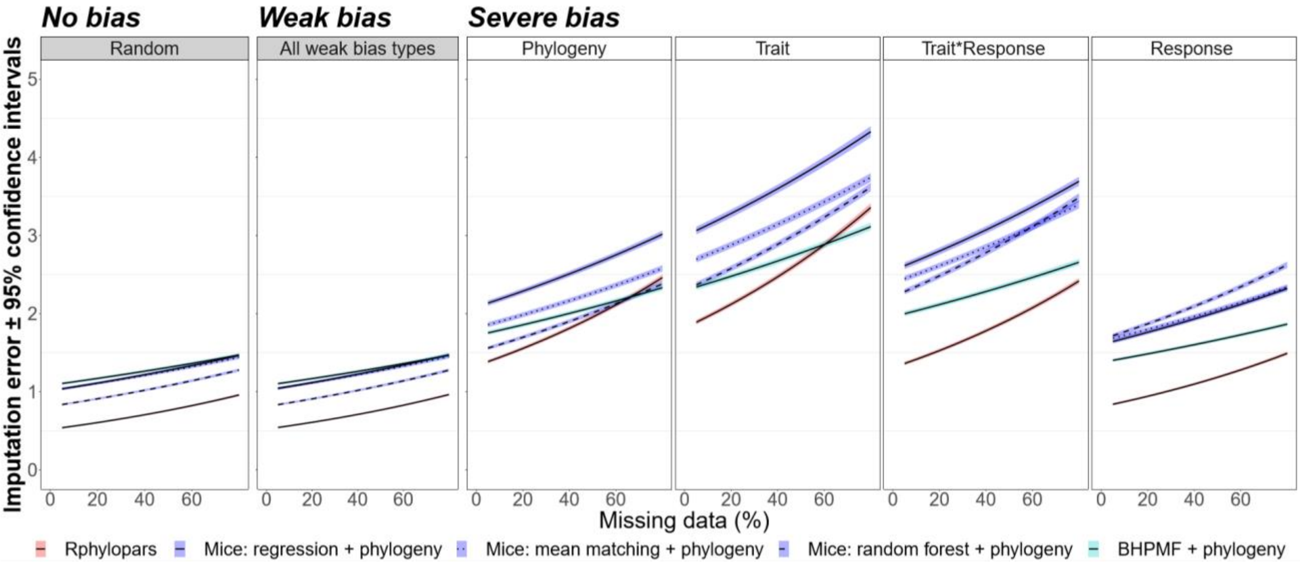


Figure 2: Difference between imputed and true trait values (RMSE, root mean square error) for five phylogenetic imputation approaches under varying percentages of missing data (*missingness*) and *bias types*. Lines depict the marginal effect of *missingness* and *bias type* from a regression model, and were averaged across other predictors: *seed, response in imputation, between-trait correlation,* and *initial slope direction.* For the equivalent plot split based on *initial slope direction* see Fig B4.1. Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%).

Rphylopars was also generally the best approach for preserving a response-trait relationship, with a significantly lower slope error than all other imputation approaches and complete-case analysis, regardless of whether the true response-trait slope was positive or there was no relationship (Table B2.2 – B2.3). However, for a severe bias on the trait or phylogeny, the best method was dependent on the true response-trait relationship: with no relationship, the *Mice* approaches performed best (Fig 3); while when the true slope was positive, complete-case was the best approach. *Rphylopars* was the fastest imputation approach (Table B8.1).

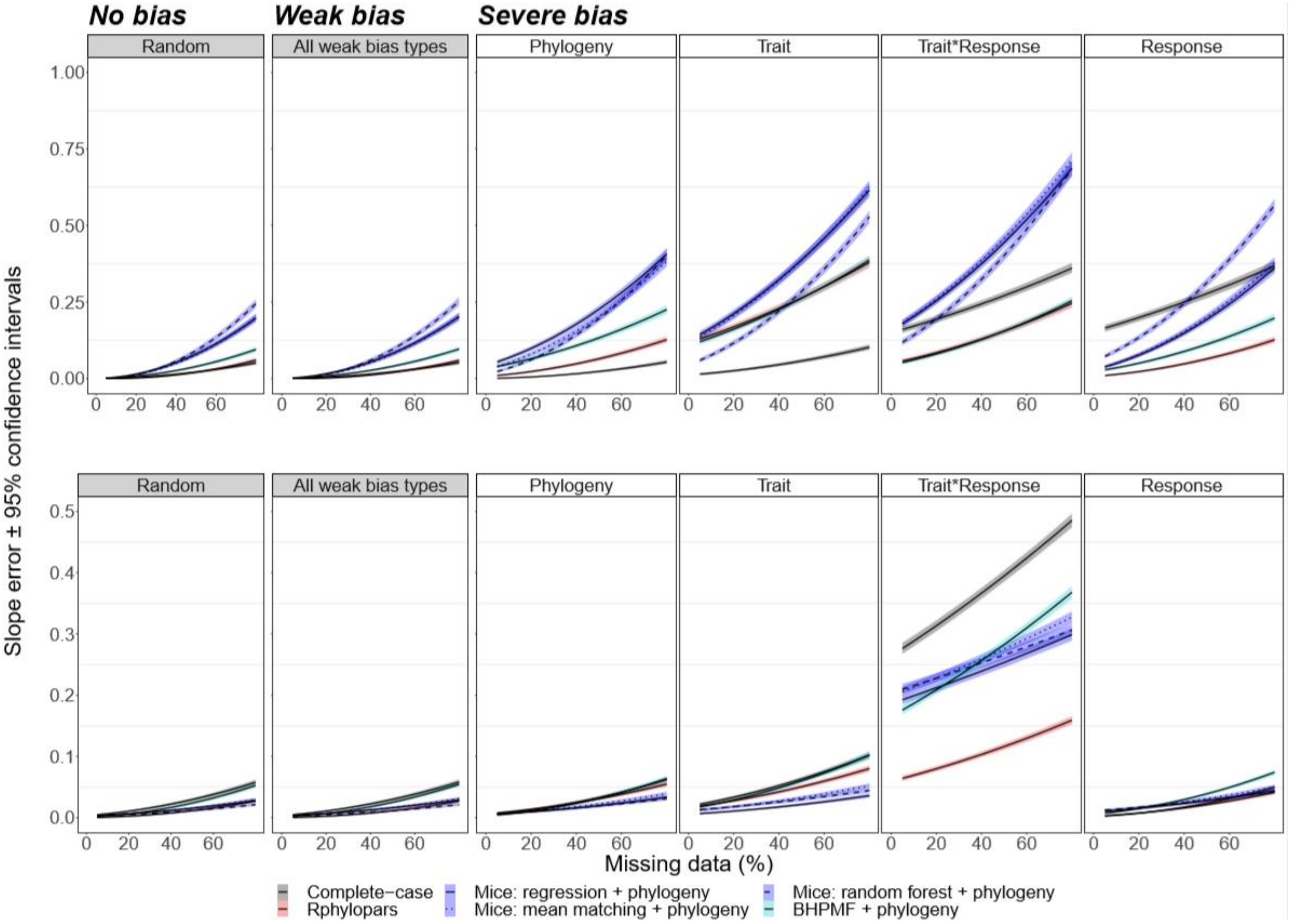


Figure 3. Absolute difference between the true response-trait slope coefficient and the slope coefficients in datasets with varying percentages of missing trait values (*missingness*) - removed under a series of *bias types*. Missing values were imputed under five phylogenetic approaches, or treated as complete-case analysis. The top row of panels show datasets in which the true slope was positive (r ~ 0.7), while the bottom row shows datasets with no relationship (r ~ 0). Lines depict the marginal effect of *missingness* and *bias type* from a regression model, and were averaged across other predictors: *seed, response in imputation,* and *between-trait correlation.* For plots split based on *response in imputation* see appendix B6. Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%). Note that the range in the y-axis differs between top and bottom panels.

Are imputed values accurate?

Imputation errors increased with the percentage of missing data: *missingness* (Est = 0.33, SE = 0.003, t = 103, p < 0.001) and were affected by *bias type* (Fig 2). Weak and stratified biases were comparable to no bias datasets, but errors were much greater when data were missing with severe bias (Appendix B4).

Imputed values were as likely to be over- as under-estimated, except when there was a severe bias on the trait (largest trait values removed), where, as expected, imputed values were primarily underestimated (Fig B5.1 – B5.5). Whilst *Rphylopars* had the smallest imputation errors, imputed values were still inaccurate - at 5% missing data, the mean difference between imputed and true values for *Rphylopars* was 0.56 ± 0.15 with no bias, 0.56 ± 0.15 in the stratified biases, 0.57 ± 0.16 in the weak biases, and 1.39 ± 0.57 in the severe biases, all increasing with *missingness* (Fig 2).

Can imputed data preserve response-trait relationships?

Slope errors increased with *missingness* (true positive relationship: Est = 0.27, SE = 0.008, t = 33, $p < 0.001$; true no relationship: Est = 0.23, SE = 0.004, t = 54, $p < 0.001$) and were affected by *bias type*, with large errors detected when data were missing with a severe bias (Fig 3). Imputed slopes were both over- and under-estimates of the true slope when there was no-true relationship (Fig B5.6 – B5.10). When the true relationship was positive, *Rphylopars*, and *BHPMF* again resulted in both over- and under-estimated slopes, but *Mice* approaches consistently underestimated the true relationship with slopes from the imputed datasets tending towards zero (Fig B5.11 – B5.15).

While imputation errors were often large, imputing missing values did not always introduce errors in the response-trait relationship. We observed low slope errors in all imputation approaches and all non-severe biases when few data were missing, but as *missingness* increased, slope error increased exponentially (Fig 3). *Rphylopars* was most robust with slope errors less than 0.05 for all levels of *missingness* in the no bias, stratified bias and weak bias datasets, regardless of the true response-trait relationship (Fig B4.6 – B4.7). However, *Rphylopars*, alongside all other approaches, had high errors under the severe biases, particularly when the bias acted on both the trait and response.

Missingness and *bias type* also influence whether slopes were significantly different from the true slope in a comparable way to slope error (Fig B4.8 – B4.11).

Should the response be included in the imputation?

When the response-trait slope was positive, including the response within the imputation decreased imputation error in all approaches and also decreased slope error substantially in *Mice*, to the point it was almost comparable with *Rphylopars* (Fig B6.1 - B6.4). Including the response in the imputation increased slope error in *Rphylopars* and *BHPMF*. When there was no relationship between the trait and response, including the response in the imputation increased imputation and slope errors in every approach, but with a small effect (Fig B6.2 - B6.6).

Can we predict when the imputed values and response-trait relationships become inaccurate?

Because *Rphylopars* was found to generally be the best method, we focused on predicting errors under this method. *Missingness*, *phylogenetic clustering*, and *change in mean* were important predictors of slope error, significant differences in slope error, and imputation error. *Change in slope* was also a relevant predictor for slope error and significant differences in slope (Fig B7.1).

Discussion

Overall, our results show that there is no single best solution to deal with missing data. *Rphylopars* was generally the best approach for predicting missing values and was consistently more accurate than *BHPMF* and *Mice* at maintaining the true response-trait relationship. However, in some scenarios, all imputation approaches were outperformed by complete-case analysis, showing that imputation is not always the best option. When using imputation, including phylogenetic information widely reduced errors in our phylogenetically derived trait datasets, but including the response during imputation had mixed effects: increased accuracy for *Mice* approaches, but decreased it for *Rphylopars* and *BHPMF*. Our results suggest researchers need to assess the available data and consider the need for imputation versus limiting the scope of the study or completing analyses for separate groups. Use of data imputation should be scrutinised, checking for changes in the data before and after imputation (which may indicate biases and assist in detection of imputation and slope errors). Table 1 summarises our findings as warnings and recommendations.

Table 1. Warnings and recommendations for handling missing trait values.

	Warnings ⚠ and recommendations ✓
⚠	Carefully select the taxonomic scope of the study, ensuring species are distributed across the phylogeny and trait space. If any clades or areas of the trait space are nearly or entirely absent, do not draw inference about these and exclude them from the study to prevent severe biases.
✓	Report what species/clades are included in the study and what species/clades have been removed to limit bias. Provide descriptive statistics or distribution plots for analysed trait values.
⚠	Every imputation approach produced inaccurate values, even with as little as 5% missing data. Slope errors consistently exceeded 0.1 when more than 40% of the values were missing, or when a severe bias was present.
⚠	Imputation is not always the best approach. Complete-case performs better than the tested imputation methods in some cases.
✓	If using imputation, *Rphylopars* is the best approach for handling missing continuous data resulting in smaller overall imputation and slope errors.
✓	If using *Rphylopars* or *BHPMF*, do not include the response in the imputation. If using *Mice*, including the response is beneficial.
✓	Include phylogenetic information when using imputation if possible. If a phylogeny is unavailable, but a taxonomy is available, use *BHPMF*. If there is no phylogeny or taxonomy information, use *Mice random forest* or the observation-only *BHPMF*.
✓	To assist in detecting biases and the subsequently high imputation and slope errors, assess phylogenetic clustering, as well as the change in the mean and change in the slope before and after imputation.
✓	Report the amount of missing information that was imputed and where this information falls on the phylogeny, trait and response (if applicable).

Which method performs best?

Rphylopars was the best overall imputation method in our study. However, we found scenarios were complete-case analysis better maintained the response-trait relationship, particularly compared to *Mice* and *BHPMF* imputation (but also outperforming *Rphylopars* under some severe biases). Our analyses, and others from the medical literature (Mukaka et al., 2016), show that imputation is not always the best solution to handle missing data. While imputation methods in ecology are not yet widely used, the use of imputation has been recommended over complete-case analysis in recent publications (Penone et al., 2014; Kim et al., 2018). Here, by expanding on the scenarios explored in previous studies, we show that imputation may lead to errors under some conditions. For example, when there was no true relationship between the response and trait, *Mice* approaches performed well. However, when there was a positive relationship, *Mice* did poorly even after the substantial improvement resulting from including the response in the imputation (Fig B6.3), with increases in missing data gradually shifting the positive response-trait relationship towards zero. Further investigation of *Mice* is required, as in this scenario we may expect inflated noise around the slope in *Mice* to cause more type-2 errors (reporting no relationship when one is present), but we would not expect *Mice* to systematically shift the slope itself.

This poor performance of *Mice* is particularly surprising as we made a concerted effort to further optimise *Mice*'s performance – see Appendix A4. However, the issues we have identified may be relevant only to our scenarios (imputing missing traits for phylogenetic comparative studies), and not reflect on *Mice* or multiple imputation as a whole - which are considered throughout the literature as the 'gold-standard' imputation approach (Van Buuren, 2012). Furthermore, despite making an effort to optimise the performance of *Mice*, there are a variety of *Mice* extensions and other multiple imputation approaches which may have fared better and could be tested in future comparisons e.g. Multilevel Joint Modelling (Quartagno et al., 2019), or *Mice:* Random indicator method for non-ignorable data (Van Buuren & Groothuis-Oudshoorn, 2011).

One particular issue with *Mice* was the way biases interacted with the phylogeny during phylogenetic eigenvector selection. As a control, we estimated the number of eigenvectors when no values were missing. In this scenario, most datasets had 6 -16 eigenvectors, but under a severe trait bias the number of eigenvectors frequently surpassed 20, and under a response bias, rarely reached 6. This discrepancy in the number of eigenvectors likely explains why incorporating phylogenetic information in *Mice* sometimes resulted in greater imputation and slope errors under a severe bias (Fig B1.1). Given these findings, we revise the advice of Penone et al. (2014) and Kim et al (2018) and suggest to include phylogenetic information in *Mice* imputation only when bias is weak or use alternate *Mice* options that allow a hierarchical structure similar to that used by *BHPMF*. Further work is needed to establish how different biases alter phylogenetic eigenvector selection and the downstream imputation and slope errors.

Unlike *Mice*, we suspect *BHPMF*s performance could be further enhanced (see Appendix A4). Most notably, as *BHPMF* does not allow imputation for species with no trait observations, we forced *BHPMF* to impute values by adding a dummy fully-populated variable. This allowed us to compare *BHPMFs* performance across all biases and levels of missing data and did not clearly affect *BHPMFs* performance (Fig A2.3). This feature of *BHPMF* could hinder the generality and taxonomic scale of studies but may also be beneficial if it deters researchers from imputing values in cases with very high missing data (where imputation errors are greatest). However, removing species with no trait values represents a form of complete-case analyses that could lead to biases and erroneous inferences.

Categorical traits are a common data type in ecological and evolutionary research, and cannot be imputed using *Rphylopars*, but can be handled by *BHPMF* and *Mice*. There has been limited assessment of categorical imputation performance, and available evaluations have delivered varied results (Stekhoven & Buhlmann, 2012; Akande et al., 2017; Kim et al., 2018). Future work exploring imputation errors and biases with categorical data would be valuable to guide researchers confronting missing data. Future work could also determine if machine learning approaches like *Mice random forest* and *BHPMF* would perform better with larger trait datasets (i.e. more than the 500 simulated species used in this study).

Are imputed values and relationships accurate?

The threshold for deciding whether imputation is accurate depends on the research question. For example, in *Rphylopars*, with 5% of data missing under no bias (best possible scenario), the mean imputation error was 0.56. If we assume the trait data have been log-transformed (base e), such error would mean that the mass of an African Elephant weighting 6,000kg ($e^{8.7}$) would be imputed with values as low as 3,430kg ($e^{8.7-0.56}$) or as high as 10,500kg ($e^{8.7+0.56}$). This error is worrying, especially considering *Rphylopars* is the most accurate imputation approach and we used the most favourable missing data scenario in this example. This finding suggests imputation is not accurate enough to estimate trait values for individual species or records. As such, any imputed values should be interpreted with great caution. Fortunately, trait values are more commonly imputed to establish relationships, in which case our results are less concerning. In linear regressions between a response and imputed traits, the difference between the *Rphylopars* slope and true slope was less than 0.05 at every level of missing data (except for severe biases). In many cases this would be deemed an acceptable amount of error and the same qualitative message, with a trend in the same direction (positive or negative) and not differing significantly from the true slope in most cases (Fig B4.8 – B4.11). However, this error would be large and could lead to qualitatively different messages in the context of debates about the true value of allometric exponents (Isaac & Carbone, 2010). Thus, unless the dataset is complete we recommend interpreting results cautiously, regardless of whether imputation or complete-case analysis are used for the analyses.

While different errors may be acceptable for different questions, our results show that analysing datasets where values are missing with a severe bias (MNAR) can lead to very wrong conclusions, especially when the bias acted on both the trait and response. This *bias type* was not tested by Penone et al*.* (2014), but is likely common in ecology and evolution, as both trait databases (González-Suárez et al., 2012) and response values are biased (Boakes et al., 2010; Troudet et al., 2017). In some cases, a severe trait and response bias shifted a positive response-trait relationship into no relationship, or even a negative relationship (Fig 3). Overall, the methods we tested are unsuitable when a severe bias is present. However, there are imputation options, beyond the scope of this study, designed specifically for severely biased MNAR data (Molenberghs, Fitzmaurice, Kenward, Tsiatis, & Verbeke, 2015). These MNAR options add a term to the imputation model to account for the bias. In common methods like selection, pattern-mixture and shared parameter models, this term describes a distribution aimed at explaining the mechanism by which data are missing. The parameters in these distributions (sometimes informed by expert opinion) can have a substantial impact on results, so sensitivity analysis becomes increasingly important. If a severe bias is suspected and the missing data mechanism cannot be accounted for by incorporating additional data (e.g. other traits, phylogeny, or other spatial or temporal information), these MNAR methods should be explored. However, the main challenge will be detecting the severe bias in the first place. Familiarity with the dataset, accompanied by careful checks of the data's distribution across space, time, the phylogeny, as well as the trait and response range, may help. Furthermore, we recommend accounting for biases in missing datasets before designing research, especially in phylogenetic comparative studies, where severe biases could simply be reduced by trimming the scale of the study and its conclusions to better represented groups.

Should the response be included in the imputation?

Including the response within the imputation substantially decreased imputation and slope errors in *Mice* (Moons et al., 2006; Sterne et al., 2009), and made its performance almost comparable to *BHPMF* and *Rphylopars*. However, for *BHPMF and Rphylopars*, including the response had little or a negative effect. We are unsure why including the response may negatively affect the performance of *BHPMF*, but for

Rphylopars we hypothesize it is due to how the phylogeny is incorporated. If the response is not associated to the phylogeny, including the response may skew the phylogenetic-trait covariance matrix affecting the performance of *Rphylopars*. In contrast, the phylogenetic eigenvectors that are appended to the *Mice* imputation act more like weakly correlated traits, so the benefit of adding a highly correlated response variable is clear. From this, it seems broadly advisable to include the response within *Mice* imputation and exclude it from *Rphylopars* and *BHPMF*. However, caution is needed as we suspect that these conclusions may contain caveats which warrant further research. For example, under the severe trait and response bias when there was no true relationship between the response and trait, imputation resulted in a significant negative slope particularly when the response was used in the imputation (Fig B6.5 – B6.6). This is evidence that including the response in the imputation of trait values, which will then be modelled back against the response, can cause a circularity problem. Nevertheless, when using *Mice,* this detrimental effect was small compared to the overall gains from incorporating the response in the imputation.

Can we predict whether imputation is advisable for a given dataset?

Within our work, we identify four ways in which data should be scrutinised before and after imputation to assist with bias detection that consider: *missingness*, *phylogenetic clustering*, *a change in mean*, and *a change in slope*. These metrics should not be used as a free-pass to claim the imputation is valid, as no method consistently detected bias e.g. finding no *change in slope* could occur if both imputation and complete-case analyses are equally wrong. Instead, these metrics should be used alongside careful scrutiny of the data, viewing the imputation process holistically.

Our proposed protocol includes four steps:

1) Explore the data to consider representation of the group of interest (both in trait and response) and assess the potential for severe bias.
2) Compare the distribution of trait data before and after imputation.
3) Use expert opinion and information on closely related species to determine if imputed values are plausible.
4) Use available tools to assess imputation results. *Rphylopars* and *BHPMF* currently lack imputation exploration functions, but custom checks can be created and adapted from the wide range offered in *Mice* (Van Buuren & Groothuis-Oudshoorn, 2011). *Rphylopars* and *BHPMF* produce uncertainty estimates for each imputed value, which could be scrutinised, or potentially added to models to inflate noise and make inference more robust in these single imputation approaches. Furthermore, if a

phylogenetic imputation approach is used, it is important to consider phylogenetic signal and branch length, otherwise the phylogeny may add noise (Molina-Venegas et al., 2018).

Notwithstanding these guidelines, gaps remain in the ecological and evolutionary literature on imputation. Three important future steps would be: 1) explore imputation methods and errors with categorical traits; 2) validate imputations with non-simulated trait datasets as they become increasingly populated; and 3) improving guidance on profiling data pre- and post-imputation. Finally, with recent reports of shifts away from fieldwork into a more quantitative and modelling based ecology (Ríos-Saldaña et al., 2018), it is important to note that the foundation for any trait-based analysis are the trait values, which can only become available from fieldwork and data compilation. There is still a critical need to go out into the field and collect data, particularly on poorly studied species, traits and regions.

Data accessibility

Code to generate data and repeat all analyses are publicly available at https://github.com/GitTFJ/Imputation

Appendix A - Supplementary methods

A1. Bias severities

The amount of bias in actual trait datasets is unknown and the only way to find out with any certainty if the data are biased is to collect the missing values – which is unfeasible for large-scale comparative studies. Without empirical examples to design our simulation we opted to treat weak and severe bias as lower- and upper-estimates, respectively, for the level of bias that may exist in actual trait datasets. The impact of these biases is then influenced by the amount missing data, creating an array of scenarios. For example, imputation errors may be similar for a dataset with 50% missing data under a weak bias, and a dataset with 10% missing data under a severe bias.

Despite not-knowing how much bias exists in trait datasets, we attempted to create a realistic scenario. We identified two parameters which could influence how much impact a weak bias may have. Firstly, we had to decide how missing data would vary across species. Penone et al. (2014) opted to evenly (50%) split species into two groups and delete a different proportion of values from each group. When we examined the distribution of missing data in common mammalian traits against a reference of body mass (the trait with fewest missing values) using data from PanTHERIA (Jones et al., 2009), we also identified that species fell into two groups:

low-mass species where traits were likely to be missing, and high-mass species were they were likely to be present (Figure A1.1). However, these groups were not evenly sized, with more species in the low-mass highly-likely-missing group. In particular, we found that the shift from "most species missing data" to "most species having data" occurred close to the 75% body mass quantile (Figure A2.1 for an example with two traits). We used this information to spilt species into unevenly sized groups (25/75%) when imposing the weak bias.

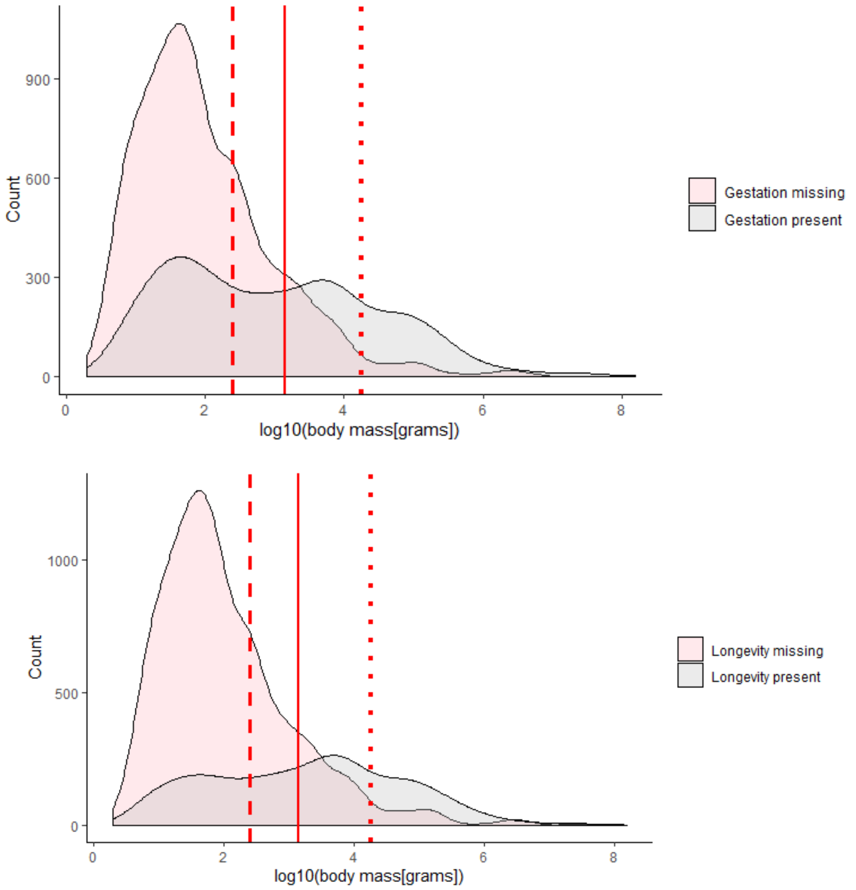


Figure A1.1. Distribution of body mass, with colour describing whether each species is missing gestation length (top), or maximum longevity (bottom) data. Red lines indicate quantile levels of the body mass distribution: dashed red line = 60%, solid red line = 75%, dotted red line = 90%. The position on the distribution where present values become more common than missing values, identifies that two groups of species are exhibiting different missing data patterns.

The second parameter to specify was how disproportionate the deletion should be within the 25/75% split. To create the lower-end estimate of bias, we wanted the trait distribution to marginally skew and shift the mean, but still contain values from across the majority of the trait range e.g. for body mass we would still have a selection of the small and large species. In pre-simulation trials we identified that sampling 4/5 of the largest 75% of values, and 1/5 of the smallest 25% of values (e.g. large values are 1.33 times more likely to be removed) was an effective choice of sampling to produce the desired weak-bias distribution. This is marginally lower than Penone et al. (2014), where missing values where 1.5 times more likely to be missing in their biased group. We also explored a more extreme deletion scenario, sampling 9/10 of the largest 75% of values, and 1/10 of the smallest 25% of values (e.g. large values are 3 times more likely to be removed), but this produced a distribution reminiscent of the severe bias scenario, so was undesirable (Figure A1.2). With more computational power, the two parameters we discuss could be flexibly adjusted to create this greater range of scenarios, but the speed of the imputation approaches was too prohibitive.

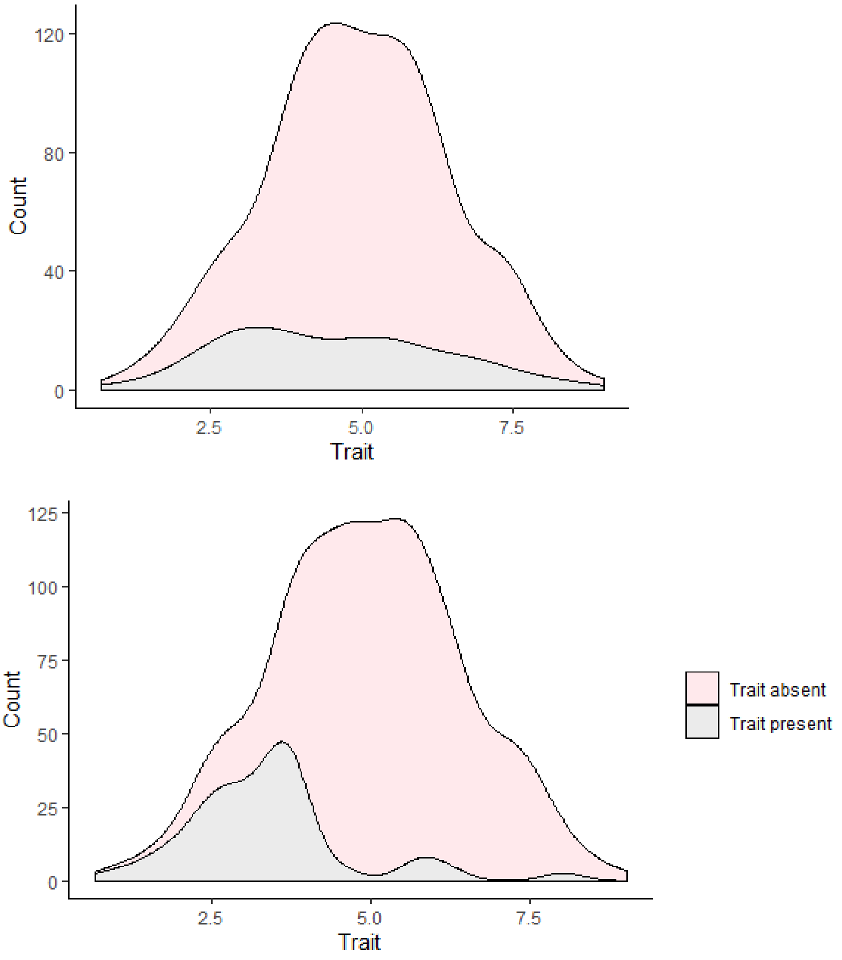
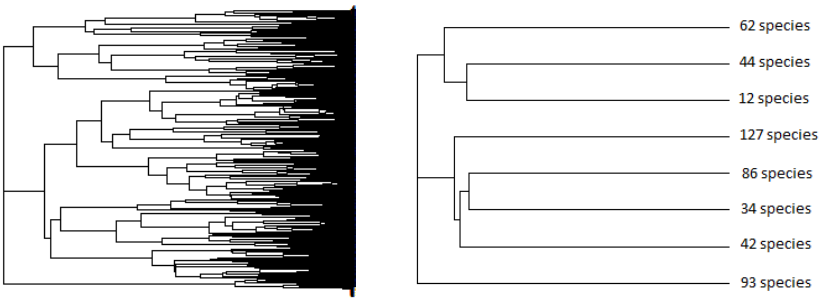


Figure A1.2. Distribution of removed trait values (pink), stacked on remaining values (grey), under two *bias types*: weak bias with 80% missing data (*missingness*) where large values are 1.33 times more likely to be missing (left), and the rejected alternative weak bias at 80% *missingness* where large values are 3 times more likely to be missing. The combined distribution (pink stacked on grey) represents the true distribution of trait values.

The weak bias groups act on the specific *bias type*, for example, under a trait bias species are split into two groups based on the trait value (e.g. smaller trait value species: 25%, larger trait value species: 75%). Whilst for a response bias, species are grouped based off the size of the response value, not the trait (e.g. smaller response value species: 25%, larger response value species: 75%).

In contrast to the weak bias, the severe bias is simpler but also less realistic, creating an extreme scenario that could occur if values are imputed without any scrutiny of the data. The exact procedure for removing values differs depending on the type of bias, but as an example, under a severe trait bias with 20% missing values, the species with the largest 20% trait values (in the trait we plan to model against the response) will have values deleted creating missing data that truncates the trait distribution. Table A1.1 provides additional examples of severe and weak bias.

Table A1.1. Description of different mechanisms of data deletion. Acronyms in the headings refer to the missing data classifications in Little & Rubin (2002). No bias (MCAR): data are 'missing completely at random'. Stratified bias: we used stratified sampling for the trait and phylogeny to improve representation of a diverse array of trait values and species. Weak bias (MAR): data are not missing at random, but the missing pattern can be explained by other available variables. In this weak bias, data are split into two groups, and values in one group are 1.33 times more likely to be removed. Severe bias (MNAR) – data are not missing at random and the missing pattern cannot be accounted for by the available variables. In this severe bias, values are systematically removed from large to small. Acronyms in square brackets [] at the end of each description denote the short-hand name for each bias type. Response refers to the outcome/dependant variable that is predicted by the 'trait of interest' which are used as predictors.

Deletion	**Deletion description**
***No bias* (MCAR)**	
Random	Removed values from each trait completely at random. [MCAR].
***Stratified* bias**	
Phylogeny	Tree was collapsed to preserve the earliest 20% of nodes. All other nodes that branched from this trimmed selection were removed and the corresponding tips were assigned to the latest linked node, creating a clade – see visual of trimming below. The number of clades in each tree ranged from 4 - 10. Trait values were deleted proportionally within each clade group, so each clade had the same proportion (approximate) of missing values (which is determined by the percentage of missing data: *missingness*). Values were removed randomly from each trait in each clade. [MOPP] 
Trait	Data removed from an approximately equal proportion (determined by *missingness* level) of species within each trait tertile (i.e, each of the three parts in an ordered distribution of trait values). Within tertiles, species were selected randomly. Proportions are approximate because the number of species per tertile was variable. [MOPT].

***Weak* bias (MAR)**

Phylogeny	Species were labelled from 1 to 500 along the phylogeny, such that closely related species received sequential number labels. Using these labels we created two groups representing 25% and 75% of the species (labels 1-125 and labels 126-500, respectively). Within each *missingness* level we then removed 1/5 of the required missing trait values from species with labels 1-125 and the remaining 4/5 of trait values from species with label 126 – 500. For example, for a 50% *missingness* level, we removed trait values from 50 species with labels 1-125, and values from 200 species with labels 126-500. Values were removed randomly within each group for each trait [WBP].
Trait	Within each *missingness* level we removed 1/5 of the required missing trait values from species in the lowest trait quartile (i.e., the lowest 25% in an ordered distribution of trait values), and the remaining 4/5 missing values from species in the remaining quartiles. For example, for 50% missing data, we removed values from 50 species in the lower quartile, and 200 species in the remaining quartiles. Species were randomly selected within groups (the lower and remaining quartiles). [WBTP]
Response	Within each *missingness* level we removed 1/5 of the required missing trait values from species in the lowest response quartile (i.e., the lowest 25% in an ordered distribution of response values), and the remaining 4/5 missing values from species in the remaining quartiles. Species were randomly selected in each group. [WBTR]
Trait*Response	For each species we calculated the product of the trait of interest value and the response value generating two groups representing the lower quartile (i.e., the lowest 25% in an ordered distribution of product values), and the remaining species. Within each *missingness* level we then removed 1/5 of trait values from species in the lowest product quartile and the remaining 4/5 from species in the remaining quartiles. Species were randomly selected in each group. [WBTPR]

***Severe* bias (MNAR)**

Phylogeny	Species were labelled from 1 to 500 along the phylogeny, such that closely related species received sequential number labels. We removed trait values systematically counting down from the species labelled 500. For example, with *missingness* level 10% species labelled 451-500 were removed. This approach resulted in situations in which entire clades were missing data, while other clades were not missing any values. [SBP]
Trait	Species were ranked based on the ordered distribution of values in the trait of interest. Values for all traits were then removed systematically down starting at the species with the highest trait value. The number of trait values removed depended on the *missingness* level, for example with 25% missing data, trait values were removed from species in the top quartile of the distribution. This approach resulted in upper truncated distributions were the largest values are missing [SBTP]
Response	Species were ranked based on the ordered distribution of values for the response. Values for all traits were then removed systematically down starting at the species with the highest response value. This approach resulted in distributions were species with the largest response values are missing trait data. [SBTR]
Trait*Response	For each species we calculated the product of the trait of interest value and the response value. Species were then ranked based on the ordered distribution of product values. Values for all traits were then removed systematically down starting at the species with the highest product value. This approach resulted in distributions were species with the combination of largest trait and response values are missing trait data. [SBTPR]

A2. Imputation methods

Mice

We evaluated the performance of three *Mice* imputation methods (from 24 available): *Mice predictive mean matching*, *Mice Bayesian linear regression*, and *Mice random forest*. The default option, *Mice predictive mean matching* imputes data by matching observed values between traits, then populates missing values in incomplete traits by adopting information from the matched species. This can be advantageous as it preserves non-linear relationships between the traits, and the imputed value is likely to be plausible as it falls within the current range of values. However, this does prevent trait values from being extrapolated beyond the available range, so should only be effective when the data available reflect the true range. One option in *Mice predictive mean matching* is to select the method for detecting matches – we used the default match type which predicts observed values and then attempts to minimise the distance between predicted values and those drawn from the approximate distribution. *Mice predictive mean matching* delivered accurate results in Penone et al. (2014) and Poyatos, Sus, Badiella, Mencuccini, & Martínez-Vilalta (2018)

Unlike *Mice predictive mean matching*, *Mice Bayesian linear regression* develops a linear model between the traits (Van Buuren & Groothuis-Oudshoorn, 2011). This allows extrapolation beyond the current range of values, but imputed values can then extend too far beyond the plausible range, with no option to constrain values without developing a customised imputation function. *Mice Bayesian linear regression* is also dependent on a linear relationship between traits, in their original scale or following transformation (Van Buuren & Groothuis-Oudshoorn, 2011).

Mice random forest uses machine learning to produce and aggregate regression trees of the observed data. This approach is robust to non-linearities in the observed data like *Mice predictive mean matching*, but can still predict beyond the current range of values like *Mice* B*ayesian linear regression* (Van Buuren & Groothuis-Oudshoorn, 2011) – incorporating the advantages of each approach. A previous case-study reported more accurate imputation in *Mice random forest* than *Mice predictive mean matching* (Shah et al., 2014). One option in *Mice random forest* is to specify the number of trees, we used the default of 10.

An important aspect of the *Mice* approaches is the ability to alter the number of times the imputation is repeated (e.g. number of chains). Repeating the imputation and modelling process multiple times, and then pooling these results reduces the type 1 error rate (Van Buuren & Groothuis-Oudshoorn, 2011). *Mice* guidelines suggest the imputation should be repeated at least five times (the default value), with more repeats as the amount of missing data increases (Van Buuren & Groothuis-Oudshoorn, 2011). As our missing data increased from 5% to 80%, we increased the number of repeats sequentially from 5 to 20. We found repeating the imputation more than 20 times was too computationally slow, and thus unfeasible for this study (we conducted thousands of separate imputations). We also opted to increase the number of iterations (the number of steps in each chain) to 10, as convergence was only marginal at the default value of 5 (Figure A2.1), and its recommended to have more iterations than variables with missing values (4 in our case) to reduce the effect of visit sequence (Van Buuren et al., 2006). We used the default visit sequence which reads variables from left to right, the data in this study were structured in the following order: response (where relevant), trait 1 (the focus of the study), trait 2 – 4, and eigenvectors (where relevant). We set the *Mice* predictor matrix to use all variables in the imputation. More details about the *Mice* imputation methods are available in the package vignette (Van Buuren & Groothuis-Oudshoorn, 2011).

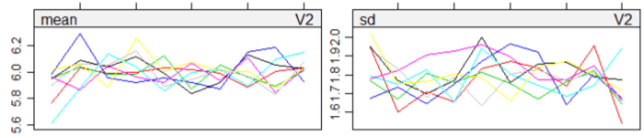
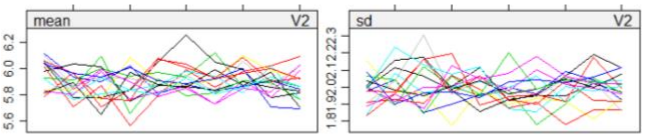
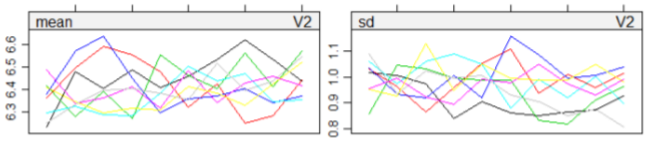
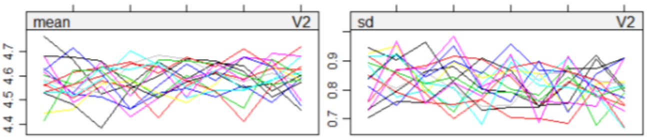
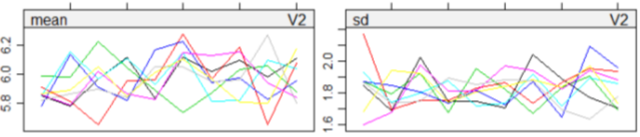
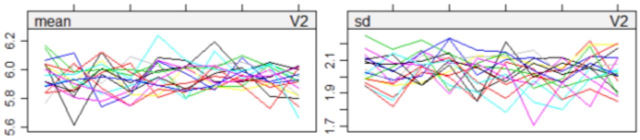
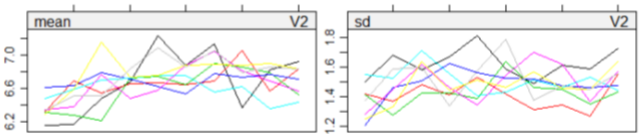
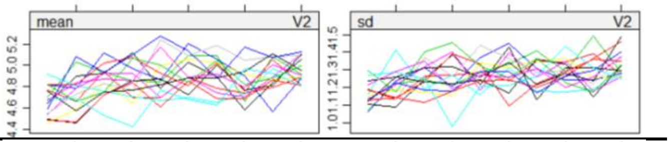
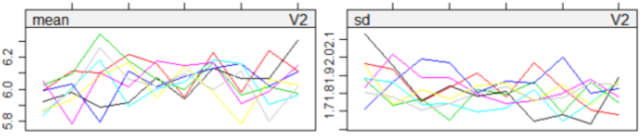
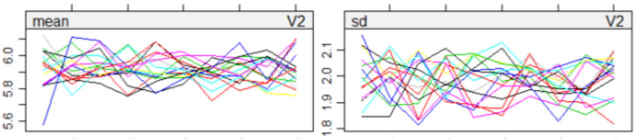
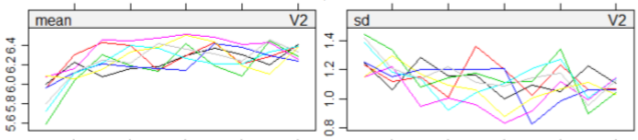
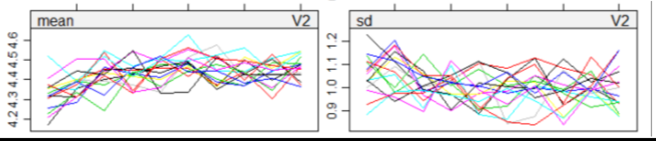
<table>
<tr><th>Method</th><th>Bias</th><th>Missing (%)</th><th>0 10 0 10</th></tr>
<tr><td rowspan="4"><i>Predictive mean matching</i></td><td>MCAR</td><td>20</td><td>
</td></tr>
<tr><td></td><td>60</td><td>
</td></tr>
<tr><td>SBTP</td><td>20</td><td>
</td></tr>
<tr><td></td><td>60</td><td>
</td></tr>
<tr><td rowspan="4"><i>Bayesian linear regression</i></td><td>MCAR</td><td>20</td><td>
</td></tr>
<tr><td></td><td>60</td><td>
</td></tr>
<tr><td>SBTP</td><td>20</td><td>
</td></tr>
<tr><td></td><td>60</td><td>
</td></tr>
<tr><td rowspan="4"><i>Random forest</i></td><td>MCAR</td><td>20</td><td>
</td></tr>
<tr><td></td><td>60</td><td>
</td></tr>
<tr><td>SBTP</td><td>20</td><td>
</td></tr>
<tr><td></td><td>60</td><td>
</td></tr>
</table>

Figure A2.1. Convergence plots for *Mice* imputation methods under a selection of *bias types* and percentage of missing data (*missingness*). Each line represents a multiple imputation chains over ten iterations. Chains show the mean and standard deviation of the target trait.

BHPMF

BHPMF is an extension of probabilistic matrix factorization, a machine learning technique that uses the correlation structure of sparse matrices to impute values. This approach shares similar advantages to *Mice random forest*, as it is robust to non-linearities in the observed data and can still predict beyond the current range of values. However, unlike *Mice random forest*, *BHPMF* is a single imputation approach. A proposed advantage to using *BHPMF* is that it incorporates correlations between traits as well as the hierarchal structure of the data to impute missing values. Incorporating hierarchical information within the imputation allows intraspecific variation, taxonomy, and spatial structure to be easily accounted for. However, we ignore the hierarchal nature of the trait data in the non-phylogenetic version of the method, only using the first level of the hierarchy (species name) within the imputation. We used the default values for all other *BHPMF* parameters.

On a few occasions, when performing *BHPMF*, the imputation ran for far longer than the average of 22 seconds, sometimes still not finishing after 40 minutes. As we needed to repeat the *BHPMF* imputation over 7000 times, we ceased any imputations that ran for longer than 3 minutes and moved onto to the next dataset. As a result, a selection of *BHPMF* imputations were excluded from the study (Figure A2.2).

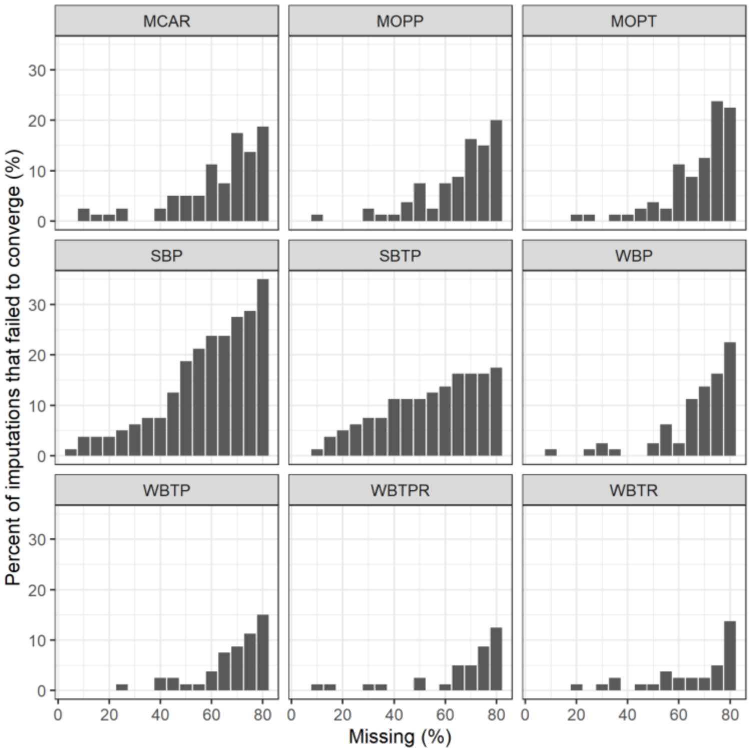


Figure A2.2. Percent of *BHPMF* imputations that over-ran the time limit across different levels of *missingness* (percentage of missing data) and *bias types*. SBTR (Severe bias response) and SBTPR (Severe bias on trait and response) did not fail on any occasions.

One obstacle with the *BHPMF* approach, is that it refuses to impute values if the species has no trait information e.g. it needs at least one trait value per species. When the response was present in the imputation, this functioned as a trait and allowed *BHPMF* to perform. However, when the response was absent, species would frequently contain no trait information, especially above 30% missing data, and in the severe biases (Figure A2.3). To provide a comparison of *BHPMF* with other imputation approaches, which allow imputation (or extrapolation) of species with no data, we included a dummy continuous variable within the trait matrix – forcing imputation and allowing comparison. This dummy variable had gaussian distribution centred at 0, with a standard deviation of 1 and an unknown degree of correlation with the other traits. This dummy trait would act as trait data, but provide little information to support *BHPMF* in estimating the missing values, essentially forcing *BHPMF* to estimate values for species with no trait data. To ensure this dummy variable did not severely detriment the performance of *BHPMF*, we compared how imputation error changed based on whether the dummy was used or not. Figure A2.3 depicts that the dummy variable showed a similar trend to the complete data (when the response and dummy where both absent).

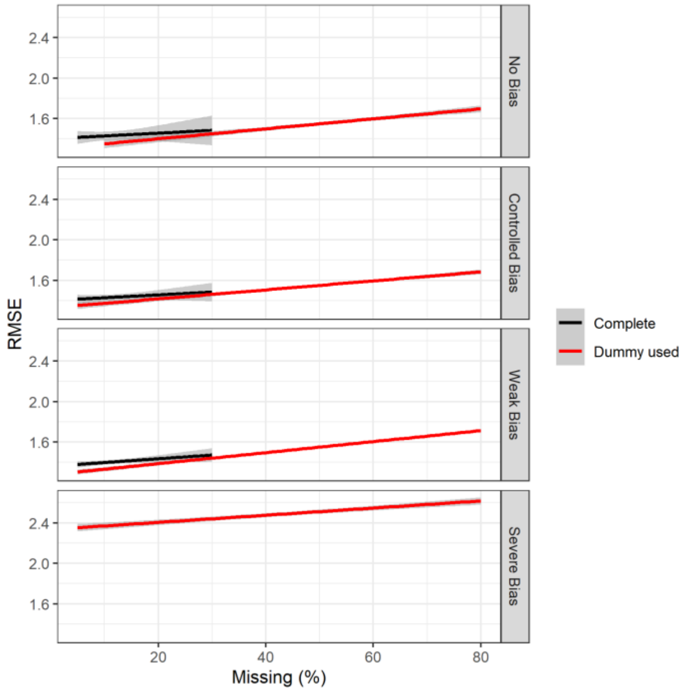


Figure A2.3. Difference between imputed and true trait values (RMSE, root mean square error) of *BHPMF* imputation when a dummy variable was present in the imputation and absent (complete – where each species has at least one trait value).

Rphylopars

Rphylopars calculates a phylogenetic predictor covariance matrix for the observed data, estimating the relationship between the traits and the phylogeny, e.g. heritable trait values for closely related species should be more similar. This phylogenetic predictor covariance matrix is then used to reconstruct the ancestral state, imputing the missing values through maximum likelihood. The phylogenetic tree is incorporated directly within *Rphylopars* through the phylogenetic predictor covariance matrix. *Rphylopars* has similar issues to *Mice* regression, with the possibility of extrapolating results too far when outliers exist in the data. *Rphylopars* does allow the user to select a minimum and maximum value, but we opted against constraining these values as in a real scenario we may not know what would classify as a reasonable boundary.

A3. Appending phylogenetic information

There is evidence that including phylogeny information can improve imputation accuracy (Penone et al., 2014; Kim et al., 2018). Unlike *Rphylopars*, which requires a phylogeny as part of the imputation, *Mice* and *BHPMF* cannot automatically incorporate a phylogeny.

Mice

For *Mice* used the MPSEM (Diniz et al., 2015) package in R to produce a selection of phylogenetic eigenvectors that best describe the relationship between the trait of interest (the trait we later model against the response) and its phylogeny.

Firstly, we fit a model between the phylogeny and the trait of interest (ignoring other traits). This model uses maximum likelihood to estimate what the phylogenies rate of evolution and steepness would need to be in order produce the available trait values. These evolution and steepness parameters are then used to create an eigenvector matrix which describes the phylogenetic distance between each species e.g. with no data missing, the matrix would contain estimates of the phylogenetic distance between each of the 500 species (500 rows/species by 500 columns/eigenvectors).

As its unfeasible to include all of these eigenvectors within the imputation model, we used forward stepwise selection (which has issues – see below) regressing the trait of interest against the eigenvectors. This identifies the eigenvectors that best explain the available trait data. When no data is missing the refined eigenvectors covered the entire phylogeny – data is available for every tip/species. However, with missing trait values, the phylogeny is trimmed and the relationship between the phylogeny and trait are based purely on available data. As a result, only species with trait values have eigenvectors. To extend the eigenvectors to every species, we used the empirically derived steepness and evolution parameters to build a new phylogenetic eigenvector matrix covering all species. From this, we pulled out the eigenvectors identified by the forward stepwise selection. This method produces a refined set of eigenvectors which describe the entire phylogeny, even in the presence of missing values. We characterise the selected eigenvectors at different levels of *missingness* and *bias type* in Figure A3.1.

Selecting eigenvectors with forward stepwise selection (significance threshold alpha = 0.01) is not desirable, as other approaches, minimising residual autocorrelation and Akaike information criterion (AIC), have been shown to be more accurate (Diniz-Filho et al., 2012). However, the minimising residual autocorrelation approach is unavailable within MPSEM, whilst the AIC approach was leaving too many eigenvectors – frequently more than 40 (sometimes up to 80), which at high levels of *missingness* well surpassed the recommended ratio of 1 variable to 3 complete cases in *Mice* imputation (Hardt et al., 2012). As a result, stepwise selection was considered the best option to refine the eigenvector matrix.

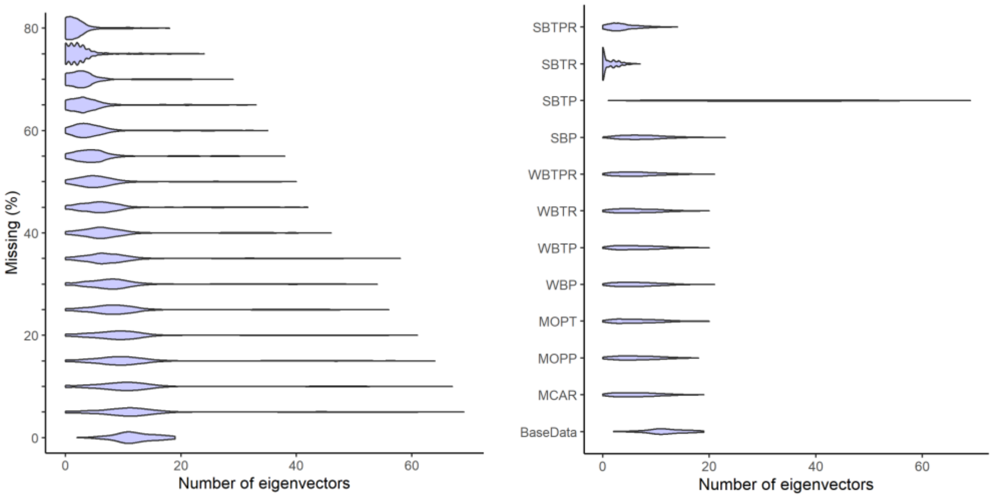


Figure A3.1. Number of eigenvectors identified as important predictors of phylogenetic signal in traits, measured at different percentages of missing data, or *missingness* (left) and *bias type* (right).

For the phylogenetic *Mice* imputations, we appended the selected eigenvectors as variables within each of the *Mice* approaches. On a few occasions we found errors within the eigenvector selection and *Mice* imputation. In three cases (0.04% of total), an error occurred when the trait data was modelled against the phylogenetic graph. We couldn’t detect the cause of this error and it seemed to derive from a package contained within MPSEM. During the analysis, these records were removed. On some occasions, *Mice* produced an error stating the imputation was ‘computationally singular’. This was caused by high multi-collinearity between eigenvectors and was particularly common when lots of eigenvectors were present (more than 25). In this circumstance, we removed 5 eigenvectors from the imputation and repeated the imputation on a loop until the imputation was no longer computationally singular. On occasion, *Mice* also offered a warning that there were too many variables to impute. To fix this, we removed the variables that *Mice* suggested were offering the least predictive power, and then repeated the imputation.

Removing eigenvectors ad-hoc is not-ideal, and in a real-scenario this could be done in a more robust way e.g. instead of arbitrarily selecting 5 as a valid number of eigenvectors to remove when *Mice* is computationally singular, we could have attempted to identify the specific eigenvectors that were causing the problem. However, as these errors and warnings were so frequent under the severe bias (Figure S4.2), it was important to implement a general rule for allowing the analysis to continue. To ensure our approaches for handling these errors were not impacting the overall results in the manuscript, we first assessed how many of the imputations required this error handling (Figure S4.2). In the non-severe biases, less than 1% of observations required eigenvectors to be removed, but the severe biases required proportionally far more error handling, with more than 90% of severe trait biases requiring eigenvectors to be trimmed. We also compared imputation error (accuracy of imputed values) when eigenvectors were removed or not (Figure A3.2). There was a small median increase in imputation error when these eigenvectors were removed, but the variance around these estimates suggest this would have little effect.

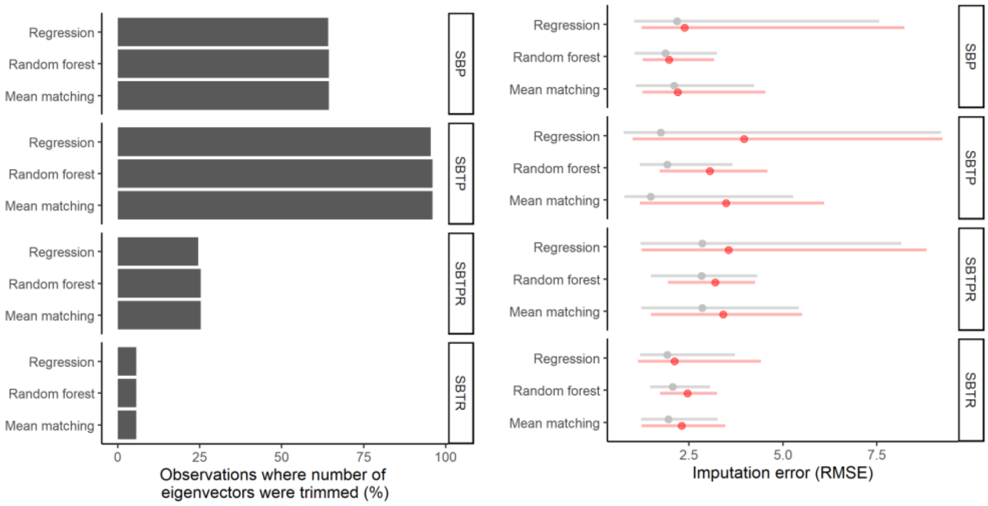
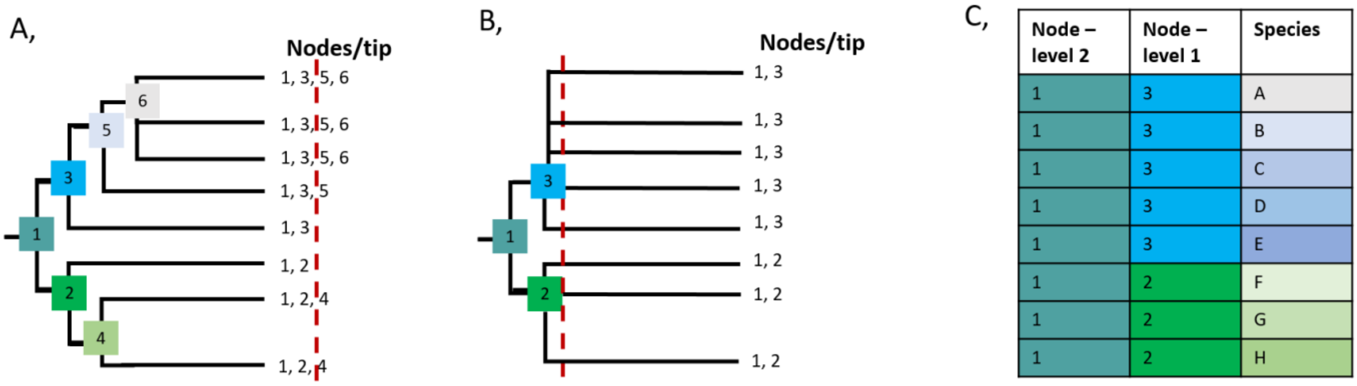


Figure A3.2. Left - Percentage of observations in *Mice* imputation methods that required eigenvectors to be removed to allow *Mice* imputations to converge or improve predictive performance. This only shows severe biases, as in the non-severe biases, eigenevtors where removed in less than 1% of cases. Right – Median Imputation error (root mean square error) with upper and lower 95% quantiles, under different imputation approaches and *bias types*. Grey depcits observations where eigenvectors where not removed. Red depicts observations with removed eigenvectors.

BHPMF

In *BHPMF*, there are two options for incorporating phylogenetic information. Firstly, the phylogenetic eigenvectors used in *Mice* could simply appended to the trait data. However, this ignores the key advantage of the *BHPMF* approach, the ability to incorporate hierarchical information. Typically, this hierarchy information would be in the form of a taxonomy, but as the data were simulated, this was not available. Instead, we utilised the phylogeny as a taxonomy, where all species are joined to one common node, which is connected to *n* more nodes, and so forth until you reach the species tip. We calculated the minimum number of nodes per species across the phylogeny. We then trimmed the tree so each tip was represented by this minimum number of nodes, keeping nodes earlier in the phylogeny. For example, if in a phylogeny the number of nodes linked to species ranges from 4 – 10, we trimmed the phylogeny, so only the four earliest nodes were represented in each species (see Figure A3.3 for an example).



Node – level 2	Node – level 1	Species
1	3	A
1	3	B
1	3	C
1	3	D
1	3	E
1	2	F
1	2	G
1	2	H

Figure A3.3. The approach for trimming a phylogeny into the hierarchal structure which can be appended to *BHPMF* imputation. A, the full phylogeny, with each coloured box representing a node. At each tip there is list of nodes that must be passed through to reach the tip. The red line indicates the minimum number of nodes across all tips (n = 2). B, the phylogeny is trimmed according to the minimum number of nodes, depicted by the red line. C, the phylogeny is described as a hierarchal matrix of nodes.

A4. Enhancing methods

Mice

We made a concerted effort to improve the performance of *Mice* relative to previous imputation comparative studies in ecology and evolution (Penone et al., 2014; Poyatos et al., 2018): 1) we used an updated and improved (albeit marginally) method for selecting phylogenetic eigenvectors for *Mice* imputation (Diniz et al., 2015); 2) included additional *Mice* imputation types that are predicted to be more accurate (Shah et al., 2014); and 3) instead of repeating each *Mice* imputation ten times before averaging, we sequentially increased the number of imputations with increased *missingness*, as recommended by the authors of the *Mice* package (Van Buuren & Groothuis-Oudshoorn, 2011).

BHPMF

Unlike *Mice*, we suspect *BHPMF*s performance could be further enhanced, as we had to deviate from the ideal *BHPMF* setup to allow comparison with the other approaches: 1) *BHPMF* does not integrate phylogenetic data but can use hierarchical taxonomic information. In a real study, taxonomies would likely be available if phylogenetic relationships are described, but for our comparison we had to convert the simulated phylogeny into a hierarchical structure (Appendix 1 – Imputation approaches) potentially adding error (although we note that the performance of *BHPMF* improved when using this phylogenetic hierarchy). 2) in some scenarios, we were forced to cease the imputation as it was failing to produce outputs within a reasonable time-frame – despite this, at least 70% of replicates were imputed in every *bias type* and level of *missingness* (Appendix 1 – Fig S5). 3) we note that *BHPMF*, and to a similar extent *Mice random forest*, are designed for imputing large datasets and it is possible increasing the number of species would decrease the discrepancy in performance between these machine learning approaches and *Rphylopars*.

A5. Alternative measures of imputation error

We selected root-mean-square-error (RMSE) to display the discrepancy between imputed and true values for two reasons: 1, RMSE is widely used and shares the same units as the data so can be easily interpreted. 2, RMSE squares the error before averaging, and then square roots the averaged error, which penalises values with large errors more than simply using the mean or median absolute errors. As a result, with RMSE any imputations that produced variable and exceptionally large difference between the imputed and true values would result in greater imputation errors. However, we also assessed imputation error using three additional metrics: mean absolute error $MeanAE$, median absolute error $MedianAE$ and R^2 between imputed and true values.

$$MeanAE = \frac{1}{N}\sum_{i=1}^{N}(y_{Im} - y_{Tr})$$

$$MedianAE = med(y_{Im} - y_{Tr})$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_{Im} - y_{Tr})^2}{\sum_{i=1}^{N}(\bar{y} - y_{Tr})^2}$$

Where N is the number of imputed values to compare, ranging from 5% of 500 (N = 25) to 80% of 500 (N = 400). y_{Im} is the imputed value for a given observation and y_{Tr} is its true value. *med* is the median value. $\bar{y}$ is the mean true value. For *Mice* imputation averaged the trait values across the repeats.

A6. Variables to model

Table A6.1. Descriptions of the variables in the models, separated into: *Descriptive* - all variables have been created explicitly by the deletion mechanism/method of imputation e.g. can answer which imputation approach is best in severe *bias types*? These variables are included in the imputation and slope error models. *Predictive* - variables to predict errors by comparing the missing values to the imputed values, without knowing anything about the *bias type* e.g. Can a big change in the mean after imputation predict a high error? These variables are included in the models to predict when imputation and slope errors might be likely based solely on the structure and characteristics of the data.

Variable	**Description**
Descriptive	
Imputation approach	Method of imputation. Categories: *Rphylopars, Mice mean matching, Mice mean matching & phylogeny, Mice regression, Mice regression & phylogeny, Mice random forest, Mice random forest & phylogeny, BHPMF, and BHPMF & phylogeny.* *Complete-case was included as a category in *slope error* analysis.
Bias type	Categories from Table 1.
Missingness	Percentage of missing values, levels varying from 5 – 80% (in 5% intervals). Modelled as a continuous variable.
Response in imputation	Whether the response was present/absent from the imputation.
Initial slope direction	Whether the response-predictor relationship is positive (slope~0.7), or has no relationship (slope~0).
Between-predictor correlation	Whether the predictors were strongly (r = 0.6) or weakly (r = 0.2) correlated.
Predictive	
Phylogenetic clustering	Distribution of predictor values across the phylogeny, from evenly distributed across the phylogeny (< 1), to all predictor values clustered within one clade (> -1). Estimated by calculating the sum of sister-clade differences (Fritz & Purvis, 2010)
Missingness	Percentage of missing values, levels varying from 5 – 80% (in 5% intervals). Modelled as a continuous variable.
Change in mean	Absolute change in the mean between complete-case analysis and imputation. Formula: Imputation mean / Pre-imputation mean. log transformed.
Change in slope	Absolute change in response-predictor slope between complete-case analysis and imputation. Formula: Absolute(Imputation range – Pre-imputation range). Square-root transformed.

Appendix B - Supplementary results

B1. Effect of including phylogeny

Including phylogenetic information reduced imputation and slope errors in all imputation methods across non-severe *bias types* (Figure B1.1). Under a severe bias in the *Mice* imputations, including the phylogeny had a mixed impact on imputation error, but increased slope error when the true relationship is positive, and decreased slope error when there is no relationship. In contrast, *BHPMF* consistently performs better with phylogenetic information across the severe biases.

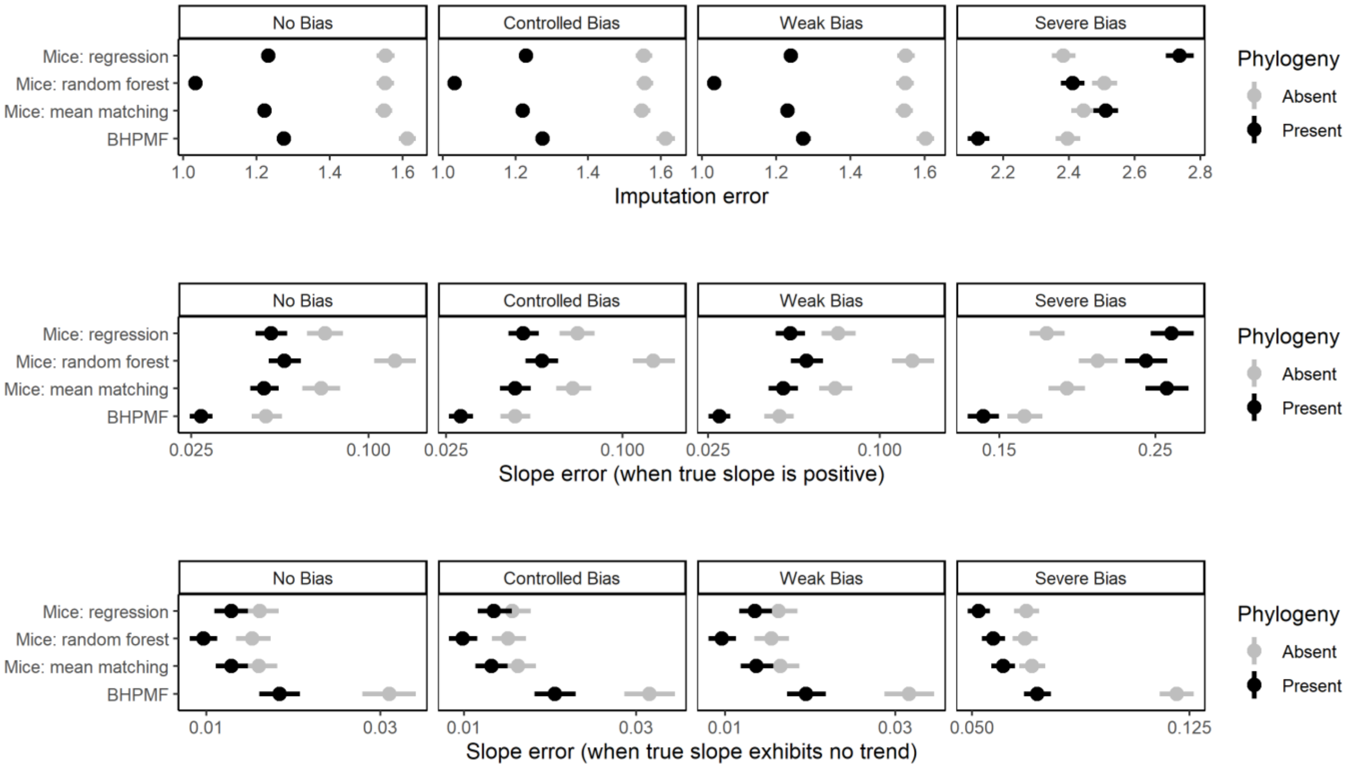


Figure B1.1. Performance of four imputation methods when phylogenetic information is present (black) or absent (grey) from the imputation. The top chart depicts the difference between imputed and true trait values (RMSE, root mean square error). The bottom two charts depict the absolute difference between the true slope and the slope after missing values are imputed, where the true relationship is positive (middle) or neural (bottom), respectively. *Rphylopars* is excluded from these plots as it was only tested when a phylogeny is present. Points show the mean error and 95% confidence intervals, all averaged over model covariates that are not displayed: *seed, response in imputation, missingness,* and *between-predictor correlation. Bias type* is averaged into the following severity levels: No bias – No averaging needed, Controlled bias – MOPT & MOPP averaged, Weak bias – WBP, WBTP, WBTR & WBTPR averaged, Severe bias – SBP, SBTP, SBTR & SBTPR averaged. In the top plot, values are also averaged over *initial-slope direction.* The x-axes differ between the plots.

B2. Model outputs

Table B2.1. Refined model output from imputation error model, showing the difference between *Rphylopars* and all other imputation approaches. Parameter estimates and confidence intervals are back-transformed from a log10 scale.

Relationship	Parameter estimate	Lower 95% confidence interval	Upper 95% confidence interval	t	p
Rphylopars Vs Mice:					
regression	2.17	2.13	2.22	77	< 0.001
regression + phylogeny	1.66	1.63	1.69	50	< 0.001
mean matching	2.24	2.20	2.29	80	< 0.001
mean matching + phylogeny	1.70	1.66	1.73	52	< 0.001
random forest	2.33	2.29	2.38	84	< 0.001
random forest + phylogeny	1.48	1.45	1.51	39	< 0.001
BHPMF	2.39	2.34	2.43	87	< 0.001
BHPMF + phylogeny	1.86	1.82	1.90	62	< 0.001

Table B2.2. Refined model output from slope error model (with a positive response-predictor relationship), showing the difference between *Rphylopars* and all other imputation approaches. Parameter estimates and confidence intervals are back-transformed from a square-root scale.

Relationship	Parameter estimate	Lower 95% confidence interval	Upper 95% confidence interval	t	p
Rphylopars Vs:					
regression	0.020	0.019	0.022	55	< 0.001
regression + phylogeny	0.022	0.020	0.023	56	< 0.001
mean matching	0.021	0.020	0.023	56	< 0.001
mean matching + phylogeny	0.020	0.019	0.023	54	< 0.001
random forest	0.036	0.037	0.038	72	< 0.001
random forest + phylogeny	0.023	0.021	0.024	59	< 0.001
BHPMF	0.011	0.01	0.012	20	< 0.001
BHPMF + phylogeny	0.003	0.002	0.003	40	< 0.001
complete-case	0.0001	0.00001	0.0002	4	< 0.001

Table B2.3. Refined model output from slope error model (with no response-predictor relationship), showing the difference between *Rphylopars* and all other imputation approaches. Parameter estimates and confidence intervals are back-transformed from a square-root scale.

Relationship	Parameter estimate	Lower 95% confidence interval	Upper 95% confidence interval	t	p
Rphylopars Vs:					
regression	0.002	0.002	0.002	31	< 0.001
regression + phylogeny	0.0005	0.0007	0.0004	17	< 0.001
mean matching	0.002	0.002	0.002	32	< 0.001
mean matching + phylogeny	0.0009	0.001	0.0007	22	< 0.001
random forest	0.002	0.002	0.001	29	< 0.001
random forest + phylogeny	0.0003	0.0004	0.0001	12	< 0.001
BHPMF	0.011	0.010	0.012	70	< 0.001
BHPMF + phylogeny	0.003	0.0025	0.003	35	< 0.001
complete-case	0.005	0.004	0.005	50	< 0.001

B3. Imputation and slope errors – phylogeny absent

Imputation error is comparable across all non-phylogenetic imputation methods (Figure B3.1). However, there is greater variation in slope error between the methods, and these results vary depending on the starting trend (Figure B3.1). When the true response-trait relationship is positive (Figure B3.2), complete-case is the best option in all but two severe bias scenarios (trait and response, and response), regardless of the percentage of missing data (*missingness*). In these two severe bias scenarios, errors are high and no one method performs best. *BHPMF* is the second-best option overall, and *Mice random forest* performs worst in the majority of scenarios. These results change dramatically when the response-trait relationship is neutral (Figure B3.3), with *Mice* consistently delivering the best results, followed by complete-case analysis, and *BHPMF.* These results are largely consistent with phylogenetic imputation method results.

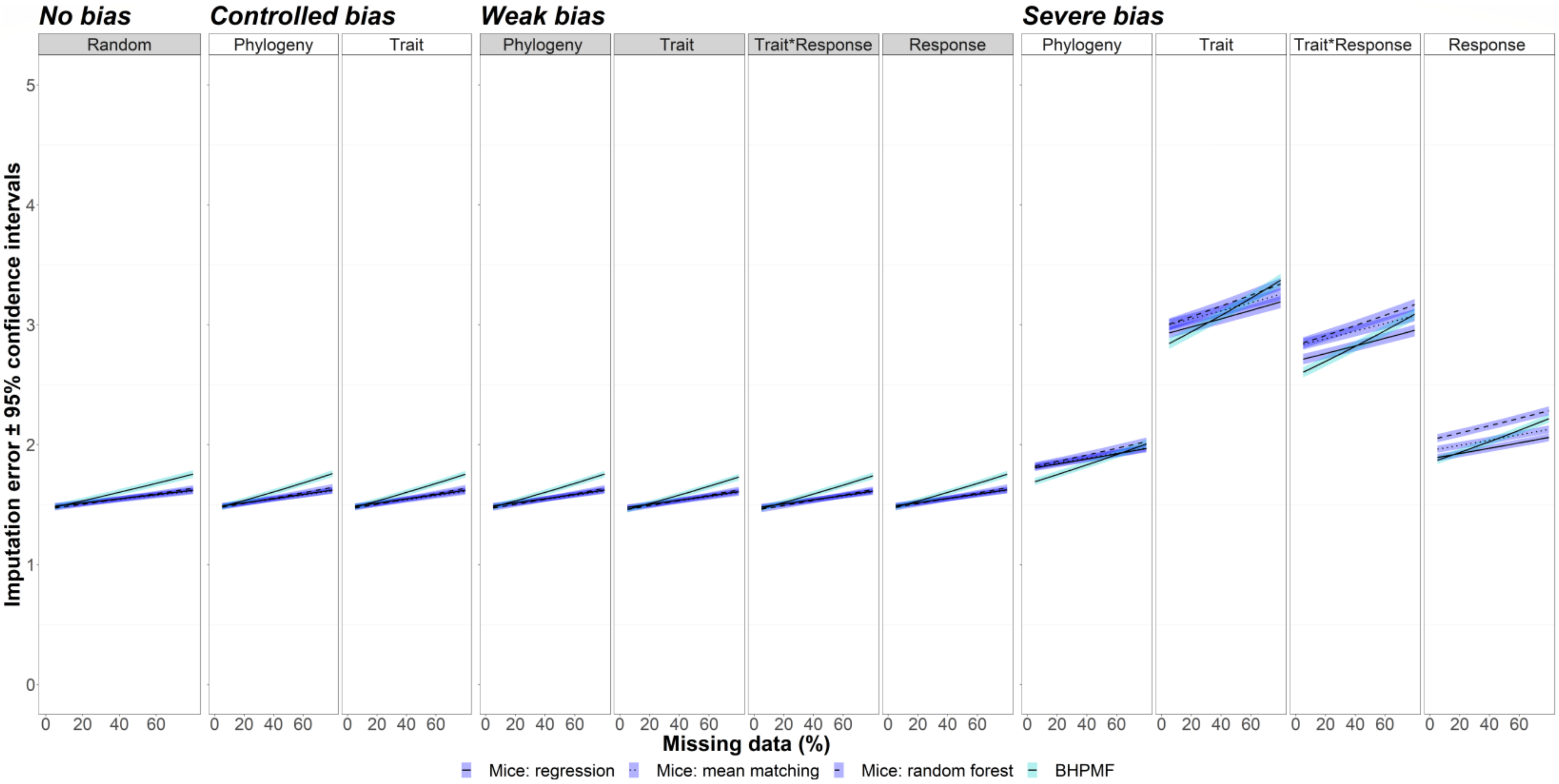


Figure B3.1: Difference between imputed and true trait values (RMSE, root mean square error) for four non-phylogenetic imputation approaches under varying percentages of missing data (*missingness*) and *bias types*. Lines depict the marginal effect of *missingness* and *bias type* from a regression model, and were averaged across other predictors: *seed, response in imputation, between-trait correlation,* and *initial slope direction*. Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%)

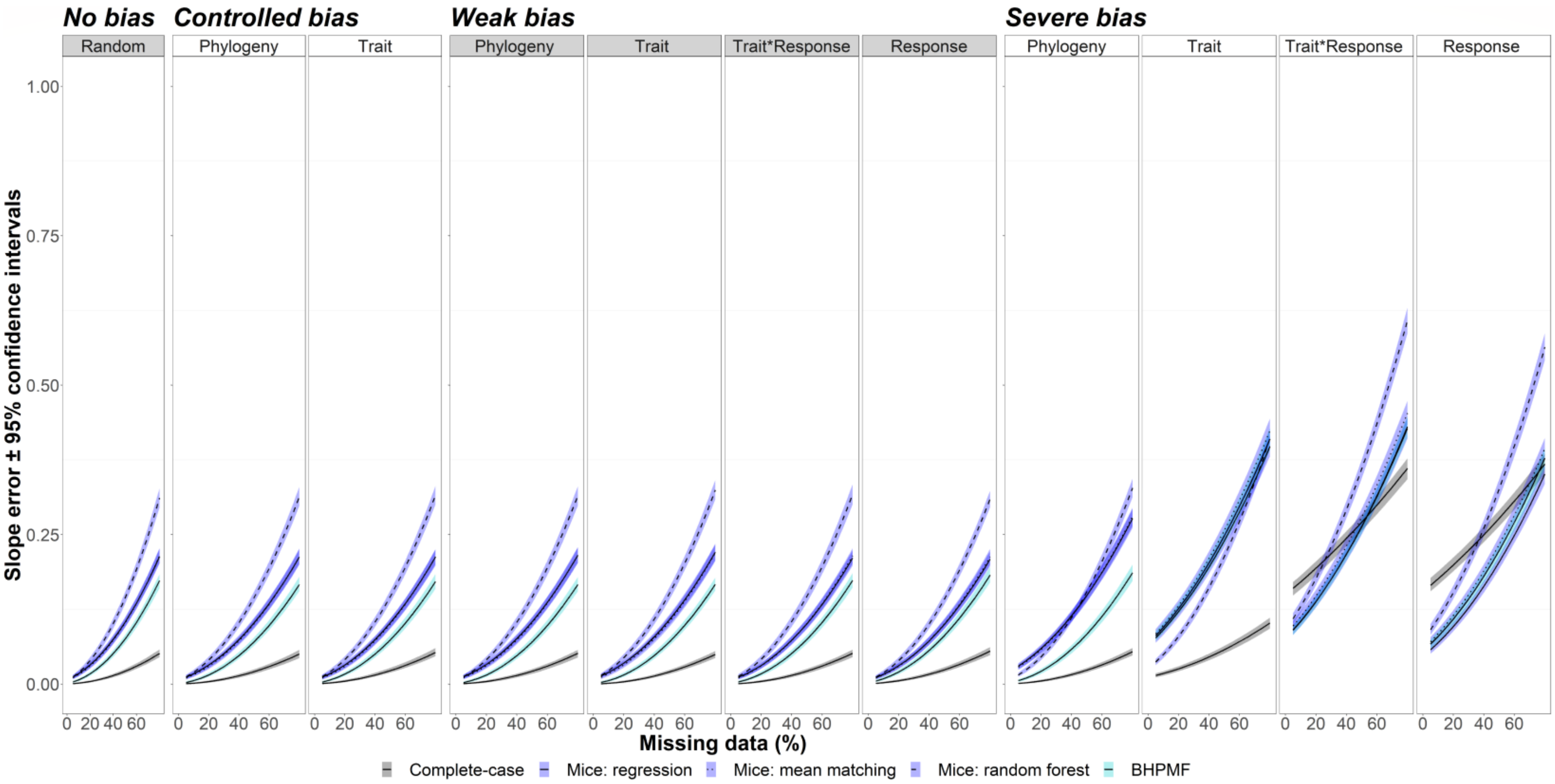


Figure B3.2. Absolute difference between the true response-trait slope coefficient and the slope coefficients in datasets with varying percentages of missing trait values (*missingness*) - removed under a series of *bias types*. Missing values were imputed under four non-phylogenetic approaches, or treated as complete-case analysis. The true slope has positive relationship between the response and trait (r ~ 0.7). Lines depict the marginal effect of *missingness* and *bias type* from a regression model, and were averaged across other predictors: *seed, response in imputation,* and *between-trait correlation.* Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%).

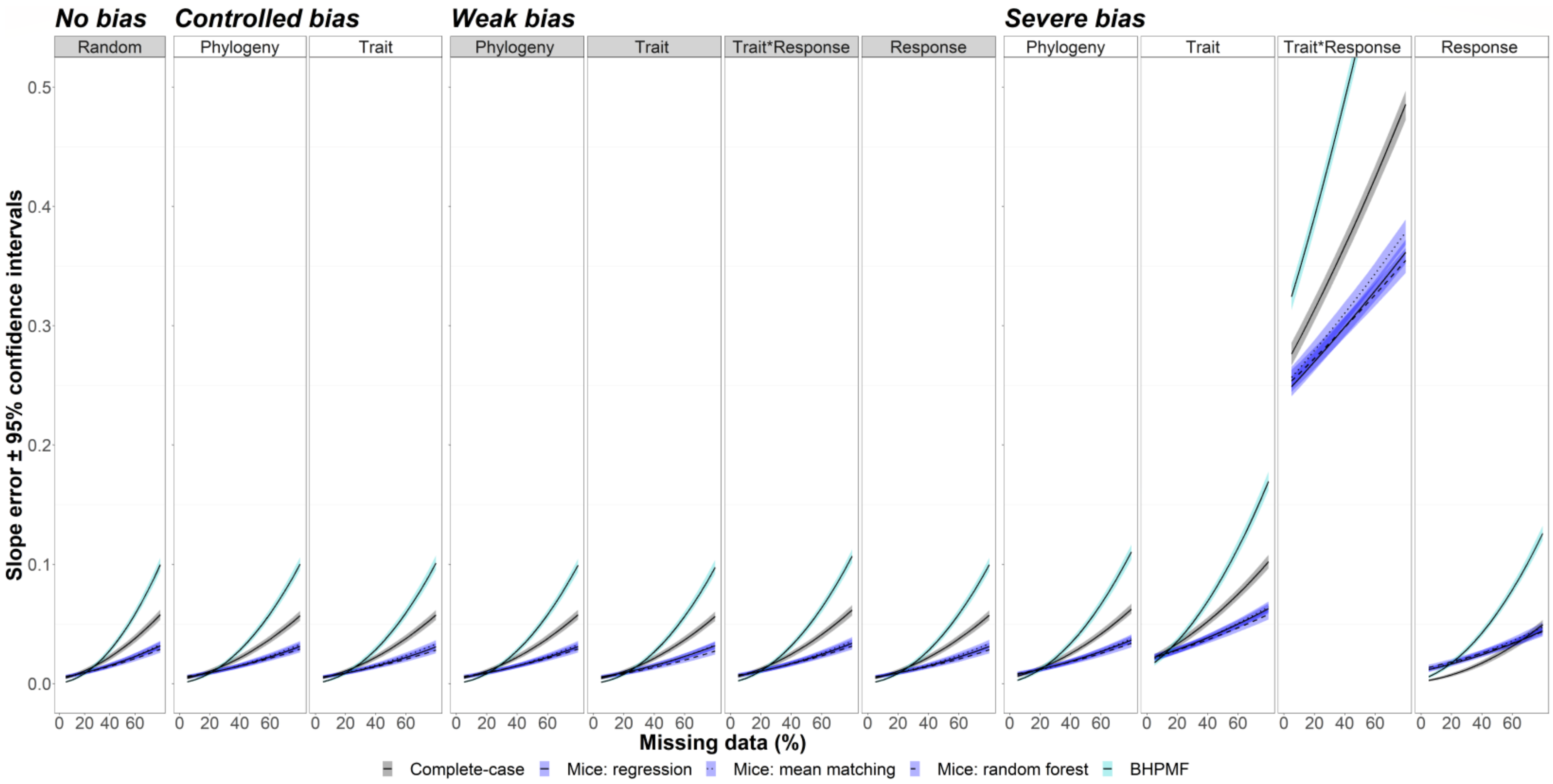


Figure B3.3. Absolute difference between the true response-trait slope coefficient and the slope coefficients in datasets with varying percentages of missing trait values (*missingness*) - removed under a series of *bias types*. Missing values were imputed under four non-phylogenetic approaches, or treated as complete-case analysis. The true slope has no relationship between the response and trait (r ~ 0). Lines depict the marginal effect of *missingness* and *bias type* from a regression model, and were averaged across other predictors: *seed, response in imputation,* and *between-trait correlation.* Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%).

B4. Imputation, slope, and inference errors – phylogeny present

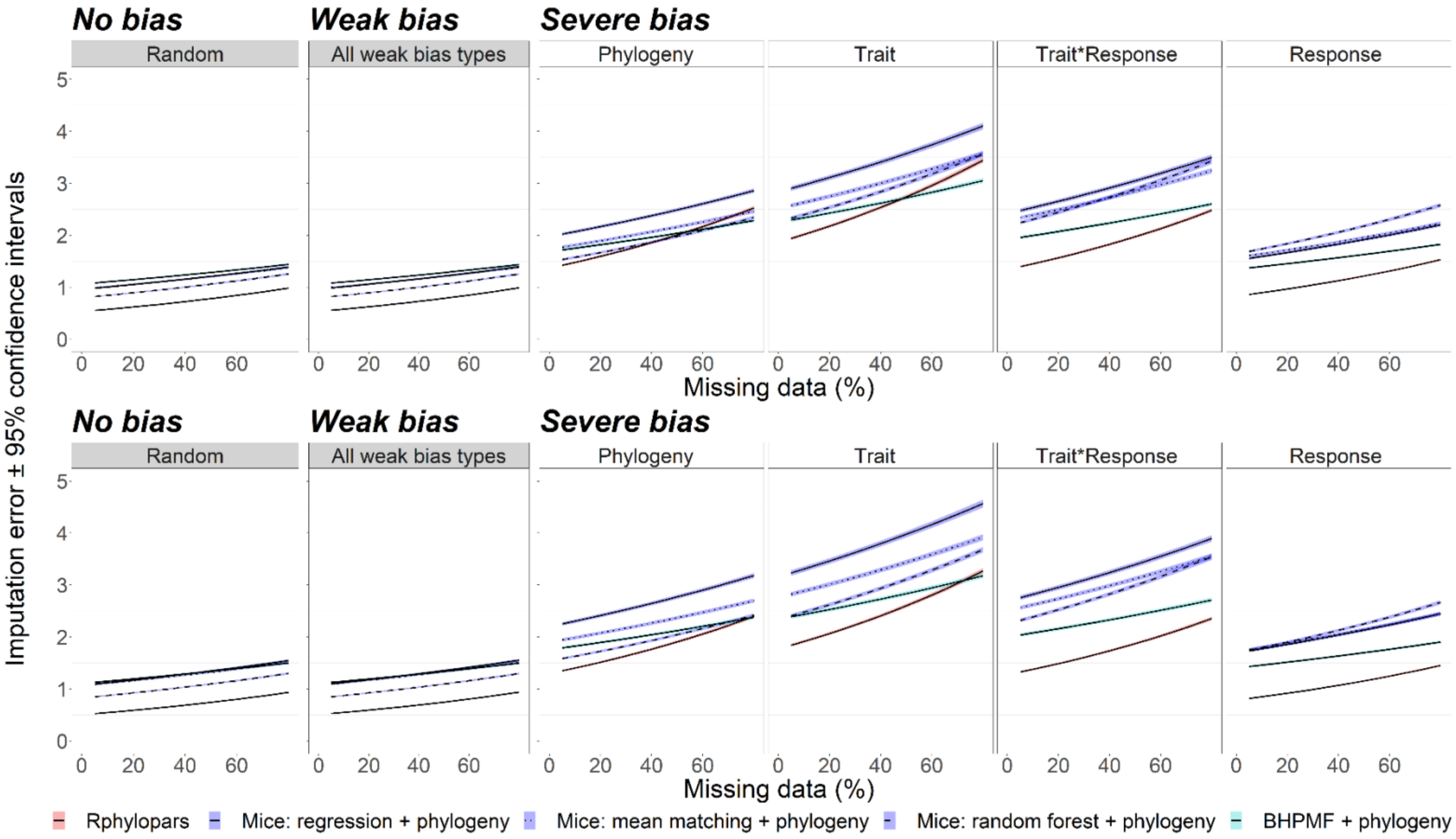


Figure B4.1: Difference between imputed and true trait values (RMSE, root mean square error) for five phylogenetic imputation approaches under varying percentages of missing data (*missingness*) and *bias types*. The top row of panels show datasets in which the true slope was positive (r ~ 0.7), while the bottom row shows datasets with no relationship (r ~ 0). Lines depict the marginal effect of *missingness* and *bias type* from a regression model, and were averaged across other predictors: *seed, response in imputation* and *between-trait correlation.* Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%)

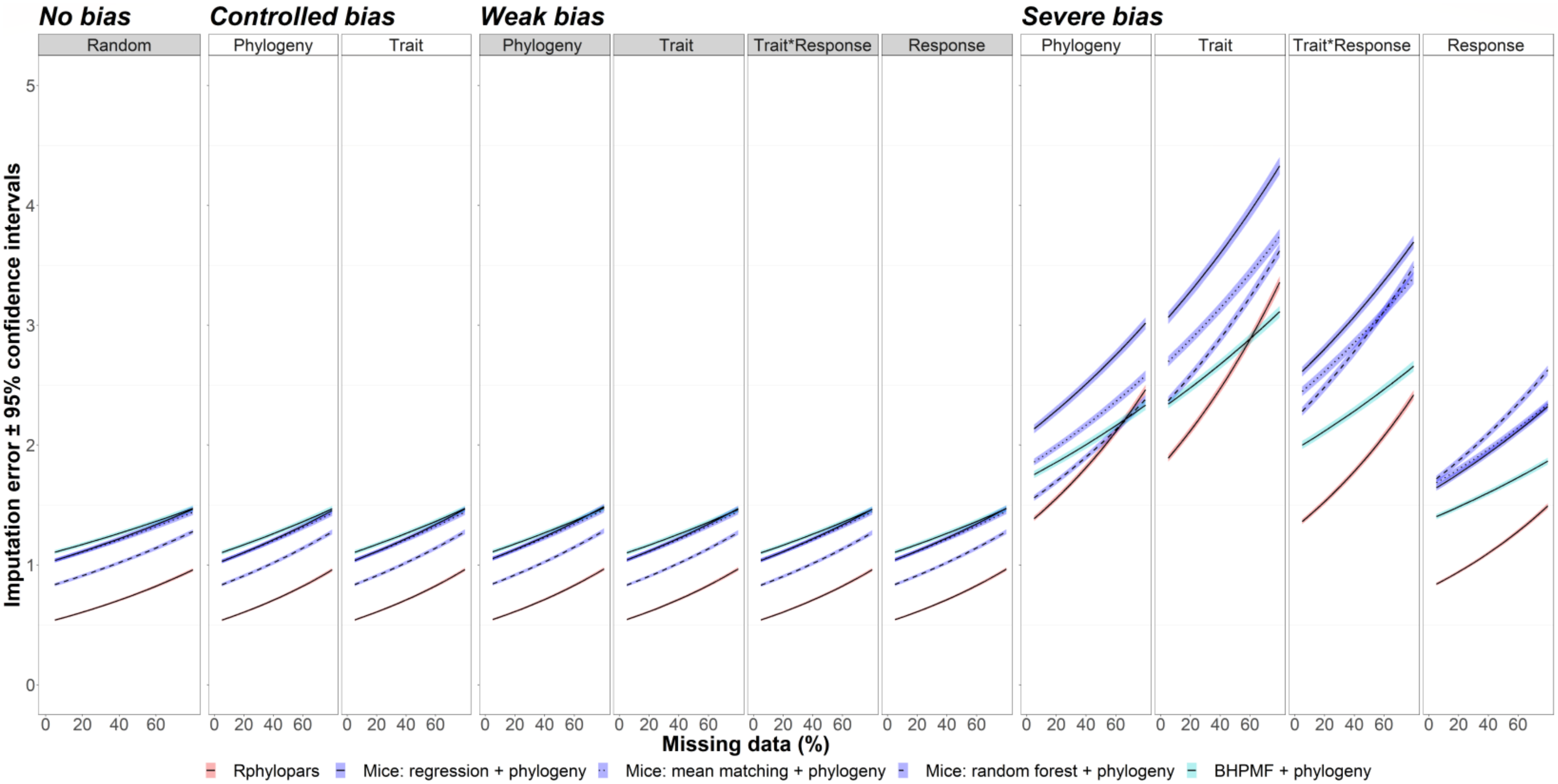


Figure B4.2: Difference between imputed and true trait values (RMSE, root mean square error) for five phylogenetic imputation approaches under varying percentages of missing data (*missingness*) and *bias types*. Lines depict the marginal effect of *missingness* and *bias type* from a regression model, and were averaged across other predictors: *seed, response in imputation, between-trait correlation,* and *initial slope direction*. Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%). This plot is an expanded version of Fig 2 in the main text.

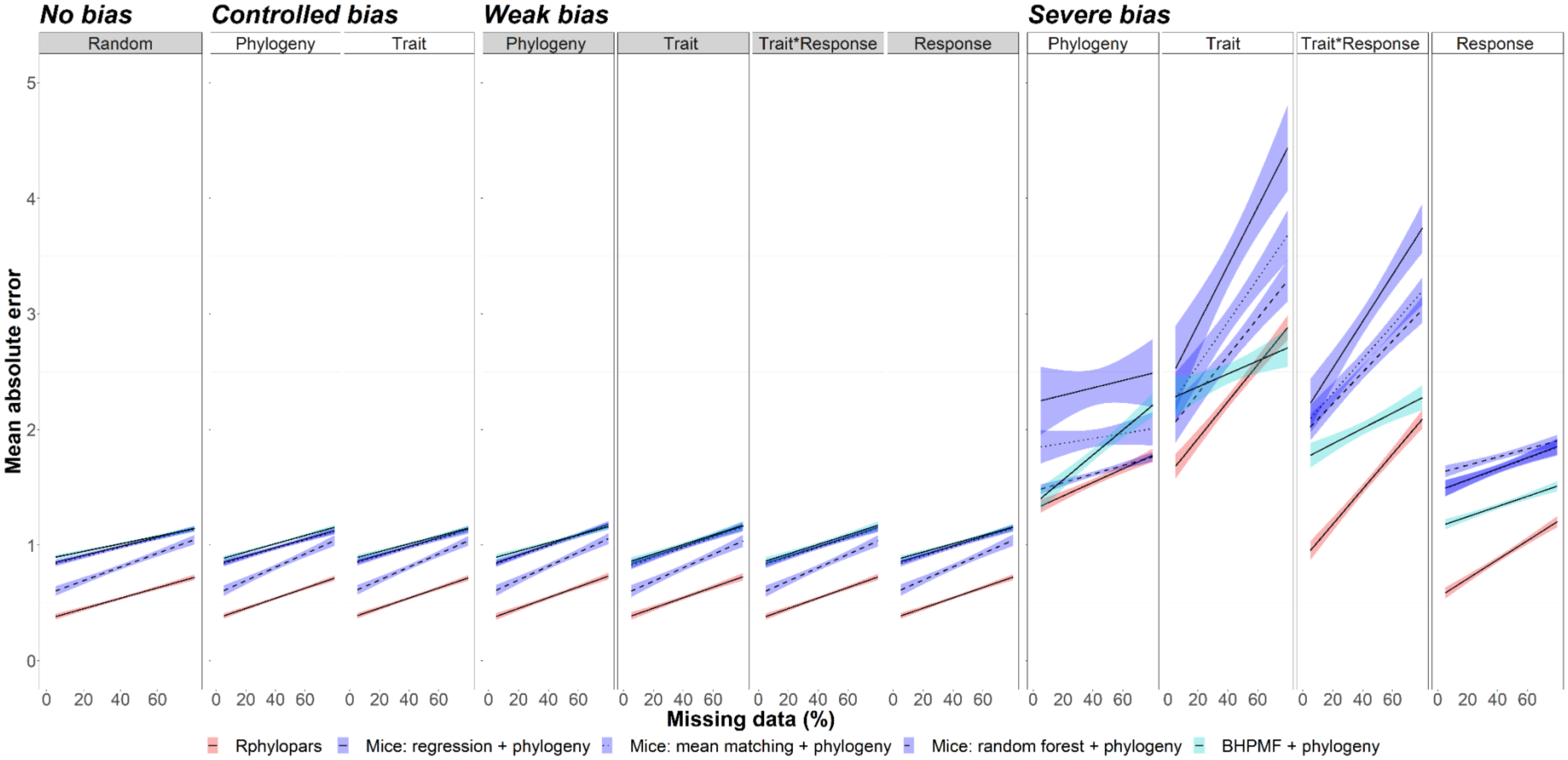


Figure B4.3: Absolute difference between imputed and true trait values (mean absolute error) for five phylogenetic imputation approaches under varying percentages of missing data (*missingness*) and *bias types*. Lines depict the marginal effect of *missingness* and *bias type* from a regression model, and were averaged across other predictors: *seed, response in imputation, between-trait correlation,* and *initial slope direction*. Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%).

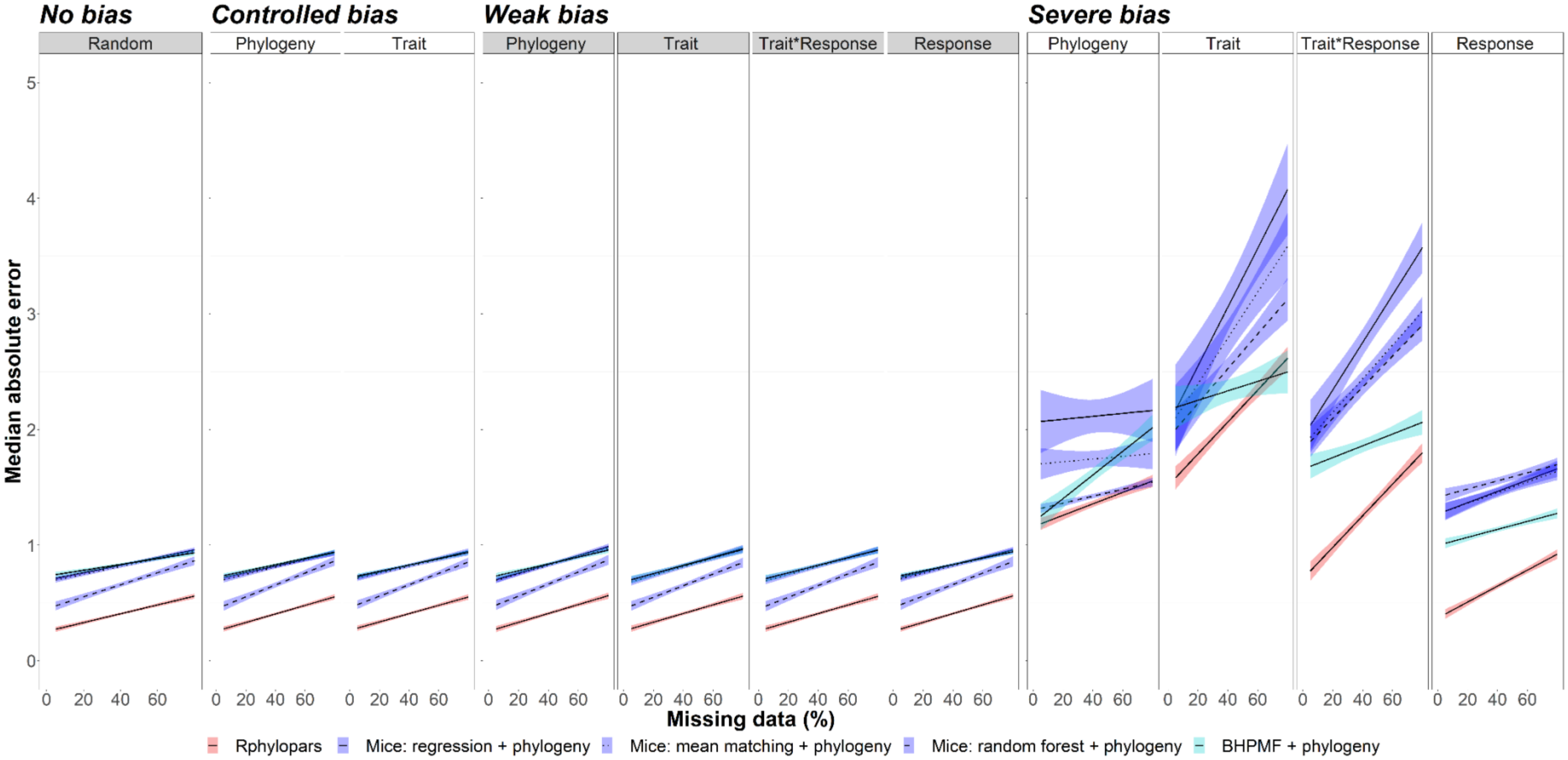


Figure B4.4: Absolute difference between imputed and true trait values (median absolute error) for five phylogenetic imputation approaches under varying percentages of missing data (*missingness*) and *bias types*. Lines depict the marginal effect of *missingness* and *bias type* from a regression model, and were averaged across other predictors: *seed, response in imputation, between-trait correlation,* and *initial slope direction*. Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%).

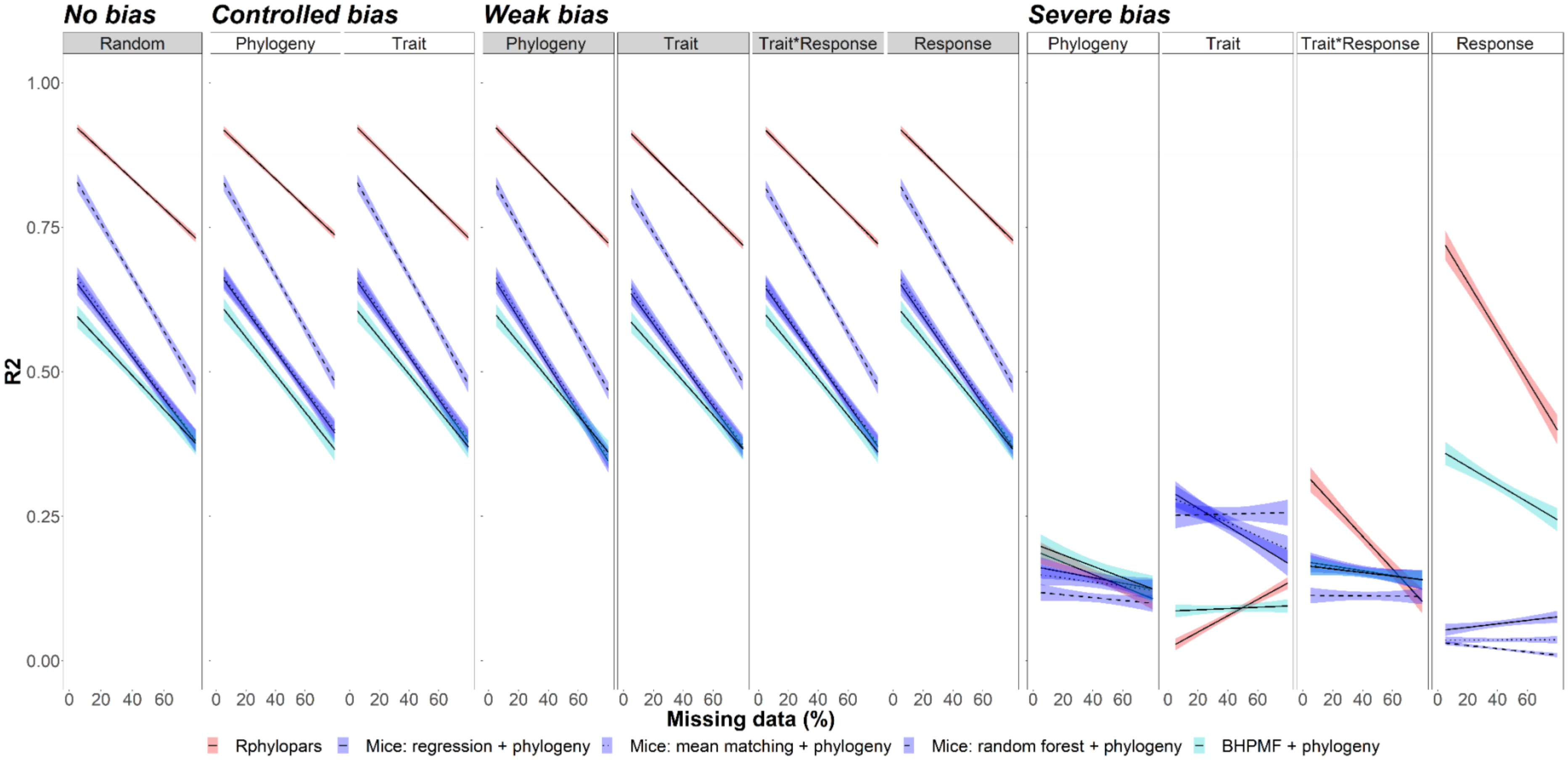


Figure B4.5: R^2 between imputed and true trait values for five phylogenetic imputation approaches under varying percentages of missing data (*missingness*) and *bias types*. Lines depict the marginal effect of *missingness* and *bias type* from a regression model, and were averaged across other predictors: *seed, response in imputation, between-trait correlation,* and *initial slope direction*. Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%).

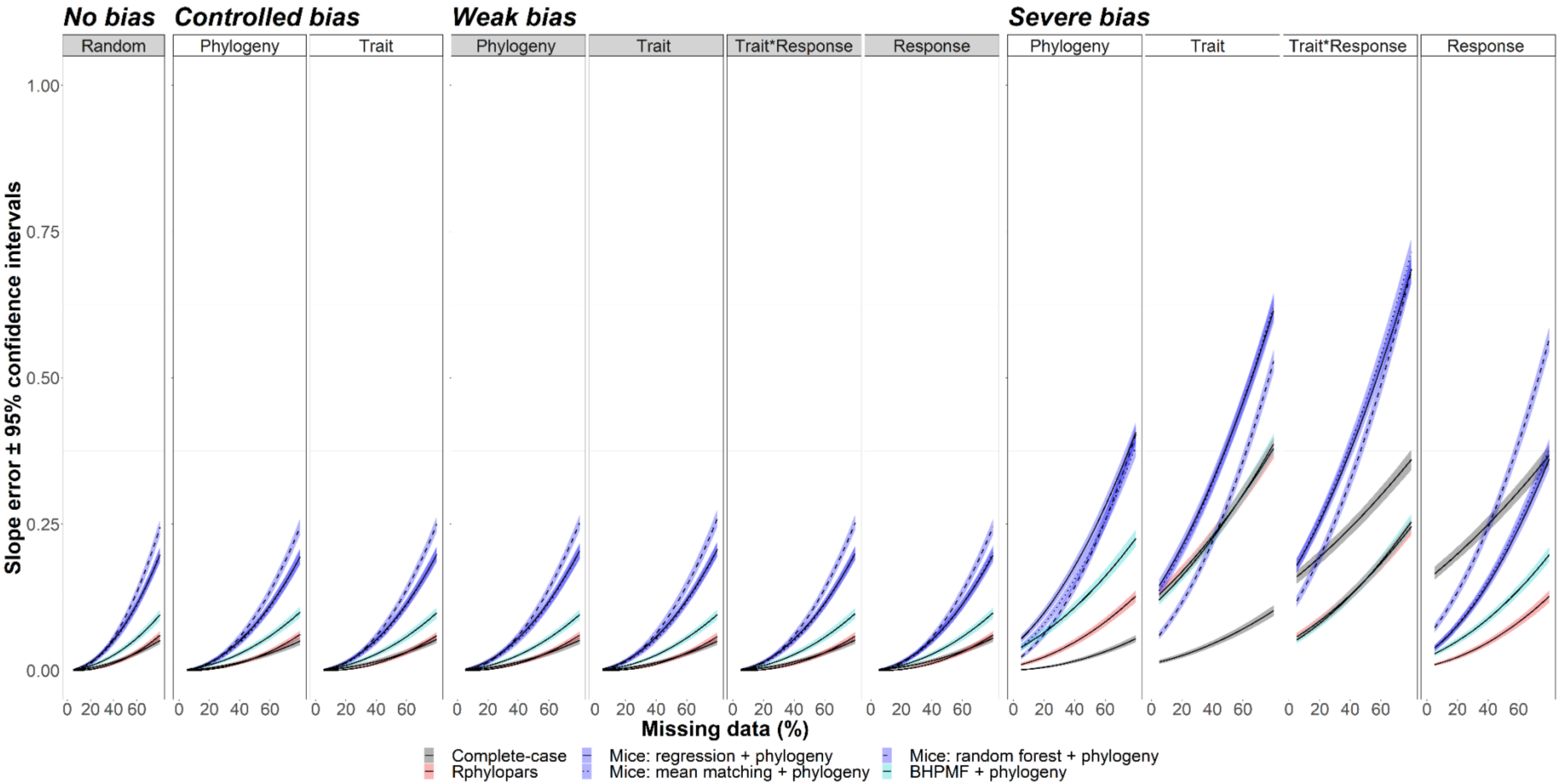


Figure B4.6. Absolute difference between the true response-trait slope coefficient and the slope coefficients in datasets with varying percentages of missing trait values (*missingness*) - removed under a series of *bias types*. Missing values were imputed under five phylogenetic approaches, or treated as complete-case analysis. The true slope was positive (r ~ 0.7). Lines depict the marginal effect of *missingness* and *bias type* from a regression model, and were averaged across other predictors: *seed, response in imputation,* and *between-trait correlation*. Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%).

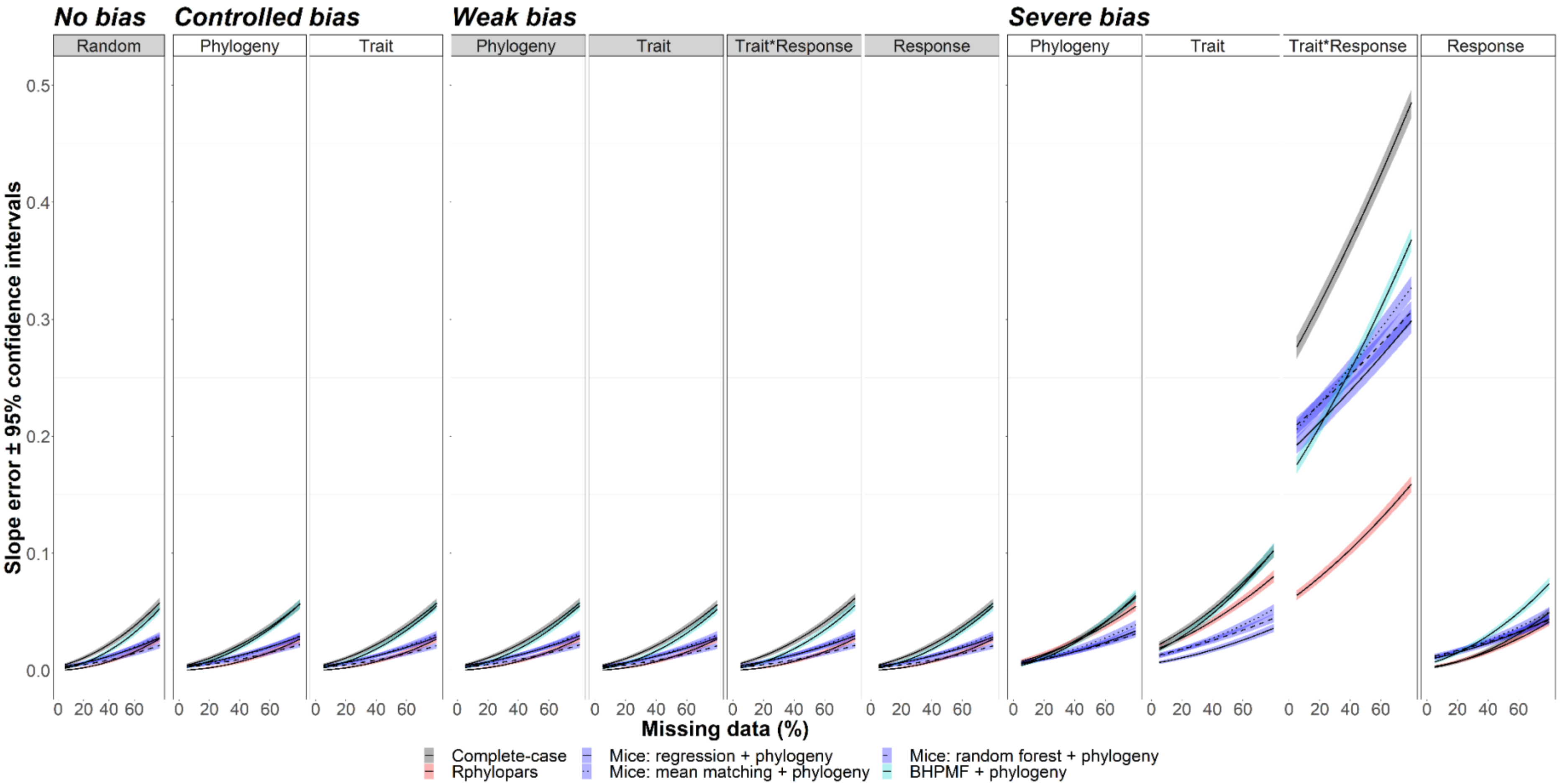


Figure B4.7. Absolute difference between the true response-trait slope coefficient and the slope coefficients in datasets with varying percentages of missing trait values (*missingness*) - removed under a series of *bias types*. Missing values were imputed under five phylogenetic approaches, or treated as complete-case analysis. The true slope had no relationship (r ~ 0). Lines depict the marginal effect of *missingness* and *bias type* from a regression model, and were averaged across other predictors: *seed, response in imputation,* and *between-trait correlation*. Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%).

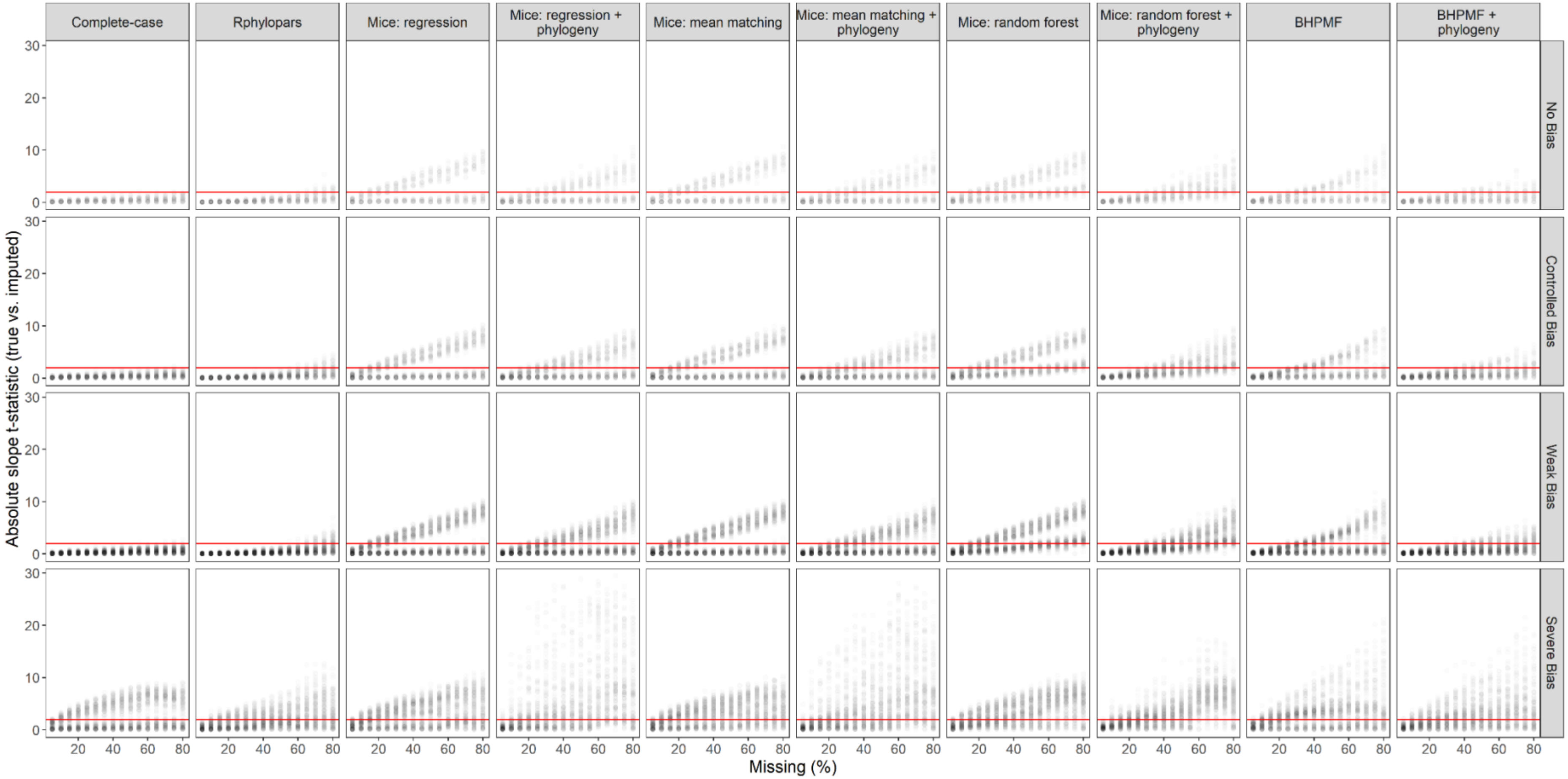


Figure B4.8. Difference (represented by the absolute t-statistic) between the true response-trait slope and the slope in datasets with varying percentages of missing trait values (*missingness*) - removed under a series of *bias types.* Missing values were imputed under nice approaches, or treated as complete-case analysis. Points above the red lines (set at 1.96) differ significantly from the true slope at the 95% confidence level. The true slope has a positive relationship (r ~ 0.7) between the response and trait, and the standard error of the slope ranges from 0.02 – 0.03. The multiple overlapping points at each level of *missingness* represent the different *bias types* within the varying bias severities, which include: Controlled bias – MOPT & MOPP, Weak bias – WBP, WBTP, WBTR & WBTPR, Severe bias – SBP, SBTP, SBTR & SBTPR.

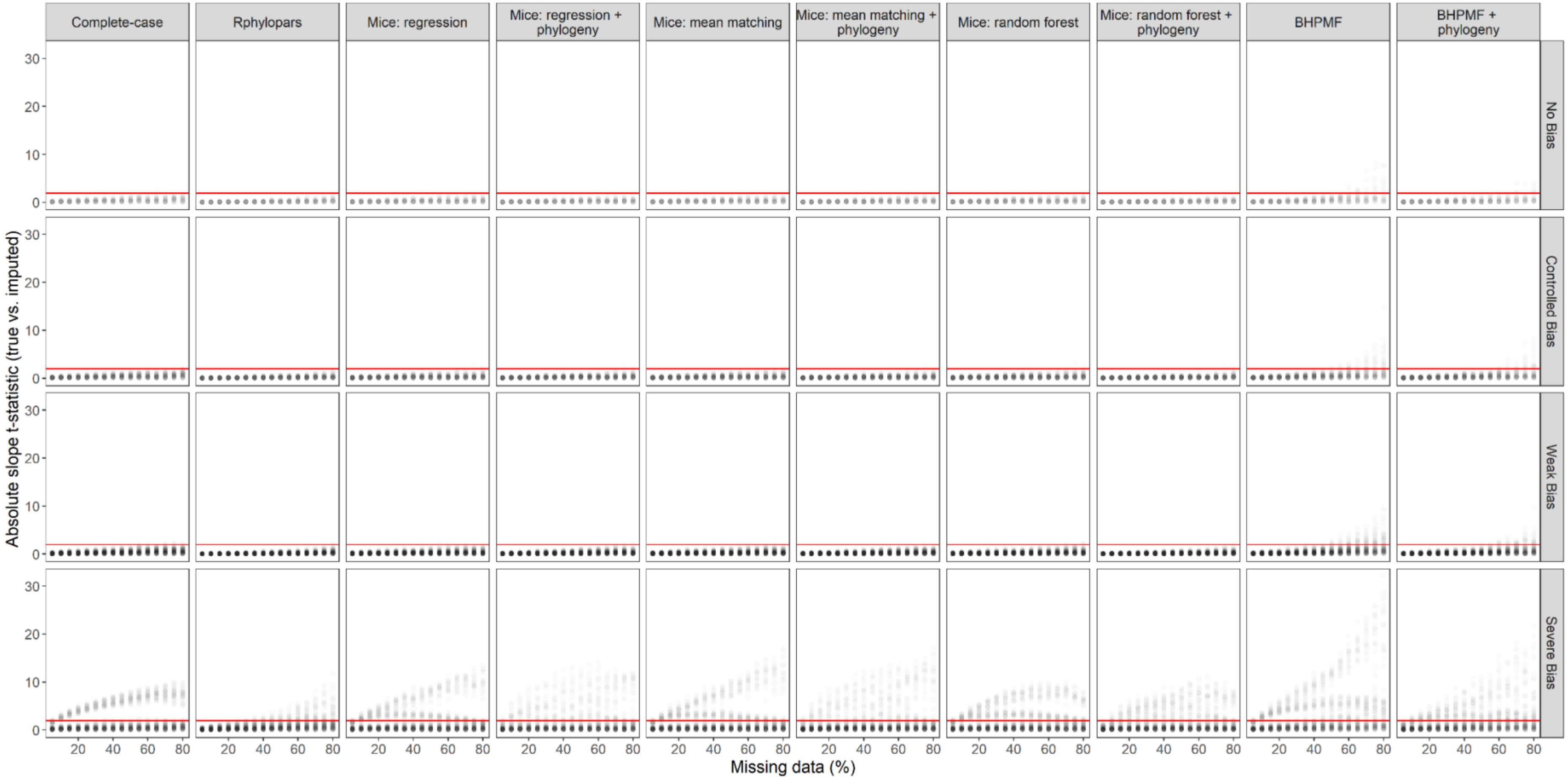


Figure B4.9. Difference (represented by the absolute t-statistic) between the true response-trait slope and the slope in datasets with varying percentages of missing trait values (*missingness*) - removed under a series of *bias types.* Missing values were imputed under nine approaches, or treated as complete-case analysis. Points above the red lines (set at 1.96) differ significantly from the true slope at the 95% confidence level. The true slope has no relationship (r ~ 0) between the response and trait, and the standard error of the slope ranges from 0.03 – 0.05. The multiple overlapping points at each level of *missingness* represent the different *bias types* within the varying bias severities, which include: Controlled bias – MOPT & MOPP, Weak bias – WBP, WBTP, WBTR & WBTPR, Severe bias – SBP, SBTP, SBTR & SBTPR.

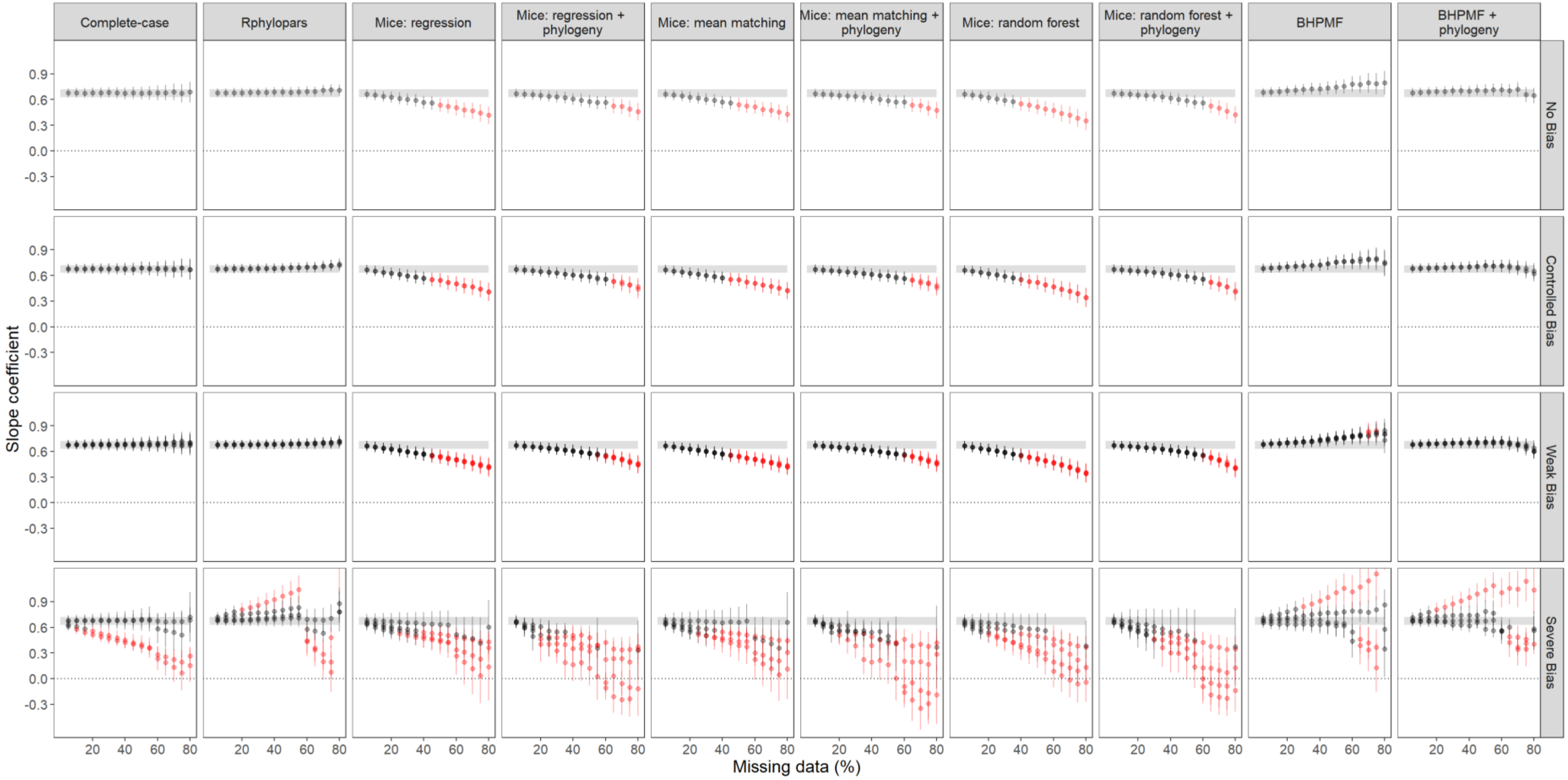


Figure B4.10. Slope coefficients (points) and 95% confidence intervals (vertical lines) under different *bias types* and levels of *missingness* for nine imputation approaches as well as complete-case analysis. The grey horizontal bar represents the lower and upper 95% confidence intervals for the true response-trait slope coefficient (positive relationship; r ~ 0.7). The dotted line indicates a slope coefficient of zero; points beneath this line would suggest the slope has the opposite relationship to the true coefficient. Black points and intervals represent an overlap between the true slope confidence intervals and that of the imputed or complete case slope confidence intervals. Red points and intervals indicate no overlap, so the imputed or complete case slope is significantly different. Slope coefficients and confidence intervals are averaged over *seed, response in imputation* and *between-predictor correlation,* but not over *bias types*, which are represented by the multiple overlapping points at each level of *missingness*, which include: Controlled bias – MOPT & MOPP, Weak bias – WBP, WBTP, WBTR & WBTPR, Severe bias – SBP, SBTP, SBTR & SBTPR.

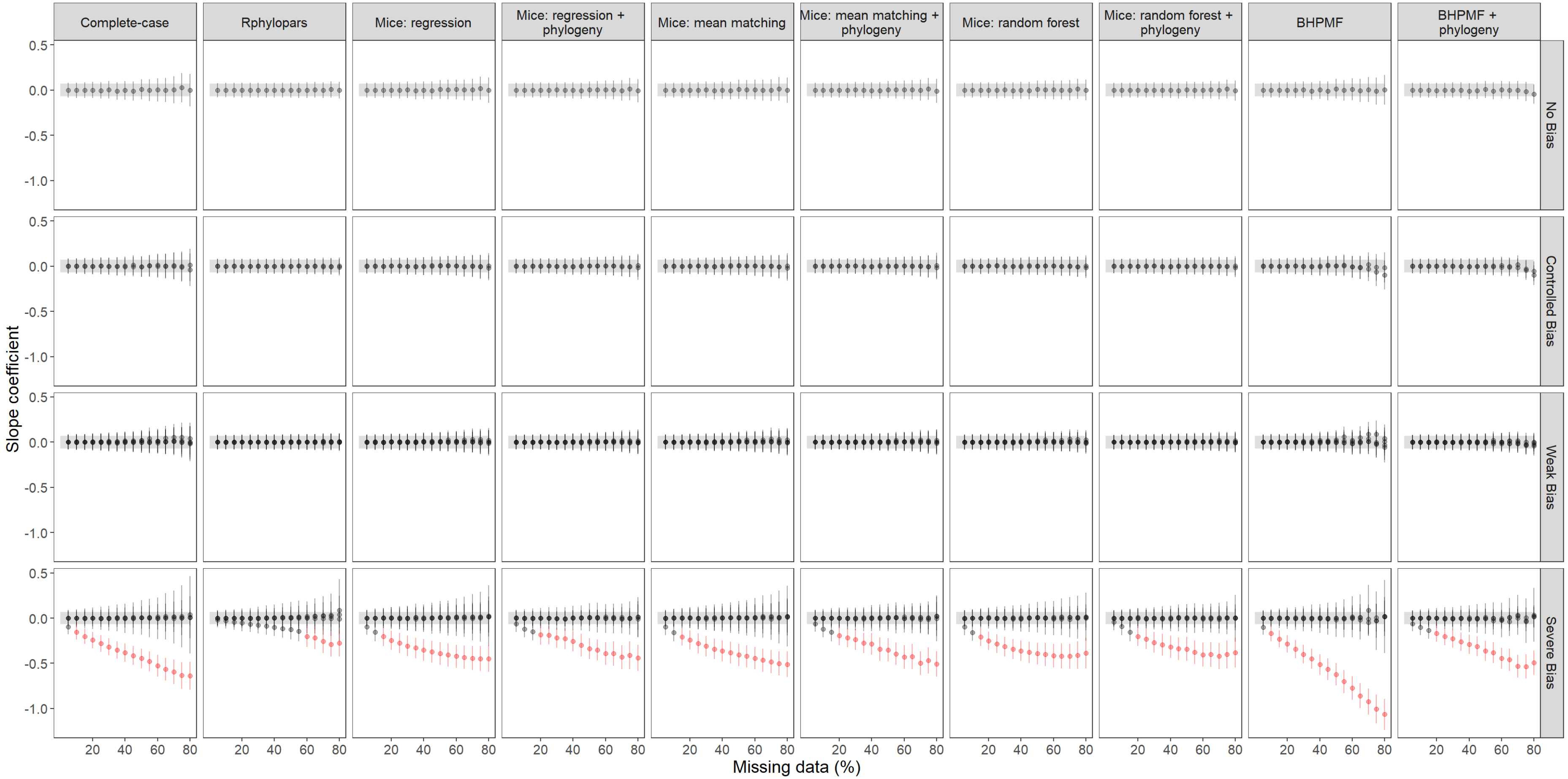


Figure B4.11. Slope coefficients (points) and 95% confidence intervals (vertical lines) under different *bias types* and levels of *missingness* for nine imputation approaches as well as complete-case analysis. The grey horizontal bar represents the lower and upper 95% confidence intervals for the true response-trait slope coefficient (no relationship; r ~ 0). Black points and intervals represent an overlap between the true slope confidence intervals and that of the imputed or complete case slope confidence intervals. Red points and intervals indicate no overlap, so the imputed or complete case slope is significantly different. Slope coefficients and confidence intervals are averaged over *seed, response in imputation* and *between-predictor correlation,* but not over *bias types*, which are represented by the multiple overlapping points at each level of *missingness*, which include: Controlled bias – MOPT & MOPP, Weak bias – WBP, WBTP, WBTR & WBTPR, Severe bias – SBP, SBTP, SBTR & SBTPR.

B5. Density plots of imputed values and slopes

We only depict imputation and slope errors under five *bias types* (random, and the four severe biases). We do not show the weak or stratified biases as their plots are very similar to the random plots.

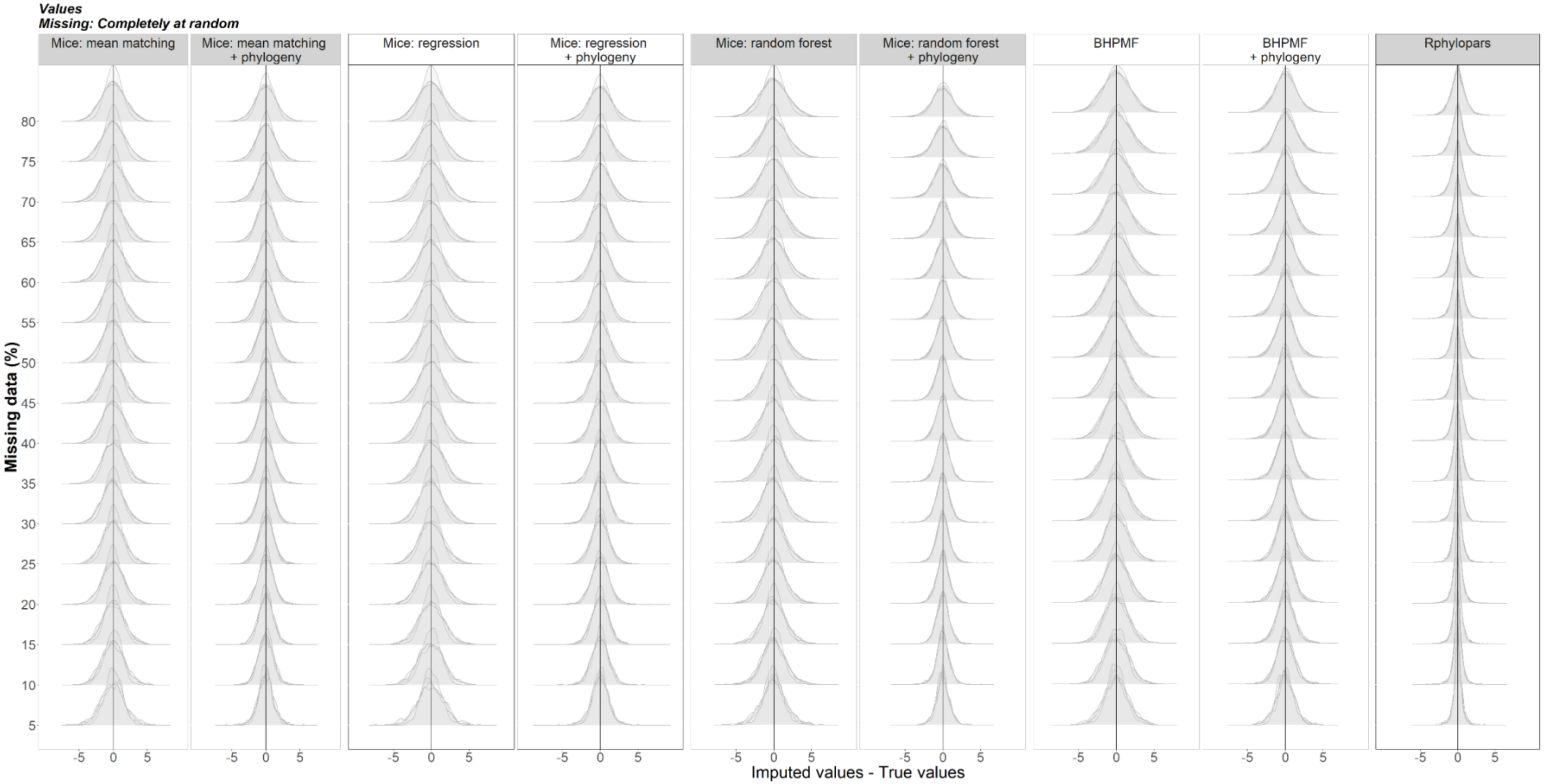


Figure B5.1. Difference between imputed and true values in nine imputation approaches at different levels of *missingness* (values removed at random). Values at zero (vertical line) have no error. To the left of the line, imputed values are underestimated, and values on the right are overestimated.

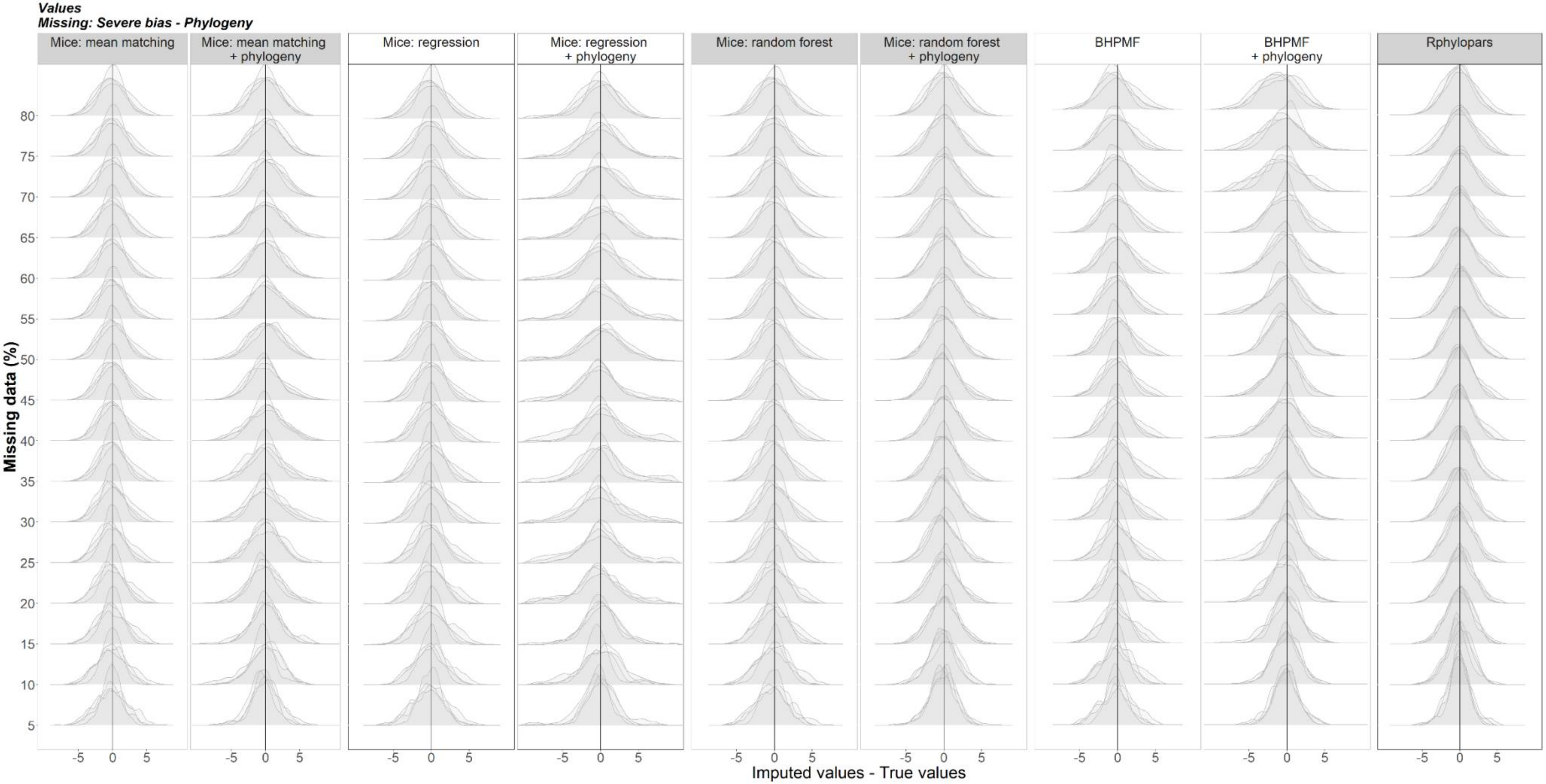


Figure B5.2. Difference between imputed and true values in nine imputation approaches at different levels of *missingness* (values removed with a severe phylogenetic bias). Values at zero (vertical line) have no error. To the left of the line, imputed values are underestimated, and values on the right are overestimated. Each of the overlapping density plots (x10) represent a different seed.

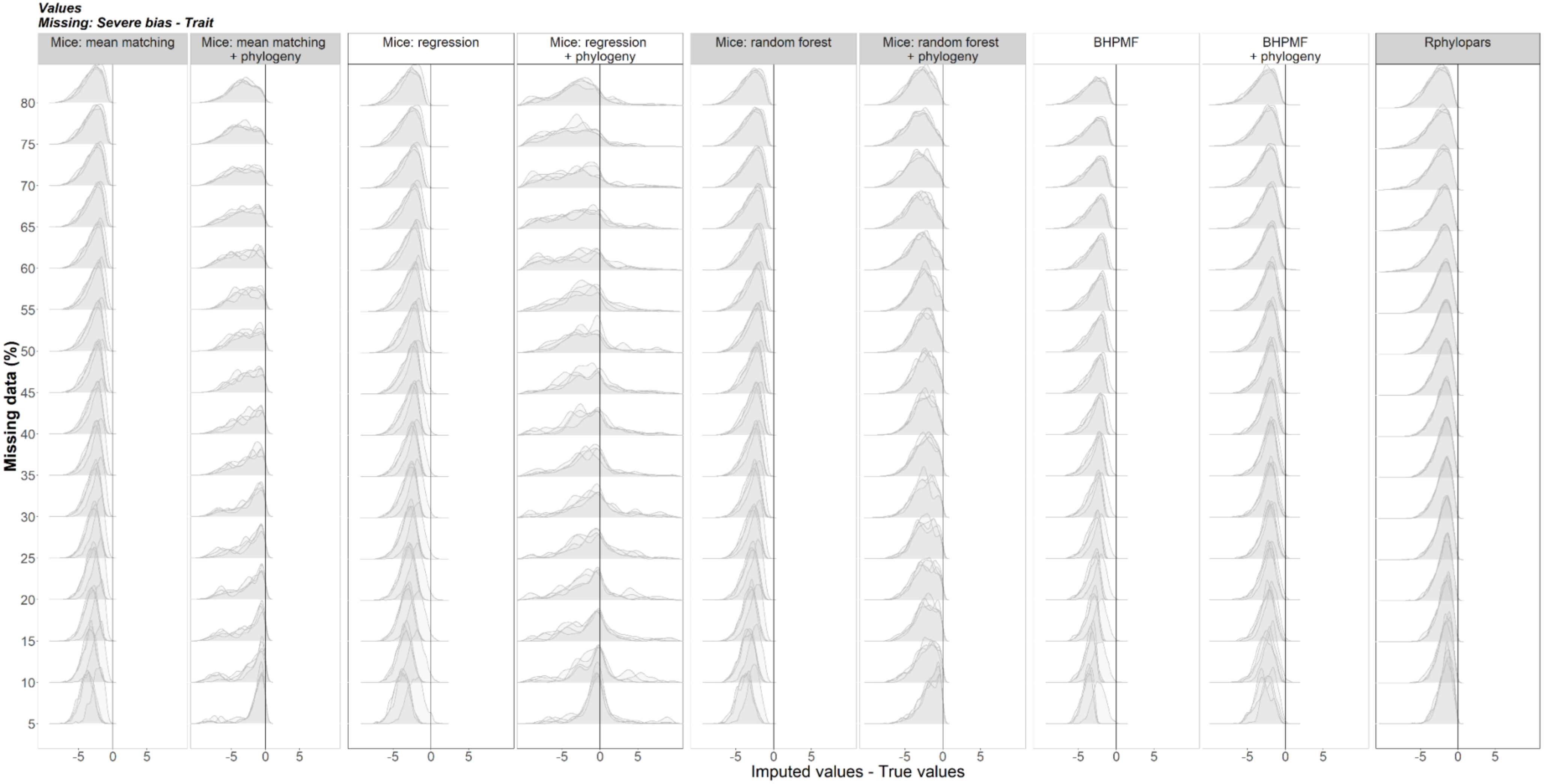


Figure B5.3. Difference between imputed and true values in nine imputation approaches at different levels of *missingness* (values removed with a severe trait bias). Values at zero (vertical line) have no error. To the left of the line, imputed values are underestimated, and values on the right are overestimated. Each of the overlapping density plots (x10) represent a different seed.

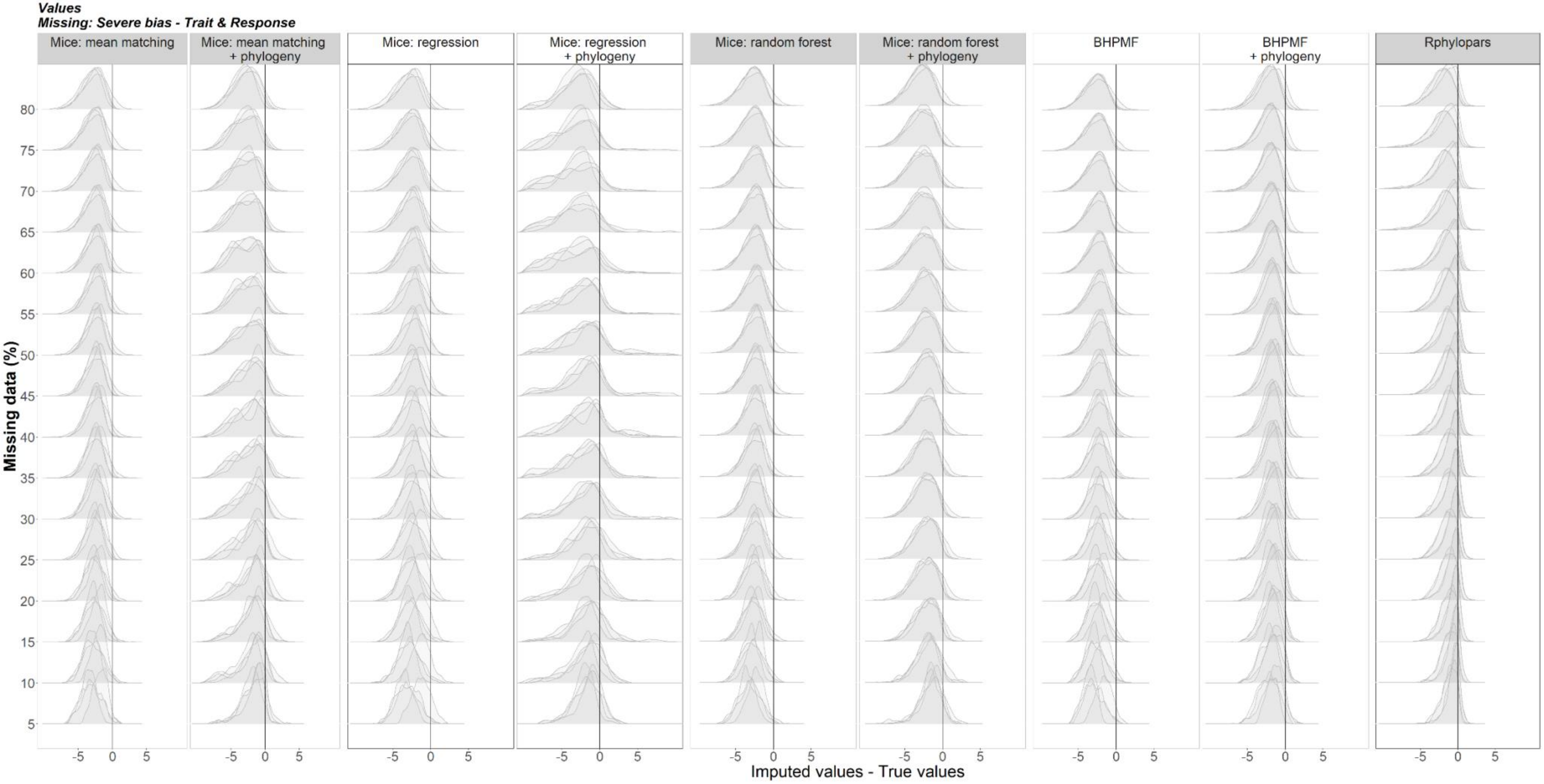


Figure B5.4. Difference between imputed and true values in nine imputation approaches at different levels of *missingness* (values removed with a severe trait and response bias). Values at zero (vertical line) have no error. To the left of the line, imputed values are underestimated, and values on the right are overestimated. Each of the overlapping density plots (x10) represent a different seed.

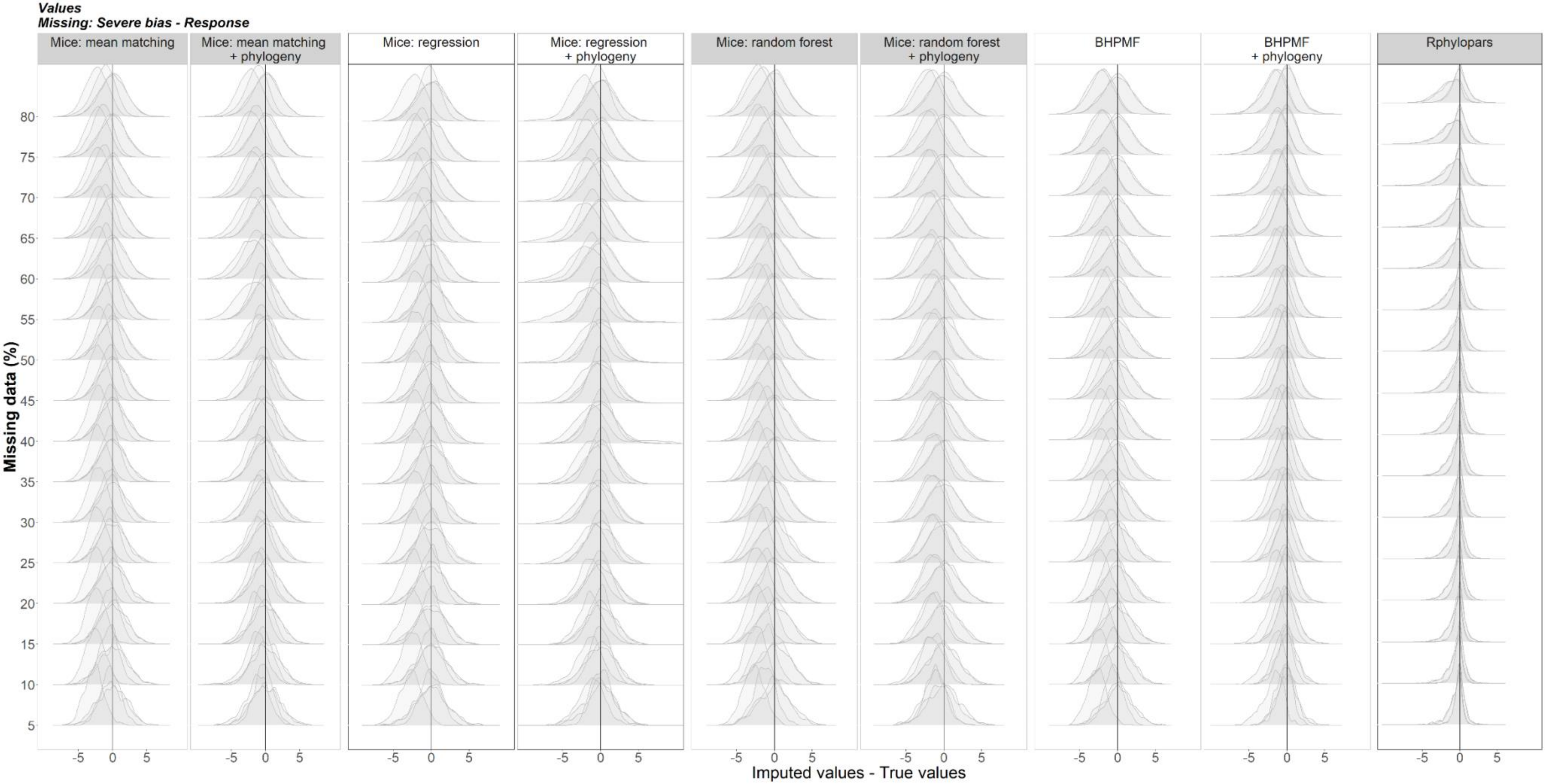


Figure B5.5. Difference between imputed and true values in nine imputation approaches at different levels of *missingness* (values removed with a severe response bias). Values at zero (vertical line) have no error. To the left of the line, imputed values are underestimated, and values on the right are overestimated. Each of the overlapping density plots (x10) represent a different seed.

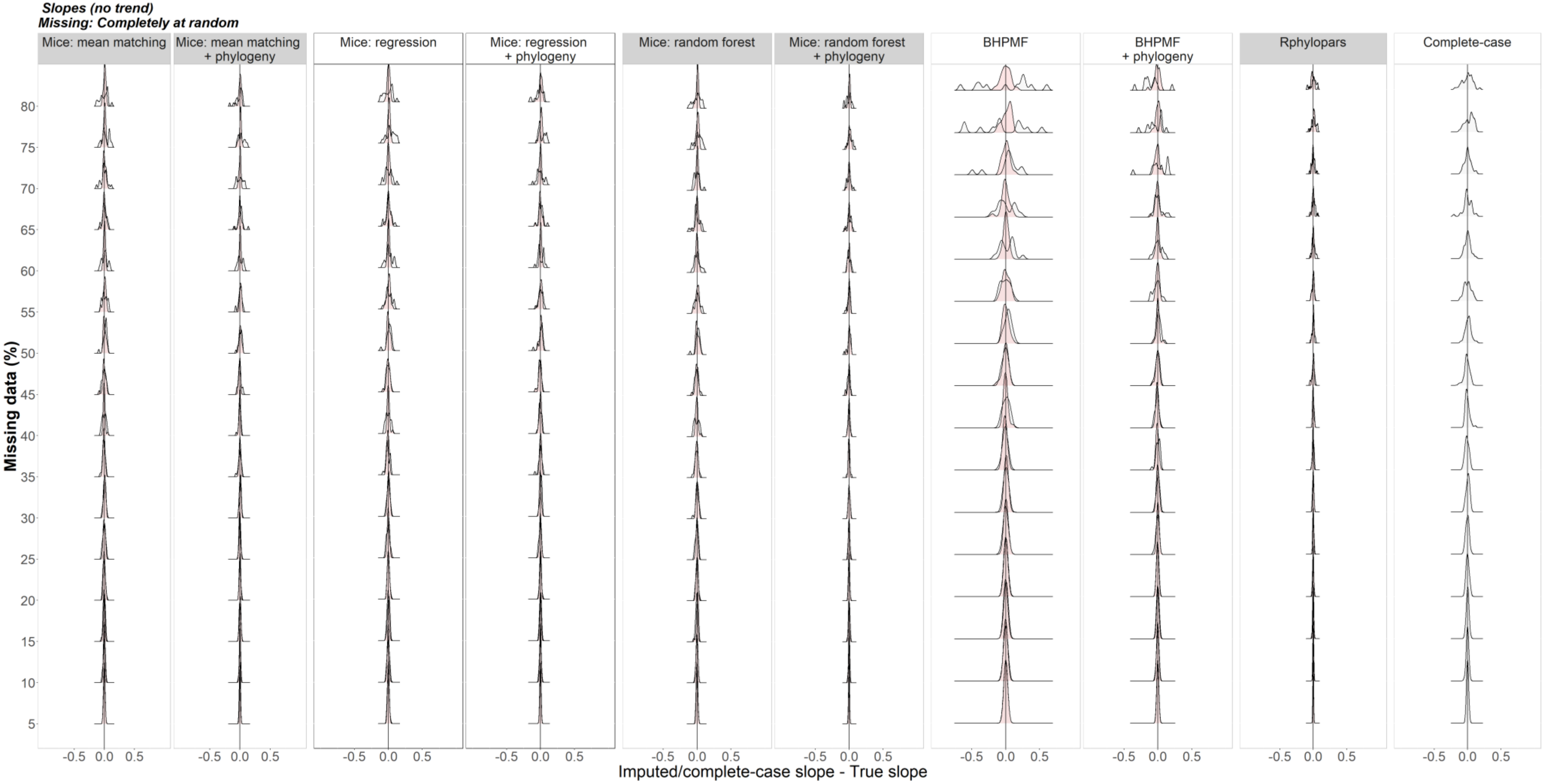


Figure B5.6. Difference between the imputed or complete-case slope, and the true slope (at approximately 0) in nine imputation approaches at different levels of *missingness* (values removed at random). Values at zero (vertical line) have no error. To the left of the line, imputed/complete-case slopes are underestimated (less than zero), and values on the right are overestimated (greater than zero). Red density plots indicate imputations with the response present, response is absent in grey density plots.

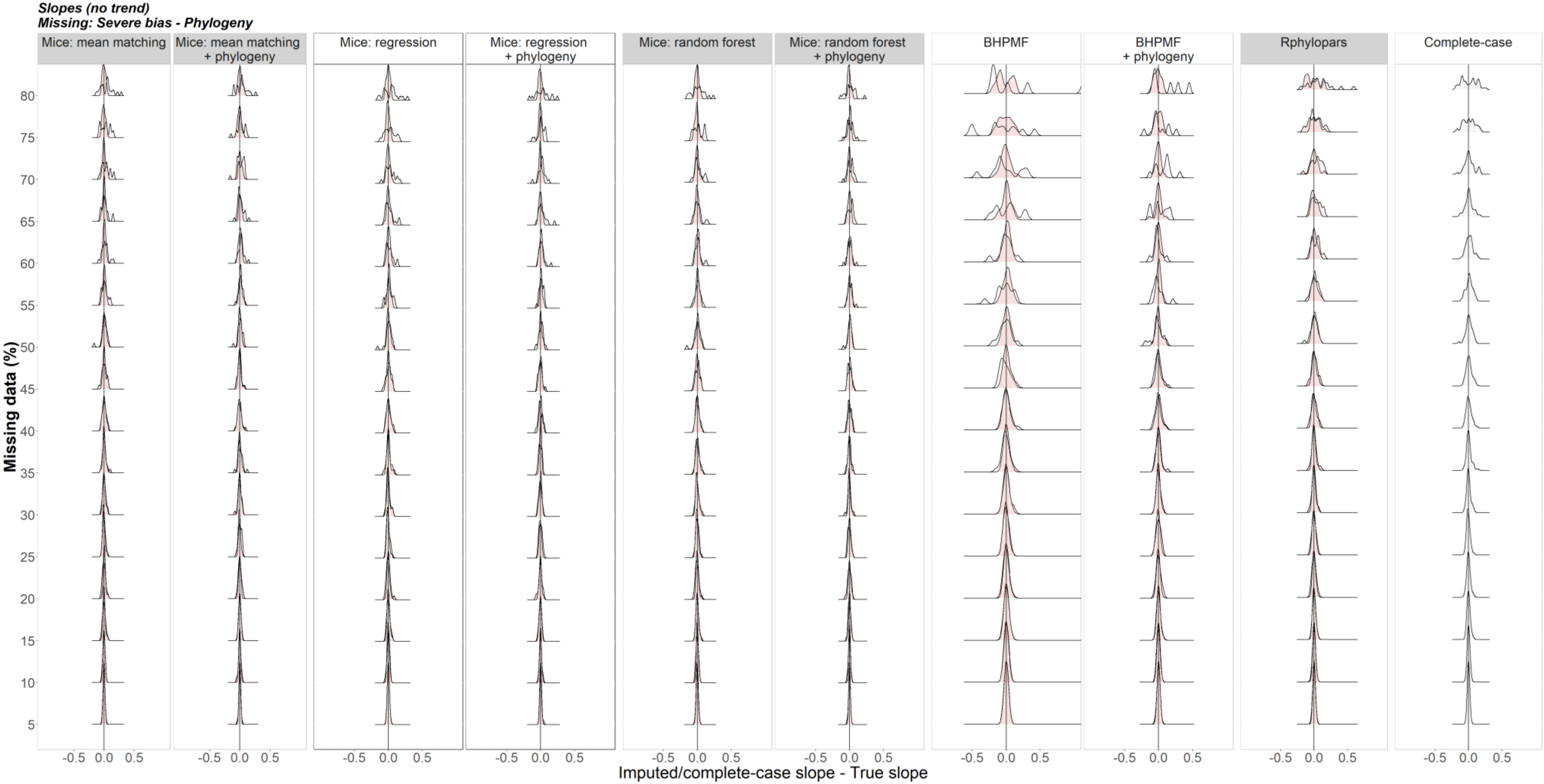


Figure B5.7. Difference between the imputed or complete-case slope, and the true slope (at approximately 0) in nine imputation approaches at different levels of *missingness* (values removed with a severe phylogenetic bias). Values at zero (vertical line) have no error. To the left of the line, imputed/complete-case slopes are underestimated (less than zero), and values on the right are overestimated (greater than zero). Red density plots indicate imputations with the response present, response is absent in grey density plots.

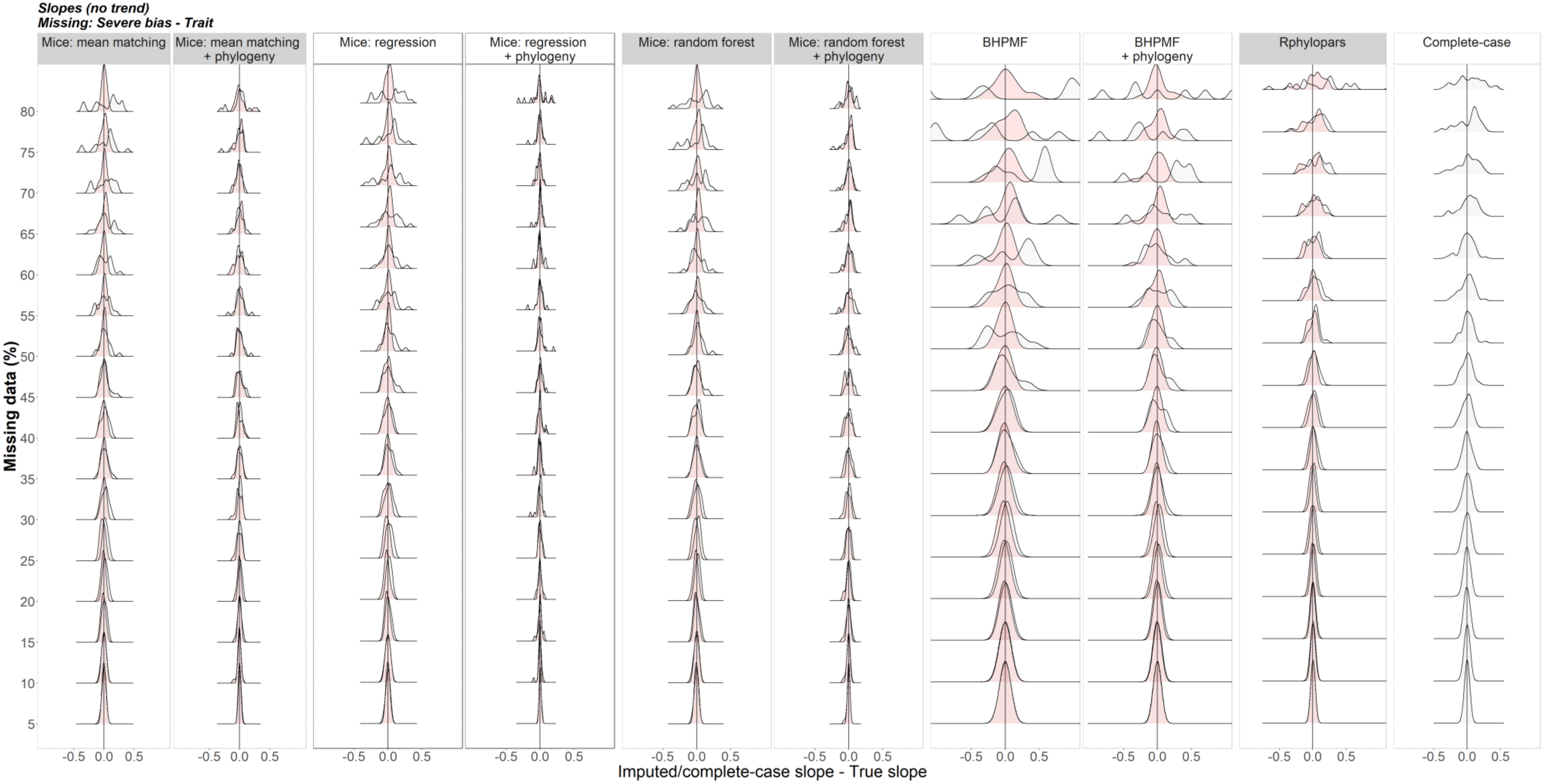


Figure B5.8. Difference between the imputed or complete-case slope, and the true slope (at approximately 0) in nine imputation approaches at different levels of *missingness* (values removed with a severe trait bias). Values at zero (vertical line) have no error. To the left of the line, imputed/complete-case slopes are underestimated (less than zero), and values on the right are overestimated (greater than zero). Red density plots indicate imputations with the response present, response is absent in grey density plots.

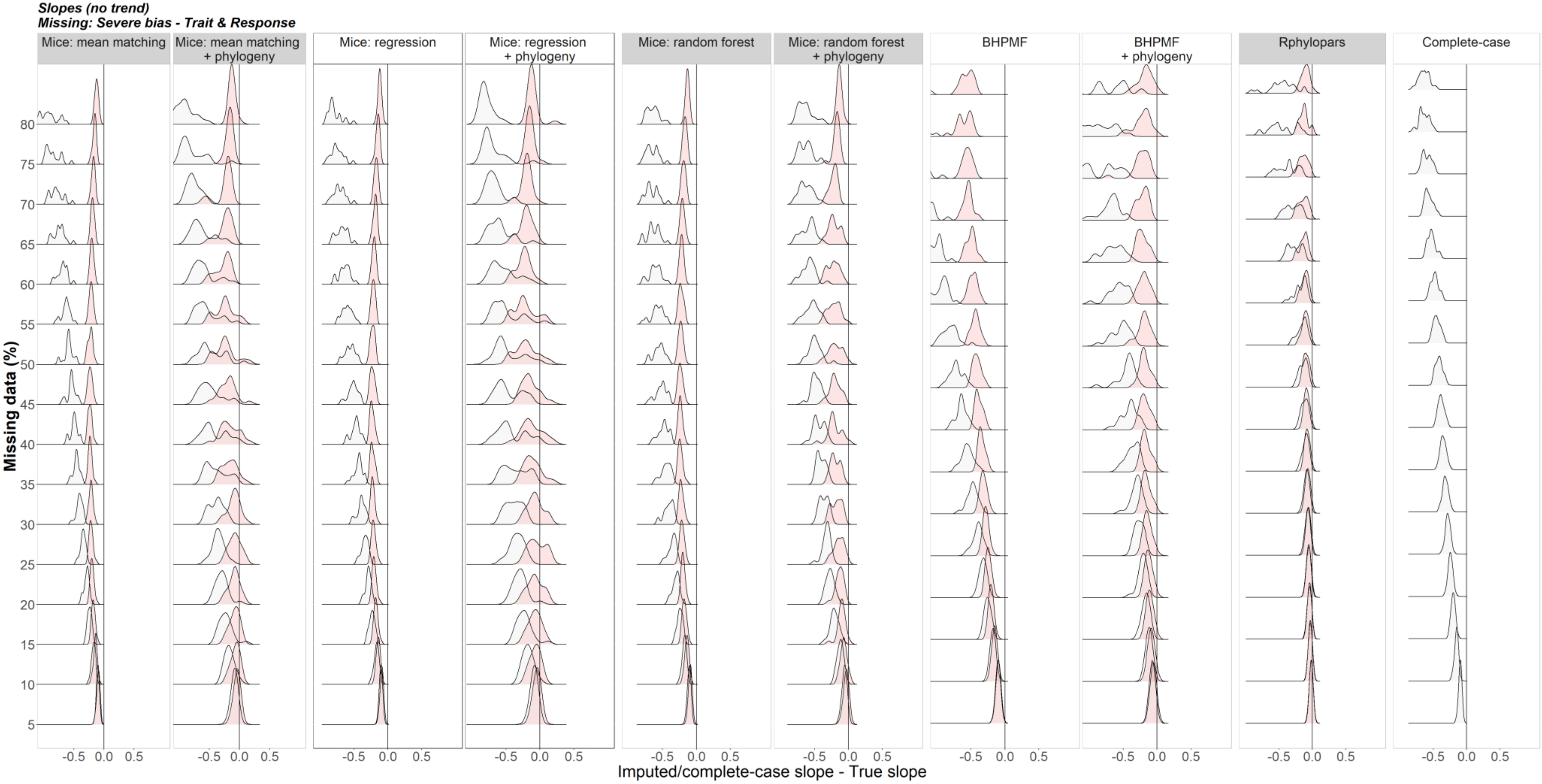


Figure B5.9. Difference between the imputed or complete-case slope, and the true slope (at approximately 0) in nine imputation approaches at different levels of *missingness* (values removed with a severe trait and response bias). Values at zero (vertical line) have no error. To the left of the line, imputed/complete-case slopes are underestimated (less than zero), and values on the right are overestimated (greater than zero). Red density plots indicate imputations with the response present, response is absent in grey density plots.

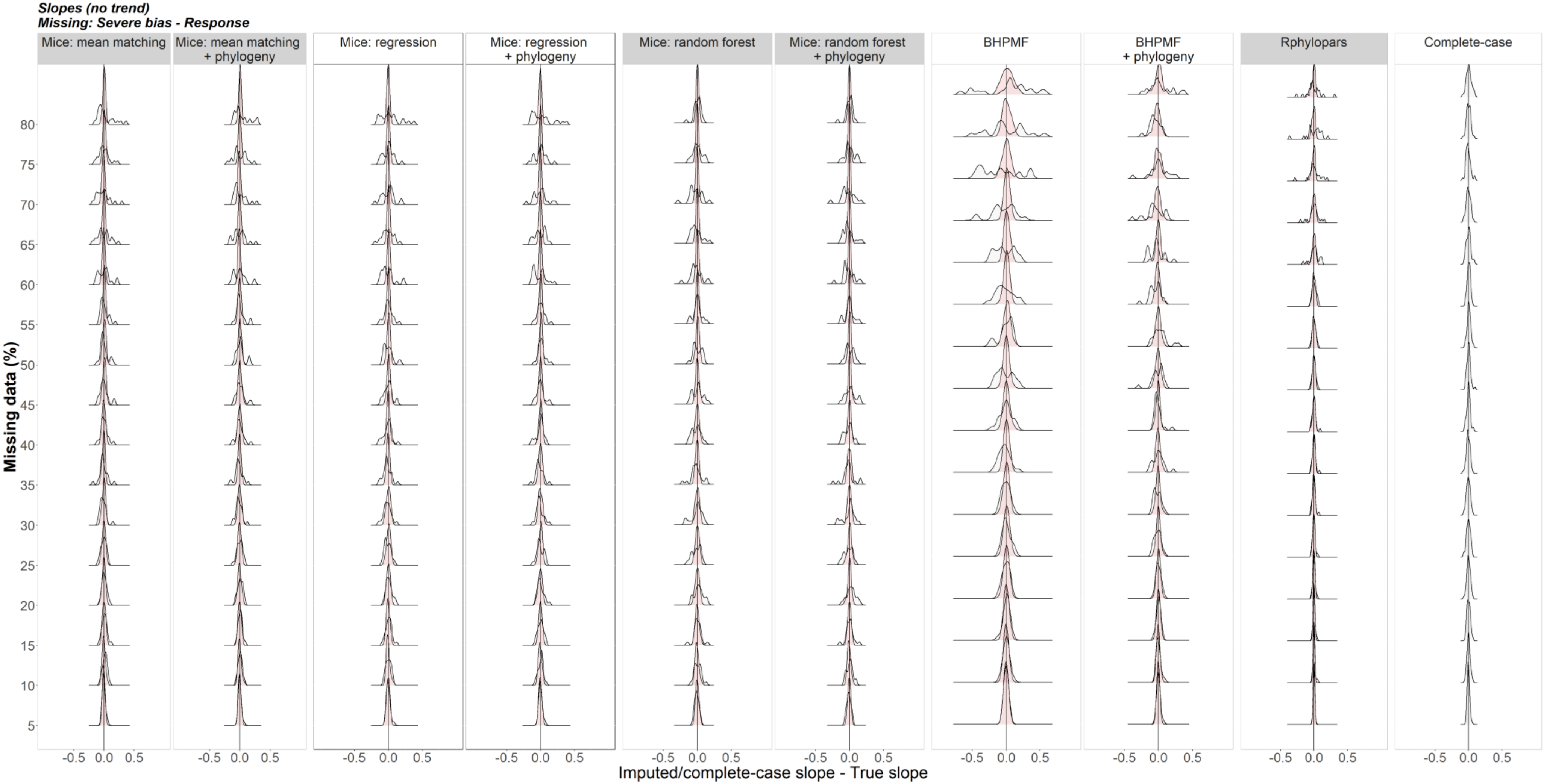


Figure B5.10. Difference between the imputed or complete-case slope, and the true slope (at approximately 0) in nine imputation approaches at different levels of *missingness* (values removed with a severe response bias). Values at zero (vertical line) have no error. To the left of the line, imputed/complete-case slopes are underestimated (less than zero), and values on the right are overestimated (greater than zero). Red density plots indicate imputations with the response present, response is absent in grey density plots.

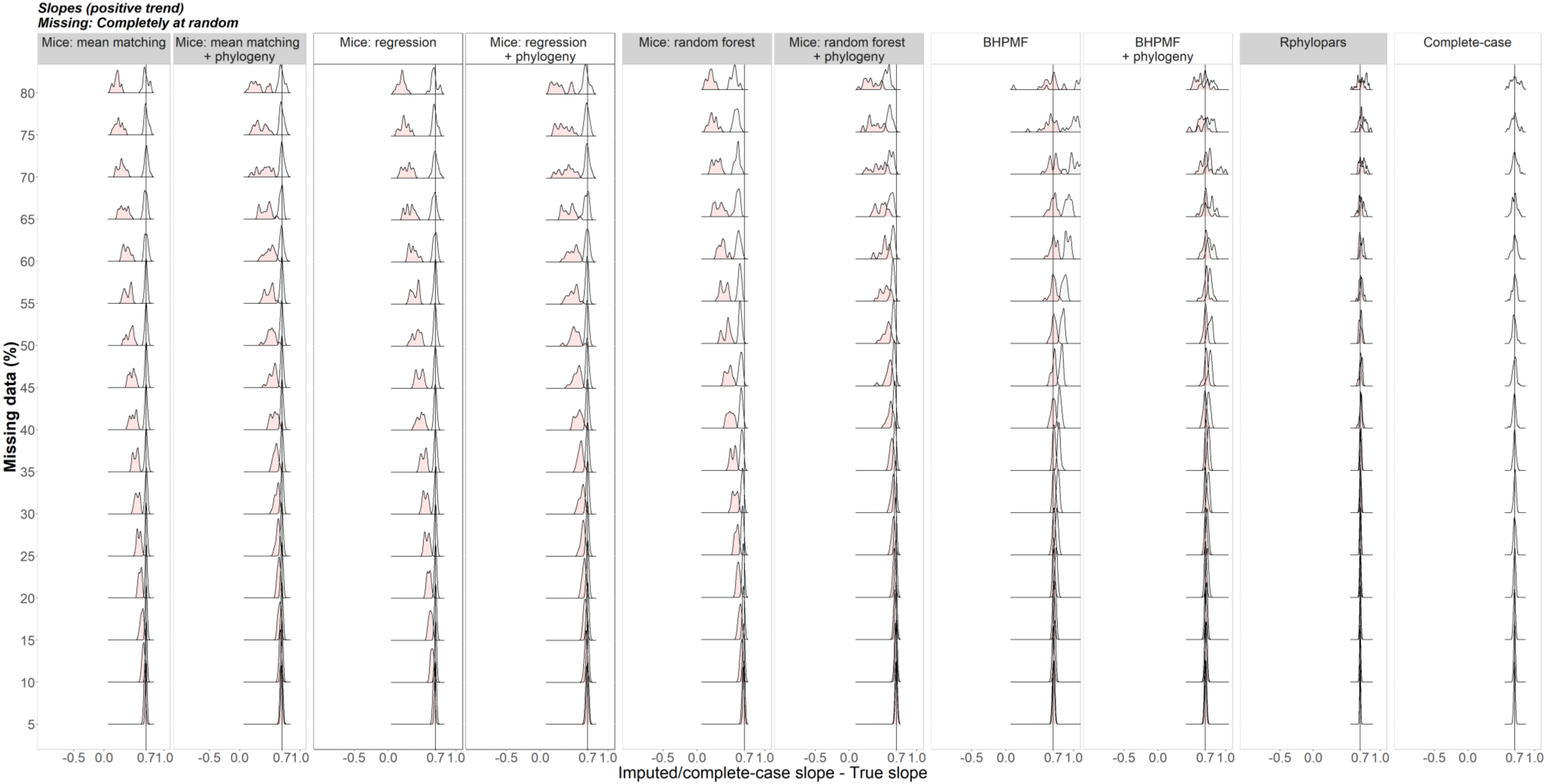


Figure B5.11. Difference between the imputed or complete-case slope, and the true slope (at approximately 0.7) in nine imputation approaches at different levels of *missingness* (values removed at random). Values at 0.7 (vertical line) have no error. To the left of the line, imputed/complete-case slopes are underestimated (less than 0.7), and values on the right are overestimated (greater than 0.7). Red density plots indicate imputations with the response present, response is absent in grey density plots.

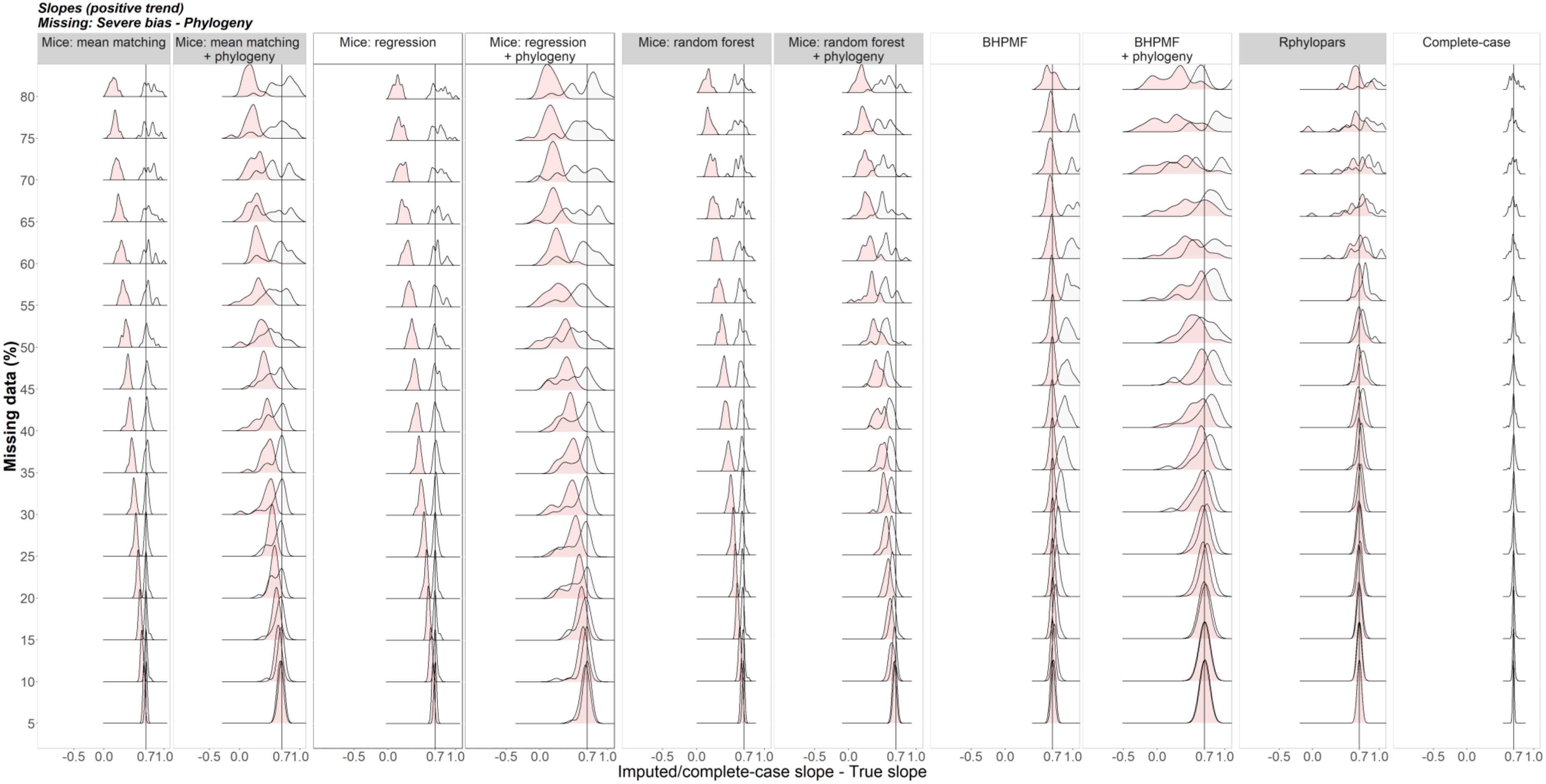


Figure B5.12. Difference between the imputed or complete-case slope, and the true slope (at approximately 0.7) in nine imputation approaches at different levels of *missingness* (values removed with a phylogenetic bias). Values at 0.7 (vertical line) have no error. To the left of the line, imputed/complete-case slopes are underestimated (less than 0.7), and values on the right are overestimated (greater than 0.7). Red density plots indicate imputations with the response present, response is absent in grey density plots.

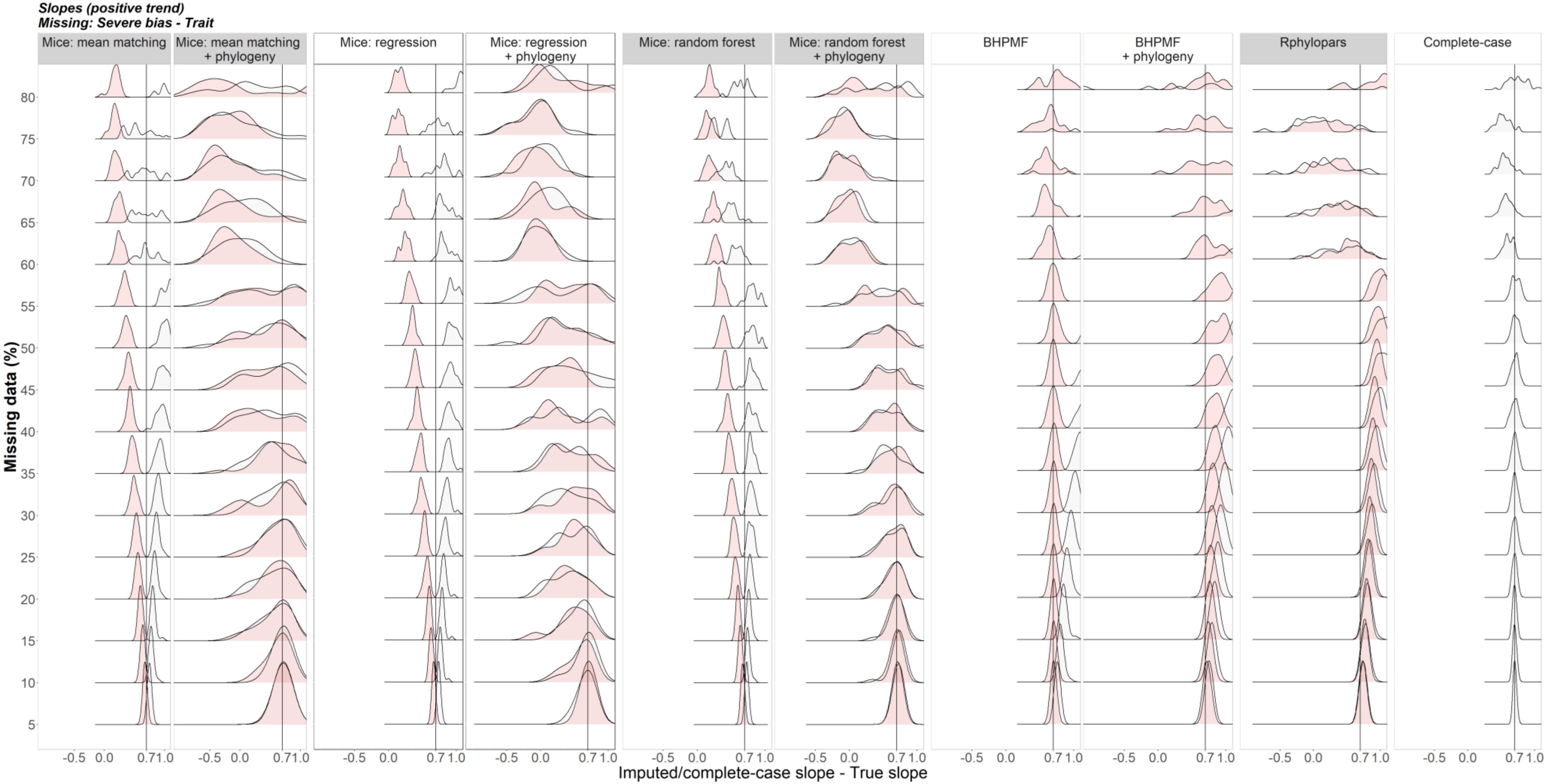


Figure B5.13. Difference between the imputed or complete-case slope, and the true slope (at approximately 0.7) in nine imputation approaches at different levels of *missingness* (values removed with a trait bias). Values at 0.7 (vertical line) have no error. To the left of the line, imputed/complete-case slopes are underestimated (less than 0.7), and values on the right are overestimated (greater than 0.7). Red density plots indicate imputations with the response present, response is absent in grey density plots.

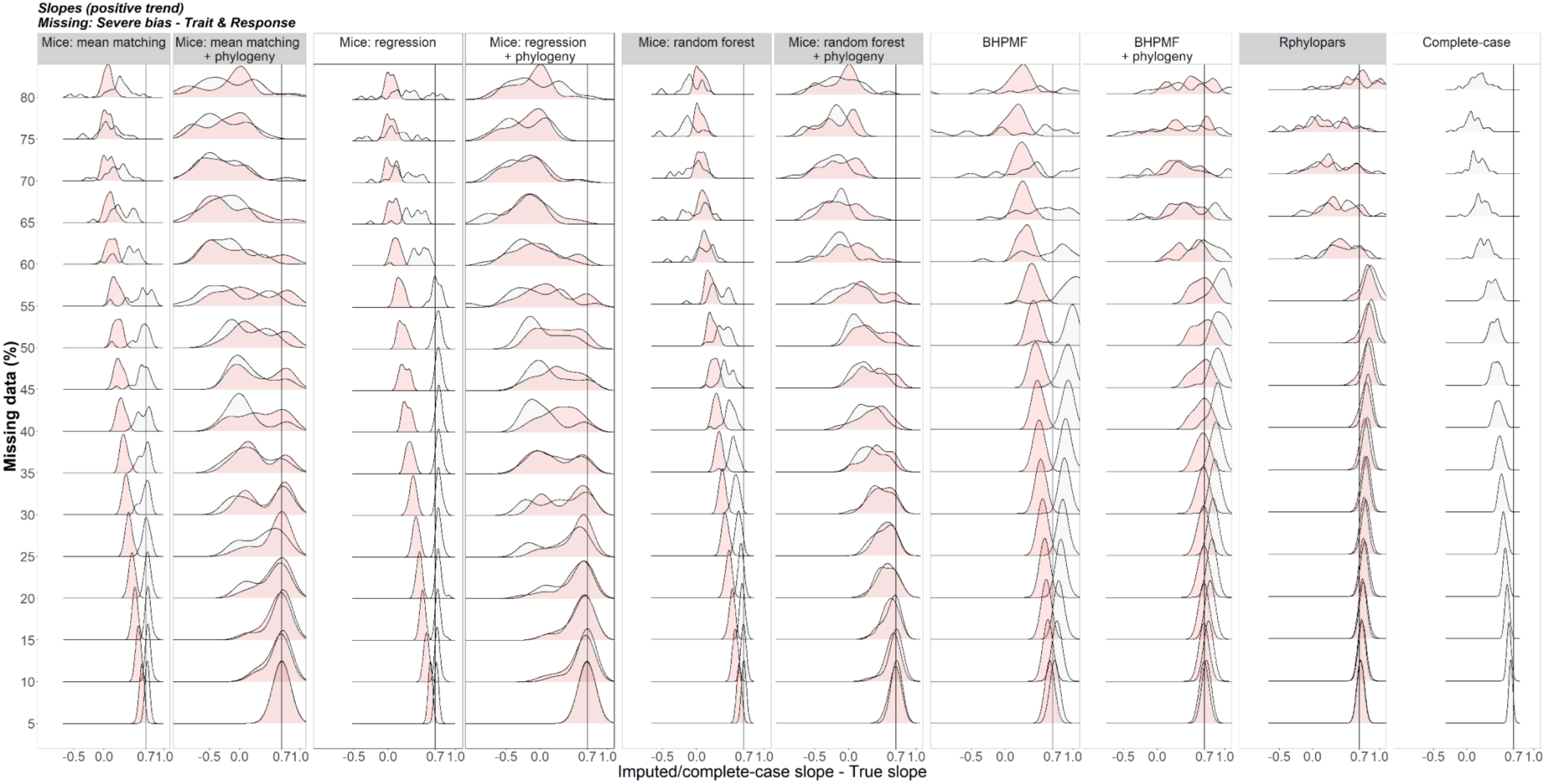


Figure B5.14. Difference between the imputed or complete-case slope, and the true slope (at approximately 0.7) in nine imputation approaches at different levels of *missingness* (values removed with a trait and response bias). Values at 0.7 (vertical line) have no error. To the left of the line, imputed/complete-case slopes are underestimated (less than 0.7), and values on the right are overestimated (greater than 0.7). Red density plots indicate imputations with the response present, response is absent in grey density plots.

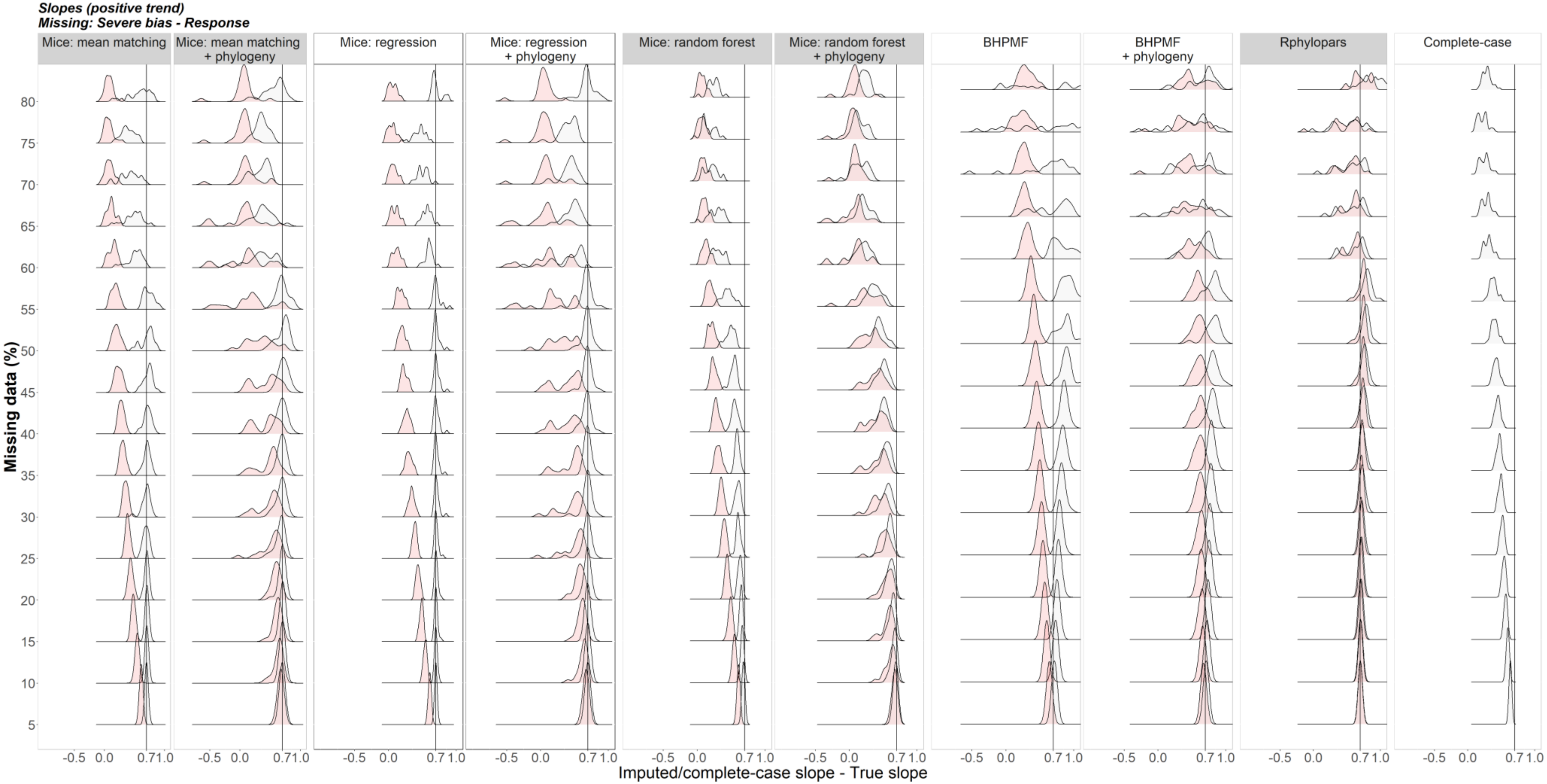


Figure B5.15. Difference between the imputed or complete-case slope, and the true slope (at approximately 0.7) in nine imputation approaches at different levels of *missingness* (values removed with a response bias). Values at 0.7 (vertical line) have no error. To the left of the line, imputed/complete-case slopes are underestimated (less than 0.7), and values on the right are overestimated (greater than 0.7). Red density plots indicate imputations with the response present, response is absent in grey density plots

B6. Impact of including the response in imputation

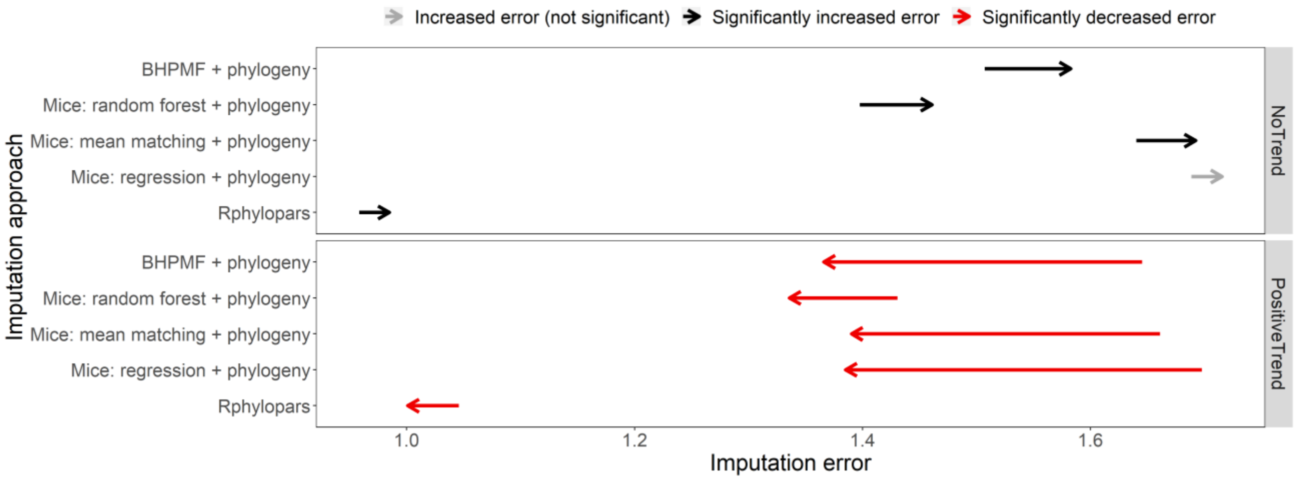


Figure B6.1. Difference between imputed and true trait values (imputation error - root mean square error) for four phylogenetic imputation approaches after including the response within the imputation. The base of the arrow represents the mean imputation error for each approach when the response is absent, and the tip represents the mean imputation error when the response is present. Larger arrows indicate a greater effect of including the response. Data is split by response-trait relationship, where a there is no relationship r~0 (top), or a positive relationship r~0.7 (bottom).

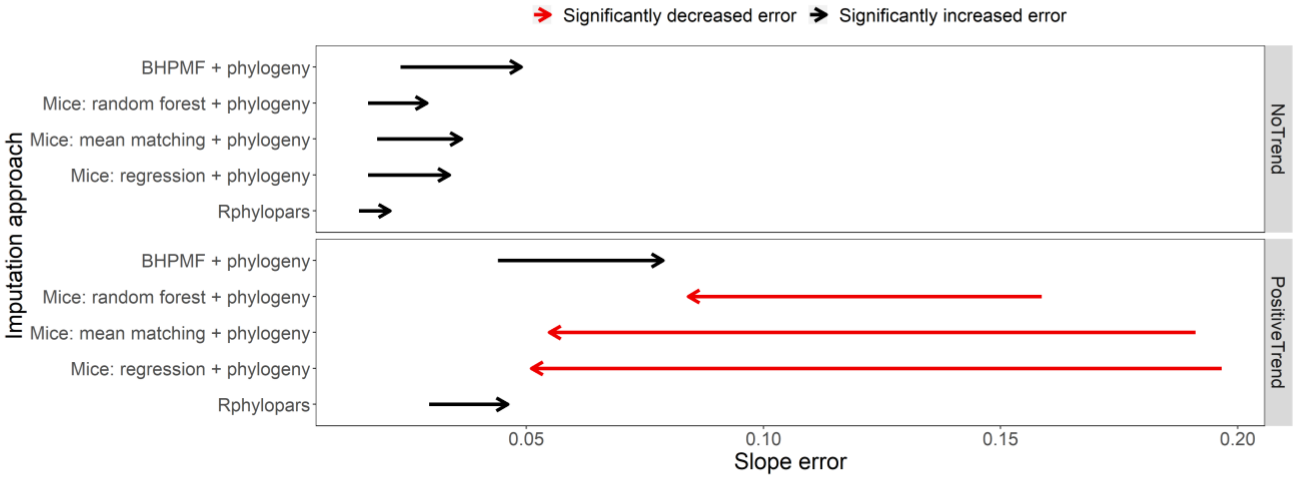


Figure B6.2. Slope error (absolute difference between the true response-trait slope and the slope after imputation) after including the response within the imputation. The base of the arrow represents the mean slope error for each approach when the response is absent, and the tip represents the mean slope error when the response is present. Larger arrows indicate a greater effect of including the response. Data is split by response-trait relationship, where a there is no relationship r~0 (top), or a positive relationship r~0.7 (bottom).

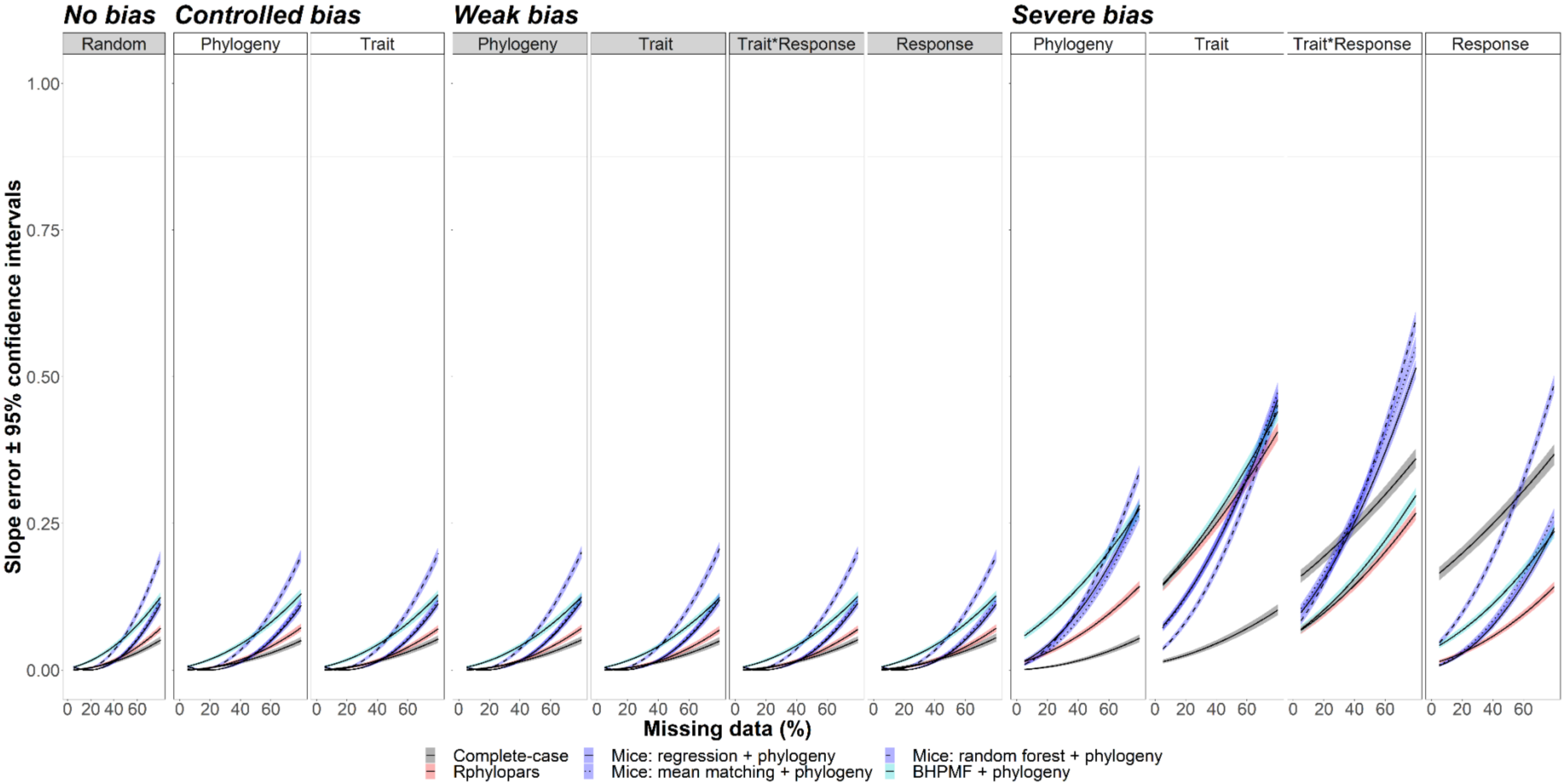


Figure B6.3. Absolute difference between the true response-trait slope coefficient and the slope coefficients in datasets with varying percentages of missing trait values (missingness) - removed under a series of bias types. Missing values were imputed under five phylogenetic approaches, or treated as complete-case analysis. The true slope was positive (r ~ 0.7) and the response was included in the imputation. Lines depict the marginal effect of missingness and bias type from a regression model, and were averaged across other predictors: seed and between-trait correlation. Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%).

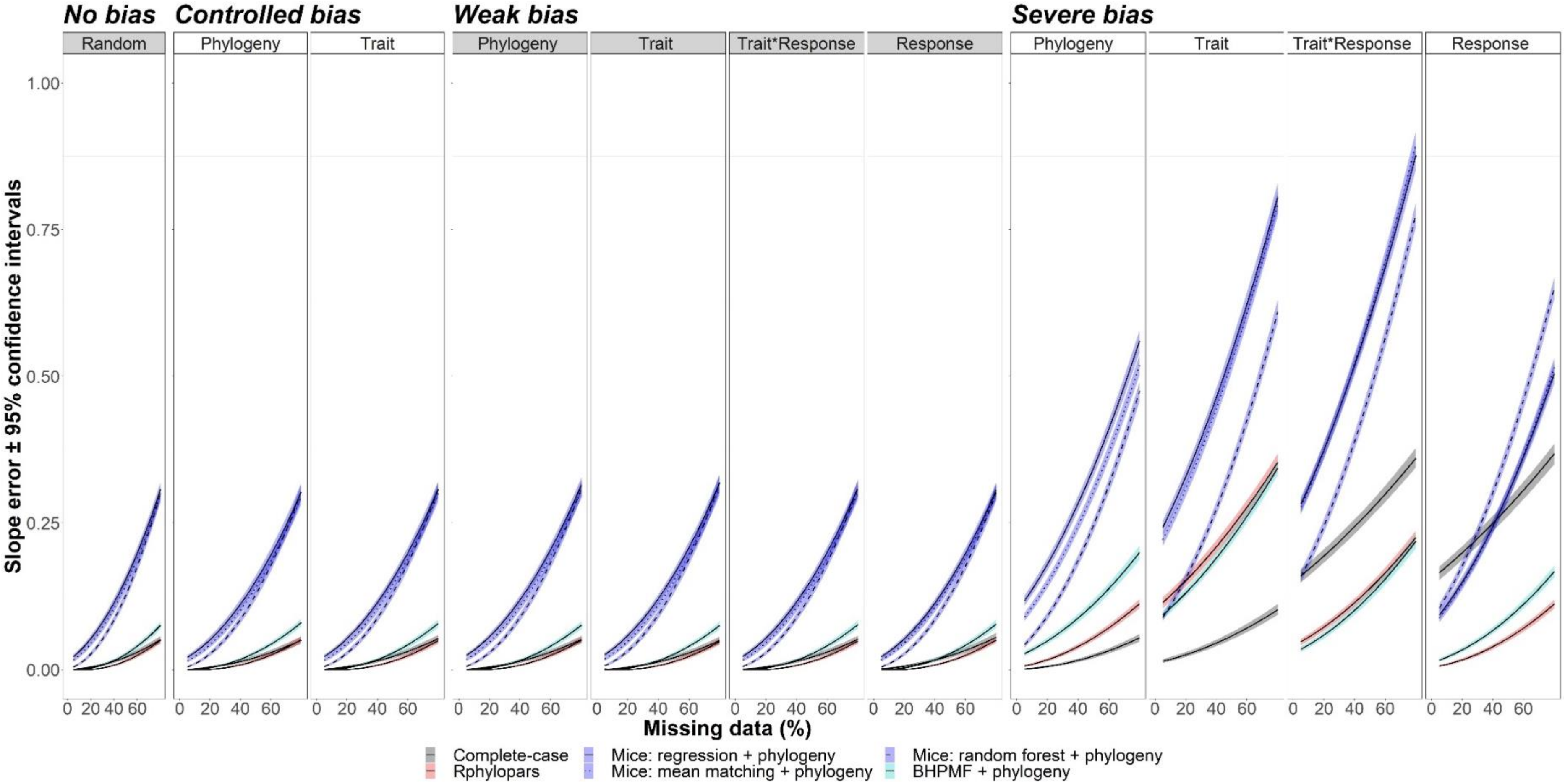


Figure B6.4. Absolute difference between the true response-trait slope coefficient and the slope coefficients in datasets with varying percentages of missing trait values (missingness) - removed under a series of bias types. Missing values were imputed under five phylogenetic approaches, or treated as complete-case analysis. The true slope was positive (r ~ 0.7) and the response was excluded from the imputation. Lines depict the marginal effect of missingness and bias type from a regression model, and were averaged across other predictors: seed and between-trait correlation. Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%).

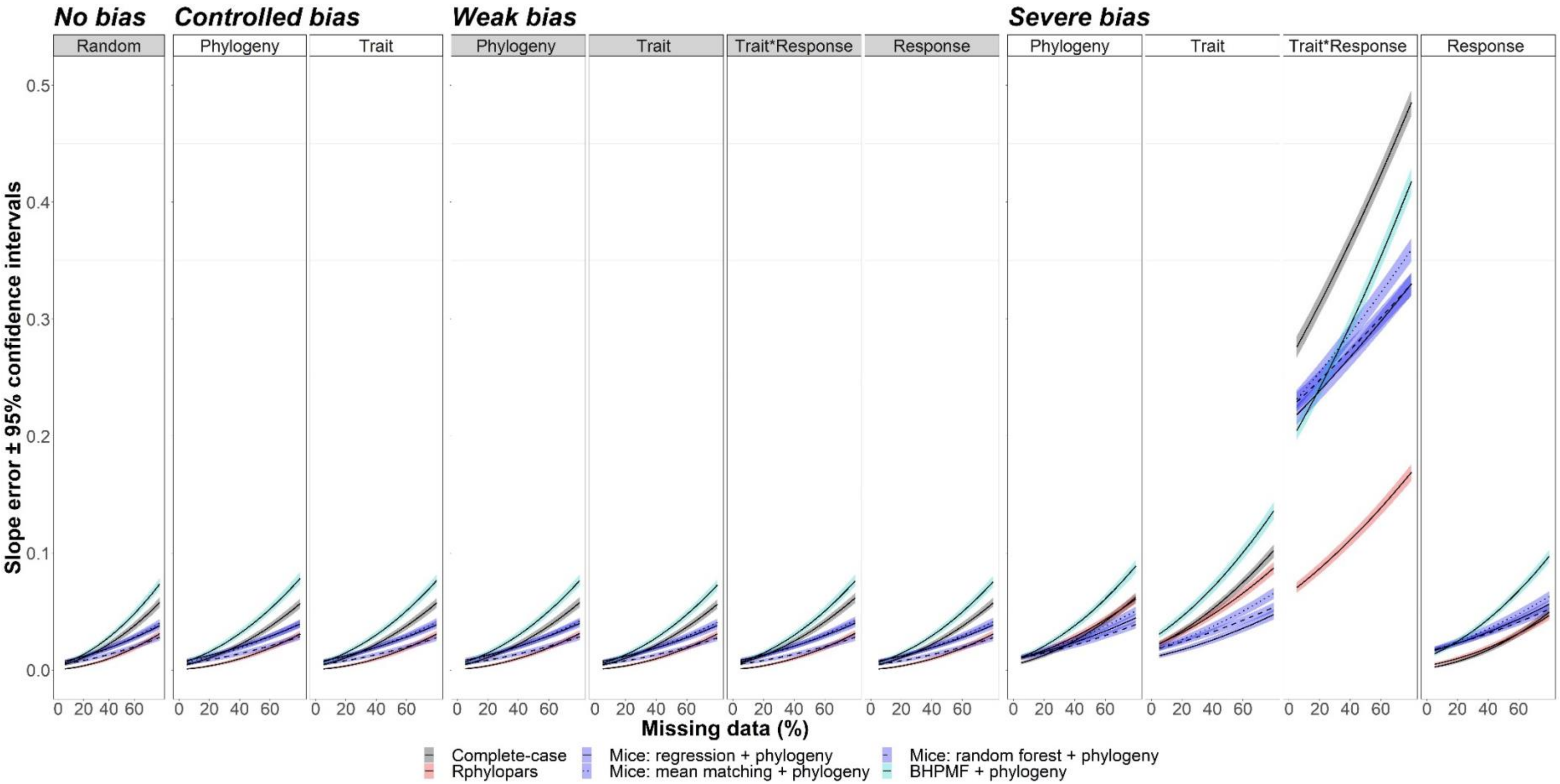


Figure B6.5. Absolute difference between the true response-trait slope coefficient and the slope coefficients in datasets with varying percentages of missing trait values (missingness) - removed under a series of bias types. Missing values were imputed under five phylogenetic approaches, or treated as complete-case analysis. The true slope had no relationship (r ~ 0) and the response was included in the imputation. Lines depict the marginal effect of missingness and bias type from a regression model, and were averaged across other predictors: seed and between-trait correlation. Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%).

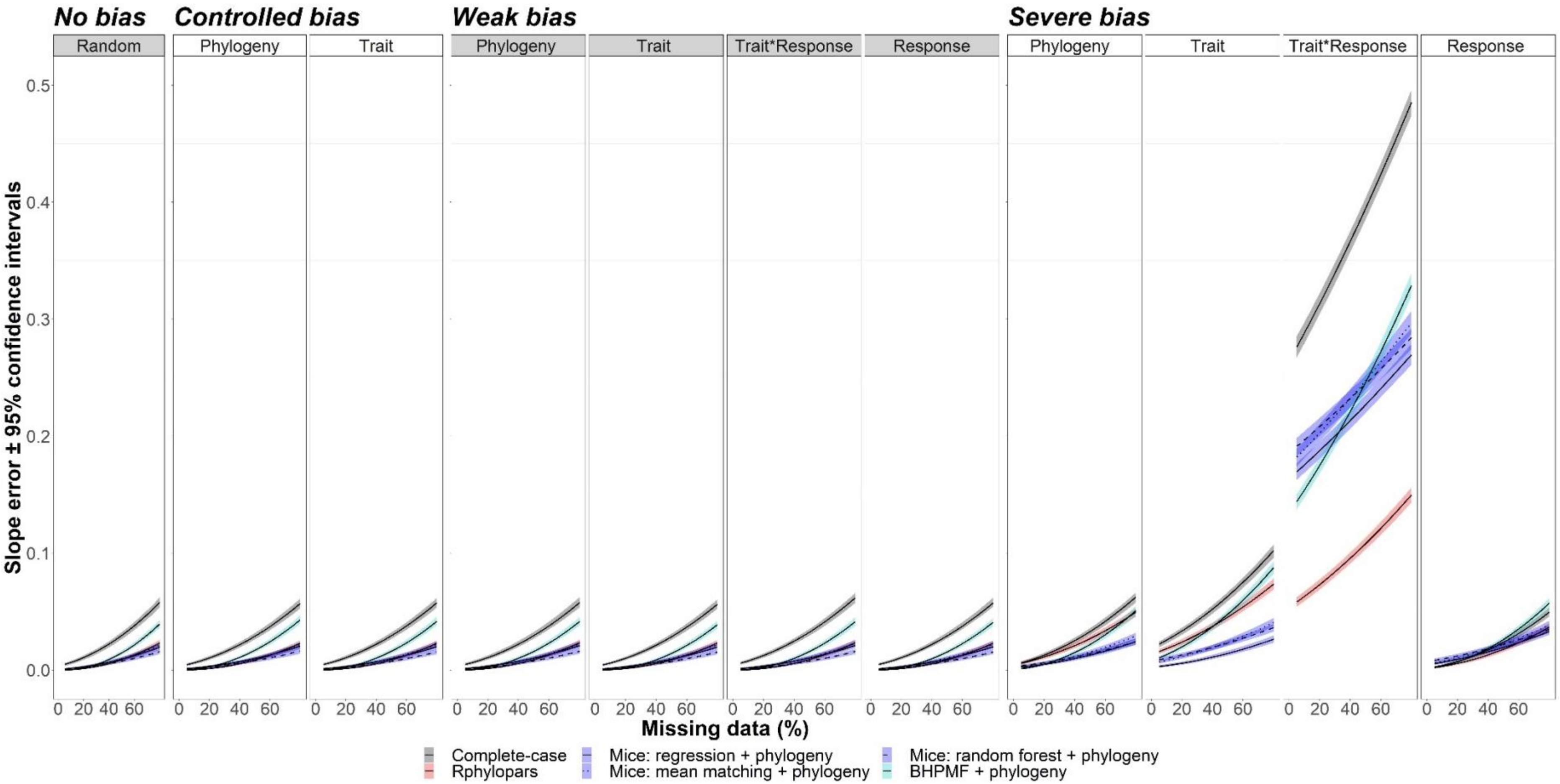


Figure B6.6. Absolute difference between the true response-trait slope coefficient and the slope coefficients in datasets with varying percentages of missing trait values (missingness) - removed under a series of bias types. Missing values were imputed under five phylogenetic approaches, or treated as complete-case analysis. The true slope had no relationship (r ~ 0) and the response was excluded from the imputation. Lines depict the marginal effect of missingness and bias type from a regression model, and were averaged across other predictors: seed and between-trait correlation. Confidence intervals were derived from 500 bootstrap simulations and depict the upper- and lower-bounds (95%).

B7. Predicting errors in imputation

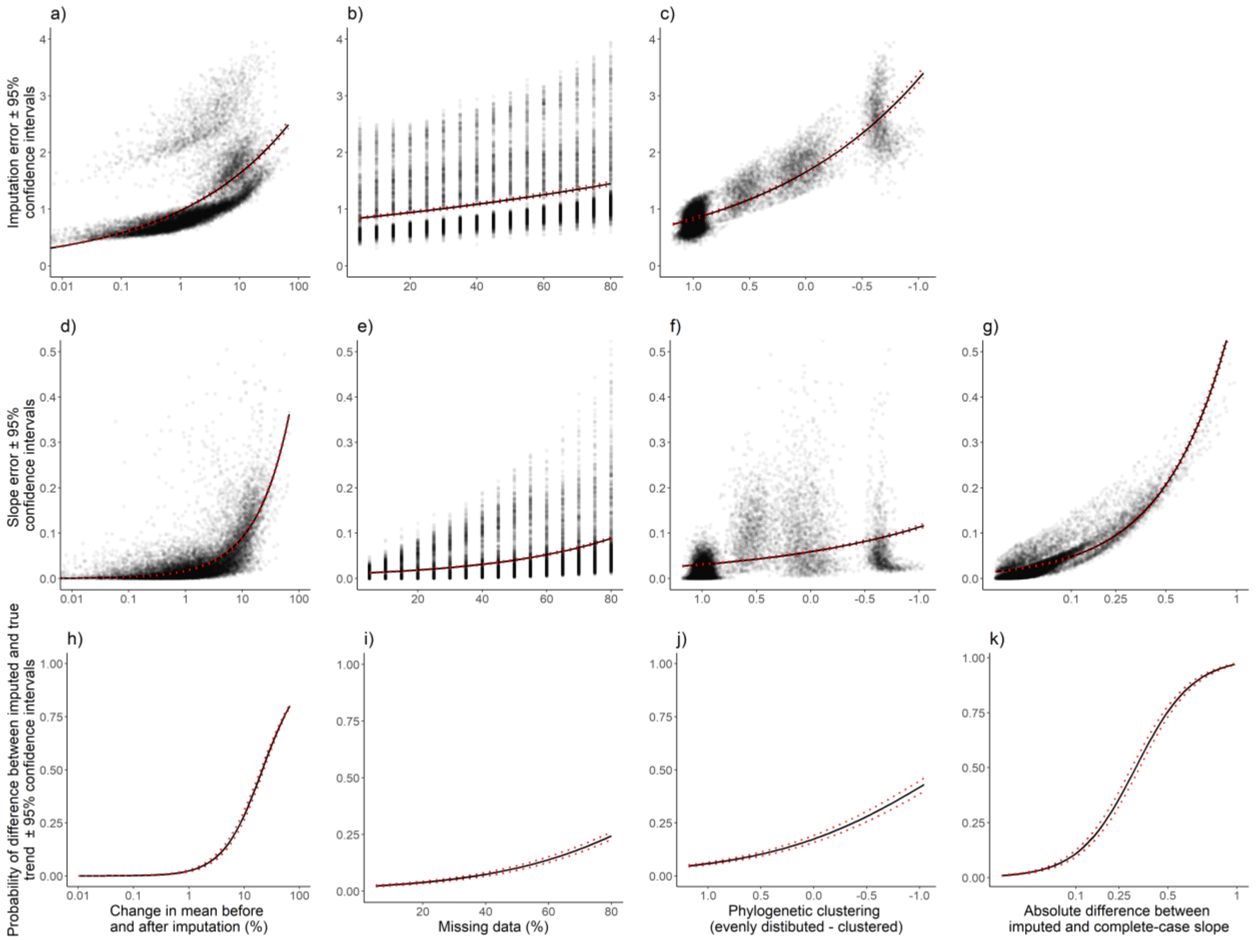


Figure B7.1. Top - Predicting imputation error (different between *Rphylopars* imputed trait values and true trait values calculated using the root mean square error), middle - slope error (difference between the true response-trait slope and the *Rphylopars* imputed slope), and bottom - the probability of the *Rphylopars* imputed slope differing significantly from the true response-trait slope, using four predictors (3 for imputation error): *change in mean, missingness, phylogenetic clustering,* and *change in slope*. Red-dotted lines depict 95% confidence intervals.

B8. Time taken for imputations

Table B8.1. Minutes taken to perform each of the phylogenetic imputation approaches. The *Mice* and *BHPMF* approaches exclude the time taken to develop eigenvectors and the phylogenetic hierarchy, respectively. Each mean and standard deviation is based off a sample size of 14080 imputations.

Imputation approaches	Minutes (mean ± standard deviation)
Rphylopars	0.12 ± 0.08
Mice: mean matching + phylogeny	0.32 ± 0.24
Mice: regression + phylogeny	0.31 ± 0.24
Mice: random forest + phylogeny	0.55 ± 0.25
BHPMF + phylogeny	0.37 ± 0.25


A Cardwell

Chapter 3: CaPTrends - a global database of population trends in large terrestrial Carnivorans

Thomas F. Johnson[1], Paula Cruz[2,3], Nick J. B. Isaac[4], Agustin Paviolo[2,3], Manuela González-Suárez[1]

1: Ecology and Evolutionary Biology, School of Biological Sciences, University of Reading, Reading, UK

2: Instituto de Biología Subtropical, CONICET-Universidad Nacional de Misiones, Bertoni 85, (N3370AIA) Puerto Iguazú, Misiones, Argentina

3: Asociación Civil Centro de Investigaciones del Bosque Atlántico, Bertoni 85, (N3370AIA) Puerto Iguazú, Misiones, Argentina

4: Biodiversity Science Area, Centre for Ecology and Hydrology, Wallingford, OX10 8BB, UK,

Corresponding author: Thomas Frederick Johnson (Thomas.frederick.johnson@outlook.com)

Author contributions: All authors contributed to project design. TFJ entered the data with support from PC. Data was validated by MGS and PC. TFJ wrote the first draft of the manuscript, all authors contributed to revisions.

Abstract

Motivation

Population trend information is key for assessing biodiversity change and has been used in global syntheses to evidence and understand declines and recoveries in wildlife – making population trend information an important indicator. Here, we present a global dataset of 1122 population trends describing changes in abundance over time in large mammals from the Order Carnivora.

Main types of variables included

The key data fields for each record are: species name, geographical location of the record, start and end date of the trend estimate, method by which abundance or density data were collected, and the approach used to estimate the overall population trend value. Population trend values are reported using quantitative metrics in 75% of records that collectively represent more than 6,500 population abundance or density estimates. The remaining 25% of records qualitatively describe population change (e.g. increase, stable, and decrease).

Spatial location and grain

Records represent locations across the globe (latitude range: -51.0 to 80.0; longitude range: -166.0 to 166.0) but more information was found within the northern temperate zone. Coordinates, as reported within the primary literature, are available for 14% of records; in the remainder we have estimated the coordinates from the locations name.

Time period and grain

Records span from 1726 to 2016, but 92% of records represent data collected after 1950.

Major taxa and level of measurement

We searched for population trend data in 87 species from four families in the order Carnivora: Canidae, Felidae, Hyaenidae and Ursidae.

Software format

.csv

Keywords: Carnivora, Canidae, Felidae, Hyaenidae, Living Planet Index, Population trends, Ursidae

Introduction

The fate of the world's biodiversity is becoming more precarious each year, with reports of massive population declines (WWF, 2016), range contractions (Wolf & Ripple, 2017) and impending extinction events (Díaz et al., 2019). However, these biodiversity changes are not happening at the same rate in all places, with the fate of populations varying across regions (Fritz et al., 2009; Polaina et al., 2016), levels of protection (Amano et al., 2018), and the intrinsic traits of the affected species (Cardillo et al., 2005; González-Suárez & Revilla, 2013; Gonzalez-Suarez et al., 2013). An example of this variability in extinction can be seen in the largest terrestrial mammals in the order Carnivora, where there is evidence for population recoveries and recolonizations (Chapron et al., 2014), alongside declines and extinctions (Ripple et al., 2014).

Currently, the largest source of mammalian population trend data is within the Living Planet Index, which has collated population time-series for more than 3,000 species (WWF, 2020a). However, by only incorporating abundance time-series and excluding studies in which only the rate of change is available, the Living Planet Index limits the available data. Here, we expand upon the Living Planet Index for four families in the order Carnivora: Canidae, Felidae, Hyaenidae and Ursidae, which represent some of the world's most charismatic and iconic fauna. For the 87 species in these families, following the IUCN taxonomy, we compiled published population trend data, including both quantitative metrics and qualitative descriptions of population change. These data provide the most comprehensive global overview of population status for these species and can be used to evaluate different factors that influence population changes.

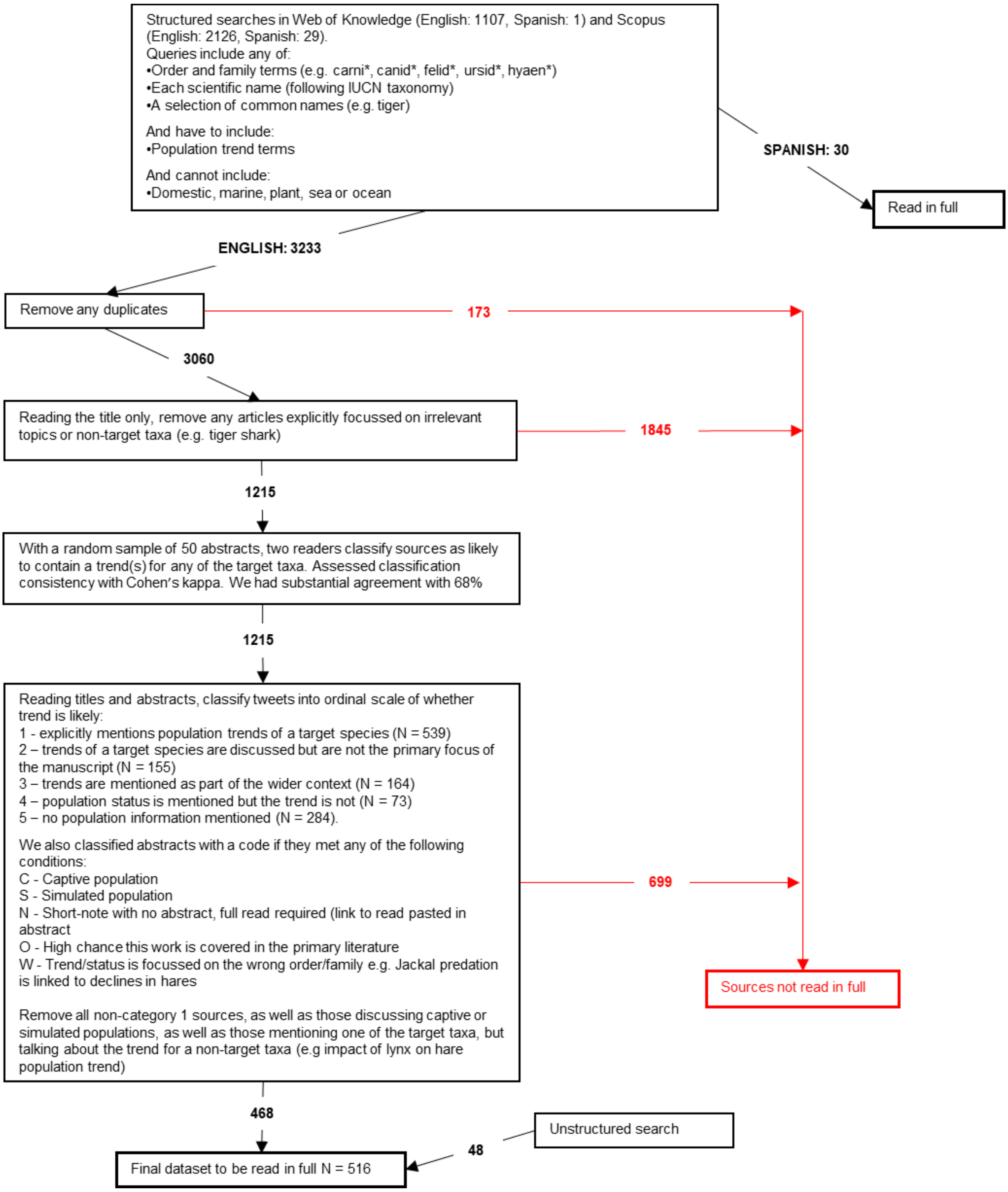
Methods

Locating population trend records

Between September 2017 and January 2018, six individuals conducted a literature search for carnivore population trends in our four target families. The purpose of this search was to ensure there was sufficient data for a large data compilation effort, and to refine the fields we wished to collect. These semi-structured searches included two terms: a reference to the taxon (e.g. common name, scientific name, family, or order) and a population trend term (e.g. population trend*, declin*, increase*, recover*, conservation status, or population growth rate). These searches were conducted within Scopus, Web of Science and Google Scholar. The six individuals identified 80 peer-reviewed publications containing population trend data.

In February 2018, confident data were abundant enough for a large compilation effort, we conducted a structured search in Scopus and Web of Science looking for terms (see full search query in Supplementary material: Box S1) in the title, abstract and keywords discussing population trends in the world's large terrestrial carnivorans. We considered the 87 species recognized in the four target families by the IUCN taxonomy (sourced February 2018). We queried the databases in English and Spanish. From this structured search, we found 3233 sources in English (reduced to 3060 after removing duplicates), and 30 sources in Spanish. Each of the Spanish sources were then read in full, but to further refine the English sources, each title was screened to remove sources not discussing the target taxa, e.g. references to '*tiger* shark' were removed, this reduced the English sources to 1215.

In the remaining 1215 English sources, two readers read the same random sample of 50 abstracts and classified them stating whether a population trend for any of the 87 species is likely present or absent. Cohen's kappa agreement between readers was substantial (68%), however it was clear some abstracts fell between the lines and could not be easily classified as containing a trend, or not. To account for this, one reader (TFJ) classified the remaining 1215 sources into five ordinal categories: 1 - explicitly mentions population trends of a target species (N = 539); 2 – trends of a target species are discussed but are not the primary focus of the manuscript (N = 155); 3 – trends are mentioned as part of the wider context (N = 164); 4 – population status is mentioned but the trend is not (N = 73); 5 – no population information mentioned (N = 284). Category 1 sources were then further refined to remove any sources discussing captive populations, simulated populations, or cases discussing trends in a non-target species (e.g. impact of lynx on hare population trends). This refining procedure left 468 category 1 sources, of which 32 had been identified in the unstructured search. After including the additional 48 sources identified in the unstructured search, we were left with 516 sources very likely to be discussing population trends in one of the target taxa. The entire refining process is summarised in Figure 1.



Structured searches in Web of Knowledge (English: 1107, Spanish: 1) and Scopus (English: 2126, Spanish: 29).
Queries include any of:
- Order and family terms (e.g. carni*, canid*, felid*, ursid*, hyaen*)
- Each scientific name (following IUCN taxonomy)
- A selection of common names (e.g. tiger)

And have to include:
- Population trend terms

And cannot include:
- Domestic, marine, plant, sea or ocean

Remove any duplicates

Reading the title only, remove any articles explicitly focussed on irrelevant topics or non-target taxa (e.g. tiger shark)

With a random sample of 50 abstracts, two readers classify sources as likely to contain a trend(s) for any of the target taxa. Assessed classification consistency with Cohen's kappa. We had substantial agreement with 68%

Reading titles and abstracts, classify tweets into ordinal scale of whether trend is likely:
1 - explicitly mentions population trends of a target species (N = 539)
2 – trends of a target species are discussed but are not the primary focus of the manuscript (N = 155)
3 – trends are mentioned as part of the wider context (N = 164)
4 – population status is mentioned but the trend is not (N = 73)
5 – no population information mentioned (N = 284).

We also classified abstracts with a code if they met any of the following conditions:
C - Captive population
S - Simulated population
N - Short-note with no abstract, full read required (link to read pasted in abstract
O - High chance this work is covered in the primary literature
W - Trend/status is focussed on the wrong order/family e.g. Jackal predation is linked to declines in hares

Remove all non-category 1 sources, as well as those discussing captive or simulated populations, as well as those mentioning one of the target taxa, but talking about the trend for a non-target taxa (e.g impact of lynx on hare population trend)

Final dataset to be read in full N = 516

Unstructured search

Read in full

Sources not read in full

Figure 1. Data flow diagram specifying steps taken to identify publications which should be read in full. Black-arrows indicate sources taken forwards to next refining stage; the associated numbers indicate the count of publications brought through each step. The red-arrows indicate sources which at different steps were deemed irrelevant and not to be read in full.

Each of the highly relevant sources (N = 516) were read in full (no other categories were read in full). A selection of these category 1 sources was secondary literature providing population trend syntheses or compilations (e.g. van de Kerk, de Kroon, Conde, & Jongejans, 2013). In these cases, we located the original primary sources described in the secondary source and added them ad-hoc to category 1, increasing the overall number of sources to 536 which were to be read in full. Of these 536, we were unable to find the full text in ten cases, in five cases the primary data had already been captured, and in four cases the text was in language other than English or Spanish. After excluding these cases, we read the remaining 517 texts in full

Extracting information from sources

When a source contained a population trend, we recorded the trend and additional metadata describing taxonomy, location, study period and methodology (Table 1). Population changes were reported in a variety of formats, but broadly fall into two groups, quantitative where the trend was described numerically (e.g. %change), and qualitative where the trend was described categorically (e.g. increase). In the quantitative category, we record the trend as presented in the original source. For studies that reported trends in multiple formats, we recorded the most informative e.g. if raw abundance data were available this would be preferred over any population change (%) estimates – see Table S1 for more information. If the population values were only reported in a graph or figure, we used a graphic digitiser (https://apps.automeris.io/wpd/) to estimate the values (Rohatgi, 2015). For population trends calculated from time-series data, we recorded the length of the time series (number of individual estimates used to derive the trend). For population trends based on matrix models and demographic parameters, we recorded the number of sampling years used to estimate the demographic parameters. For estimates of annual rates of change (λ and *r*) derived from three or more data points, we also noted any available estimate of dispersion (e.g., variance) and test-statistic values. For the qualitative descriptions of trends, we inferred the meaning of each description and placed the trend into one of the following 4 categories: increase – the populations increased; stable – the populations stable or unchanged; decrease – the populations decreased; varied – the population has increased and decreased over the monitoring period.

For each trend we recorded the binomial species name following the IUCN taxonomy – we identify discrepancies between the IUCN taxonomy and the Wilson & Reeder (2005) reference taxonomy in Table S2. In cases where the species name in the primary literature did not match the IUCN taxonomy, we referred to the IUCN taxonomy synonyms to locate the accepted IUCN species name. Subspecies names were also available in some primary sources, and we noted these as recorded in the primary source. For location, we recorded the name of the study site given in the primary source, whether the site was described as a protected area, and the country or countries it overlapped. If provided, we recorded the study site's coordinates (minimum, mid-point and maximum) converted into decimal degrees. Coordinate precision was likely variable among studies and is overall unknown. If studies did not report coordinates, we used the name given to the study site and location country to populate the coordinates using OpenCage (Salmon, 2018). OpenCage provides coordinates and a degree of confidence in the estimate, where 1 is low and 9 is high. For all coordinates were the confidence level fell below 7, we manually checked and if needed amended coordinates. When reported in the primary source, we also recorded the area (size) of the study site. For the study period in each record, we noted the start and end date of the population monitoring, and if available the corresponding population sizes at these dates. We captured the data collection and analysis methods from each source using several descriptors (Table 1). For studies that combined multiple methods, we precautionarily recorded the least robust approach. If we could not identify the method, the record was assigned 'undefined'.

Causes of change

Some sources tested or discussed the role of distinct factors to explain observed population changes. We recorded these factors reclassified into a modified version of the IUCN standardized classification schemes for Threats (v3.2) and Conservation Actions (v2.0), see Table S6. For each recorded factor we noted its effect (associated to increase or to decline) and how this influence was determined. It is important to note that effects were not always negative for threat scheme or positive for the conservation actions scheme. For example, urbanisation is listed under the threat scheme but has led to population increases in red fox *Vulpes vulpes* (Gloor et al., 2001). Finally, we note that factors not listed for a given record do not imply a threat or conservation action was not important or did not occur in that population, but simply that the factor was not mentioned in the source.

Validating records

The first author TJF read all sources and entered all data. To validate the records and ensure quality control, 31 (10%) of the category one sources TFJ identified as having data were also reviewed by an additional author (either PC or MGS). We selected the 10% sample with a random stratified approach to ensure each of the different formats of trends were reviewed e.g. percentage change, population time-series, and qualitative descriptions. TFJ then further scrutinised the 31 sources to detect errors in TFJs original work, that of that second readers, and identify causes of discrepancies in data entry.

Results

From the 542 sources read in full, 232 did not contain the population trend information we required and so were excluded from the dataset. Trends were excluded for a variety of reasons, examples include: the trend was simulated (N = 23), the trend referred to primary sources already captured in the dataset (N = 20), the trend described range change instead of abundance change (N = 6).

We identified and recorded 1122 population trends from the remaining 310 sources. These represented 50 (57%) of the studied species covering the four taxonomic families and 25 (69%) out of 36 genera (Figure 2). Some species had a single trend estimate, while we compiled 621 trend estimates for the top five species: gray wolf (*Canis lupus*), brown bear (*Ursus arctos*), grizzly bear (*Ursus americanus*), lion (*Panthera leo*) and eurasian lynx (*Lynx lynx*). Many of the records occurred within the northern hemisphere (Figure 3), particularly in Europe (N = 384) and North America (N = 415), but there was also a cluster of records in East and Southern Africa (N = 170) – with records in 86 countries in total. We located very few records in Central, North and West Africa, Central and South America, or Northern Asia. The dataset includes records extending from 1726-2017 (Figure 4), with the vast majority (90%) starting after 1950.

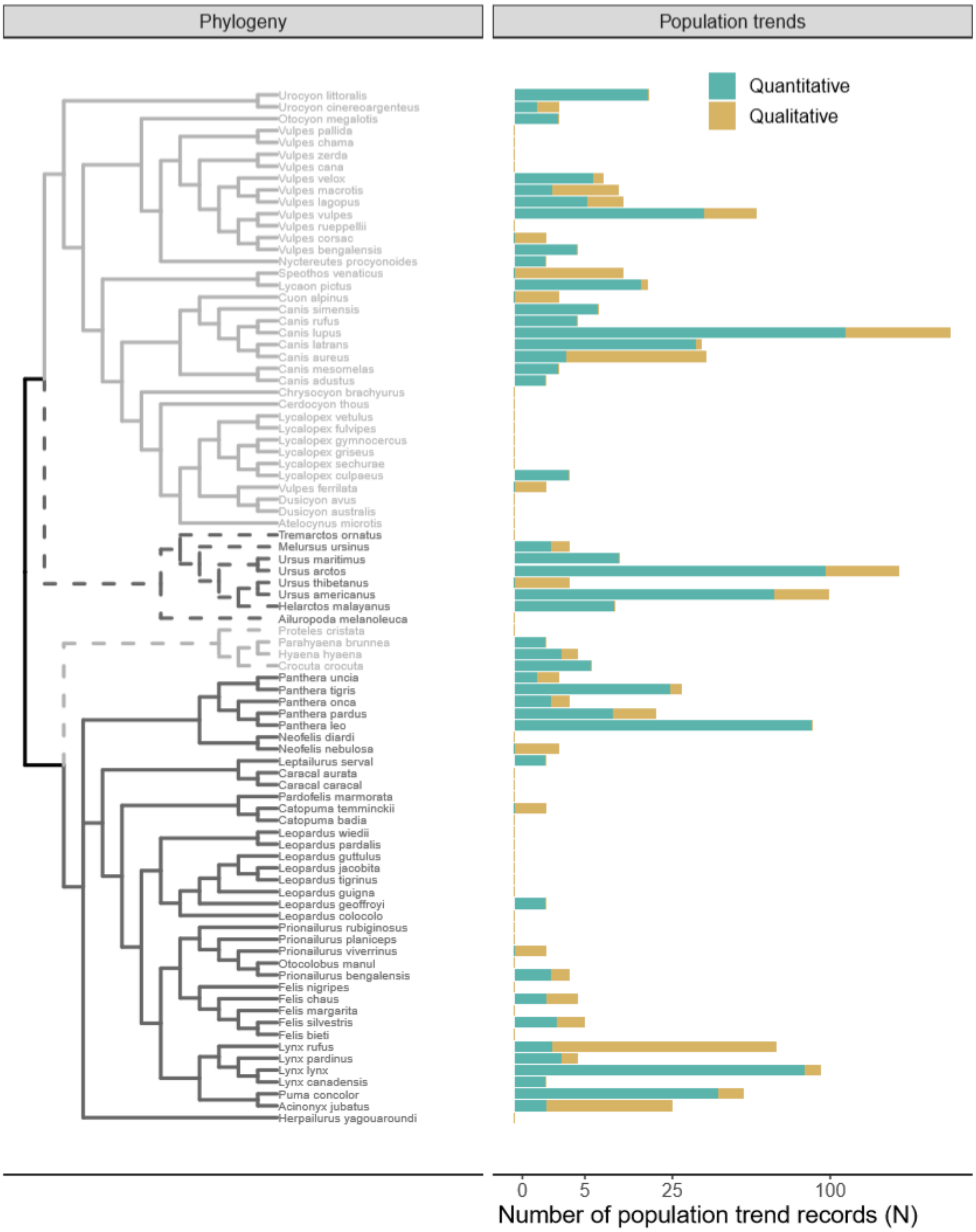


Figure 2. Number of population trend records per studied species, shown across the phylogeny. The tree represents four taxonomic families: Canidae (light grey – solid line), Ursidae (light grey – dotted line), Hyaenidae (dark grey – dotted line) and Felidae (dark grey – solid line). We shows records for both quantitative (teal) and qualitative (gold) trends.

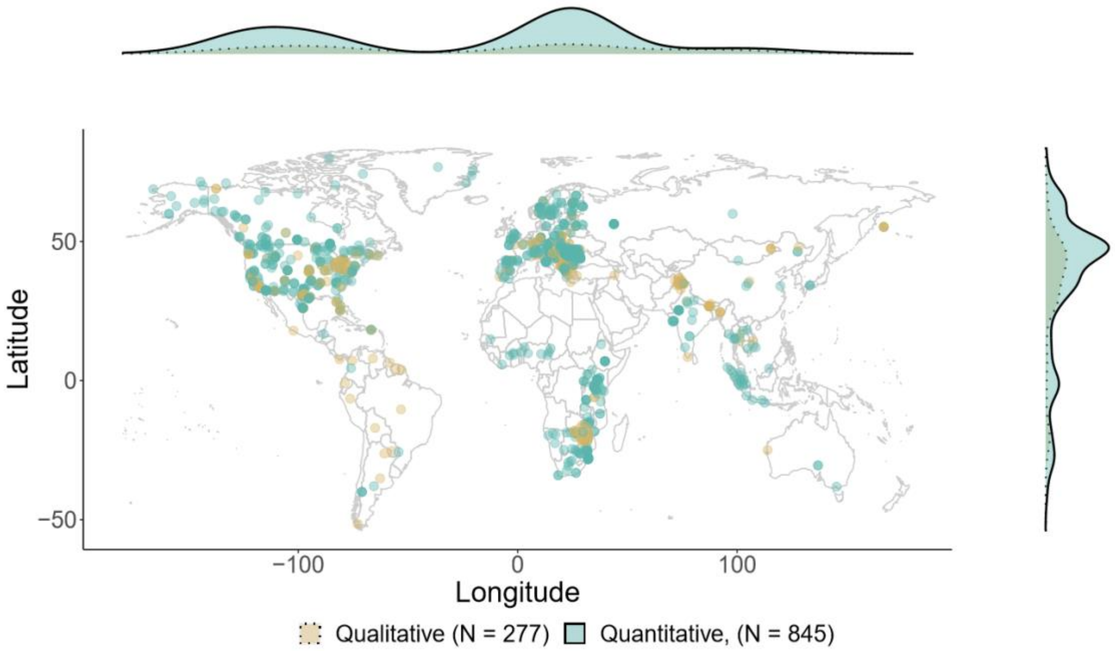


Figure 3. Location of study population from which we compiled quantitative (teal) and qualitative (gold) population trend records. Density plots indicate the frequency of the data points at varying latitudes and longitudes. Coordinates are decimal degrees.

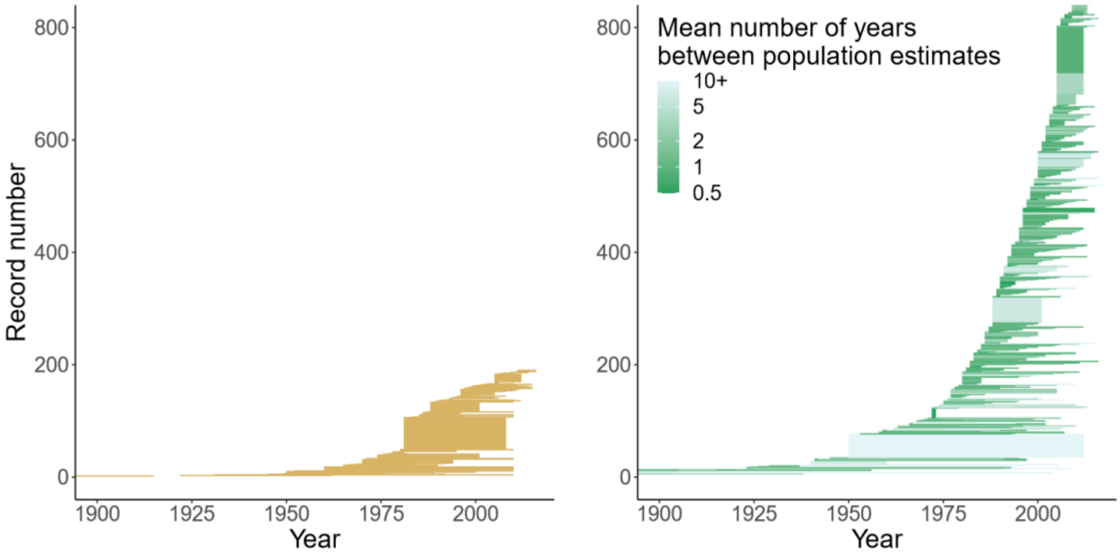


Figure 4. Distribution of qualitative (gold) and quantitative (green) population trend records between 1900-2015. Start and end date of each population trend record, ranked in ascending order of study start. For the quantitative plot, we display the mean number of years between population estimates in each trend as a proxy for sampling effort.

Most of the 1122 population trends represent quantitative estimates (N = 845), with a quarter (N = 277) providing only qualitative descriptors. The quantitative records collectively represent 6597 population size estimates. Most of the quantitative trends are recorded as a time-series of abundance values (63.9%), followed by population lambdas (17.4%), percentage change (7.5%), fold change (5.8%), and annual slope coefficients (5.4%).

Validating records

Across the 31 sources that underwent validation, TFJ and the second readers identified 46 population trends. The second readers located 40 and missed six. TFJ located 45 and missed one, which suggests that across all of the evaluated sources TFJ could have failed to detect ~2% of population trends. Further discrepancies were identified when TFJ re-scrutinised the 31 sources, and compared the original data entry to that of the second readers. Tthe second readers misclassified more of the values than TFJ (7.4% vs 0%), produced more missing values (6.4% vs. 0.3%), and made more typos (0.5% vs. 0.3%). Despite these discrepancies, the results were qualitatively very similar in those trends identified by both TFJ and the second readers, with TFJ and the second readers producing the same trend value, same species, and similar locations e.g. TFJ and the second readers identified the same state or region in all cases. Furthermore, both TFJ and the second readers identified the same trends that should be treated cautiously and flagged with a warning in the dataset. All of this considered, the classification protocol was considered robust as TFJ, who entered the values in the full dataset, produced relatively few errors.

Discussion

We searched the literature to retrieve population trend records for 87 species of large carnivorans and located 1122 estimates of population change representing 50 species. These records cover a wide temporal window (1726-2017) and represent diverse locations around the globe, although, there is temporal and spatial heterogeneity with more records in recent years and temperate areas of the Northern hemisphere. Our effort substantially expanded the previous dataset for these species (the Living Planet Index includes 367 trends, as of 2019) and thus, CaPTrends provides a valuable resource to address ecological questions, complete a more comprehensive assessment of population status for these species, and explore potential predictors of observed population changes.

Our dataset located additional time-series records not reported in the Living Planet Index, but also added less precise and qualitative descriptors which need to be interpreted with caution. For example, we found that studies that provided summarised quantitative metrics (e.g. annual population growth/lambda) did not always offer estimates of their error and thus, we could not extract uncertainty around the trend in all cases. This issue is even more emphasised in the qualitative descriptions (e.g. stable, increase, or decrease), where both the error and magnitude of the trend are unknown. However, if used cautiously, the lower resolution metrics could be important in addressing data gaps for species and locations for which high resolution population trend records are not available (WWF, 2016). This is particularly important, as these data gaps are most prevalent in biodiverse regions (WWF, 2016), which are experiencing the greatest negative-change in human footprint (Venter et al., 2016). Incorporating lower resolution metrics into models of biodiversity change could reduce some of these biases - providing a robust modelling approach is used.

Whilst CaPTrends, and especially the inclusion of the qualitative data, increases the taxonomic, spatial, and temporal coverage of the publicly available Carnivoran population trend data, biases are still present. Across the taxonomy, the Lycalopex and Leopardus genera are particularly data poor, and generally, it appears smaller species occurring outside of North America and Europe are the most poorly represented. These spatial biases do persist for a whole array of species though, with records largely absent from South and Central America, West and North Africa, Eastern Europe, as well as North and West Asia. For the temporal biases, records are very sparse before 1950, and sparse as well as generally lower quality (few abundance observations per trend) before 1980, suggesting this data is likely unfit for use preceding the 1970 baseline used in the Living Planet index (WWF, 2020b).

Within the data, there is a risk that the records are prone to publication bias, whereby records are more likely to be available within the literature for certain species and locations, and even for certain trend types (e.g. perhaps stable trends are less likely to be published compared to decreasing trends). Publication bias is likely widespread within ecological datasets compiled through systematic reviews (Møllerand & Jennions, 2001; Jennions & Møller, 2002), and when this data is analysed, the likely publication bias warrants careful consideration and methodological nuance. However, relative to many other ecological datasets, we would expect the degree of this publication bias to be far smaller, as Carnivorans are often viewed as charismatic, and so even stable trends outside of the global north may be viewed as attractive by journals, and so would be detected by the systematic search.

Future work could expand on this dataset by exploring other languages, beyond the English and Spanish that are already included, or even capturing qualitative data across a wider array of species. However, compilation efforts like this are very time-consuming, and so future work could also explore options for automating this data collection using text analysis tools (Cornford et al., 2021; Johnson et al., 2021c)

Usage notes

CaPTrends is presented as a relational dataset (Figure 5). The main file ‘captrends.csv’ includes all master data (e.g. unique id, species, location and time-frame), as well as all population data, except the population time-series. Time-series of population abundances and population changes are located in ‘ts_abundance.csv’ and ‘ts_change.csv’, respectively, both of which are linked to ‘captrends.csv’ through the ‘DataTableID’ field. ‘direction.csv’ also links to ‘captrends.csv’ through ‘DataTableID’ and describes positive and negative influences of each trend. Finally, ‘sources.csv’ links to ‘captrends.csv’ through ‘Citation_key’ and contains information on where the trend was sourced from e.g. the tile of the publication. Comprehensive metadata is available for each of these datasets in the supplementary material.

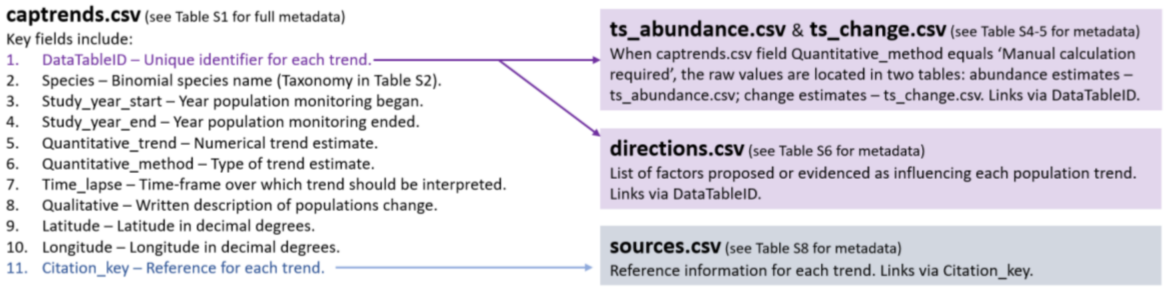


Figure 5. Diagram depicting relational database, including each datasets contents, and how each dataset is linked (arrows).

To support the use of this dataset, each population trend record has been annotated and labelled (Table S1). Much of this information would be helpful in filtering the dataset to exclude trends that are deemed of low quality or irrelevant to a given research question. For example, for investigating extinction risk, one may opt to remove data for invasive populations.

This dataset may be analysed focusing on different descriptors. Including qualitative descriptors provides the most records but highest uncertainty and involves setting thresholds to assign quantitative metrics into categories which is not always straightforward. Focusing only on quantitative records reduces the scope and increase biases (not all species and areas are equally like to have quantitative records as shown in figure 2). Approaches like data integration (Isaac et al., 2020) which can incorporate both data types are likely to be least biased (spatially, temporally, and taxonomically).

Another consideration when using this data is to account for differences in population sampling effort over time. In the majority of records, abundance estimates are scaled relative to effort, or provide an empirically derived abundance or density estimate. However, in nearly 30% of observations, the reported trend, or described timeseries is either insufficiently described to know whether the population sampling effort has been account for in the trend and/or timeseries, or the paper has specifically described that the abundance estimates are not relative to effort – see captrends.csv metadata for column Modelling_method. When using this dataset, its essential that these less robust estimates are treated cautiously.

Acknowledgements

Thanks to Ella Coley, Jasmine Ashley, Jessica Marshall, Matthew Bemment, Monty Jefferson, Sarah Granger that assisted in data collection, and Julia Martínez Pardo that helped in project design. This work was funded by a NERC (Natural Environment Research Council) Centre for Doctoral Training studentship (J71566E).

Supplementary material and dataset metadata

Box S1. Structured search queries in Web of Science and Scopus. Queries were developed and searched in both English and Spanish

Web of Science (English): TS=(("carnivor*" OR "felid*" OR "canid*" OR "ursid*" OR "hyaen*" OR "Atelocynus microtis" OR "Canis adustus" OR "Canis aureus" OR "Canis latrans" OR "Canis lupus" OR "Canis mesomelas" OR "Canis rufus" OR "Canis simensis" OR "Cerdocyon thous" OR "Chrysocyon brachyurus" OR "Cuon alpinus" OR "Dusicyon australis" OR "Dusicyon avus" OR "Lycalopex culpaeus" OR "Lycalopex fulvipes" OR "Lycalopex griseus" OR "Lycalopex gymnocercus" OR "Lycalopex sechurae" OR "Lycalopex vetulus" OR "Lycaon pictus" OR "Nyctereutes procyonoides" OR "Otocyon megalotis" OR "Speothos venaticus" OR "Urocyon cinereoargenteus" OR "Urocyon littoralis" OR "Vulpes bengalensis" OR "Vulpes cana" OR "Vulpes chama" OR "Vulpes corsac" OR "Vulpes ferrilata" OR "Vulpes lagopus" OR "Vulpes macrotis" OR "Vulpes pallida" OR "Vulpes rueppellii" OR "Vulpes velox" OR "Vulpes vulpes" OR "Vulpes zerda" OR "Acinonyx jubatus" OR "Caracal aurata" OR "Caracal caracal" OR "Catopuma badia" OR "Catopuma temminckii" OR "Felis bieti" OR "Felis chaus" OR "Felis margarita" OR "Felis nigripes" OR "Felis silvestris" OR "Herpailurus yagouaroundi" OR "Leopardus colocolo" OR "Leopardus geoffroyi" OR "Leopardus guigna" OR "Leopardus guttulus" OR "Leopardus jacobita" OR "Leopardus pardalis" OR "Leopardus tigrinus" OR "Leopardus wiedii" OR "Leptailurus serval" OR "Lynx canadensis" OR "Lynx lynx" OR "Lynx pardinus" OR "Lynx rufus" OR "Neofelis diardi" OR "Neofelis nebulosa" OR "Otocolobus manul" OR "Panthera leo" OR "Panthera onca" OR "Panthera pardus" OR "Panthera tigris" OR "Panthera uncia" OR "Pardofelis marmorata" OR "Prionailurus bengalensis" OR "Prionailurus planiceps" OR "Prionailurus rubiginosus" OR "Prionailurus viverrinus" OR "Puma concolor" OR "Crocuta crocuta" OR "Hyaena hyaena" OR "Parahyaena brunnea" OR "Proteles cristata" OR "Ailuropoda melanoleuca" OR "Helarctos malayanus" OR "Melursus ursinus" OR "Tremarctos ornatus" OR "Ursus americanus" OR "Ursus arctos" OR "Ursus maritimus" OR "Ursus thibetanus" OR "Canis thous" OR "Canis brachyurus" OR "Canis alpinus" OR "Canis australis" OR "Pseudalopex culpaeus" OR "Pseudalopex fulvipes" OR "Pseudalopex griseus" OR "Pseudalopex gymnocercus" OR "Pseudalopex sechurae" OR "Pseudalopex vetulus" OR "Canis procyonoides" OR "Canis cinereoargenteus" OR "Vulpes littoralis" OR "Alopex lagopus" OR "Vulpes rueppelli " OR "Canis vulpes" OR "Fennecus zerda" OR "Felis jubata" OR "Profelis aurata" OR "Felis badia" OR "Pardofelis badia" OR "Felis temminckii" OR "Pardofelis temminckii" OR "Felis silvestris" OR "Felis yagouaroundi" OR "Herpailurus yagouaroundi" OR "Puma yagouaroundi" OR "Lynchailurus colocolo" OR "Oncifelis colocolo" OR "Oncifelis geoffroyi" OR "Oncifelis guigna" OR "Felis jacobita" OR "Oreailurus jacobita" OR "Oreailurus jacobitus" OR "Oreailurus jacobitus" OR "Caracal serval" OR "Felis nebulosa" OR "Felis manul" OR "Felis leo" OR "Felis onca" OR "Felis pardus" OR "Felis tigris" OR "Felis uncia" OR "Uncia uncia" OR "Felis concolor" OR "Hyaena brunnea" OR "Proteles cristatus" OR "Ursus melanoleucus" OR "Ursus malayanus" OR "Bradypus ursinus" OR "Ursus ornatus" OR "Thalarctos maritimus" OR "Aardwolf" OR "Fox" OR "Wolf" OR "Lynx" OR "Lion" OR "Leopard" OR "Caracal" OR "Bear" OR "Jackal" OR "Bobcat" OR "Puma" OR "Cougar" OR "Dhole" OR "Hyaena" OR "Cheetah" OR "Fennec" OR "Cat" OR "Dog" OR "Panda" OR "Margay" OR "Ocelot" OR "Tiger") AND ("population trend*" OR "population dec*" OR "population increas*" OR "population recover*" OR "conservation status" OR "population growth") NOT ("domestic" OR "marine" OR "plant" OR "sea" OR "ocean"))

Scopus (English): TITLE-ABS-KEY ("carnivor*" OR "felid*" OR "canid*" OR "ursid*" OR "hyaen*" OR "Atelocynus microtis" OR "Canis adustus" OR "Canis aureus" OR "Canis latrans" OR "Canis lupus" OR "Canis mesomelas" OR "Canis rufus" OR "Canis simensis" OR "Cerdocyon thous" OR "Chrysocyon brachyurus" OR "Cuon alpinus" OR "Dusicyon australis" OR "Dusicyon avus" OR "Lycalopex culpaeus" OR "Lycalopex fulvipes" OR "Lycalopex griseus" OR "Lycalopex gymnocercus" OR "Lycalopex sechurae" OR "Lycalopex vetulus" OR "Lycaon pictus" OR "Nyctereutes procyonoides" OR "Otocyon megalotis" OR "Speothos venaticus" OR "Urocyon cinereoargenteus" OR "Urocyon littoralis" OR "Vulpes bengalensis" OR "Vulpes cana" OR "Vulpes chama" OR "Vulpes corsac" OR "Vulpes ferrilata" OR "Vulpes lagopus" OR "Vulpes macrotis" OR "Vulpes pallida" OR "Vulpes rueppellii" OR "Vulpes velox" OR "Vulpes vulpes" OR "Vulpes zerda" OR "Acinonyx jubatus" OR "Caracal aurata" OR "Caracal caracal" OR "Catopuma badia" OR "Catopuma temminckii" OR "Felis bieti" OR "Felis chaus" OR "Felis margarita" OR "Felis nigripes" OR "Felis silvestris" OR "Herpailurus yagouaroundi" OR "Leopardus colocolo" OR "Leopardus geoffroyi" OR "Leopardus guigna" OR "Leopardus guttulus" OR "Leopardus jacobita" OR "Leopardus pardalis" OR "Leopardus tigrinus" OR "Leopardus wiedii" OR "Leptailurus serval" OR "Lynx canadensis" OR "Lynx lynx" OR "Lynx pardinus" OR "Lynx rufus" OR "Neofelis diardi" OR "Neofelis nebulosa" OR "Otocolobus manul" OR "Panthera leo" OR "Panthera onca" OR "Panthera pardus" OR "Panthera tigris" OR "Panthera uncia" OR "Pardofelis marmorata" OR "Prionailurus bengalensis" OR "Prionailurus planiceps" OR "Prionailurus rubiginosus" OR "Prionailurus viverrinus" OR "Puma concolor" OR "Crocuta crocuta" OR "Hyaena hyaena" OR "Parahyaena brunnea" OR "Proteles cristata" OR "Ailuropoda melanoleuca" OR "Helarctos malayanus" OR "Melursus ursinus" OR "Tremarctos ornatus" OR "Ursus americanus" OR "Ursus arctos" OR "Ursus maritimus" OR "Ursus thibetanus" OR "Canis thous" OR "Canis brachyurus" OR "Canis alpinus" OR "Canis australis" OR "Pseudalopex culpaeus" OR "Pseudalopex fulvipes" OR "Pseudalopex griseus" OR "Pseudalopex gymnocercus" OR "Pseudalopex sechurae" OR "Pseudalopex vetulus" OR "Canis procyonoides" OR "Canis cinereoargenteus" OR "Vulpes littoralis" OR "Alopex lagopus" OR "Vulpes rueppelli " OR "Canis vulpes" OR "Fennecus zerda" OR "Felis jubata" OR "Profelis aurata" OR "Felis badia" OR "Pardofelis badia" OR "Felis temminckii" OR "Pardofelis temminckii" OR "Felis silvestris" OR "Felis yagouaroundi" OR "Herpailurus yagouaroundi" OR "Puma yagouaroundi" OR "Lynchailurus colocolo" OR "Oncifelis colocolo" OR "Oncifelis geoffroyi" OR "Oncifelis guigna" OR "Felis jacobita" OR "Oreailurus jacobita" OR "Oreailurus jacobitus" OR "Oreailurus jacobitus" OR "Caracal serval" OR "Felis nebulosa" OR "Felis manul" OR "Felis leo" OR "Felis onca" OR "Felis pardus" OR "Felis tigris" OR "Felis uncia" OR "Uncia uncia" OR "Felis concolor" OR "Hyaena brunnea" OR "Proteles cristatus" OR "Ursus melanoleucus" OR "Ursus malayanus" OR "Bradypus ursinus" OR "Ursus ornatus" OR "Thalarctos maritimus" OR "Aardwolf" OR "Fox" OR "Wolf" OR "Lynx" OR "Lion" OR "Leopard" OR "Caracal" OR "Bear" OR "Jackal" OR "Bobcat" OR "Puma" OR "Cougar" OR "Dhole" OR "Hyaena" OR "Cheetah" OR "Fennec" OR "Cat" OR "Dog" OR "Panda" OR "Margay" OR "Ocelot" OR "Tiger") AND ("population trend*" OR "population dec*" OR "population increas*" OR "population recover*" OR "conservation status" OR "population growth") AND NOT ("domestic" OR "marine" OR "plant" OR "sea" OR "ocean")

Web of Science (Spanish): TS = (("carnivor*" OR "felid*" OR "canid*" OR "ursid*" OR "hyaen*" OR "Atelocynus microtis" OR "Canis adustus" OR "Canis aureus" OR "Canis latrans" OR "Canis lupus" OR "Canis mesomelas" OR "Canis rufus" OR "Canis simensis" OR "Cerdocyon thous" OR "Chrysocyon brachyurus" OR "Cuon alpinus" OR "Dusicyon australis" OR "Dusicyon avus" OR "Lycalopex culpaeus" OR "Lycalopex fulvipes" OR "Lycalopex griseus" OR "Lycalopex gymnocercus" OR "Lycalopex sechurae" OR "Lycalopex vetulus" OR "Lycaon pictus" OR "Nyctereutes procyonoides" OR "Otocyon megalotis" OR "Speothos venaticus" OR "Urocyon cinereoargenteus" OR "Urocyon littoralis" OR "Vulpes bengalensis" OR "Vulpes cana" OR "Vulpes chama" OR "Vulpes corsac" OR "Vulpes ferrilata" OR "Vulpes lagopus" OR "Vulpes macrotis" OR "Vulpes pallida" OR "Vulpes rueppellii" OR "Vulpes velox" OR "Vulpes vulpes" OR "Vulpes zerda" OR "Acinonyx jubatus" OR "Caracal aurata" OR "Caracal caracal" OR "Catopuma badia" OR "Catopuma temminckii" OR "Felis bieti" OR "Felis chaus" OR "Felis margarita" OR "Felis nigripes" OR "Felis silvestris" OR "Herpailurus yagouaroundi" OR "Leopardus colocolo" OR "Leopardus geoffroyi" OR "Leopardus guigna" OR "Leopardus guttulus" OR "Leopardus jacobita" OR "Leopardus pardalis" OR "Leopardus tigrinus" OR "Leopardus wiedii" OR "Leptailurus serval" OR "Lynx canadensis" OR "Lynx lynx" OR "Lynx pardinus" OR "Lynx rufus" OR "Neofelis diardi" OR "Neofelis nebulosa" OR "Otocolobus manul" OR "Panthera leo" OR "Panthera onca" OR "Panthera pardus" OR "Panthera tigris" OR "Panthera uncia" OR "Pardofelis marmorata" OR "Prionailurus bengalensis" OR "Prionailurus planiceps" OR "Prionailurus rubiginosus" OR "Prionailurus viverrinus" OR "Puma concolor" OR "Crocuta crocuta" OR "Hyaena hyaena" OR "Parahyaena brunnea" OR "Proteles cristata" OR "Ailuropoda melanoleuca" OR "Helarctos malayanus" OR "Melursus ursinus" OR "Tremarctos ornatus" OR "Ursus americanus" OR "Ursus arctos" OR "Ursus maritimus" OR "Ursus thibetanus" OR "Canis thous" OR "Canis brachyurus" OR "Canis alpinus" OR "Canis australis" OR "Pseudalopex culpaeus" OR "Pseudalopex fulvipes" OR "Pseudalopex griseus" OR "Pseudalopex gymnocercus" OR "Pseudalopex sechurae" OR "Pseudalopex vetulus" OR "Canis procyonoides" OR "Canis cinereoargenteus" OR "Vulpes littoralis" OR "Alopex lagopus" OR "Vulpes rueppelli " OR "Canis vulpes" OR "Fennecus zerda" OR "Felis jubata" OR "Profelis aurata" OR "Felis badia" OR "Pardofelis badia" OR "Felis temminckii" OR "Pardofelis temminckii" OR "Felis silvestris" OR "Felis yagouaroundi" OR "Herpailurus yagouaroundi" OR "Puma yagouaroundi" OR "Lynchailurus colocolo" OR "Oncifelis colocolo" OR "Oncifelis geoffroyi" OR "Oncifelis guigna" OR "Felis jacobita" OR "Oreailurus jacobita" OR "Oreailurus jacobitus" OR "Oreailurus jacobitus" OR "Caracal serval" OR "Felis nebulosa" OR "Felis manul" OR "Felis leo" OR "Felis onca" OR "Felis pardus" OR "Felis tigris" OR "Felis uncia" OR "Uncia uncia" OR "Felis concolor" OR "Hyaena brunnea" OR "Proteles cristatus" OR "Ursus melanoleucus" OR "Ursus malayanus" OR "Bradypus ursinus" OR "Ursus ornatus" OR "Thalarctos maritimus" OR "Borochi" OR "Chacal" OR "Zorro" OR "Gato" OR "León" OR "Manigordo" OR "Manul" OR "Mitzli" OR "Ocelote" OR "Onza" OR "Oso" OR "Pacha" OR "Panda" OR "Pantera" OR "Perro" OR "Renard" OR "Tigre" OR "Tigrillo" OR "Tirica" OR "Umba" OR "Yaguar*" OR "Jaguar" OR "Lince" OR "Lobo" OR "Aguará") AND ("tendencia poblacional" OR "declinación poblacional" OR "incremento poblacional" OR "recuperación poblacional" OR "estado de conservación" OR "tasa de incremento poblacional") NOT ("Doméstic*" OR "Marin*" OR "Planta" OR "Océano"))

Scopus (Spanish): TITLE-ABS-KEY ("carnivor*" OR "felid*" OR "canid*" OR "ursid*" OR "hyaen*" OR "Atelocynus microtis" OR "Canis adustus" OR "Canis aureus" OR "Canis latrans" OR "Canis lupus" OR "Canis mesomelas" OR "Canis rufus" OR "Canis simensis" OR "Cerdocyon thous" OR "Chrysocyon brachyurus" OR "Cuon alpinus" OR "Dusicyon australis" OR "Dusicyon avus" OR "Lycalopex culpaeus" OR "Lycalopex fulvipes" OR "Lycalopex griseus" OR "Lycalopex gymnocercus" OR "Lycalopex sechurae" OR "Lycalopex vetulus" OR "Lycaon pictus" OR "Nyctereutes procyonoides" OR "Otocyon megalotis" OR "Speothos venaticus" OR "Urocyon cinereoargenteus" OR "Urocyon littoralis" OR "Vulpes bengalensis" OR "Vulpes cana" OR "Vulpes chama" OR "Vulpes corsac" OR "Vulpes ferrilata" OR "Vulpes lagopus" OR "Vulpes macrotis" OR "Vulpes pallida" OR "Vulpes rueppellii" OR "Vulpes velox" OR "Vulpes vulpes" OR "Vulpes zerda" OR "Acinonyx jubatus" OR "Caracal aurata" OR "Caracal caracal" OR "Catopuma badia" OR "Catopuma temminckii" OR "Felis bieti" OR "Felis chaus" OR "Felis margarita" OR "Felis nigripes" OR "Felis silvestris" OR "Herpailurus yagouaroundi" OR "Leopardus colocolo" OR "Leopardus geoffroyi" OR "Leopardus guigna" OR "Leopardus guttulus" OR "Leopardus jacobita" OR "Leopardus pardalis" OR "Leopardus tigrinus" OR "Leopardus wiedii" OR "Leptailurus serval" OR "Lynx canadensis" OR "Lynx lynx" OR "Lynx pardinus" OR "Lynx rufus" OR "Neofelis diardi" OR "Neofelis nebulosa" OR "Otocolobus manul" OR "Panthera leo" OR "Panthera onca" OR "Panthera pardus" OR "Panthera tigris" OR "Panthera uncia" OR "Pardofelis marmorata" OR "Prionailurus bengalensis" OR "Prionailurus planiceps" OR "Prionailurus rubiginosus" OR "Prionailurus viverrinus" OR "Puma concolor" OR "Crocuta crocuta" OR "Hyaena hyaena" OR "Parahyaena brunnea" OR "Proteles cristata" OR "Ailuropoda melanoleuca" OR "Helarctos malayanus" OR "Melursus ursinus" OR "Tremarctos ornatus" OR "Ursus americanus" OR "Ursus arctos" OR "Ursus maritimus" OR "Ursus thibetanus" OR "Canis thous" OR "Canis brachyurus" OR "Canis alpinus" OR "Canis australis" OR "Pseudalopex culpaeus" OR "Pseudalopex fulvipes" OR "Pseudalopex griseus" OR "Pseudalopex gymnocercus" OR "Pseudalopex sechurae" OR "Pseudalopex vetulus" OR "Canis procyonoides" OR "Canis cinereoargenteus" OR "Vulpes littoralis" OR "Alopex lagopus" OR "Vulpes rueppelli " OR "Canis vulpes" OR "Fennecus zerda" OR "Felis jubata" OR "Profelis aurata" OR "Felis badia" OR "Pardofelis badia" OR "Felis temminckii" OR "Pardofelis temminckii" OR "Felis silvestris" OR "Felis yagouaroundi" OR "Herpailurus yagouaroundi" OR "Puma yagouaroundi" OR "Lynchailurus colocolo" OR "Oncifelis colocolo" OR "Oncifelis geoffroyi" OR "Oncifelis guigna" OR "Felis jacobita" OR "Oreailurus jacobita" OR "Oreailurus jacobitus" OR "Oreailurus jacobitus" OR "Caracal serval" OR "Felis nebulosa" OR "Felis manul" OR "Felis leo" OR "Felis onca" OR "Felis pardus" OR "Felis tigris" OR "Felis uncia" OR "Uncia uncia" OR "Felis concolor" OR "Hyaena brunnea" OR "Proteles cristatus" OR "Ursus melanoleucus" OR "Ursus malayanus" OR "Bradypus ursinus" OR "Ursus ornatus" OR "Thalarctos maritimus" OR "Borochi" OR "Chacal" OR "Zorro" OR "Gato" OR "León" OR "Manigordo" OR "Manul" OR "Mitzli" OR "Ocelote" OR "Onza" OR "Oso" OR "Pacha" OR "Panda" OR "Pantera" OR "Perro" OR "Renard" OR "Tigre" OR "Tigrillo" OR "Tirica" OR "Umba" OR "Yaguar*" OR "Jaguar" OR "Lince" OR "Lobo" OR "Aguará") AND ("tendencia poblacional" OR "declinación poblacional" OR "incremento poblacional" OR "recuperación poblacional" OR "estado de conservación" OR "tasa de incremento poblacional") AND NOT ("Doméstic*" OR "Marin*" OR "Planta" OR "Océano") AND (LIMIT-TO (LANGUAGE , "Spanish"))

Table S1. Description of fields in the table captrends.csv. This is the core dataset with all master information and is the foundation that links to the other relational databases: timeseries.csv, direction.csv, and sources.csv. 'Data type' describes the format of the data, for categorical fields the selection options are underlined with its description italicised.

Field	Description	Data type
DataTableID	Unique numerical code for each population trend record. Matches with tables: ts_abundance.csv, ts_change.csv, direction.csv.	Character
Species	Binomial species name following IUCN taxonomy	Character [populated from Table S2]
Sub_species	Subspecies as listed within the source	Character
Citation_key	Unique alphanumerical code for each source to match with table sources.csv	Character
Spatial_locality	If papers have trends split into different sites, each site is given its own spatial unique numerical code	Numeric
Temporal_locality	If papers have trends split into different time points (e.g. 1980 - 1990, and 1990 - 2000), each consecutive time series is given its own temporal numerical code	Numeric
Locality_name	Name of study site as described in the primary source	Character
Singular_country	Country where studied population occurs following ISO3166 naming standards as of 2018 (e.g. source mentions Soviet Union and coordinates indicate Russia, Russia was recorded).	Character [populated from Table S3]
Multiple_countries	When studied population overlaps multiple countries, each country is included in a list separated with semi-colons. Country names follow ISO3166 standards.	Character [populated from supplementary: Country list]
Wider_population	Further information about the study site e.g. name of the region, state or national park.	Character
Locality_area	Numeric estimate of the study site area	Numeric
Locality_area_units	Units in which area of study site 'Locality_area' is reported. Categories: Hectares: *Area where the population was studied (recorded in hectares)* Km^2: *Area where the population was studied (recorded in square kilometres)*	Categorical
Study_year_start	Year of first population size estimate	Numeric
Study_year_end	Year of final population size estimate	Numeric
N_observations	Number of population size estimates used to derive the trend. For quantitative population trends, the minimum value is 2. For records which include the complete time-series the value is missing but can be extracted from the ts_abundance.csv and ts_change.csv tables. For qualitative trends this value regularly equals one or zero, as there are many cases in the qualitative trends where only one or zero population estimates are made, and the assessment of the trend is more subjective. For matrix models, this value represents the number of sampling years, rather than the number of population size estimates.	Numeric
Field_method	Field method for deriving population size estimates or demographic information. Categories: Individuals identified: *All individuals of a population were identified.* Systematic – direct: *Monitoring approach is systematic (not-opportunistic), is not clearly prone to spatial or temporal bias, and*	Categorical

	involves direct observations of the animal (either alive or dead) e.g. through camera-trap grids or road-transects. *Systematic – indirect: Monitoring approach is systematic (not-opportunistic), is not clearly prone to spatial or temporal bias, and involves indirect observations of the animal e.g. footprint, audio calls, fur traps.* *Systematic -undefined: Monitoring approach is systematic (not-opportunistic) and is not clearly prone to spatial or temporal bias but the actual method of making observations is unclear or a mix of direct and indirect.* *Unsystematic – direct: Monitoring approach is opportunistic or not completely systematic and is at least partially prone to spatial or temporal bias; also involves direct observations of the animal (either alive or dead) e.g. through camera-trap grids or road-transects.* *Unsystematic – indirect: Monitoring approach is opportunistic or not completely systematic and is at least partially prone to spatial or temporal bias; also involves indirect observations of the animal e.g. footprint, audio calls, fur traps.* *Unsystematic - undefined: Monitoring approach is opportunistic or not completely systematic and is at least partially prone to spatial or temporal bias; also the actual method of making observations is unclear or a mix of direct and indirect.* *Undefined: Population monitoring method poorly defined or does not meet one of the above criteria*.	
Modelling_method	Analysis method for deriving population estimates or demographic information. Categories: *Model derived abundance/density: Statistical model used to convert field data into population abundance or density estimates.* *Model occupancy: Statistical model used to convert occupancy field data into population abundance or density estimates.* *Matrix modelling: Statistical model to estimate population change using demographic parameters.* *Total count: Total population size is known, no need for statistical inference of abundance.* *Relative abundance: Statistical approach to control for different sampling effort in detection events e.g. relative abundance.* *Field values: Raw field data presented, no statistical modelling used to control for differences in sampling effort, observers etc.* *Undefined: Approach for estimating population size is unclear or not explained, or does not clearly fall into any other category.*	Categorical
Population_metric	Type of population size measurement. Categories: *Abundance: Estimates of the number of individuals in the population.* *Density: Estimate of the number of individuals per unit of area. Units defined by variable 'Density_scale'.* *Other: Estimate of the population size in alternate units e.g. relative abundance.*	Categorical
Density_scale	Units of population_metric when reported as *Density*.	Character
Population_start	Population size estimate in the first recorded year (as listed in field ‘Study_year_start’). Type of estimate described in field ‘Population_metric ‘	Numeric

PS_dispersion_estimate	Estimate of dispersion or uncertainty in the population size value provided in field ‘Population_start’. Values entered here when they are provided as single estimate (e.g., SE or SD) Type of estimate described in field ‘PS_PE_dispersion_description’	Numeric
PS_dispersion_lower	Estimate of dispersion or uncertainty in the population size value provided in field ‘Population_start’. Values entered here when they are provided as a lower bounded estimate (e.g., range or confidence intervals) Type of estimate described in field ‘PS_PE_dispersion_description’	Numeric
PS_dispersion_upper	Estimate of dispersion or uncertainty in the population size value provided in field ‘Population_start’. Values entered here when they are provided as an upper bounded estimate (e.g., range or confidence intervals) Type of estimate described in field ‘PS_PE_dispersion_description’	Numeric
Population_end	Population size estimate in the last recorded year (as listed in field ‘Study_year_end’). Type of estimate described in field ‘Population_metric ‘	Numeric
PE_dispersion_estimate	Estimate of dispersion or uncertainty in the population size value provided in field ‘Population_end’. Values entered here when they are provided as single estimate (e.g., SE or SD) Type of estimate described in field ‘PS_PE_dispersion_description’	Numeric
PE_dispersion_lower	Estimate of dispersion or uncertainty in the population size value provided in field ‘Population_end’. Values entered here when they are provided as a lower bounded estimate (e.g., range or confidence intervals) Type of estimate described in field ‘PS_PE_dispersion_description’	Numeric
PE_dispersion_upper	Estimate of dispersion or uncertainty in the population size value provided in field ‘Population_end’. Values entered here when they are provided as an upper bounded estimate (e.g., range or confidence intervals) Type of estimate described in field ‘PS_PE_dispersion_description’	Numeric
PS_PE_dispersion_description	Type of dispersion or uncertainty estimate(s) in population size values. Categories: *SD: Standard deviation.* *SE: Standard error.* *Range: Minimum and maximum estimates.* *90% CI: 90% confidence intervals.* *95% CI: 95% confidence intervals.* *Bayesian 90% CI: 90% credible intervals derived through Bayesian sampling.*	Categorical
Quantitative_trend	Numerical estimate of change in population size. Type of estimate described in field ‘Quantitative_method’.	Numeric
Quantitative_method	Type of population trend metric provided in field ‘Quantitative_trend’. Categories: *Manual calculation required: complete time series available in the table [timeseries.csv]. Data fall into two categories: 1) estimates of abundance at different time points. 2) Estimates of change in abundance (e.g. population lambda, or percent change) at different time points. See metadata: timeseries.csv for more detail.* *Lambda: finite rate of population change (lambda=1 represents a stable trend). Lambdas were estimated using different methods including ratio of abundance between two time intervals (N_{t+1}/N_t), different demographic models, or as the exponential of an R-trend coefficient.*	Categorical

	R-trend: instantaneous rate of population change. Values were calculated with different methods but most frequently using a log-regression model of population size (R-trend = 0 represents a stable trend). *Percentage change: change in population size between two time points (100 is stable) [formula = $(N_{t+1}/N_t) * 100$].* *Qualitative only: only a verbal description of population change was available*.	
Other_quantitative_descriptor	Additional notes and comments about the quantitative descriptor extracted during data compilation to explain less-clear cases.	Character
Dispersion_description	Type of estimate of dispersion or uncertainty provided for the population trend metric. Estimate of dispersion provided in field ‘Dispersion_estimate’. Categories: *VAR: Variance.* *SD: Standard deviation.* *SE: Standard error.* *Range: Minimum and maximum estimates.* *90% CI: 90% confidence intervals.* *95% CI: 95% confidence intervals.* *Bayesian 90% CI: 90% credible intervals derived through Bayesian sampling*.	Categorical
Dispersion_estimate	Estimate of dispersion or uncertainty for population trend (provided in field ‘Quantitative_trend field’). Type of uncertainty/dispersion described in field ‘PS_PE_dispersion_description’	Numeric
Dispersion_lower	Estimate of lower bound dispersion or uncertainty (e.g., confidence intervals or range) for population trend (provided in field ‘Quantitative_trend’). Type of uncertainty/dispersion described in field ‘Dispersion_description’	Numeric
Dispersion_upper	Estimate of upper bound dispersion or uncertainty (e.g., confidence intervals or range) for population trend (provided in field ‘Quantitative_trend’). Type of uncertainty/dispersion described in field ‘Dispersion_description’	Numeric
Significance_reported	Descriptor of whether statistical significance in population trend was tested. Categories: *NA: not reported or not relevant.* *Yes: test statistic and/or significance level reported.*	Categorical
Test_statistic	Value of the statistic (e.g. z, t, or F value) used to describe significance in population trend when available.	Numeric
Significance	P-value associated to the ‘Test_statistic’ used to describe significance in population trend when available.	Numeric
Significant_trend	Binary descriptor of whether, if statistically tested, the population trend was found to be significantly increasing or declining. Categories: *TRUE: trend was significant.* *FALSE: trend was not-significant*	Categorical
Time_lapse	Timeframe (in years) at which Quantitative_trend should be interpreted e.g. a 10-year study may describe the annual finite rate of change (lambda), as its annual the Time_lapse would equal 1. However, some lambda’s are measured at 0.5 year or 10 year scale, so the metric is used to scale the Quanittative_trend to a	Numeric

	standard time-frame. This value equals NA when the Quantitative_method is Qualitative only or a Manual trend estimate.	
Qualitative	Verbal description of population change as provided by the sources. Categories: *Increase: trend described as increasing, or recovering, or something synonymous.* *Stable: trend described as stable or exhibiting no population change, or something synonymous.* *Decrease: trend as described decreasing, declining, or reducing, or something synonymous.* *Varied: trend described as showing both increases and decreases at different time periods, but crucially, the first and the last population estimates are similar.*	Category
Other_driver_of_trend	Factors described in source as influencing population trends but which could not be captured by threat or conservation actions schema	Character
Comment	Additional notes and comments extracted during data compilation.	Character
Possible_issues	Description of issues that may limit use or interpretation of the trend e.g. author may describe the trend estimate as inaccurate.	Character
Genetic_data	Binary descriptor of whether the population trend was derived from genetic information. Categories: *1: yes* *NA: no*	Numeric-binary
Harvest_data	Binary descriptor of whether the population trend was derived from harvest information e.g. number of individuals hunted. Categories: *1: yes* *NA: no*	Numeric-binary
Invasive_species	Binary descriptor of whether the studied population was non-native to the study site. Categories: *1: yes* *NA: no*	Numeric-binary
Record_labelled_inaccurate	Binary descriptor of whether the population trend was described as inaccurate in the source. Categories: *1: yes* *NA: no*	Numeric-binary
Asymptotic_growth	Binary descriptor of whether the population trend described asymptotic or observed growth. Categories: *1: yes* *NA: no*	Numeric-binary
Metric_unusual	Binary descriptor of whether the population trend was reported in an unconventional way. Categories: *1: yes* *NA: no*	Numeric-binary
Peer_review	Binary descriptor of whether the source has been published after peer-reviewed. Categories: *1: no*	Numeric-binary

	NA: yes	
Date_missing	Binary descriptor of whether any of the date values are missing (Study_year_start or Study_year_end). Categories: *1: yes* *NA: no*	Numeric-binary
Latitude	Latitudinal centroid in decimal degrees of the study site/population	Numeric
Longitude	Longitudinal centroid in decimal degrees of the study site/population	Numeric
Source	Source of the coordinates. Categories: *Georeferenced – automatically: obtained from OpenCage georeferencer using locality name and country from the source.* *Georeferenced - manually adjusted: obtained from OpenCage georeferencer using locality name and country, but coordinates were inaccurate so were manually corrected.* *Within study - calculated centroid: Coordinates included in the source as extent ranges from which the centroid was calculated.* *Within study - reported centroid: centroid reported in the source.*	Categorical
Coordinate_comment	Process for reviewing coordinates that were georeferenced. Categories: *Checked - location is approximate: georeferenced coordinates were checked and the precise location could not be found. Coordinates approximated manually.* *Checked - Location refined: georeferenced coordinates were checked and the deemed inaccurate, so were manually adjusted.* *Checked - Original is robust: georeferenced coordinates were checked and deemed robust.* *Not checked - Record appears robust: georeferenced coordinates had high a confidence value (greater than or equal to 7) and so were not checked.* *NA – coordinates not checked as they were extracted from the primary source.*	Categorical

Table S2. Reference table for captrends.csv ‘Species’ field. Includes binomial species names for four target families (Canidae, Felidae, Hyaenidae, and Ursidea) within the order Carnivora. These species names follow the IUCN species list/taxonomy (downloaded in 2018), but we also provide comparison to the common mammalian reference taxonomy of Wilson & Reeder (2005). The comment column describes any discrepancies in these taxonomies to facilitate future dataset use.

Family	Species (IUCN)	Species (WR2005)	Comment
CANIDAE	*Atelocynus microtis*	*Atelocynus microtis*	
	Canis adustus	*Canis adustus*	
	Canis aureus	*Canis aureus*	
	Canis latrans	*Canis latrans*	
	Canis lupus	*Canis lupus*	
	Canis mesomelas	*Canis mesomelas*	
	Canis rufus	*NA*	Is a sub-species of Canis lupus in WR2005
	Canis simensis	*Canis simensis*	
	Cerdocyon thous	*Cerdocyon thous*	
	Chrysocyon brachyurus	*Chrysocyon brachyurus*	
	Cuon alpinus	*Cuon alpinus*	
	Dusicyon australis	*Dusicyon australis*	
	Dusicyon avus	*NA*	No record in WR2005
	Lycalopex culpaeus	*Lycalopex culpaeus*	
	Lycalopex fulvipes	*Lycalopex fulvipes*	
	Lycalopex griseus	*Lycalopex griseus*	
	Lycalopex gymnocercus	*Lycalopex gymnocercus*	
	Lycalopex sechurae	*Lycalopex sechurae*	
	Lycalopex vetulus	*Lycalopex vetulus*	
	Lycaon pictus	*Lycaon pictus*	
	Nyctereutes procyonoides	*Nyctereutes procyonoides*	
	Otocyon megalotis	*Otocyon megalotis*	
	Speothos venaticus	*Speothos venaticus*	
	Urocyon cinereoargenteus	*Urocyon cinereoargenteus*	
	Urocyon littoralis	*Urocyon littoralis*	
	Vulpes bengalensis	*Vulpes bengalensis*	
	Vulpes cana	*Vulpes cana*	
	Vulpes chama	*Vulpes chama*	
	Vulpes corsac	*Vulpes corsac*	
	Vulpes ferrilata	*Vulpes ferrilata*	
	Vulpes lagopus	*Vulpes lagopus*	
	Vulpes macrotis	*Vulpes macrotis*	
	Vulpes pallida	*Vulpes pallida*	
	Vulpes rueppellii	*Vulpes rueppellii*	
	Vulpes velox	*Vulpes velox*	
	Vulpes vulpes	*Vulpes vulpes*	
	Vulpes zerda	*Vulpes zerda*	
FELIDAE	*Acinonyx jubatus*	*Acinonyx jubatus*	
	Caracal aurata	*Profelis aurata*	
	Caracal caracal	*Caracal caracal*	
	Catopuma badia	*Catopuma badia*	
	Catopuma temminckii	*Catopuma temminckii*	
	Felis bieti	*Felis bieti*	
	Felis chaus	*Felis chaus*	
	Felis margarita	*Felis margarita*	

	Felis nigripes	*Felis nigripes*	
	Felis silvestris	*Felis silvestris*	
	Herpailurus yagouaroundi	*Puma yagouaroundi*	Assigned genus Puma in WR2005
	Leopardus colocolo	*Leopardus colocolo*	
	Leopardus geoffroyi	*Leopardus geoffroyi*	
	Leopardus guigna	*Leopardus guigna*	
	Leopardus guttulus	*NA*	Is a sub-species of Leopardus tigrinus in WR2005
	Leopardus jacobita	*Leopardus jacobita*	
	Leopardus pardalis	*Leopardus pardalis*	
	Leopardus tigrinus	*Leopardus tigrinus*	
	Leopardus wiedii	*Leopardus wiedii*	
	Leptailurus serval	*Leptailurus serval*	
	Lynx canadensis	*Lynx canadensis*	
	Lynx lynx	*Lynx lynx*	
	Lynx pardinus	*Lynx pardinus*	
	Lynx rufus	*Lynx rufus*	
	Neofelis diardi	*NA*	Is a sub-species of Neofelis nebulosi in WR2005
	Neofelis nebulosa	*Neofelis nebulosa*	
	Otocolobus manul	*Felis manul*	Assigned genus Felis in WR2005
	Panthera leo	*Panthera leo*	
	Panthera onca	*Panthera onca*	
	Panthera pardus	*Panthera pardus*	
	Panthera tigris	*Panthera tigris*	
	Panthera uncia	*Uncia uncia*	Assigned genus Uncia in WR2005
	Pardofelis marmorata	*Pardofelis marmorata*	
	Prionailurus bengalensis	*Prionailurus bengalensis*	
	Prionailurus planiceps	*Prionailurus planiceps*	
	Prionailurus rubiginosus	*Prionailurus rubiginosus*	
	Prionailurus viverrinus	*Prionailurus viverrinus*	
	Puma concolor	*Puma concolor*	
	NA	*Felis catus*	Not considered species by IUCN
	NA	*Leopardus braccatus*	Not considered species by IUCN
	NA	*Leopardus pajeros*	Not considered species by IUCN
	NA	*Prionailurus iriomotensis*	Not considered species by IUCN
HYAENIDAE	*Crocuta crocuta*	*Crocuta crocuta*	
	Hyaena hyaena	*Hyaena hyaena*	
	Parahyaena brunnea	*Hyaena brunnea*	Assigned genus Hyaena in WR2005
	Proteles cristata	*Proteles cristata*	
URSIDAE	*Ailuropoda melanoleuca*	*Ailuropoda melanoleuca*	
	Helarctos malayanus	*Helarctos malayanus*	
	Melursus ursinus	*Melursus ursinus*	
	Tremarctos ornatus	*Tremarctos ornatus*	
	Ursus americanus	*Ursus americanus*	
	Ursus arctos	*Ursus arctos*	
	Ursus maritimus	*Ursus maritimus*	
	Ursus thibetanus	*Ursus thibetanus*	

Table S3 Reference table for captrends.csv ‘Singular_country’ and ‘Multiple_countries’ fields. Country names follow ISO 3166 country name and two-character code standards. If sources described the global population trend, we added one row titled ‘GLOBAL’

Two-character code	Country
AF	Afghanistan
AX	Åland Islands
AL	Albania
DZ	Algeria
AS	American Samoa
AD	Andorra
AO	Angola
AI	Anguilla
AQ	Antarctica
AG	Antigua and Barbuda
AR	Argentina
AM	Armenia
AW	Aruba
AU	Australia
AT	Austria
AZ	Azerbaijan
BS	Bahamas
BH	Bahrain
BD	Bangladesh
BB	Barbados
BY	Belarus
BE	Belgium
BZ	Belize
BJ	Benin
BM	Bermuda
BT	Bhutan
BO	Bolivia, Plurinational State of
BQ	Bonaire, Sint Eustatius and Saba
BA	Bosnia and Herzegovina
BW	Botswana
BV	Bouvet Island
BR	Brazil
IO	British Indian Ocean Territory
BN	Brunei Darussalam
BG	Bulgaria
BF	Burkina Faso
BI	Burundi
KH	Cambodia
CM	Cameroon
CA	Canada
CV	Cape Verde
KY	Cayman Islands
CF	Central African Republic
TD	Chad
CL	Chile

CN	China
CX	Christmas Island
CC	Cocos (Keeling) Islands
CO	Colombia
KM	Comoros
CG	Congo
CD	Congo, The Democratic Republic of the
CK	Cook Islands
CR	Costa Rica
CI	Côte D'Ivoire
HR	Croatia
CU	Cuba
CW	Curaçao
CY	Cyprus
CZ	Czech Republic
DK	Denmark
DJ	Djibouti
DM	Dominica
DO	Dominican Republic
EC	Ecuador
EG	Egypt
SV	El Salvador
GQ	Equatorial Guinea
ER	Eritrea
EE	Estonia
ET	Ethiopia
FK	Falkland Islands (Malvinas)
FO	Faroe Islands
FJ	Fiji
FI	Finland
FR	France
GF	French Guiana
PF	French Polynesia
TF	French Southern Territories
GA	Gabon
GM	Gambia
GE	Georgia
DE	Germany
GH	Ghana
GI	Gibraltar
GLOBAL	GLOBAL
GR	Greece
GL	Greenland
GD	Grenada
GP	Guadeloupe
GU	Guam
GT	Guatemala
GG	Guernsey
GN	Guinea

GW	Guinea-Bissau
GY	Guyana
HT	Haiti
HM	Heard Island and McDonald Islands
VA	Holy See (Vatican City State)
HN	Honduras
HK	Hong Kong
HU	Hungary
IS	Iceland
IN	India
ID	Indonesia
IR	Iran, Islamic Republic of
IQ	Iraq
IE	Ireland
IM	Isle of Man
IL	Israel
IT	Italy
JM	Jamaica
JP	Japan
JE	Jersey
JO	Jordan
KZ	Kazakhstan
KE	Kenya
KI	Kiribati
KP	Korea, Democratic People's Republic of
KR	Korea, Republic of
KW	Kuwait
KG	Kyrgyzstan
LA	Lao People's Democratic Republic
LV	Latvia
LB	Lebanon
LS	Lesotho
LR	Liberia
LY	Libya
LI	Liechtenstein
LT	Lithuania
LU	Luxembourg
MO	Macao
MK	Macedonia, The former Yugoslav Republic of
MG	Madagascar
MW	Malawi
MY	Malaysia
MV	Maldives
ML	Mali
MT	Malta
MH	Marshall Islands
MQ	Martinique
MR	Mauritania
MU	Mauritius

YT	Mayotte
MX	Mexico
FM	Micronesia, Federated States of
MD	Moldova, Republic of
MC	Monaco
MN	Mongolia
ME	Montenegro
MS	Montserrat
MA	Morocco
MZ	Mozambique
MM	Myanmar
NA	Namibia
NR	Nauru
NP	Nepal
NL	Netherlands
NC	New Caledonia
NZ	New Zealand
NI	Nicaragua
NE	Niger
NG	Nigeria
NU	Niue
NF	Norfolk Island
MP	Northern Mariana Islands
NO	Norway
OM	Oman
PK	Pakistan
PW	Palau
PS	Palestine, State of
PA	Panama
PG	Papua New Guinea
PY	Paraguay
PE	Peru
PH	Philippines
PN	Pitcairn
PL	Poland
PT	Portugal
PR	Puerto Rico
QA	Qatar
RE	Reunion
RO	Romania
RU	Russian Federation
RW	Rwanda
BL	Saint Barthélemy
SH	Saint Helena, Ascension and Tristan Da Cunha
KN	Saint Kitts and Nevis
LC	Saint Lucia
MF	Saint Martin (French Part)
PM	Saint Pierre and Miquelon
VC	Saint Vincent and the Grenadines

WS	Samoa
SM	San Marino
ST	Sao Tome and Principe
SA	Saudi Arabia
SN	Senegal
RS	Serbia
SC	Seychelles
SL	Sierra Leone
SG	Singapore
SX	Sint Maarten (Dutch Part)
SK	Slovakia
SI	Slovenia
SB	Solomon Islands
SO	Somalia
ZA	South Africa
GS	South Georgia and the South Sandwich Islands
SS	South Sudan
ES	Spain
LK	Sri Lanka
SD	Sudan
SR	Suriname
SJ	Svalbard and Jan Mayen
SZ	Swaziland
SE	Sweden
CH	Switzerland
SY	Syrian Arab Republic
TW	Taiwan, Province of China
TJ	Tajikistan
TZ	Tanzania, United Republic of
TH	Thailand
TL	Timor-Leste
TG	Togo
TK	Tokelau
TO	Tonga
TT	Trinidad and Tobago
TN	Tunisia
TR	Turkey
TM	Turkmenistan
TC	Turks and Caicos Islands
TV	Tuvalu
UG	Uganda
UA	Ukraine
AE	United Arab Emirates
GB	United Kingdom
US	United States
UM	United States Minor Outlying Islands
UY	Uruguay
UZ	Uzbekistan
VU	Vanuatu

VE	Venezuela
VN	Viet Nam
VG	Virgin Islands, British
VI	Virgin Islands, U.S.
WF	Wallis and Futuna
EH	Western Sahara
YE	Yemen
ZM	Zambia
ZW	Zimbabwe

Table S4. Description of fields in the ts_abundance.csv table which provides the time series of population abundance estimates. The ‘Data type’ column describes the format of the data.

Field	Description	Data type
DataTableID	Unique numerical code for each source to match with table captrends.csv	Character
Value	Time-series value representing population abundance or density.	Numeric
Year	Time point of population abundance estimate (in years)	Numeric

Table S5. Description of fields in the ts_change.csv table which provides the time series of population change estimates. The ‘Data type’ column describes the format of the data, for categorical fields the selection options are underlined and each options description is italicised.

Field	Description	Data type
DataTableID	Unique numerical code for each source to match with table captrends.csv	Character
Type_of_measure	Descriptor of the type of estimate in the time series, values presented in the ‘Value’ field. Categories: *<u>Lambda</u>: estimate of the finite rate of population change between two time periods (represented by fields Year1 and Year2). 1 is stable* *<u>Percentage change</u>: estimate of the percentage change in population size between two time periods (represented by fields Year1 and Year2). 100 is stable [formula = (N_{t+1}/N_t) * 100].*	Categorical
Value	Time-series value representing change in populations size. Interpreted alongside the Type_of_measure, Year1, and Year2 fields	Numeric
Year1	Reference time-point (in years) e.g. date of first population estimate.	Numeric
Year2	Change time-point (in years) e.g. date of second population estimate used to derive population change.	Numeric

Table S6. Description of fields in the direction.csv table, which contains information on influences of the population trend, including whether the influence is positive or negative. This dataset uses existing classification schemes described in Table S6. The 'Data type' column describes the format of the data, for categorical fields the selection options are underlined and each options description is italicised.

Field	**Description**	**Data type**
DataTableID	Unique numerical code for each source to match with table captrends.csv	Character
Code	Amended threat or conservation action category described by the source as influencing the population trend. If an influencing factor could not be matched to a category, the driver of the trend, as described by the primary source, was entered as free text in the field 'other_drivers_ot_trend' in table captrends.csv	Categorical [calls on Table S7]
Direction	Binary descriptor of whether the factor was described by the source as potentially or actually having resulted, or being expected to result in a population increase (recorded as "Positive") or in a population decline (recorded as "Negative"). The degree of influence on the trend is informed by the Key_driver field.	Categorical
Measured	Descriptor of the evidence provided by the source to support the link between a named factor and changes in population trend. Categories: <u>Not explained</u>*: sources mentioned potentially important factor but did not provide information on how it may affect population trend* <u>Speculated:</u> *source speculated about a link between the factor and the population trend* <u>Proxy-estimate</u>*: source provided some evidence for how the factor influenced the population trend* <u>Quantified</u>*: source presented evidence that a factor has impacted the population trend*	Categorical
Key_driver	Descriptor of whether the factor was likely to be a strong driver of the observed population trend (recorded as "Positive"). Categories: <u>No:</u> *Factor not considered an important driver of the trend according to primary source.* <u>Yes</u>*: Factor considered an important driver of the trend according to primary source.*	Categorical

	<u>Unknown:</u> *Primary source did not describe impact of the factor, or described the impact as unknown.*	
Comment	Any additional notes regarding how the trend is influenced by the factor	Character

Table S7. Reference table from the ‘Code’ field in the direction.csv file. When entering influences of population trends, the most detailed code possible was used. For example, if the source describes the trend as being influenced by small-scale fragmentation, 0.2.1 (Small-scale fragmentation) was selected. However, if the source describes the trend as being influenced by fragmentation, 0.2 (General fragmentation) was selected. Each ‘Code’ also falls within a higher level ‘Category’ which aggregates codes into broader groups. For each code the ‘Scheme description’ column provide the matching classifications in the IUCN Threats (scheme v3.2) and Conservation Actions (scheme v2.0) with the scheme specified in the ‘Scheme’ column as Threats or Conservation. Some threats and actions listed by sources were not well-matched to existing scheme categories, we created new Codes which are briefly described in the ‘Scheme description’ column and labelled as Added in the ‘Scheme’ column. Some categories from the IUCN schemes were not mentioned by the reviewed sources and were not used in the database. These are indicated with a ‘-‘ in the ‘Code’ column, and their scheme name is followed by an ‘X’.

Category	Code	Scheme description	Scheme
Habitat altered	0.1 (habitat altered - not targeted restoration)	0.1 Habitat alteration but not restoration	Added
	0.1.1 (Small-scale habitat altered - not targeted restoration)	0.1.1 Small scale alteration	Added
	0.1.2 (Large-scale habitat altered - not targeted restoration)	0.1.2 Large-scale alteration	Added
	0.2 (General fragmentation)	0.2 Fragmentation	Added
	0.2.1 (Small-scale fragmentation)	0.2.1 Little fragmentation	Added
	0.2.2 (Large-scale fragmentation)	0.2.2 Large-scale fragmentation	Added
Residential & commercial development	1.1 (Habitat urbanised)	1.1 Housing & urban areas	Threats
	1.2 (Habitat industrialised)	1.2 Commercial & industrial areas	Threats
	1.3 (Habitat made available for recreation)	1.3 Tourism & recreation areas	Threats
Agriculture & aquaculture	2.1 (Habitat altered for farming)	2.1 Annual & perennial non-timber crops	Threats
	-	2.1.1 Shifting agriculture	Threats X
	2.1.2 (Habitat altered for small-scale farming)	2.1.2 Small-holder farming	Threats
	2.1.3 (Habitat altered for large-scale farming)	2.1.3 Agro-industry farming	Threats
	2.2 (Habitat altered for plantations)	2.2 Wood & pulp plantations	Threats
	2.2.1 (Habitat altered for small-scale plantations)	2.2.1 Small-holder plantations	Threats
	2.2.2 (Habitat altered for large-scale plantations)	2.2.2 Agro-industry plantations	Threats
	-	2.2.3 Scale unknown/Unrecorded	Threats X
	2.3 (Habitat altered for ranching)	2.3 Livestock farming & ranching	Threats
	2.3.1 (Habitat altered for nomadic ranching)	2.3.1 Nomadic grazing	Threats

	2.3.2 (Habitat altered for small-scale ranching)	2.3.2 Small-holder grazing, ranching or farming	Threats
	2.3.3 (Habitat altered for large-scale ranching)	2.3.3 Agro-industry grazing, ranching or farming	Threats
	-	2.3.4 Scale unknown/Unrecorded	Threats X
	-	2.4 Marine & freshwater aquaculture	Threats X
	-	2.4.1 Subsistence/artisanal aquaculture	Threats X
	-	2.4.2 Industrial aquaculture	Threats X
	-	2.4.3 Scale unknown/Unrecorded	Threats X
Energy production & mining	-	3.1 Oil & gas drilling	Threats X
	-	3.2 Mining & quarrying	Threats X
	-	3.3 Renewable energy	Threats X
Transportation & service corridors	4.1 (Road & railroads generally)	4.1 Roads & railroads	Threats
	4.1.1 (Developing roads & rails)	4.1.1 Roads & railroads development	Added
	4.1.2 (Vehicle collisions)	4.1.2 Roads & railroads vehicle collisions	Added
	-	4.2 Utility & service lines	Threats X
	-	4.3 Shipping lanes	Threats X
	-	4.4 Flight paths	Threats X
Human intrusions & disturbance	6.1 (Disturbance from recreational activities)	6.1 Recreational activities	Threats
	6.2 (Disturbance from war)	6.2 War, civil unrest & military exercises	Threats
	6.3 (Disturbance from people working)	6.3 Work & other activities	Threats
	-	6.4 Other disturbance	Threats X
Natural system modifications	7.1 (System altered by excess fire)	7.1 Fire & fire suppression	Threats
	7.1.1 (System altered by fire shortage)	7.1.1 Increase in fire frequency/intensity	Threats
	-	7.1.2 Suppression in fire frequency/intensity	Threats X
	-	7.1.3 Trend unknown/Unrecorded	Threats X
	7.2 (System altered by water shortage/dams)	7.2 Dams & water management/use	Threats
	7.3 (System altered by ecosystem modifications)	7.3 Other ecosystem modifications	Threats
Invasive & other problematic species, genes & diseases	8.1 (Population effected by invasive disease)	8.1 Invasive non-native/alien species/diseases	Threats
	-	8.1.1 Unspecified species	Threats X
	-	8.1.2 Named species	Threats X
	8.2 (Population effected by native disease)	8.2 Problematic native species/diseases	Threats
	-	8.2.1 Unspecified species	Threats X

	-	8.2.2 Named species	Threats X
	8.3 (Population effected by introduced genes)	8.3 Introduced genetic material	Threats
	-	8.4 Problematic species/diseases of unknown origin	Threats X
	-	8.4.1 Unspecified species	Threats X
	-	8.4.2 Named species	Threats X
	-	8.5 Viral/prion-induced diseases	Threats X
	-	8.5.1 Unspecified species	Threats X
	-	8.5.2 Named species	Threats X
	8.6 (Population effected by unknown disease)	8.6 Diseases of unknown cause	Threats
	-	8.7 General disease	Threats X
Pollution	9.1 (Population/Habitat effected by domestic waste)	9.1 Domestic & urban waste water	Threats
	9.2 (Population/Habitat effected by industrial waste)	9.2 Industrial & military effluents	Threats
	-	9.3 Agricultural & forestry effluents	Threats X
	-	9.4 Garbage & solid waste	Threats X
	-	9.5 Air-borne pollutants	Threats X
	-	9.6 Excess energy	Threats X
	9.6.1 (Population/Habitat effected by light pollution)	9.6.1 Light pollution	Threats
	9.6.2 (Population/Habitat effected by thermal pollution)	9.6.2 Thermal pollution	Threats
	9.6.3 (Population/Habitat effected by noise pollution)	9.6.3 Noise pollution	Threats
	-	9.6.4 Type unknown/unrecorded	Threats X
Geological events	10.1 (Population/Habitat effected by volcanoes)	10.1 Volcanoes	Threats
	10.2 (Population/Habitat effected by earthquakes/tsunamis)	10.2 Earthquakes/tsunamis	Threats
	10.3 (Population/Habitat effected by avalanches/landslides)	10.3 Avalanches/landslides	Threats
Climate change & severe weather	11.1 (Habitat shifts from climate change)	11.1 Habitat shifting & alteration	Threats
	11.2 (Population/Habitat effected by drought)	11.2 Droughts	Threats
	11.3 (Population/Habitat effected by temperature extremes)	11.3 Temperature extremes	Threats
	11.4 (Population/Habitat effected by storms/flooding)	11.4 Storms & flooding	Threats
	11.5 (Population/Habitat effected by unspecified climate change)	11.5 Other impacts	Threats

Biological resource use – adapted from section 5 of threats v3.2 to make all actions legal	13.1 (General legal hunting)	5.1 Hunting & collecting terrestrial animals	Threats
	13.1.1 (Legal hunting of carnivore)	5.1.1 Intentional use (species being assessed is the target)	Threats
	13.1.2 (Indirect effect from legal hunting)	5.1.2 Unintentional effects (species being assessed is not the target)	Threats
	13.1.3 (Legal persecution/control of carnivore)	5.1.3 Persecution/control	Threats
	-	5.1.4 Motivation unknown/Unrecorded	Threats X
	13.2 (Indirect effect of gathering plants)	5.2 Gathering terrestrial plants	Threats
	-	5.2.1 Intentional use (species being assessed is the target)	Threats X
	-	5.2.2 Unintentional effects (species being assessed is not the target)	Threats X
	-	5.2.3 Persecution/control	Threats X
	-	5.2.4 Motivation unknown/Unrecorded	Threats X
	-	5.3 Logging & wood harvesting	Threats X
	-	5.3.1 Intentional use: subsistence/small scale (species being assessed is the target [harvest]	Threats X
	-	5.3.2 Intentional use: large scale (species being assessed is the target)[harvest]	Threats X
	-	5.3.3 Unintentional effects: subsistence/small scale (species being assessed is not the target)[harvest]	Threats X
	-	5.3.4 Unintentional effects: large scale (species being assessed is not the target)[harvest]	Threats X
	-	5.3.5 Motivation unknown/Unrecorded	Threats X
	13.4 (Indirect effect of fishing)	5.4 Fishing & harvesting aquatic resources	Threats
	-	5.4.1 Intentional use: subsistence/small scale (species being assessed is the target)[harvest]	Threats X
	-	5.4.2 Intentional use: large scale (species being assessed is the target)[harvest]	Threats X
	13.4.3 (Indirect effect of small-scale fishing)	5.4.3 Unintentional effects: subsistence/small scale (species being assessed is not the target) [harvest]	Threats
	13.4.4 (Indirect effect of large-scale fishing)	5.4.4 Unintentional effects: large scale (species being assessed is not the target) [harvest]	Threats
	-	5.4.5 Persecution/control	Threats X
	-	5.4.6 Motivation unknown/Unrecorded	Threats X
	13.5 (Legal poisoning of carnivore)	5.5 Poisoning	Added
	13.5.1 (Legal targeting poison towards carnivore)	5.5.1 Intentional use (species being assessed is the target)	Added
	13.5.2 (Legal indirect poison of carnivore)	5.5.2 Unintentional effects (species being assessed is not the target)	Added
	14.1 (General illegal hunting/poaching)	5.1 Hunting & collecting terrestrial animals	Threats

Biological resource use – adapted from section 5 of threats v3.2 to make all actions illegal	14.1.1 (Illegal hunting of carnivore/poaching)	5.1.1 Intentional use (species being assessed is the target)	Threats
	14.1.2 (Indirect effect from illegal hunting/poaching)	5.1.2 Unintentional effects (species being assessed is not the target)	Threats
	14.1.3 (Illegal persecution/control of carnivore)	5.1.3 Persecution/control	Threats
	-	5.1.4 Motivation unknown/Unrecorded	Threats X
	14.2 (Indirect effect of gathering plants)	5.2 Gathering terrestrial plants	Threats
	-	5.2.1 Intentional use (species being assessed is the target)	Threats X
	-	5.2.2 Unintentional effects (species being assessed is not the target)	Threats X
	-	5.2.3 Persecution/control	Threats X
	-	5.2.4 Motivation unknown/Unrecorded	Threats X
	-	5.3 Logging & wood harvesting	Threats X
	-	5.3.1 Intentional use: subsistence/small scale (species being assessed is the target [harvest]	Threats X
	-	5.3.2 Intentional use: large scale (species being assessed is the target)[harvest]	Threats X
	-	5.3.3 Unintentional effects: subsistence/small scale (species being assessed is not the target)[harvest]	Threats X
	-	5.3.4 Unintentional effects: large scale (species being assessed is not the target)[harvest]	Threats X
	-	5.3.5 Motivation unknown/Unrecorded	Threats X
	14.4 (Indirect effect of illegal fishing)	5.4 Fishing & harvesting aquatic resources	Threats
	-	5.4.1 Intentional use: subsistence/small scale (species being assessed is the target)[harvest]	Threats X
	-	5.4.2 Intentional use: large scale (species being assessed is the target)[harvest]	Threats X
	14.4.3 (Indirect effect of small-scale illegal fishing)	5.4.3 Unintentional effects: subsistence/small scale (species being assessed is not the target) [harvest]	Threats
	14.4.4 (Indirect effect of large-scale illegal fishing)	5.4.4 Unintentional effects: large scale (species being assessed is not the target) [harvest]	Threats
	-	5.4.5 Persecution/control	Threats X
	-	5.4.6 Motivation unknown/Unrecorded	Threats X
	14.5 (Illegal poisoning of carnivore)	5.5 Poisoning	Added
	14.5.1 (Illegal targeting poison towards carnivore)	5.5.1 Intentional use (species being assessed is the target)	Added
	14.5.2 (Illegal indirect poison of carnivore)	5.5.2 Unintentional effects (species being assessed is not the target)	Added
	15.1 (General competition)	15.1 Competition	Added

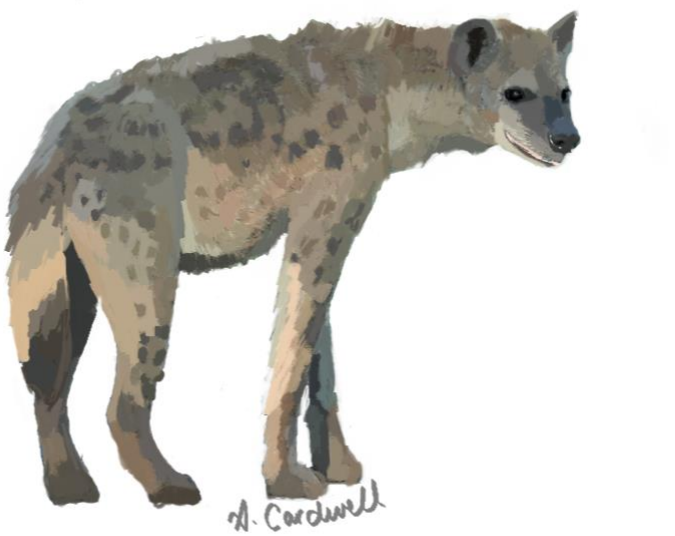
Biological population drivers	15.1.1 (Low competition - incomplete guild)	15.1.1 Low inter-specific competition – incomplete carnivore guild	Added
	15.1.2 (Low competition - no reason)	15.1.2 Low competition (reason not described)	Added
	15.1.3 (High competition - within guild)	15.1.3 High inter-specific competition within guild	Added
	15.1.5 (High competition - no reason)	15.1.5 High competition (reason not described)	Added
	15.1.6 (Low prey availability)	15.1.6 Low prey availability	Added
	15.1.7 (High prey availability)	15.1.7 High prey availability	Added
	15.1.8 (Competition from invasive species)	15.1.8 Invasive non-native/alien species/diseases	Added
	15.2 (Carnivore predated)	15.2 Predation	Added
	15.2.1 (Low predation risk - unbalanced guild)	15.2.1 Low predation risk – unbalanced guild	Added
	15.2.2 (Low predation risk - no reason)	15.2.2 Low predation risk (reason not described	Added
	15.2.3 (High predation risk - unbalanced guild)	15.2.3 High predation risk – unbalanced guild	Added
	15.2.4 (High predation risk - no reason)	15.2.4 High predation risk (reason not described)	Added
	15.2.5 (Predation from invasive species)	15.2.5 Invasive effects predation	Added
	15.3 (Below Minimum Viable Population)	15.3 Population at minimum level	Added
	15.4 (High immigration/emigration)	15.4 Population open	Added
	15.4.1 (High emigration)	15.4.1 Emigration out of population	Added
	15.4.2 (High immigration)	15.4.2 Immigration into population present	Added
	15.4.3 (Population expanding/recolonising areas)	15.4.3 Range expansion/natural recolonization.	Added
	15.5 (Population closed/isolated)	15.5 Population closed	Added
	15.5.1 (Low connectivity in population)	15.5.1 Low connectivity	Added
	15.5.2 (Low connectivity - Inbreeding possible)	15.5.2 Low connectivity – inbreeding possible	Added
	15.5.3 (Low connectivity - Inbreeding present)	15.5.3 Low connectivity – inbreeding present	Added
	15.5.4 (No connectivity in population)	15.5.4 No connectivity	Added
	15.5.5 (No connectivity - Inbreeding possible)	15.5.5 No connectivity – inbreeding possible	Added
	15.5.6 (No connectivity - Inbreeding present)	15.5.6 No connectivity – inbreeding present	Added
	15.5.7 (Unspecified genetic threat)	15.5.7 Unspecified genetic threat	Added
Land/water protection	1.1 (Protected area)	1.1 Site/area protection	Conservation
	1.2 (Protected habitat)	1.2 Resource & habitat protection	Conservation
	1.3 (Habitat developed over - considering sustainability)	1.3 Resource & habitat protection	Conservation

Land/water management	2.1 (Site managed)	2.1 Site/area management	Conservation
	2.2 (Control of problematic species)	2.2 Invasive/problematic species control	Conservation
	2.3 (Habitat restoration)	2.3 Habitat & natural process restoration	Conservation
Species management	3.1 (Species managed)	3.1 Species management	Conservation
	3.1.1 (Harvest managed)	3.1.1 Harvest management	Conservation
	3.1.2 (Trade managed)	3.1.2 Trade management	Conservation
	3.1.3 (Population growth managed - culling)	3.1.3 Limiting population growth	Conservation
	3.2 (Action to enable population recovery)	3.2 Species recovery	Conservation
	3.3 (Re-introduction)	3.3 Species re-introduction	Conservation
	-	3.3.1 Reintroduction	Conservation X
	3.3.2 (Benign introduction)	3.3.2 Benign introduction	Conservation
	3.4 (Ex-situ - captive breeding/artificial propagation)	3.4 Ex-situ conservation	Conservation
	-	3.4.1 Captive breeding/artificial propagation	Conservation X
	-	3.4.2 Genome resource bank	Conservation X
Education & awareness	4.1 (Formal education)	4.1 Formal education	Conservation
	4.2 (Train practitioners)	4.2 Training	Conservation
	4.3 (Educate public)	4.3 Awareness & communications	Conservation
Law & policy	5.1 (General protective legislation)	5.1 Legislation	Conservation
	5.1.1 (International legislation)	5.1.1 International level	Conservation
	5.1.2 (National legislation)	5.1.2 National level	Conservation
	5.1.3 (Regional legislation)	5.1.3 Sub-national level	Conservation
	-	5.1.4 Scale unspecified	Conservation X
	5.2 (Protective policy)	5.2 Policies and regulations	Conservation
	-	5.3 Private Sector Standards & Codes	Conservation X
	5.4 (Enforcing general policy/legislation)	5.4 Compliance and enforcement	Conservation
	5.4.1 (Enforcing international policy/legislation)	5.4.1 International level	Conservation
	5.4.2 (Enforcing National policy/legislation)	5.4.2 National level	Conservation
	5.4.3 (Enforcing Regional policy/legislation)	5.4.3 Sub-national level	Conservation
	-	5.4.4 Scale unspecified	Conservation X
Livelihood, economic & other incentives	6.1 (Communities livelihood linked to species success))	6.1 Linked enterprises & livelihood alternatives	Conservation
	6.2 (Substitute carnivore for sustainable alternative)	6.2 Substitution	Conservation

6.3 (Carnivore market managed e.g. hunting levy)	6.3 Market forces	Conservation
6.4 (Compensation schemes)	6.4 Conservation payments	Conservation
6.5 (Utilise spiritual/religious connections for management)	6.5 Non-monetary values	Conservation

Table S8. Description of fields in the sources.csv table, which contains information on all reviewed sources. The 'Data type' column describes the format of the data, for categorical fields the selection options are underlined and each options description is italicised.

Field	**Description**	**Data type**
Year	Year of source publication	Numeric
Title	Title of source	Character
Citation_key	Unique alphanumerical identifier for each source which corresponds to captrends.csv	Character
Category	Category describing how the source was processed. Categories: *Read – Data available: Population trend information was available within the source and extracted* *Read – No Data available: Population trend information was unavailable within the source* *NA: Source could not be accessed to assess if trend information was available.*	Categorical
From_Syst/Unsyst_search	Descriptor of how the source was found. Categories: *1: Source identified through the unstructured search* *2: Source identified through the structured search* *3: Source identified through other sources (e.g. when reading category 1 or 2 sources, which mentioned other population trend values)*	Categorical
Comment	Any additional information regarding the source e.g. why trend data was not extracted.	Character


A. Cardwell

Chapter 4: A global assessment of large carnivores

Authors: Thomas Frederick Johnson[1], Nick J. B. Isaac[2], Agustin Paviolo[3,4], Manuela González-Suárez[1]

Corresponding author: Thomas Frederick Johnson (Thomas.frederick.johnson@outlook.com)

1: Ecology and Evolutionary Biology, School of Biological Sciences, University of Reading, Reading, UK

2: Biodiversity Science Area, Centre for Ecology and Hydrology, Wallingford, OX10 8BB, UK,

3: Instituto de Biología Subtropical, CONICET-Universidad Nacional de Misiones, Bertoni 85, (N3370AIA) Puerto Iguazú, Misiones, Argentina

4: Asociación Civil Centro de Investigaciones del Bosque Atlántico, Bertoni 85, (N3370AIA) Puerto Iguazú, Misiones, Argentina

Author contribution statement: TFJ and MGS conceived the idea, with support from NI. TFJ led the analysis with support from MGS and NI. TFJ wrote the first draft, with critique from all authors.

Abstract

Global assessments of biodiversity change have already identified that stressors like land-use and climate change are linked to wildlife population declines. Yet research into features that may promote population recoveries, and will help bend the biodiversity curve, have been relatively neglected. Here, focussing on 87 terrestrial species from the order Carnivora, we explore how a wide variety of features influence population declines and recoveries across the planet and further forecast changes by 2050. Specifically, we investigate how impacts can be magnified or mitigated depending on species traits, the quality of national-level governance, and the interactions between drivers. We show that populations decline when primary land is lost, regardless of what the land is converted into or the species niche breadth. Climate change impacts are more complex, with detrimental effects only for large species and populations outside protected areas. We also find that societal growth, through human development, is linked to sharp population declines, but once human development is high, carnivore populations have the potential to recover. Considering anticipated changes in human development and environmental stressors, we then predict species abundances and the status of the carnivore guild up to 2050. These predictions offer a glimpse of hope forecasting likely recoveries of many of these culturally and ecologically important large carnivore.

Main text

Humans are transforming the planet, driving species to extinction and altering ecosystems (IPBES, 2019) - pushing biodiversity closer to its planetary limit (Newbold et al., 2016). However, whilst extinction rates suggest biodiversity is extremely threatened, there is a lack of consensus across the biodiversity change literature, with recent work showing that many populations are not declining (Daskalova et al., 2020; Leung et al., 2020; van Klink et al., 2020). Instead, environmental change causes a flux in wildlife population trends across ecological communities, with some populations declining whilst others prosper (Dornelas et al., 2019); reshaping community structure and possibly altering ecosystem functions (Zavaleta et al., 2010; Hautier et al., 2018). One challenges for understanding biodiversity change, is to identify which species will decline under environmental change, and which will prosper. This is not an easy task as data describing population trends are often noisy, and the factors that influence population dynamics can be numerous and difficult to measure (WWF, 2020b). This, in-part, explains the low predictive accuracy of macro-scale biodiversity change models e.g. Soroye et al., (2020) & Spooner et al., (2018).

Previous work has focussed on understanding how land-use change (Newbold et al., 2015; Polaina et al., 2019), forest-loss (Daskalova et al., 2020), climate change (Spooner et al., 2018; Soroye et al., 2020; Trisos et al., 2020), and in some cases a combination of all of these, influences biodiversity change. However, focussing primarily on environmental change ignores other factors that can mitigate or magnify environmental change impacts. For example, populations are more likely to increase in areas with strong governance (Amano et al., 2018), and population declines can be exacerbated by armed conflicts and species life-history traits (Braga-Pereira et al., 2020). To effectively detect the signal of environmental change impacts, it is important to consider the multidimensional context and diversity of potential factors that influence population dynamics. Here, we take this comprehensive approach to explore how land-use change, climate change, and governance, impact population trends of large terrestrial carnivores globally. We also explore how species life-history traits can mitigate or magnify these trends, specifically focussing on species from the families Canidae, Felidae, Hyaenidae and Ursidae of the order Carnivora, which include the largest terrestrial carnivores on the planet.

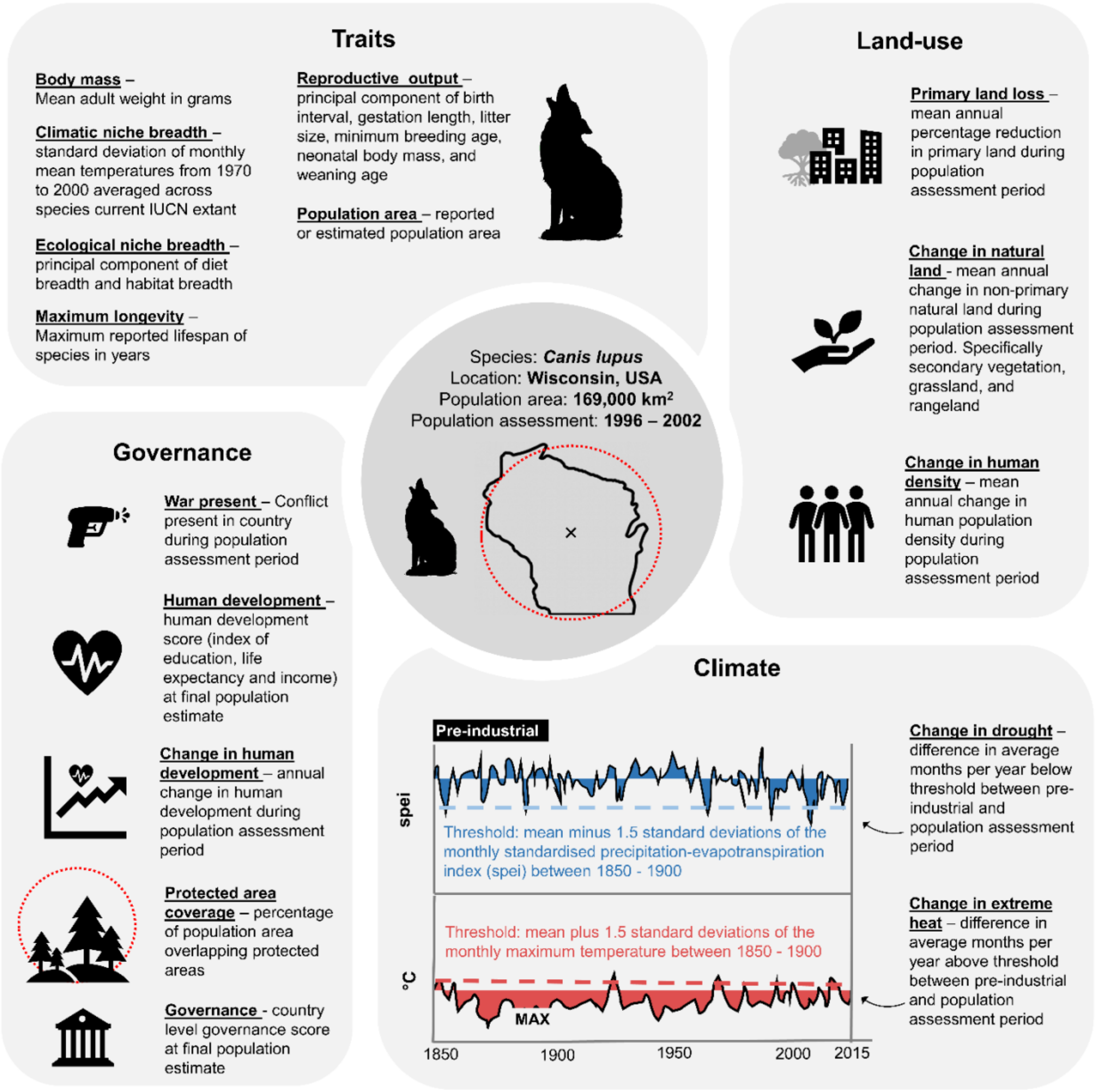
Large carnivores (such as lions, tigers, and wolves) are an important focal group to study as they are amongst the most culturally important fauna (Clucas et al., 2008), are essential for regulating ecosystem function (Ripple et al., 2014), and can act as indicator species of the overall status of biodiversity within an area (Sergio et al., 2008b). Large carnivores are also generally well-studied taxa with abundant trend datasets available from the primary literature across a wide spectrum of environmental change and governance scenarios. The morphology, ecology and behaviour of these taxa is also generally well described (González-Suárez et al., 2012), allowing us to evaluate these factors without being impacted by poor inference from missing data (Johnson et al., 2021b). Finally, despite being well-studied, the population status of large carnivores is unclear, with reports of devastating declines (Ripple et al., 2014) contrasted with remarkable recoveries (Chapron et al., 2014). Studying these differences in responses can provide insight beyond these taxa, revealing strategies and scenarios that could help bend the biodiversity curve (Leclère et al., 2020).

Influences of population change

To determine how land-use change, climate change, and governance influence population trends in large carnivores, we developed a censored weighted hierarchical Bayesian linear model (see Supplementary methods). By using a censored response-term in the model, we incorporated two types of response data: 1) quantitative estimates of change which be converted into annual rates of change, and 2) qualitative descriptions of change which we treated as censored (partially known) rates of change. By including these qualitative records, we increased the sample size, as well as the taxonomic and spatial representativeness of the trend data. In total, we compiled trends for 1127 populations, sourced from 7352 abundance estimates compiled in the CaPTrends (Johnson et al., 2021a) and Living Planet databases (WWF, 2020a). Rates of change were available for 50 of 87 species in our focal group, with locations representing 75 countries, all human-inhabited continents, and variable time periods between 1970 – 2015. For each rate of change, we estimated several metrics describing land-use, climate, and governance features, each of which we expected to influence population trends (Figure 1). Models included sixteen covariates, as well as six interactions, which we specified as we anticipated the impacts of environmental change could depend on species traits (González-Suárez & Revilla, 2013), the quality of national-level governance (Amano et al., 2018), and the interactions between different types of environmental change. For example, specialist species are likely to experience greater declines under land-use and climate change (Pacifici et al., 2017).

Land-use

As predicted and shown in other taxa (Newbold et al., 2015), we found that large carnivore populations declined when primary land was lost. However, we expected declines to be more extreme when land was replaced by people, relatively mitigated when replaced by semi-natural land, and highly dependent on the species ecological niche breadth (Figure 2), but we found no evidence supporting these interactions. Our findings suggest that all populations will decline in the immediate aftermath of primary land loss, regardless of the species ecological niche breadth and what the land is replaced with. However, these factors could influence responses in the longer term, e.g. a lag may occur before new generalist species are able to colonise an area and show increasing population trends.



Body mass – Mean adult weight in grams

Climatic niche breadth – standard deviation of monthly mean temperatures from 1970 to 2000 averaged across species current IUCN extant

Ecological niche breadth – principal component of diet breadth and habitat breadth

Maximum longevity – Maximum reported lifespan of species in years

Reproductive output – principal component of birth interval, gestation length, litter size, minimum breeding age, neonatal body mass, and weaning age

Population area – reported or estimated population area

Primary land loss – mean annual percentage reduction in primary land during population assessment period

Change in natural land - mean annual change in non-primary natural land during population assessment period. Specifically secondary vegetation, grassland, and rangeland

Change in human density – mean annual change in human population density during population assessment period

Species: ***Canis lupus***
Location: **Wisconsin, USA**
Population area: **169,000 km^2**
Population assessment: **1996 – 2002**

War present – Conflict present in country during population assessment period

Human development – human development score (index of education, life expectancy and income) at final population estimate

Change in human development – annual change in human development during population assessment period

Protected area coverage – percentage of population area overlapping protected areas

Governance - country level governance score at final population estimate

Threshold: mean minus 1.5 standard deviations of the monthly standardised precipitation-evapotranspiration index (spei) between 1850 - 1900

Threshold: mean plus 1.5 standard deviations of the monthly maximum temperature between 1850 - 1900

Change in drought – difference in average months per year below threshold between pre-industrial and population assessment period

Change in extreme heat – difference in average months per year above threshold between pre-industrial and population assessment period

Figure 1. Sixteen covariates with a proposed effect on carnivore population trends highlighted in bold and underlined. Covariates fall in four groups: Traits, Land-use, Climate, and Governance. Text alongside covariates briefly explains how the variable was derived, whilst full explanation and justifications for inclusion are available in Supplementary methods: Covariates.

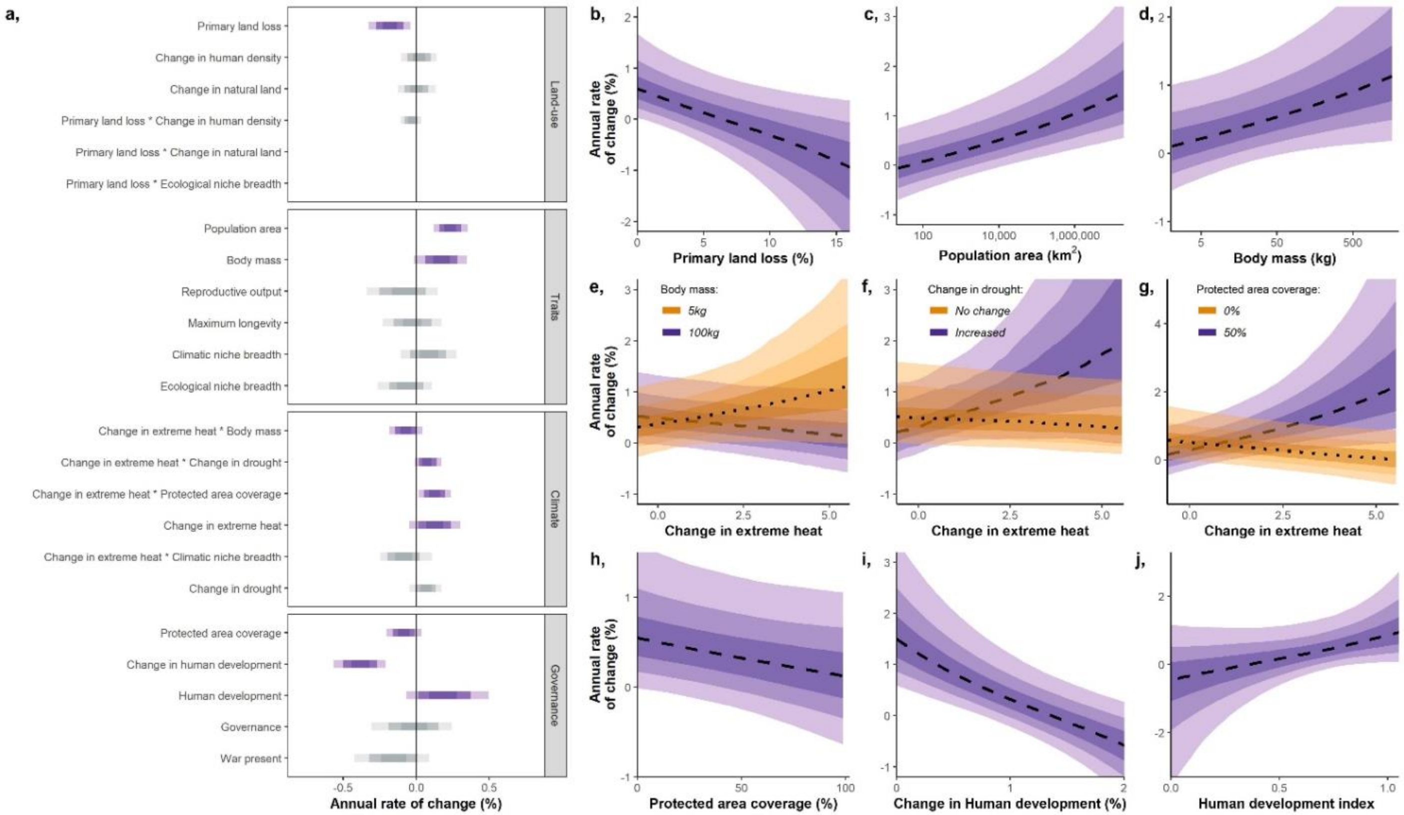


Figure 2. a) Annual rate of change coefficients for fixed effect parameters in a hierarchical Bayesian regression, with 50%, 80%, and 95% credible intervals. Effects for two primary land loss interaction parameters are not shown as they occurred in too few models (via model selection) – see Supplementary methods. All covariates were z-transformed prior to modelling and the response was back-scaled from an inverse hyperbolic sine transformation. b-j) Marginal effects for relevant covariates showing median predicted annual rate of change (error ribbons represent the 50%, 80%, and 95% credible intervals: mean annual primary land loss over the population monitoring period (b); area of population buffer zone on the $\log_{10}$ scale (c); species body mass on the $\log_{10}$ scale (d); mean number of months per year where the average temperature of the population monitoring period exceeded the mean plus 1.5 standard deviations of the average temperature of the baseline period (1901 – 1920), interacting with body mass (e); protected area coverage (g); and change in drought (f). Change in drought is measured as the mean number of months per year where the average standardised precipitation-evapotranspiration index (spei) of the population monitoring period was lower than the mean minus 1.5 standard deviations of the average spei of the baseline period, ‘No change’ equals zero months and ‘Increased’ equals two months. We also depict the marginal effect of: protected area coverage as a non-interactive effect (h); annual change in human development over the population monitoring period (i); and human development at the final year of population monitoring (j). All covariates were back scaled from any transformations.

Climate change

Larger bodied species exhibited greater population growth (Figure 2). The largest carnivores have historically been, and continue to be, persecuted and feared for carrying a threat to human lives and livelihoods (Broekhuis et al., 2020). Yet, these larger species are also amongst the highest profile taxa (Macdonald et al., 2017), likely receiving more conservation funds and attention which may explain the greater population growth. Similarly, public interest in wildlife is increasing (Millard et al., 2021), and so its plausible that people may now be more tolerant towards these carnivores, and willing to coexist. This change in perceptions could (at least partly) explain the population recoveries of the largest carnivores. However, our analyses reveal a potential increasing threat to these species, as the rising frequency of extreme heat from climate change is associated with population declines in large species (and increased growth in smaller carnivores). Prior work has shown that large-bodied birds may be more susceptible to rising temperatures from climate change (Prokosch et al., 2019), and if this result holds true for other vertebrates, it may further threaten large flagship megafauna (e.g. polar bear, tiger and lion). However, this was not our only complex climate change result; we also detected a positive synergistic effect of heat and drought, with populations showing higher growth in the presence of increased extreme heat and drought. This may at first appear surprising, but we hypothesize it reflects a trophic effect. Drier and hotter conditions could reduce plant biomass (Eziz et al., 2017), and in turn, reduce the fitness of the herbivorous species (Duncan et al., 2012) that large carnivores often prey on. This decline in prey fitness could be advantageous for large-predatory carnivores leading to increased hunting success. However, such benefits would likely be short-term; if climate change continues to impact prey fitness and densities, carnivore populations would eventually crash. Extreme heat and drought events have become increasingly common in recent years, so our data may have captured that initial stage of large carnivores benefitting from declines in prey fitness.

Our results also show another complex synergy between climate change and conservation measures, with protected areas potentially mitigating the impacts of extreme heat. As protected areas are amongst the least impacted fragments of land on the planet, they may naturally offer features that buffer extreme temperatures e.g. micro-climates from canopy cover (Suggitt et al., 2018; Davis et al., 2019). Increased population densities could also reflect increased use and movement towards these protected areas from less suitable habitat in the short term These results support previous work on birds (Lehikoinen et al., 2019) and suggest the expansion of protected areas could be an effective approach to support future carnivore populations in the face of climate change. Although, the size of the protected area is an important consideration here; if individuals do seek refuge in protected areas, we would expect wildlife population increases to be greater in smaller protected areas– our data is unsuitable for exploring this nuance

Generally, protected areas cannot promote population growth in the long term, as the area and resources inside them are finite. In fact, our results show overall stable populations in fully protected areas with population likely at carrying capacity, and more positive growth rates outside. These population increases outside protected areas could reflect spill-over effects (Di Lorenzo et al., 2020), or the ability of some carnivore species to adapt to humanized environments (Santini et al., 2019). Our work reveals interesting and complex relationships between climate variables and habitat protection that need to be further investigated in other taxa and monitored over time to detect and respond to changes.

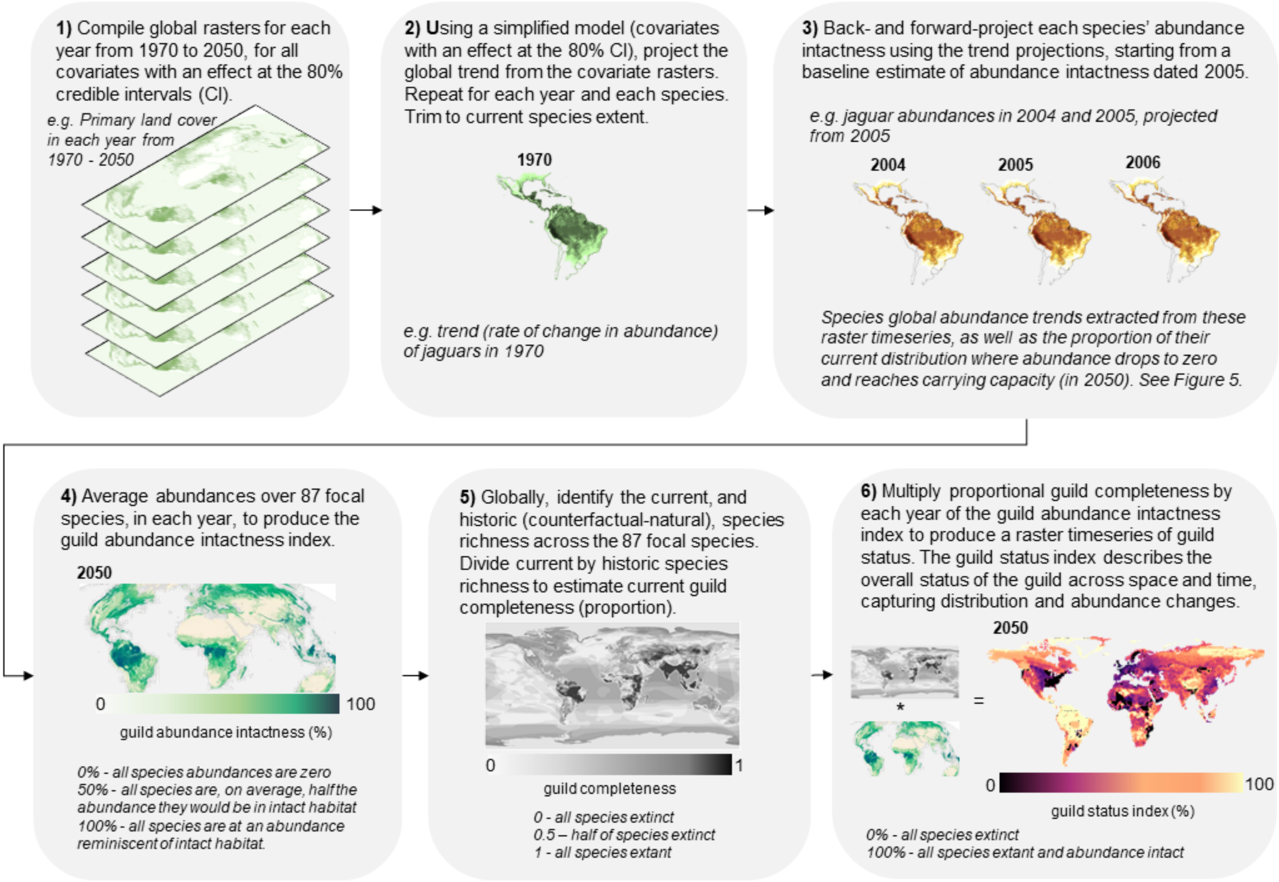
Governance

Beyond land protection, we found no effect of governance on carnivore populations, although populations were more likely to increase in areas with high human development scores (Figure 2). Human characteristics associated with development (i.e. improved quality of human life) appear to be more important for large carnivores, than higher-level governance characteristics (i.e. rule of law and legislation). This finding differs from previous work that has suggested declines in birds are greatest in areas with weak governance (Amano et al., 2018) and has partly attributed the recoveries of Europe’s carnivores to the strong governance of the European Union (Chapron et al., 2014). Wildlife can receive regulatory protection through governance, but protection may also be achieved by improving human quality of life, and in turn, tolerance for wildlife. For example, if a carnivore kills livestock, tolerance would be likely lower if the owner lives in extreme poverty (low human development) and more likely to lead to retaliation and human-wildlife conflict regardless of the species regulatory protection.

Whilst carnivore population increases are most likely in areas with high human development, there is a cost to reaching this development level, as we found rapid human development growth was associated with carnivore population declines (Figure 2). While initially we may think this relationship could reflect underlying factors stimulating human development growth (e.g. detrimental land-use changes, or natural resource extraction for economic growth), these factors were directly modelled and were less clearly associated (despite being measured at more relevant spatial-scales). We hypothesize human development change captures a more holistic snapshot of environmental and societal transformation, including underlying factors as well as changes in culture and relationships with wildlife. For example, development in Kenya has seen increases in urbanisation and infrastructure as well as changes in people's relationships with nature (Fernández-Llamazares et al., 2020). While we cannot establish the mechanisms by which rapid human development can lead to population declines, a trade-off between healthy carnivore populations and human development could lead to conflicts for the achievement of the UN sustainable development goals in developing countries. For example, improvements in health, education and equality of income (SDG 1-5) could negatively impact large carnivores (and biodiversity as a whole), hindering progress towards SDG 15.

Projecting the status of large carnivores

We produced global projections describing changes in species abundance intactness and overall carnivore guild status from 1970 to 2050 (Figure 3). To calculate species abundance intactness, we first generated spatially-explicit timeseries of annual rates of population change for each species across their current distribution (IUCN, 2020b) using a simplified version of our regression model (including effects at the 80% credible interval) and raster timeseries' of covariates. Using these change rates and an estimate of 2005 baseline abundance intactness (Newbold et al., 2016), we then calculated for each species the spatially-explicit and total (average) estimates of yearly abundance relative to that expected in intact habitat (100% equals abundance reminiscent of intact areas which we consider a proxy for carrying capacity hereafter, 0% indicates local extinction). These values were used to estimate historic (1970-2020) and future (2020-2050) annual rates of change in abundance intactness for each species. Second, we calculated guild abundance intactness and completeness estimates that combined describe the overall carnivoran carnivore guild status over space and time, capturing both distribution and abundance changes. To calculate guild abundance intactness we aggregated the spatially-explicit abundance intactness of all species co-occurring in each location. Guild completeness was calculated as the proportion of currently extant species relative to extant and recently, within the past few hundred years, extinct species (Faurby et al., 2018).



1) Compile global rasters for each year from 1970 to 2050, for all covariates with an effect at the 80% credible intervals (CI).

e.g. Primary land cover in each year from 1970 - 2050

2) Using a simplified model (covariates with an effect at the 80% CI), project the global trend from the covariate rasters. Repeat for each year and each species. Trim to current species extent.

e.g. trend (rate of change in abundance) of jaguars in 1970

3) Back- and forward-project each species' abundance intactness using the trend projections, starting from a baseline estimate of abundance intactness dated 2005.

e.g. jaguar abundances in 2004 and 2005, projected from 2005

Species global abundance trends extracted from these raster timeseries, as well as the proportion of their current distribution where abundance drops to zero and reaches carrying capacity (in 2050). See Figure 5.

4) Average abundances over 87 focal species, in each year, to produce the guild abundance intactness index.

0% - all species abundances are zero
50% - all species are, on average, half the abundance they would be in intact habitat
100% - all species are at an abundance reminiscent of intact habitat.

5) Globally, identify the current, and historic (counterfactual-natural), species richness across the 87 focal species. Divide current by historic species richness to estimate current guild completeness (proportion).

0 - all species extinct
0.5 – half of species extinct
1 - all species extant

6) Multiply proportional guild completeness by each year of the guild abundance intactness index to produce a raster timeseries of guild status. The guild status index describes the overall status of the guild across space and time, capturing distribution and abundance changes.

0% - all species extinct
100% - all species extant and abundance intact

Figure 3. Procedure for estimating species and guild abundance intactness, as well as guild status. Models were developed using a simplified form of the regression model (see Figure 2), as well as the baseline abundance intactness from 2005 (Newbold et al., 2016). Current species richness was estimated using IUCN current range maps (IUCN, 2020b), and historic species richness was derived from counterfactual-natural ranges (Faurby et al., 2018). See the supplementary material for a more detailed description.

Species abundance intactness

Abundance intactness varied substantially across time and the phylogeny (Figure 4a). In the historical period (1970 – 2020), the average abundance intactness (across species) increased (Figure 4b), with a mean annual rate of change of 0.27%, and values for the 85 non-extinct species ranging from -1.17% to 2.05%. Thirty species were projected to have declined over this period, with others stable or increasing, a balance of winners and losers – as found in other taxa (Dornelas et al., 2019). In the future (2020 to 2050) average abundance intactness was projected to stabilise with a mean annual rate of change of 0.18%, and species values ranging from -0.66% to 0.78%. Stabilisation occurred as populations approached carrying capacity across their range (Figure 4c).

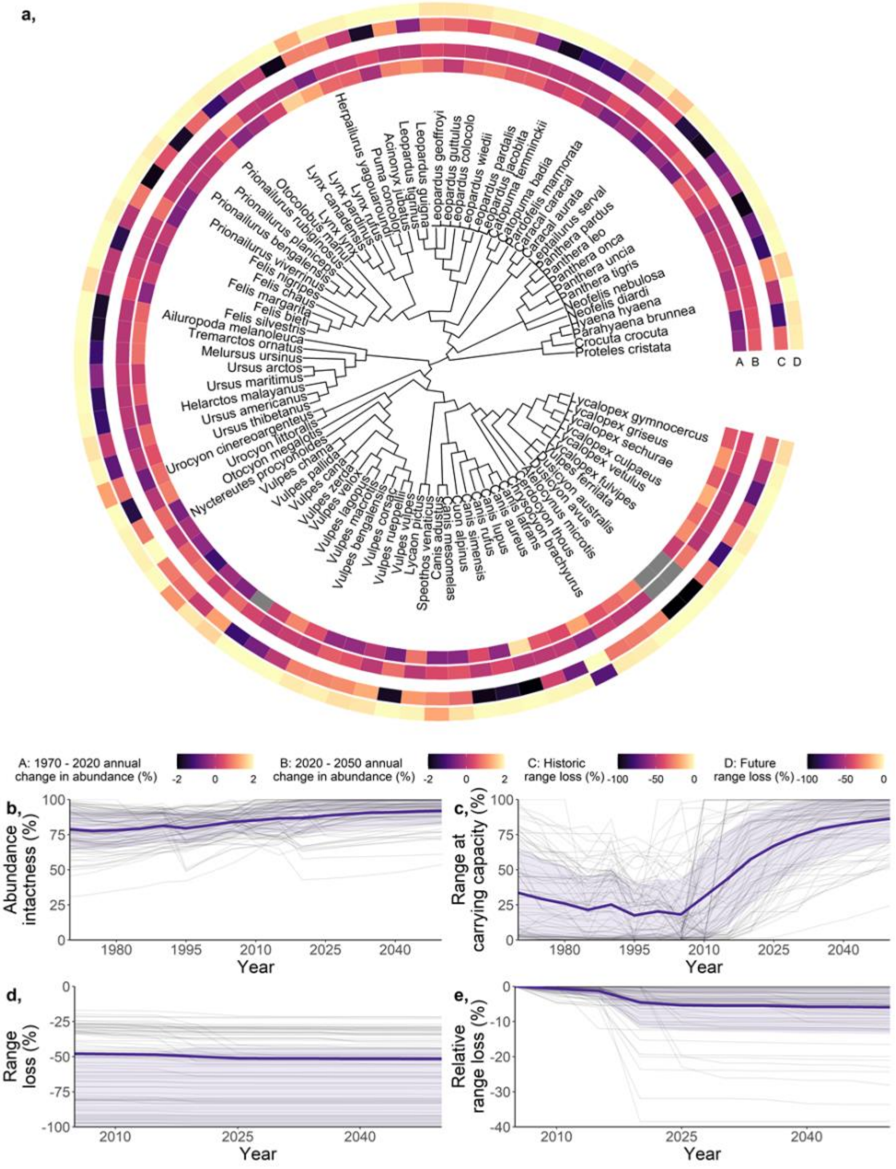


Figure 4. Cladogram of 87 species in the orders Canidae, Felidae, Hyaenidae and Ursidae, following the IUCN taxonomy and using an amended Nyakatura and Bininda-Emonds (2012) phylogeny. From in to out, the rings indicate: A - the annual rate of change in abundance (%) between 1970 and 2020; B - the annual rate of change in abundance (%) between 2020 and 2050; C - the percentage decrease in the species current range relative to its counterfactual natural range, as defined by Faurby et al. (2018). Two species in this ring, *Canis latrans* and *Nyctereutes procyonoides*, had range expansions instead of losses, but for visualisation, we set range losses to 0%; D - describes additional future range losses from abundance crashes comparing the current range (baseline period of 2005) to 2050. b) Mean abundance intactness between 1970 and 2050). c) Mean percentage of species' current range at or approaching carrying capacity, which we define as cases where species abundance intactness is greater than 95% (essentially intact). d) Mean percentage of species' range to become extirpated compared to species' counterfactual natural range. e) Mean percentage of species' distribution to become extirpated compared to species' current range at the baseline period of 2005. For panels b-e, the thicker lines represent the average across all species, with shading showing ± 1 standard deviation. Grey lines show individual species.

Whilst the projected recoveries in abundance intactness are encouraging for the status of large carnivores, it is important to note that these species have already experienced large range losses (Figure 4a,d). Specifically, current IUCN range extents are approximately half the size of the historic counterfactual-natural ranges. Yet, much of these range losses likely occurred before 1970, and our projections suggest future range losses will be minimal, relative to their historic ranges (Figure 4d). Even comparing species current to future ranges, losses are projected to average less than 4% across all species by 2050 (Figure 4e). Furthermore, our projections do not consider potential range expansions, some of which are already happening (Cimatti et al., 2021). If our projections are correct and many of these carnivores approach carrying capacity across their range, further expansions seem likely.

Carnivore guild status

At a global scale, we project that carnivore guild status has been largely stable since 1970, and stability will continue into the future (Figure 5). However, guild status varied across regions, as above we identify winners and losers. Most notably, in South America, the region with the most favourable carnivore guild status, we project increases from 1980 to 2010, which then plateaus until 2050. In contrast, we project guild status has declined in Africa over the last 50 years - driven by the predicted rapid rate of human development, primary land loss, and large species suffering under extreme heat – but guild status then begins to recover in 2020. By 2050, guild status will only be relatively intact in the Amazon, protected areas of Africa, high altitude areas (e.g. Himalayas and Andes), and high latitude regions.

Our approach relies on land-use, climate and human development projection data and would be affected if these projections are incorrect. Land-use and climate projections are based on well-established sources (Lange, 2019a, 2019b; Cucchi et al., 2020) averaged over a variety of shared socioeconomic and relative concentration pathways to capture projection uncertainty (see Supplementary methods). However, available human development projections are crude and highly uncertain, largely assuming development will continue at a similar trajectory. This may not be a reasonable assumption: climate catastrophes are expected to slow human development growth in low human development countries (UNDP, 2020). Less growth could be good news for biodiversity, but countries could attempt to escape from the economic toll of a climate catastrophe by extracting natural resources and placing added pressure on large carnivores. As improved projections of human development become available, we could refine our biodiversity predictions evaluating potential scenarios.

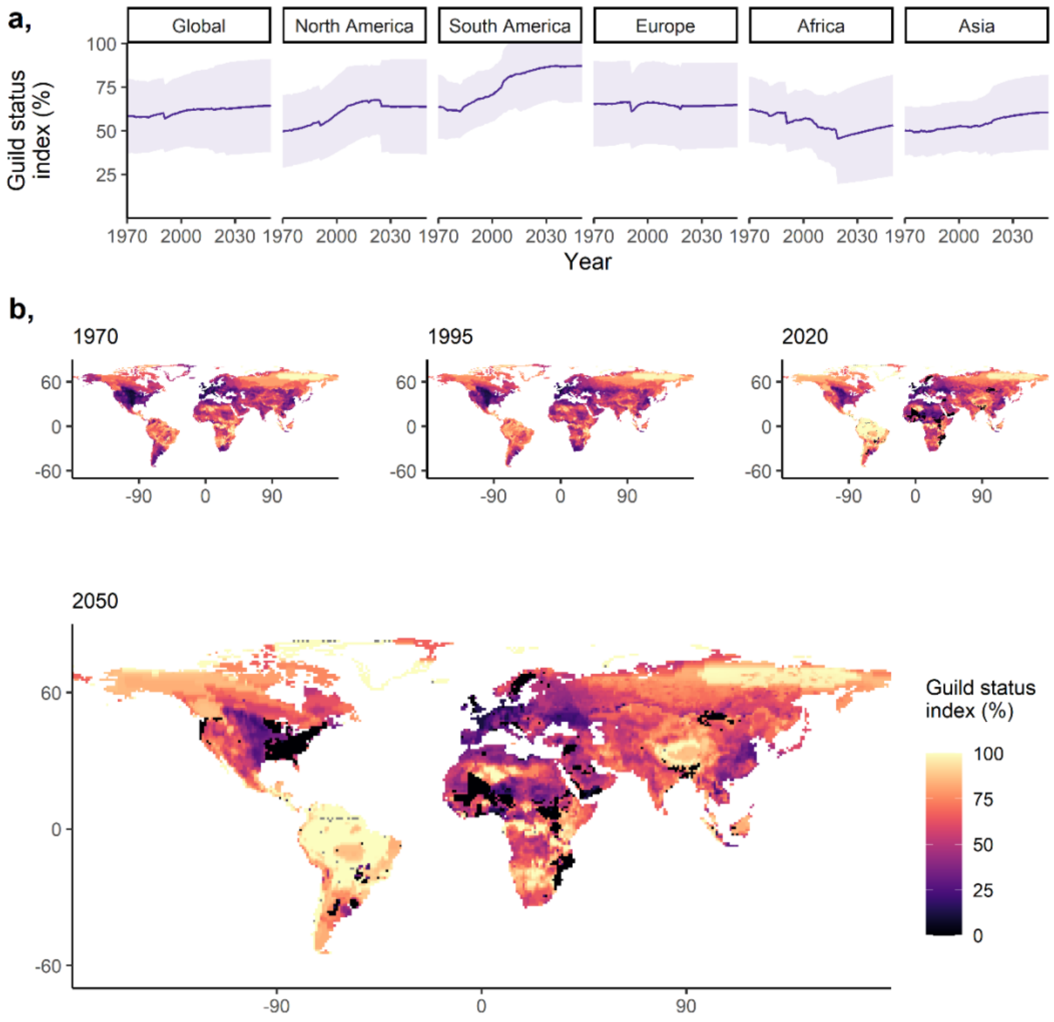


Figure 5. a) Mean annual carnivore guild status index (± 1 standard deviation) for the entire globe, as well as each of the five continents where the study species occur. b) Carnivore guild status index displayed at 1-degree resolution across the globe at four time points: 1970, 1995, 2020, and 2050. Carnivore guild status index is derived by multiplying the guild-averaged abundance intactness (scaled from 0 to 100) by guild completeness (locally extant species divided by extant and recently extinct species), where 0 indicates all species have become locally extinct, whilst 100 indicates the guild is complete and all species' populations are at carrying capacity.

Our projections are also limited by the biases and predictive accuracy of our population trend model. Whilst we attempted to minimise biases in the trend data by using a novel censored and weighted regression, which allowed us to increase the taxonomic and spatial scale of the work relative to previous analyses, biases are certainly still present and the impact of these on our inference remains unclear. Future work could also utilise approaches like state-space trend modelling to better capture uncertainty within these trend estimates. Encouragingly, sensitivity analysis show the same covariates and interactions were supported across different dataset and model structures (Supplementary results: Sensitivity analysis). The model's predictive accuracy was intermediate (conditional R^2 ~0.5). We explored a broader array of factors (e.g. climate, land-use, governance and traits) than previous work, but we still lack data on likely important aspects (e.g. poaching, persecution, culling, and the conservation benefits of being flagship species). Capturing this multitude of different influences has also allowed us to make projections of population change, but these warrant careful critique. Future work could validate these projections and develop more effective solutions for capturing projection uncertainty e.g. hindcasting. Until this point, these projections should only be viewed as a possible scenario, rather than an actionable piece of evidence.

Our study offers a new comprehensive analysis of global population trends for large mammalian carnivores. We show that abundance trends have been influenced by stressors like land-use and climate change and that their effects have been magnified and mitigated depending on the species traits, the protected status of the land, and some interactions between the stressors. Large carnivores do not conform to the biodiversity decline narratives (Díaz et al., 2019), but highlight a scenario of winners and losers under environmental change (Dornelas et al., 2019). Our study also reveals the importance of accounting for human population characteristics (i.e. human development) which call for a wider socio-ecological scope when evaluating biodiversity changes. Focussing solely on stressors like land-use and climate change may be effective at identifying causes of declines, but provides few opportunities to identify the features that support recovery, and these feature may hold the key to bending the biodiversity curve. In large carnivores, recoveries could already be underway, but new challenges are on the horizon. We predict many large carnivore populations could reach carrying capacity within their current ranges in the next 30 years, which could lead to range expansions. Range expansion will bring political and societal challenges: governments and the people living alongside these large carnivores will need to decide how willing they are to coexist with these species. Decisions like this will shape the future status of large carnivores, and biodiversity more generally.

Code and data availability

Code and data available upon request and will be made openly available prior to publication.

Acknowledgements

Funding to TFJ from NERC (Natural Environment Research Council) Centre for Doctoral Training studentship (J71566E), and to MGS from The Royal Society (IE160539).

Supplementary methods

Population trends

We sourced population trend information for all species in the families Canidae, Felidae, Hyaenidae, and Ursidae of the order Carnivora from two large trend datasets: CaPTrends (Johnson et al., 2021a) and the Living Planet Database (WWF, 2020a). CaPTrends contributed 1,220 trends, and the Living Planet Database contributed 350, combining to produce a cumulative 1,474 unique (non-duplicated) trends. In the Living Planet Database, and for most records in CaPTrends, trends are reported as a timeseries of abundance (or density) estimates. We modelled these timeseries with log-linear regressions, where abundance (the response) was $\log_e$ transformed, and year of abundance estimates was selected as the predictor. We extracted the slope coefficient which represents the annual instantaneous rate of change, sometimes called the population growth rate (r_t). There are also other formats of quantitative trends in CaPTrends which fall into three broad datatypes, all of which we converted into an annual instantaneous rate of change (r_t):

a) Finite rate of change

$$r_t = log_e(\lambda)$$

Where λ represents the mean annual finite rate of change.

b) Estimates of relative abundance change between two points in time (e.g. percentage or fold change in the past 10 years)

$$r_t = \frac{log_e(1 + (P/100))}{N}$$

Where P represents the additive percentage change (e.g. a population doubling in size = 100%), and N is the difference in time (in years) between the two estimates of abundance. For fold changes, we first converted the fold change into an additive percentage change.

c) Timeseries of population change estimates, reported as either population lambdas or percentage changes e.g. in year 1 the population doubled ($\lambda = 2$) and in year 2 it halved ($\lambda = 0.5$). We back-converted the change estimates into abundance estimates against a constant value of 100. We then fitted log-linear regressions with abundance and year, as in the abundance timeseries.

We converted all annual instantaneous rates of change into an annual rate of change percentage to improve interpretability. These rates of change ranged from -75% to 68%, but the majority of values fell within -10% to 10% (Figure S1a).

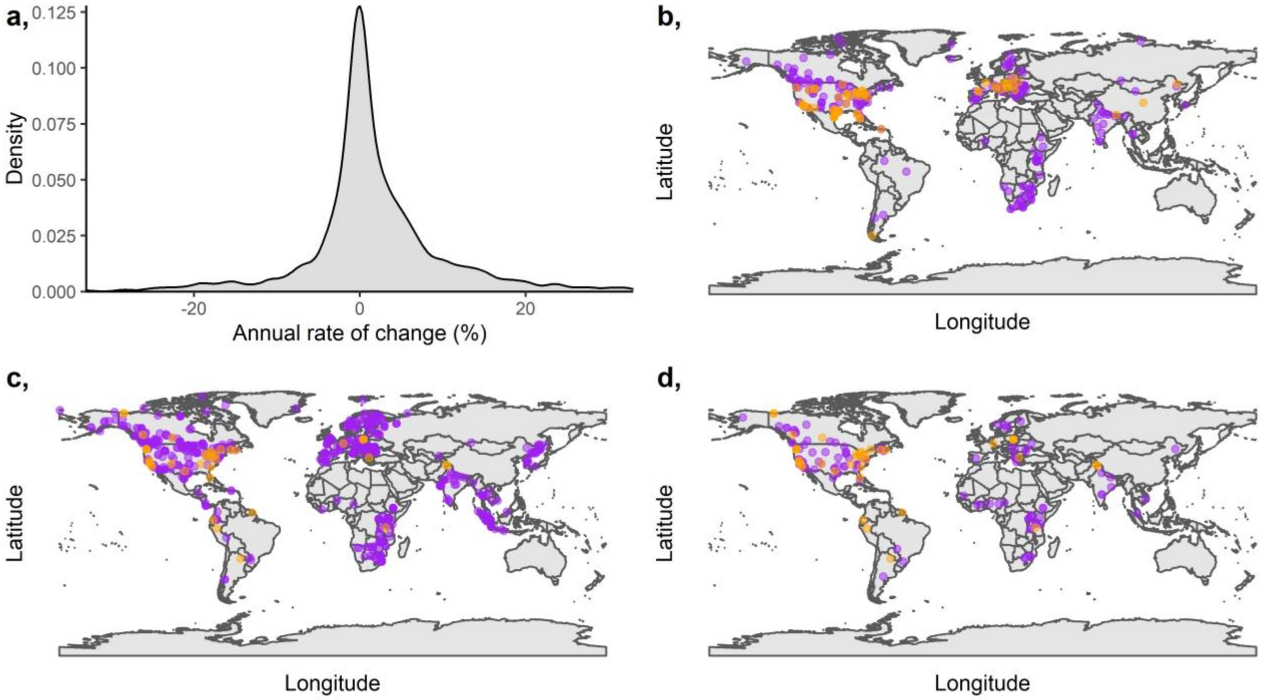
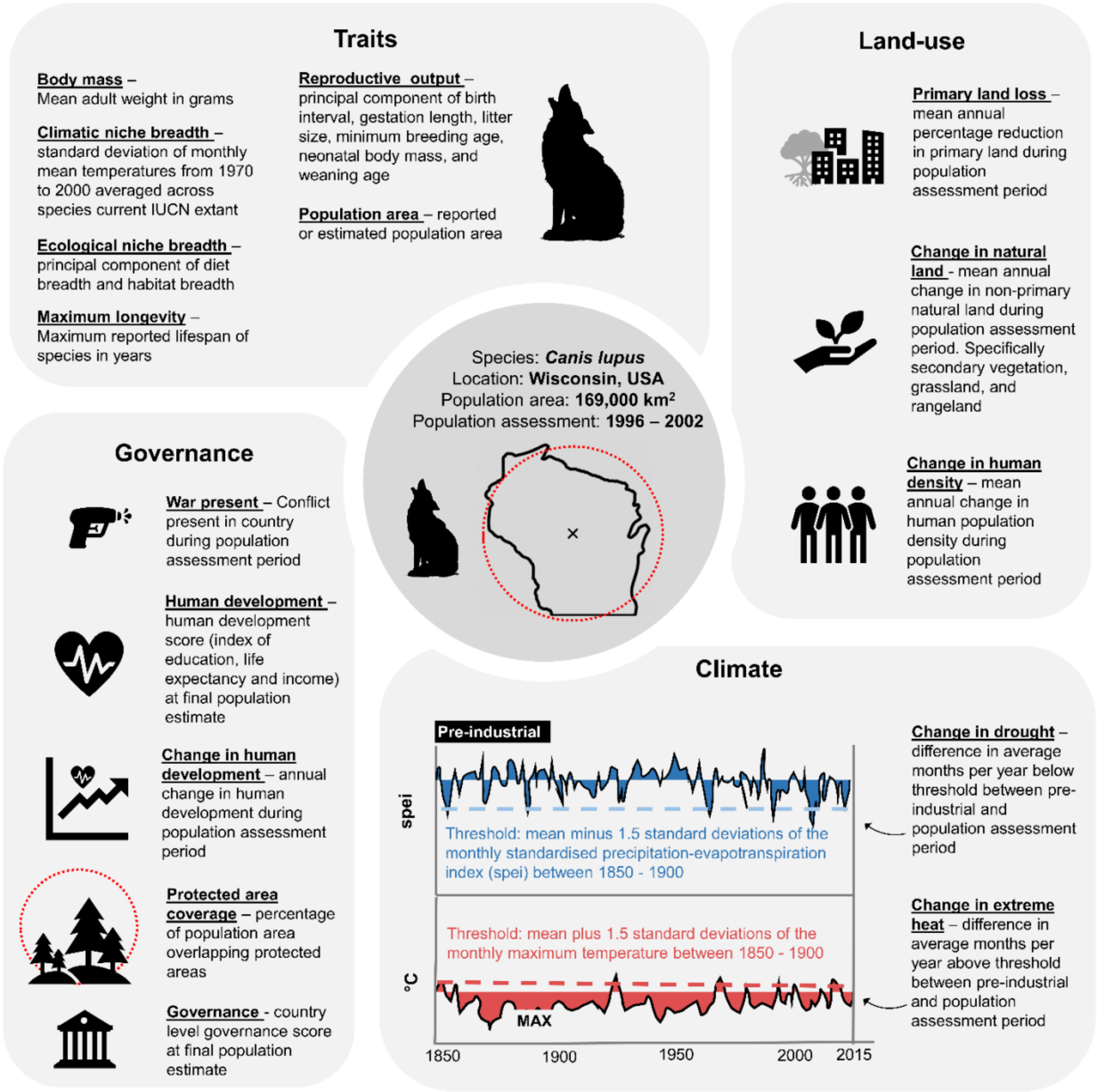


Figure S1. a) Distribution of quantitative records (represented as annual instantaneous rates of change); rates of change range from -75% to 68%, but axes have been trimmed to more clearly represent the bulk of the data. b-d) Spatial representation of quantitative (purple) and qualitative (orange) records, split into increasing (b), stable (c), and decreasing (d) trends. For the quantitative records, trends exceeding an annual rate of change of 5% were classed as increasing, between -5% and 5% were classed as stable, and less than 5% were classed as decreasing. The qualitative records fell naturally into the increasing, stable, and decreasing categories.

Alongside the quantitative records, 138 populations in the CaPTrends dataset were only described qualitatively with categories: increase, stable, and decrease. These records were more common for populations located in traditionally poorer-sampled countries (e.g. with lower human development), so whilst they are less informative (only describing the direction and not the magnitude), we deem them important to reduce known biases (Figure S1b-d). As a result, we used a combination of annual rate of change (%) and qualitative categories as our responses in our inference model – see below.

Covariates

Our covariates fall into four categories: land-use, climate, governance, and traits (Figure S2). One of the challenges in identifying how covariates impact population trends is matching the spatial scale of the covariate with the population i.e. how much of the population is affected by the covariate. To tackle this problem, we used data on the area of extent of each population to generate a circular distribution zones around the population's coordinate centroid. We refer to this as the 'population area' hereafter. In populations without a reported extent (N = 347), we searched the locality and location description online to identify the approximate size of the population area. For example, the top result in a Google search for ‘Serengeti area’ described the location as 30,000km^2. If we could not find an area, we assigned the population as one of the following categories: small locations (e.g. towns and counties): 1,000km^2 [N = 123], medium locations (e.g. regions and states): 10,000km^2 [N = 151], large locations (e.g. countries): 100,000km^2 [N = 73].



Traits

Body mass – Mean adult weight in grams

Climatic niche breadth – standard deviation of monthly mean temperatures from 1970 to 2000 averaged across species current IUCN extant

Ecological niche breadth – principal component of diet breadth and habitat breadth

Maximum longevity – Maximum reported lifespan of species in years

Reproductive output – principal component of birth interval, gestation length, litter size, minimum breeding age, neonatal body mass, and weaning age

Population area – reported or estimated population area

Land-use

Primary land loss – mean annual percentage reduction in primary land during population assessment period

Change in natural land - mean annual change in non-primary natural land during population assessment period. Specifically secondary vegetation, grassland, and rangeland

Change in human density – mean annual change in human population density during population assessment period

Species: ***Canis lupus***
Location: **Wisconsin, USA**
Population area: **169,000 km^2**
Population assessment: **1996 – 2002**

Governance

War present – Conflict present in country during population assessment period

Human development – human development score (index of education, life expectancy and income) at final population estimate

Change in human development – annual change in human development during population assessment period

Protected area coverage – percentage of population area overlapping protected areas

Governance - country level governance score at final population estimate

Climate

Change in drought – difference in average months per year below threshold between pre-industrial and population assessment period

Change in extreme heat – difference in average months per year above threshold between pre-industrial and population assessment period

Threshold: mean minus 1.5 standard deviations of the monthly standardised precipitation-evapotranspiration index (spei) between 1850 - 1900

Threshold: mean plus 1.5 standard deviations of the monthly maximum temperature between 1850 - 1900

Figure S2. Sixteen covariates with a proposed effect on carnivore population trends highlighted in bold and underlined. Covariates fall in four groups: Traits, Land-use, Climate, and Governance. Text alongside covariates briefly explains how the variable was derived, whilst full explanation and justifications for inclusion are available in Supplementary methods: Covariates.

Given population areas regularly exceeded 10,000km^2 (Figure S3a), it was not computationally feasible to extract covariates over the entire area; thus, we sampled from a random selection of points within each population area, sampling more frequently in larger areas (range: 13 – 295 sampling points, Figure S3b). Random sampling was only used for land-use and climate covariates, as governance covariates are measured at the national level, and all traits (except for population area itself) are measured at the species level. The population areas and corresponding sampling points were developed with a Mollweide equal-area projection, but we transformed these areas and points back into a WGS84 projection to match all covariate rasters (see below). In all covariates, 'population monitoring period' refers to the period (start and end year) the population was monitored for.

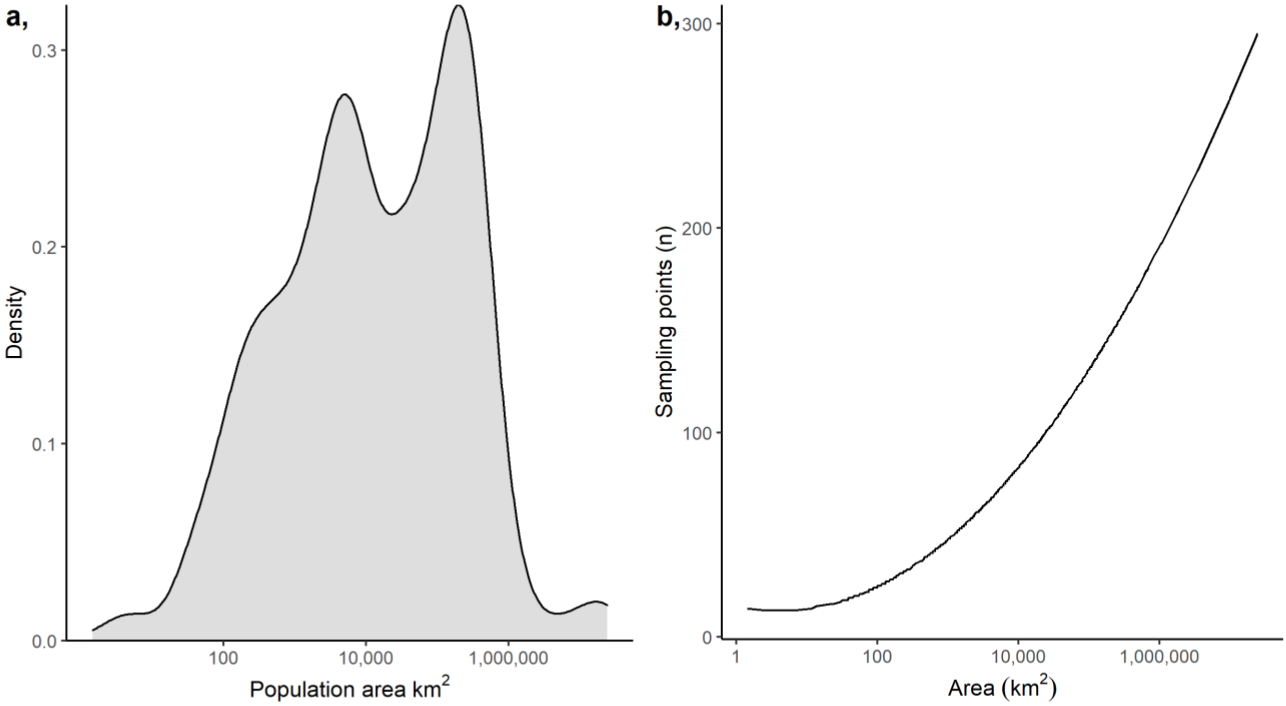


Figure S3. a) Distribution of population areas, the area of extent of population monitoring, for 1,127 quantitative population trends extracted from CaPTrends (Johnson et al., 2021a) and the Living Planet Database (WWF, 2020a). b) Frequency of covariate sampling points relative to population area size, where populations occurring over larger areas receive more covariate sampling. Area, the x-axis, is displayed on the $\log_{10}$ scale.

Land-use

We extracted three land-use covariates: Primary land loss, Change in natural land, and Change in human density. Primary land loss and Change in natural land were derived from the land-use harmonization dataset (Hurtt et al., 2020), which reports the annual proportional coverage of 11 land-use types between 1850 and 2015AD, at a 0.25° spatial resolution. To make the land-use types more biologically relevant to predators, we amalgamated a selection of the 11 types into two summary-types: primary land – the sum of 'forested primary' and 'non-forested primary'; and natural land – the sum of 'potentially forested secondary', 'potentially non-forested secondary’, ‘managed pasture’ and ‘rangeland’. To estimate Primary land loss we found the mean primary land across sampling points in each population area for each year in the population monitoring period. We then estimated the rate of loss in primary land over time by dividing the rate of loss in each year by the previous year, and then converted this to a percentage loss. We defined the mean Primary land loss (%) for each population area as the average across this timeseries of loss rates. We followed an identical procedure for Change in natural land. Importantly, primary land cannot be restored, so primary land only decreases or remains stable. Whilst natural land can fluctuate up and down.

We estimated the Change in human density using the Global human settlement human population raster (Florczyk et al., 2019), which describes the human density per km^2 for four years: 1975, 1990, 2000, 2015. For each year, we mean aggregated the original data to reduce the spatial resolution to 0.1°. In order to estimate the Change in human density for each population area’s monitoring period, we had to estimate missing human density values (years) in each population area. To do this, we first extracted the mean human density across each population area, in all four of the available years. In each population area, we then used a log-linear regression to predict human density (base log transformed) by year between 1960 and 2015. As human density change was non-linear we modelled year (predictor) with a cubic fit. We then extracted the back-transformed predicted values of human density for all years in each population area. As we were only working with four data points, model predictions were highly uncertain. This uncertainty was included by resampling our model with 100 bootstrap iterations. For each population monitoring period and iteration, we extracted the predicted human densities and estimated the rate of change (%) as calculated for the other land-use covariates. Finally, we calculated the mean human density rate of change (%) across all iterations, as well as the standard deviation, which was used to represent uncertainty in the values within the inference model (see below).

Climate

Our two climatic covariates, Change in extreme heat and Change in drought, describe how the number of months exceeding an extreme heat or drought threshold (respectively) changed between a pre-industrial period (1850 – 1900) and the population monitoring period (Figure S2). To derive our Change in extreme covariates, we compiled a raster timeseries of bias-corrected daily maximum near-surface air temperature from 1850 to 2014, at a 0.5° resolution (Lange, 2019a, 2019b; Cucchi et al., 2020). We then averaged the daily maximum temperature for each month, in each year, creating a monthly raster timeseries of the mean maximum temperature from 1850 to 2015. From this, we calculated the mean and standard deviation of the monthly maximum temperature in the pre-industrial period, and defined the extreme heat threshold as the mean + 1.5 standard deviations of the mean in each pixel. We selected 1.5 standard deviations as the measurement of noise as this limited the number of months that exceed the threshold to between 5 and 10% within the pre-industrial period i.e. meaning extreme heat events were rare but happened. Next, we found the mean number of months per year in the pre-industrial period that exceeded this threshold for every pixel, as well as the number of months to exceed this threshold in all years between 1960 and 2015. We then subtracted the number of threshold-exceeding months in each year (1960-2015) from the mean threshold-exceeding months in the pre-industrial period, creating a raster timeseries describing the difference in threshold-exceeding months e.g. how do the number of months exceeding the threshold in 1970 differ to the average across the pre-industrial period? Finally, for each sampling point in each population area, we found the mean difference across the population monitoring period, and then averaged this difference across all sampling points to produce a population area estimate of the Change in extreme heat.

To derive our Change in drought covariate, we compiled two raster timeseries' describing the bias-corrected daily near-surface air temperature and daily bias-corrected precipitation, both from 1850 to 2014, at a 0.5° spatial resolution (Lange, 2019a, 2019b; Cucchi et al., 2020). We then averaged the daily temperature rasters and summed the daily precipitation rasters for each month, in each year, creating two monthly raster timeseries, describing monthly mean temperatures and total precipitation (in mm) from 1850 to 2015. For temperature, we calculated Thornthwaite's evapotranspiration across the raster timeseries, which uses the mean temperature, latitudinal position and number of daylight hours to estimate the evapotranspiration rate (Thornthwaite, 1948). Next, we subtracted this monthly evapotranspiration estimate from the monthly precipitation (mm) estimate to produce Thornthwaite's standardised precipitation-evapotranspiration index (spei), a standard metric used to describe water availability (Vicente Serrano et al., 2010). We then proceeded to estimate a spei threshold and the mean difference in months overlapping the threshold (pre-industrial vs. population monitoring period) in an identical way to how monthly maximum temperature is treated in the Change in extreme heat covariate.

Governance

We identified five governance covariates that we considered important to large predator population trends, four of which were measured at the country-level: War present, Governance, Human development, and Change in human development. We used three datasets to populate these covariates: 1) For War-present, we used the

UCDP/PRIO Armed conflict dataset (Pettersson, 2019), which lists conflicts (between 1946-2019) where fatalities exceeded 25 per year and at least one of the parties is governmental. We summarised this dataset into a timeseries that describes whether a war was present in each country's territory in each year between 1960 and 2016. 2) For Governance, we extracted the world governance indicator metrics (Kaufmann et al., 2011), which present six annual governance timeseries for each country between 1996 and 2016. 3) Finally, for Human development we sourced the UN human development index (UNDP, 2021), which provides an annual timeseries describing life expectancy, education level, and income per capita between 1990 and 2016 for 189 countries.

As the governance and human development indicator data only stretch back until 1996 and 1990, respectively, some of the trend data preceded the indicator values. We imputed missing values through a multiple imputation chained equations (MICE) framework (Van Buuren & Groothuis-Oudshoorn, 2011). We used a hierarchical normal (2l.pan function) imputation model, where observations are nested into the different countries, and included: the year of the observation, the six governance indicator metrics, the human development index, whether war was present in that year (yes or no), as well as the country's gross domestic product (log 10 transformed). MICE imputations are stochastic and repeated numerous times, creating an approximate distribution for each missing value. We imputed missing values for each variable between 1960 and 2016 in each country, and repeated the imputation 100 times with a 50 iteration burn in – all variables showed convergence (Figure S4).

Using the imputed datasets, we extracted the mean value across the six governance indicators in each country, year, and imputation chain. We then calculated the mean and standard deviation of this combined governance across the imputation chains to produce an annual governance timeseries (and associated error) for each country. For Human development, we averaged over the 50 stored imputation chains to calculate the mean and associated standard deviation for each country and year. We assessed if imputed values were plausible (Figure S5).

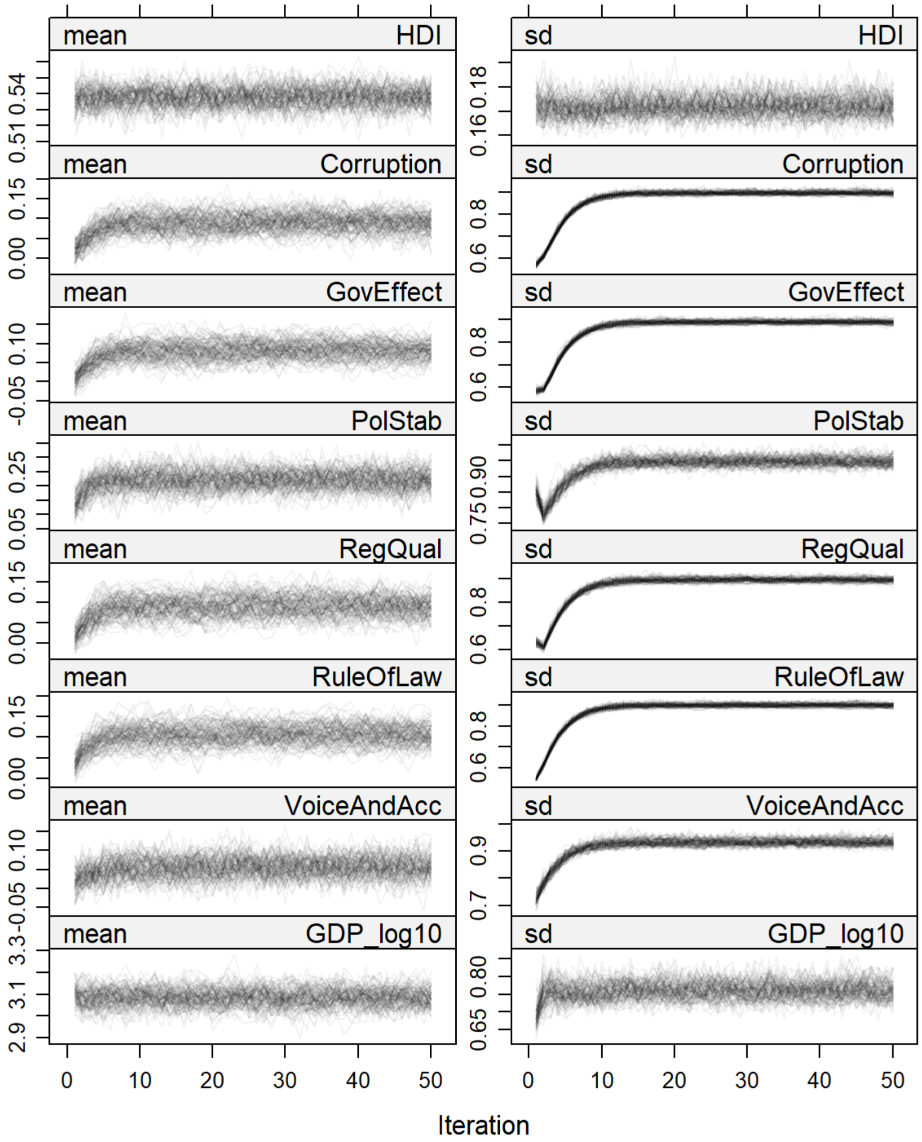


Figure S4. Convergence of mean (left) and standard deviation (right) of variables with missing values in the imputation model: HDI – human development index, Corruption – control of corruption, GovEffect – government effectiveness, PolStab – political stability and absence of violence, RegQual – regulatory quality, RuleOfLaw – rule of law, VoiceAndAcc – voice and accountability, and GDP_log10 – gross domestic product (log 10 transformed). Convergence ran with 50 iterations and 100 chains.

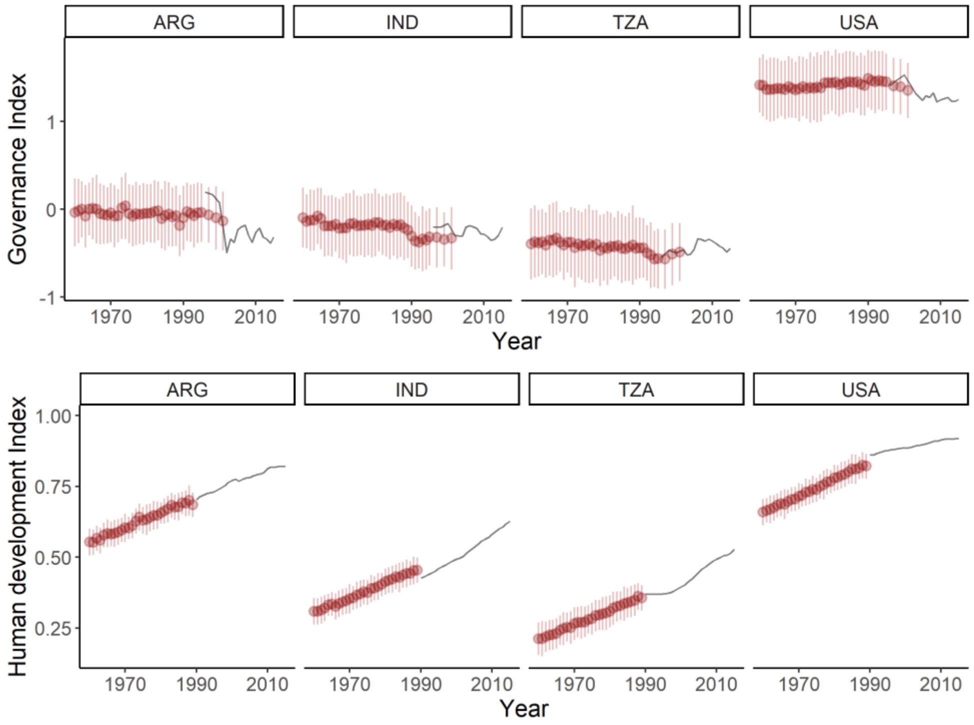


Figure S5. Governance (top) and human development (bottom) index scores for Argentina (ARG), India (IND), Tanzania (TZA), and the United States of America (USA). True values are depicted with the black line, whilst the mean imputed values (point) and associated 95% confidence intervals (bars) are depicted in red.

After we derived the governance and human development timeseries, we began extracting the covariates. For War present, we created a binary variable that described whether war(s) had occurred in the country where the population is located, at any point during the population monitoring period. For Governance and Human development, we extracted the mean scores per country, and associated standard deviations from the final year of the population monitoring period. For Change in human development, we extracted all human development values across the population monitoring period, and divided each value by the value in the previous year to produce a timeseries describing the annual changes in human development. We then averaged these values and converted the average into a percentage which describes the annual rate of change (%) in human development.

Our only governance covariate not-measured at the country-scale is Protected area coverage. For this variable, we compiled the annual timeseries of protected areas polygons covering the period 1960 to 2020 from the World Database of Protected Areas (UNEP-WCMC & IUCN, 2021). In each year, we converted the polygons into a 0.1° resolution raster describing the proportional cover of protected areas in each pixel. In the final year of the population monitoring period, we calculated the mean coverage of protected areas across pixels within the population area.

Traits

We identified five species traits which may influence population trends in large predators: Body mass, Maximum longevity, Climatic niche breadth, Ecological niche breadth, and Reproductive output. Body mass describes the average body weight of an adult of the species in grams ($\log_{10}$ transformed), Maximum longevity describes the maximum lifespan of species in years (log 10 transformed), and Climatic niche breadth describes the standard deviation of the mean monthly temperatures across the species current IUCN range, calculated using WorldClim 2.1 (Fick & Hijmans, 2017). The other two trait-covariates, Ecological niche breadth and Reproductive output, are principal components of a larger array of traits. Specifically, Ecological niche breadth captures habitat and diet breadth. Habitat breadth is defined as the frequency of different IUCN habitat classifications (IUCN, 2020a) the species occurs in. Diet breadth is defined as the number of different food-types the species has been recorded consuming (or with evidence of consuming through faecal or stomach content analysis), from the following 12 options: mammals, birds, reptiles and amphibians, fish, invertebrates, fruit, pollen and nectar, leaves and branches, seeds, grass, roost and tubers, and carrion – sourced from an unpublished trait dataset (González-Suárez, 2014). Our Reproductive output trait is a principal component of the following traits (all log 10 transformed): inter birth interval, gestation length, litter size, minimum breeding age, neonatal body mass, and weaning age, each of which describe a different characteristic of each species breeding biology, and generally highly correlated. As a result, our five traits of interest were reliant on collecting values for 12 common traits.

We sourced values for our traits from three different trait datasets: PanTHERIA (Jones et al., 2009) AnAge (De MagalhÃes & Costa, 2009), and an unpublished large predator trait dataset (González-Suárez, 2014). We used multiple trait datasets to populate missing values at the species level. However, the values sometimes differed between the trait datasets, and in these cases, we created multiple records for the species to capture this uncertainty in the trait value. As a result, many species had more than one value for a given trait. However, despite using multiple trait datasets, values were still missing for some species in some traits (Table S1), and so we imputed missing trait values with Rphylopars (Goolsby et al., 2017). Rphylopars outperforms MICE imputation (as used in the governance covariates above) as it uses both the trait values and species' phylogeny to estimate missing values – Rphylopars is considered one of the best imputation methods (Johnson et al., 2021b). In our Rphylopars model, we trialled the three Nyakatura & Bininda-Emonds (2012) Carnivora phylogenies to ensure the imputations did not drastically change depending on the phylogeny (Figure S6). Once we confirmed the choice of phylogeny had little impact, we proceeded with the rest of the analysis only using the phylogeny considered ‘best’ by Nyakatura & Bininda-Emonds (2012). We also included all 12 traits mentioned above in the imputation model, and six other traits to attempt to account for biases in the imputation model, specifically: species area of occurrence, minimum absolute latitude species occurs at, maximum absolute latitude species occurs at, difference in maximum and minimum latitude, maximum mean monthly temperature, and minimum mean monthly temperature. As the phylogenies we used were not perfectly matched to the CaPTrends and Living Planet Index taxonomies, we corrected synonymous species names in the phylogeny, and where species included in the taxonomy were absent from the phylogeny, we appended the species to a node that could be a close ancestor (based on taxonomy).

Table S1. Percentage of values missing in each trait.

Trait	**Missing trait values (%)**
Body mass	8.0
Maximum longevity	12.6
Climatic niche breadth	0.0
Habitat breadth (part of ecological niche breadth)	0.0
Diet breadth (part of ecological niche breadth)	23.0
Age of sexual maturity (part of reproductive output)	24.1
Litter size (part of reproductive output)	10.3
Gestation length (part of reproductive output)	9.2
Weaning age (part of reproductive output)	27.6
Interbirth interval (part of reproductive output)	12.6
Neonatal body mass (part of reproductive output)	28.7

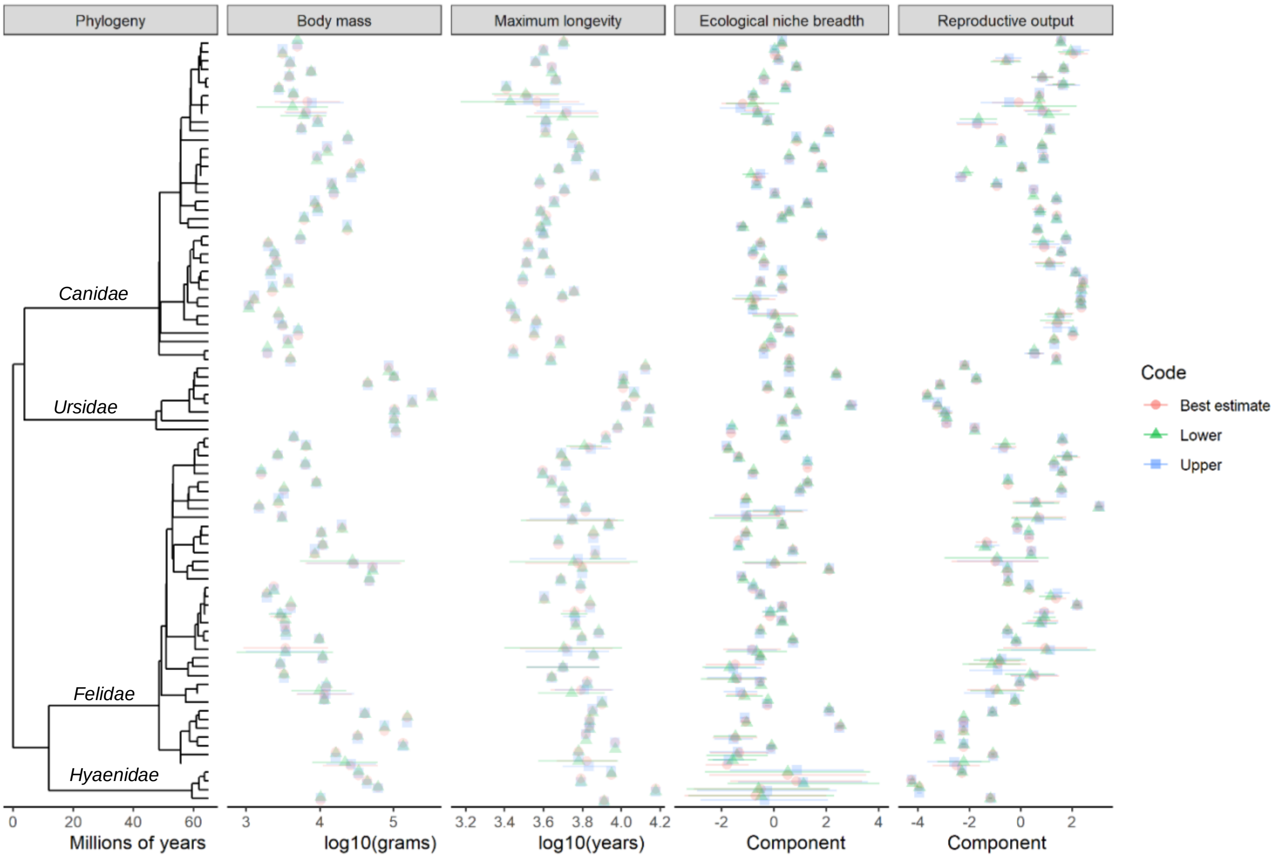


Figure S6. Trait values represented on the phylogeny; climatic niche breadth is excluded as it had no missing values. Error bars represent the 95% confidence around the mean imputed values. If observations were complete (i.e. not missing values) the standard deviation around the observation was zero and so there are no confidence intervals. We include the three phylogeny types in Nyakatura & Bininda-Emonds (2012).

An advantage to using Rphylopars is that it provides an estimate of the standard deviation around the missing trait values, which is something we wanted to capture in our modelling (see Inference model below). Three of our traits were going to be used as covariates directly within the modelling (Body mass, Maximum longevity, and Climatic niche breadth), so required no further manipulation as their associated standard deviations were available from the imputation. However, Ecological niche breadth and Reproductive output required dimension reduction through principal component analysis (PCA), with the number of variables shifting from 2 to 1, and 6 to 1, respectively. Performing PCA on the mean values would fail to capture trait uncertainty, and so instead we developed normal distributions for each species' trait value using their mean, and an error of one standard deviation. We then sampled from each distribution 100 times, and each time conducted a PCA on the trait to develop an eigenvector. We saved the eigenvector values on each of the 100 repeats, and once the repeats were complete, we calculated the mean and standard deviation for each species across the eigenvector values. This PCA sampling procedure was performed separately on the Ecological niche breadth and Reproductive output components. We examined the trait values to ensure they were plausible (Figure S6), and also checked between-trait correlations were acceptable i.e. sufficient variance in the correlation (Figure S7).

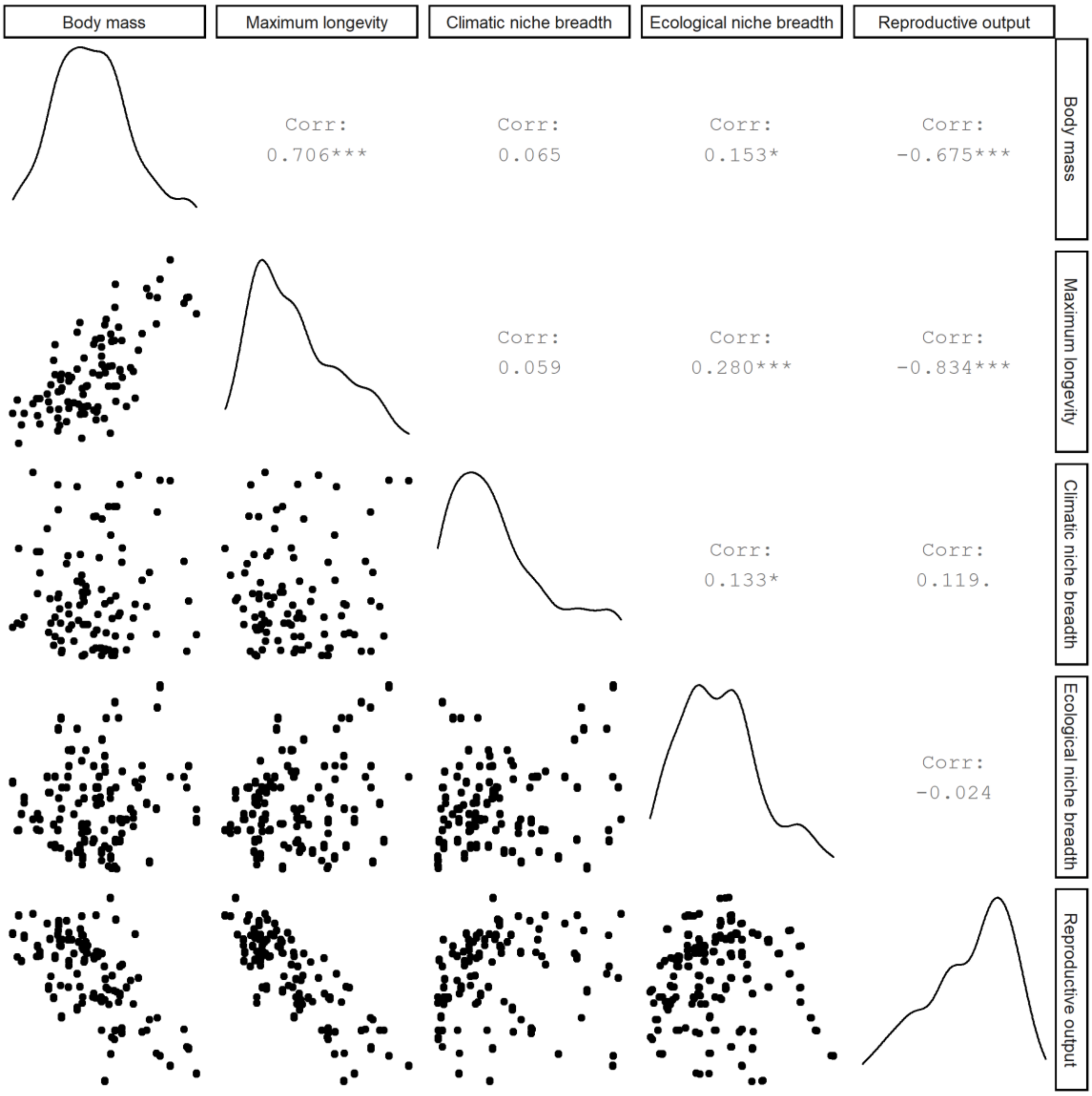


Figure S7. Distributions of traits are represented on the diagonal. The person correlations between traits are represented above the diagonal with varying levels of statistical significance (p-value): '***' when p<0.001, '**' when p<0.01, '*' when p<0.05, and '.' when p<0.10. The scatter of observations (one point per species) below the diagonal graphically represent these correlations

Temporal lag

A challenge in identifying how covariates impact population trends, is matching the temporal scale of the covariate with the population i.e. how long is the lag between the covariates impact and a change in the population. This lag period is likely variable across covariates (i.e. it could be different with land-use and climate features) and species traits. For example, recent work has suggested population change in large mammals peaks at approximately 10 years after forest loss (Daskalova et al., 2020). As a result, we conduct sensitivity analysis (see Sensitivity analysis below) to determine how model fit was influenced by lag selection, considering three options: 1) No lag, so covariate changes are measured between the start and end year of each population monitoring period. 2) Five-year lag, where covariate changes are measured between the five-years prior to the start of each population monitoring period, and run to the end of each period. 3) Ten-year lag, where covariate changes are measured between the ten-years prior to the start of each population monitoring period, and run to the end of each period.

Cleaning data

We opted to remove a selection of the population trend and covariate data as the values were deemed unreliable or unsuitable. Specifically, we removed any population trend records beginning before 1970 or after 2016 (N = 11), where governance data was largely incomplete. We also removed records overlapping multiple countries (N = 10), and any population trends with an excessively large population buffer-area (N = 40) – we set the threshold at 2 million km^2 which could accommodate state and small-country level estimates, but would exclude large countries. For example, the largest population area in the dataset covered all of Russia (~21 million km^2). Any population trends discussing non-native species were removed (N = 6), as well as records not overlapping any land (N = 4) e.g. *Ursus maritimus* populations occurring exclusively on sea-ice. We also removed any population trends where the population had either recolonised an area or become locally extinct (N = 80), which represent an extreme form of population change that could skew our inference. After excluding records, we were left with 989 estimates of annual rate of change, and 138 qualitative descriptions of change.

Modelling

Inference model

We fitted a hierarchical linear model (Figure S8) to determine the effect of a combined 23 covariates and interactive effects on the rate of change in large predator populations. Our model development falls into seven compartments: censoring, random intercepts, coefficients and covariates, imputation uncertainty, weighted error, confirming parameters, and model running. The model was written in BUGS language and implanted in JAGS 4.3.0 (Plummer, 2016) via R 4.0.3 (R Development Core Team, 2020).

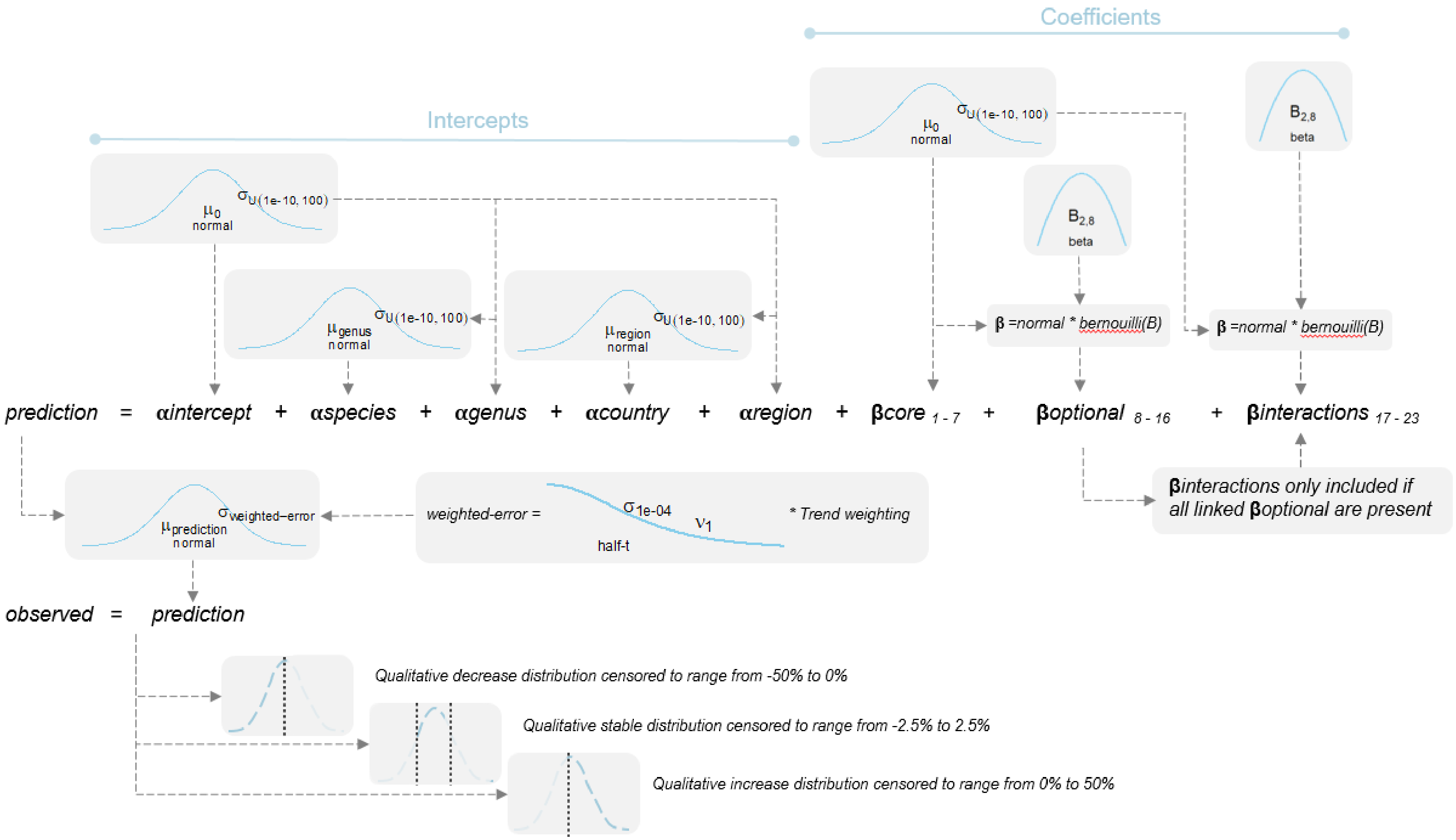


Figure S8. Model structure of hierarchical linear model, describing distributions of priors and hyperpriors, as well as the process for incorporating overall error, imputation error, trend weights, and censoring within the model. We use five distributions (parameters described in brackets) within the model: normal (μ = mean, σ = standard deviation), beta (shape1, shape 2), half-t distribution (μ = mean, σ = standard deviation, df = degrees of freedom), U/uniform (minimum, maximum), and Bernoulli (B = probability).

Censoring

We included the two different population trend datatypes within the model: quantitative annual rates of change and qualitative descriptions of change. We modelled both of these datatypes with a normal error prior setting the standard deviation of this prior as a half-t (also known as a half-cauchy) distribution hyperprior centred at zero, with a standard deviation of 0.001 and one degree of freedom. However, to deal with the different data types, and the unknown values of the qualitative descriptions, we censored the qualitative records to indicate that the true value is unknown, but it occurs within a specified range. We specified these annual rate change ranges as -50% to 0%, -2.5 to 2.5%, and 0% to 50% within the decrease, stable and increase categories, respectively. Many of these qualitative records occur in less-well represented regions, species, and time-periods, so their inclusion addresses known data biases (Figure S1). However, these lower quality records will also be more prone to error. As a result, we conduct sensitivity analysis (see Sensitivity analysis below) to assess how including censored observation altered model fit, compared to only using quantitative, and high-quality quantitative (derived from at least three abundance observations), trends. Finally, in the modelling, we identified that the annual rates of change exhibited a heavy tailed t-distribution, we transformed our responses into a gaussian distribution with an inverse-hyperbolic sine transformation.

Random intercepts

We used a hierarchical model structure to account for phylogenetic and spatial non-independence in the data, including species as a random intercept nested with genus, and country as a random intercept nested within sub-regions, as defined by the United Nations (https://www.un.org/about-us/member-states). These parameters were fit with a normal distribution centred at zero and their error terms were given a vague uniform hyper prior, with a standard deviation ranging from $1e^{-10}$ to 100.

Coefficients and covariates

With a combined 23 covariates and interactive effects, we were conscious of overparameterizing the model. As a result, we split these parameters into three groups: 1) core parameters – which included main effects that have previously been reported as influential, are expected to be influential, or control for other parameters and methodological features; 2) optional parameters – which included main effects we considered interesting but with little evidence to-date of any influence on trends; and 3) interactive parameters – which includes all interactive effects. Core parameters included: Change in human density, Primary land loss, Population area, Body mass, Change in extreme heat, Governance, and Protected area coverage. These core parameters were included in every model, but we used Kuo and Mallick variable selection (Kuo & Mallick, 1998) to identify important parameters from the optional and interactive groups, where variables were only included in an iteration if they were selected from Bernoulli priors. Our optional parameter group was assigned a Bernoulli prior, which sampled from a beta hyperprior (a = 2, β = 8), such that approximately 20% of optional effects would be included in any iteration, on average, but this could range from 0 – 100%. The interactive parameter group had an identical, but separate prior setup. Crucially, this interaction prior was only activated if both main effect parameters were present in the model. For example, for the Change in extreme heat and Change in drought interaction to be selected, it would require Change in drought to be selected from the optional Bernoulli prior, and then the interaction itself would need to be selected from the interactive Bernoulli prior. As variable selection can be highly influenced by the standard deviation of the parameter slope coefficients, we specified the slope standard deviation as a vague uniform hyperprior ranging from $1e^{-10}$ to 100. As this variable selection could lead to thousands of model combinations, each receiving different levels of support, we only selected well supported iterations - specifically, the most frequent parameter combinations accounting for up to 80% of all iterations.

Imputation uncertainty

Six of the covariates in the model contained missing values that were filled using imputation (see *Land-use*, *Traits* and *Governance* within the *Covariates* section above). To improve the robustness of our model inference, we accounted for uncertainty in the imputed estimates by treating imputed values of the covariates as distributions instead of point estimates, where each imputed value was assigned a normal distribution centred at the mean imputed estimate and with an error varying by the imputed observation standard deviation. As we z-transformed all of our covariates to standardise coefficients, except ‘War present’ which is a categorical variable, we also had to rescale the associated imputation standard deviations. As standard deviations cannot be rescaled in the same way as the imputed estimates, we first converted the standard deviation into confidence intervals, we then z-transformed the intervals using the mean and standard deviation of the covariate, and then back calculated the standard deviation from these intervals.

Weighted error

When developing the model, we were conscious that all not rates of change should contribute equally to the fit. For example, whilst including the censored records could decrease taxonomic and spatial biases in the data, they may also introduce error, as these censored records are unlikely to be as accurate as the quantitative trends. As a result, we included a weight term to inflate the uncertainty in these lower quality records, where the half-t hyperprior discussed above is multiplied by a weight term: the inverse of the estimated error in the rate of change. This weight term was developed through simulation (see below), and these simulated error weights inflated the variance around the trend in all low-quality observations, not just the qualitative ones.

When simulating the trend weights, we considered our real trend data to be estimates of true trends with some degree of error. This error would be influenced by the certainty of the population abundance estimates, the sampling intensity (e.g. is the population sampled every year or only in 50% of years), and the sampling duration (e.g. is the trend based on 2 or 20 years). As a result, we developed a simulated trend dataset comprised of ‘true’ trends where abundance values are known (not estimates) and complete, and an edited trend dataset where abundances are uncertain and missing as expected in a real scenario and observed in our trend dataset.

For the ‘true’ trend dataset, we simulated 6000 timeseries’ of abundances which varied in duration (from 2 to 20 years), with an estimate of abundance in all years throughout that duration. We then calculated the true trend for each timeseries by modelling abundances (response) against year in a log-linear regression, and converted the slope estimate into an annual rate of change (%). Abundance values exhibit a normal distribution ranging from approximately 0 to 500.

For the edited trend dataset, we altered two parameters in each of the 6000 timeseries’ of abundances generated above. Firstly, for each abundance estimate in each timeseries, we developed a random normal distribution, centred on the true abundance value, but with varying levels of error (coefficient of variation from 0.02 to 0.2). For example, with a true abundance of 100, a low error of 0.02 would produce a range of abundance estimates from approximately c.95 to c.105, whilst the abundance would range from c.50 to c.150 with an error of 0.2 (Figure S9). We sampled from these newly created abundance distributions to produce new error-prone abundance estimates, reminiscent of real uncertainty in abundance estimation. Secondly, we removed a random sample (between 0% and 90%) of the observations in each timeseries, producing timeseries' with varying levels of completeness. We then re-calculated the annual rate of change (%) as in the true trend to produce the estimated trend. The distribution of the estimated trend was largely similar to the true trend obtained from the complete dataset (Figure S10).

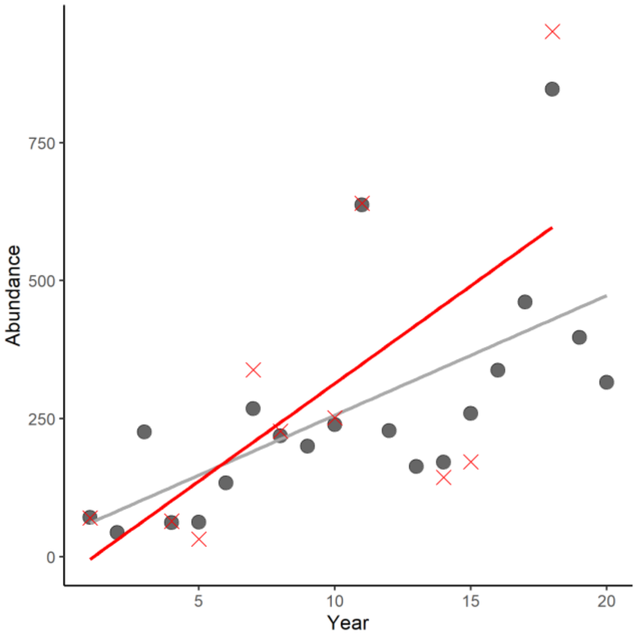


Figure S9. Impact of adding noise to abundance values and removing abundance values on the population trend, with the true trend (derived from known and complete abundance values) in grey, and the estimated ones in red. In this example, the coefficient of variation equals 0.2, with a sampling intensity of 50% i.e. half the years in the population monitoring period have abundance values.

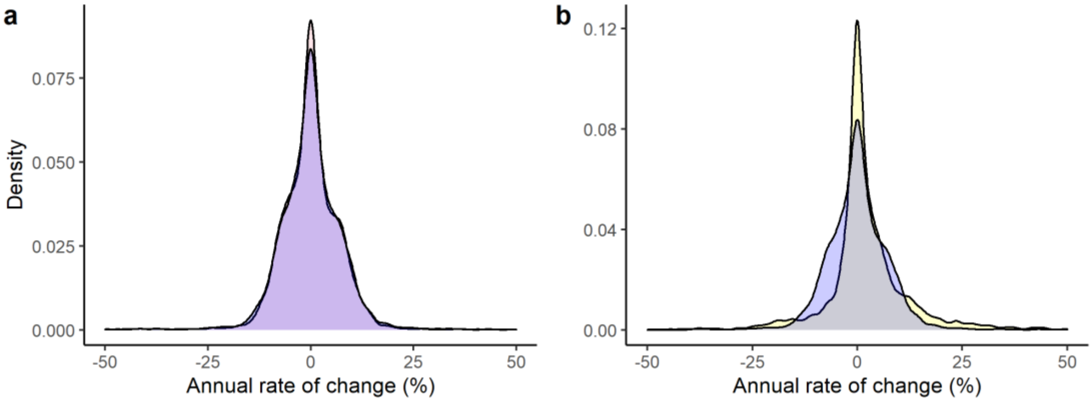


Figure S10. a) Distribution of simulated true trend values (pink) and simulated estimated trend values (blue). b) Distribution of simulated estimated trend values (blue) and real trend values compiled from CaPTrends (Johnson et al., 2021a) and the Living Planet Index (WWF, 2020a) in yellow.

We extracted the absolute error (difference) in the annual rate of change of the true and estimated trends, and modelled this error (as the response) against sampling intensity (what percentage of years have observations), the coefficient of variation around the abundance estimates, and the duration of the trend, all in a log-linear regression. Trends with a higher sampling intensity (coef = -1.59, CI: -1.72, -1.47), lower coefficient of variation (coef = 4.09, CI: 3.68, 4.49), and longer duration of the trend (coef = -0.12, CI: -0.13, -0.11), had lower errors (Figure S11). We used this model based on simulated data to predict the likely error in the real data. For sampling intensity, we calculated the percentage of abundance values used to calculate the trend relative to the trend duration. For the trend duration, we calculated the number of years in each population monitoring period. Unfortunately, in most cases the estimates of uncertainty around the raw abundance values were unavailable, so we were unable to directly calculate the coefficient of variation for each trend. However, we did have data describing the quality of the sampling and modelling which could act as a proxy for the accuracy of the abundance values. Specifically, we scored trends separately in three areas (Table S2), where trends could only be assigned one category per area; we then added the score across the three areas: *Sampling* – how systematically was the population sampled? *Modelling* – how robust was the approach for modelling abundance values? *Low-quality record* – does the record meet any of the criteria for being considered low quality? For example, a trend with systematic population sampling (+0.04), where sampling effort is accounted for (+0.04), meeting none of the low-quality criteria (+0), would be given a coefficient of variation score of 0.08. For an abundance value of 100, this coefficient of variation score would allow the abundance to vary between 75 and 125.Admittedly, our scoring criteria here is arbitrary, simply deigned to add uncertainty around trends that used less robust methods, rather than, say, describe the true uncertainty in the trend. However, as these arbitrary values only contribute one feature of three in the weighting system, there impact is likely minimal, and is tested in sensitivity analysis regardless (see supplementary results).

After predicting the error in the real trend data using the simulated weight model, we scaled and flipped the values so that 1 indicates low error and 0.0001 indicates high error. These values had to be flipped, as the weight term in our hierarchical linear model (Figure S8) is multiplied by the precision (i.e. uncertainty) around each trend observations, in which a precision would then be deflated (i.e. uncertainty inflated) if multiplied by a high error trend. For example, for observation A with a low error of 0.9, a precision of 10 would be deflated to 6, whilst for observation B with a high error of 0.1, a precision would be deflated to 1, so A would receive 6 times more weight than B.

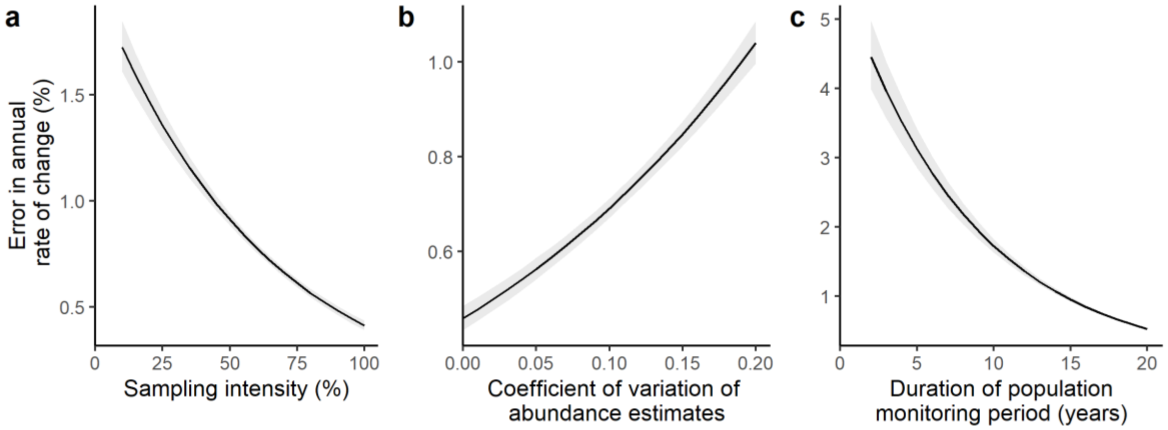


Figure S11. Marginal effect of sampling intensity (a), coefficient of variation around abundance values (b), and trend duration (c) on the absolute error in the annual rate of change (%), comparing the simulated-true to the estimated trend. Sampling intensity describes the percentage of years with abundance values in the population monitoring period. Trend duration describes the length of the population monitoring period e.g. 1990 – 1992 equals three years.

Table S2. Scoring criteria used to define a coefficient of variation (CV), uncertainty, in abundances.

	Description	**CV**
Sampling		
	Method of population sampling is not described or is unsystematic/biased.	0.08
	Method of population sampling is systematic.	0.04
	All individuals in the population identified.	0.01
Modelling		
	Method of deriving abundance from population sampling is not described or values are just reported in their raw format.	0.08
	Sampling effort accounted for in abundance estimates.	0.04
	Abundance derived through complex modelling, or total abundance known.	0.01
Low quality record		
	Abundance values derived from genetic or harvest data; or the trend is labelled as inaccurate within the primary literature; or trend describes asymptotic instead of observed growth; or trend metric is unconventional.	0.04

To ensure our weight term benefitted the model fit, we conducted sensitivity analysis to compare the model fit under four options: 1) the simulated error weight (described above); 2) weighting by trend sample size, whereby trends derived from more abundance observation are given more weight; and 3) unweighted i.e. all observations are treated equally.

Sensitivity analysis

We conducted sensitivity analysis to test how the different weighting, censoring, and temporal lag options influenced our model results, with the aim of selecting parameters which maximised model marginal and conditional R^2, whilst also balancing this decision against potential risks. For example, including censored observations may reduce model fit but this could still be worthwhile if it reduces taxonomic and spatial biases. For weighting, we ran models separately under each of the three options, including censored observations and a 5-year lag on all covariates in all cases. After identifying the simulated error weighting as the best option for maximising fit and minimising bias (see Supplementary results) we tested the censoring options, again holding all covariates at the 5-year lag. Including censored observations was valuable, so we included the censored observations when assessing the different temporal lag models, from which we identified that using a 10-year lag improved model fit. In each case, we ran the model through two chains, each with 10,000 iterations and discarding the first 5,000. We thinned the complete chains to store every other iteration (thinning factor of 2). We monitored convergence of key parameters within each model, specifically: standard deviation of the model intercept, standard deviation of beta coefficients, standard deviation of each random effect (regions, countries, genus, and species), standard deviation of the overall model error, the optional parameter beta hyperprior, and the interactive parameter beta hyperprior. We ensured the multivariate potential scale reduction factor was less 1.1 across all models in the sensitivity analysis.

Model running

After selecting the simulated error weighting, censored observations, and a 10-year lag from the from the sensitivity analysis (see Sensitivity analysis in the supplementary results), we ran the full model through three chains, each with 120,000 iterations. The first 20,000 iterations in each chain were discarded, and we only stored every 10th iteration along the chain (thinning factor of 10). We opted for a large chain and burn-in due to the model complexity, and to allow a broad selection of parameter combinations to be tested under variable selection. We assessed convergence of the full model on all parameters monitored in the sensitivity analysis, as well as the model intercept, and all 23 main and interactive effect slope coefficients. We checked the standard assumptions of a mixed effect linear model (normal residuals and heterogeneity of variance) and tested the residuals to ensure no spatial (Moran’s test) or phylogenetic (Pagel’s lambda) autocorrelation. We also conducted posterior predictive checks to ensure independently simulated values were broadly reminiscent of model predicted values.

After selecting the iterations of the most common (80%) model combinations (see *Coefficients and covariates* above) we calculated how frequently, as a proportion, each of the 23 main and interactive effects occurred within the iterations. For the optional parameters, this was derived by dividing the frequency of occurrence by the total count of iterations. For the interactive parameters, whose inclusion was dependent on the frequency by which their derivative main effects were selected, we divided the frequency of occurrence by the total count of iterations where both derivative main effects were present. Using the selected 80% of iterations, we report the median slope coefficient and associated credible intervals for each of the main and interactive effects. We also produce marginal effect plots for each parameter with an effect either above or below zero at the 80% credible interval threshold. These marginal effects hold all other covariates at zero. We also display the distribution of the random intercepts e.g. for each region, country, genus, and species. Throughout all these plots (but the not the plots in the *Projections* section below), we have propagated uncertainty through the model. For instance, uncertainty within the missing trait and governance values, as well as the less robust trend values is fed forwards through this analysis and is represented within the model posteriors. As a result, all effects and results capture multiple facets of uncertainty within the model.

Projections

Our inference model (above) describes how the annual rate of change (%) in abundance is influenced by a series of covariates, and we endeavoured to use these covariate coefficients to project how abundance in large predators had changed in the past (from 1970) and will continue to change in the future (up to 2050). We developed these projections through six steps: compiling covariates, projecting trends, and calculating: abundance intactness, guild abundance intactness, guild completeness, and finally, guild status (Figure S12).

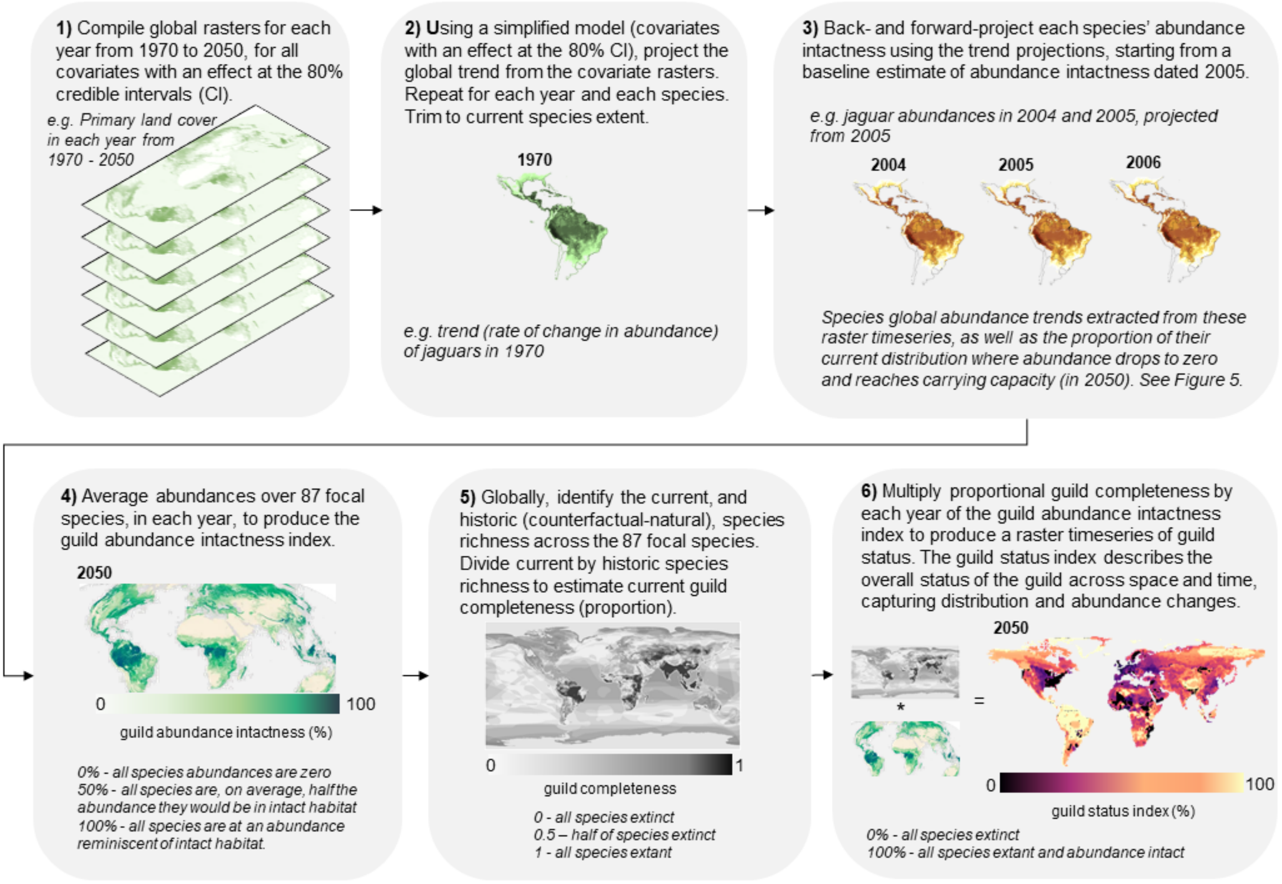


Figure S12. Procedure for estimating species and guild abundance intactness, as well as guild status. Models were developed using a simplified form of the linear regression in the inference model, as well as the baseline abundance intactness from 2005 (Newbold et al., 2016). Current species richness was estimated using IUCN current range maps (IUCN, 2020b), and historic species richness was derived from counterfactual-natural ranges (Faurby et al., 2018). See the supplementary material for a more detailed description.

Compiling covariates

In order to project trends into the future for each species, we needed to compile an annual timeseries of spatially explicit data for all covariates deemed important in the inference model, up to the year 2050 - specifically: Primary land loss, Change in extreme heat, Change in drought, Human development, Change in human development, and Protected area coverage. All this data had already been compiled up to 2015 and used in the earlier model, but to project up to 2050, we sourced dataset extensions that facilitate future projections. However, as future projections are uncertain and influenced by representative concentration pathways (RCP) and shared socioeconomic pathways (SSP), amongst other things, we opted to consider three different scenarios to capture projection uncertainty. Specifically, for the Primary land loss, Change in extreme heat and Change in drought, we downloaded the Land-use harmonization and UK Earth system model future projection datasets for SSP1 – RCP2.6, SSP3 – RCP7.0, and SSP5 – RCP8.5 (Lange, 2019a, 2019b; Cucchi et al., 2020; Hurtt et al., 2020). The three SSP options propose different socioeconomic pathways; the ones we selected describe sustainability (SSP1), regional rivalry (SSP3), and fossil-fueled development (SSP5). The RCP options describe varying projected increases of greenhouse gas concentrations ranging from low (RCP2.6) to high (RCP8.5). Incorporating the new projection scenarios split each covariate spatial timeseries into three with pathways diverging from 2015.

We were unable to find pathway projections, or any future projections at all, for Human development, Change in human development and Protected area coverage covariates. For the human development variables, we opted to solve this by developing our own crude projections up to 2050. We conducted a beta regression for each country with the mean human development index (from the imputation) as the response and year of the estimate as the predictor (quadratic fit). We then rasterised the predicted values for each country to produce a spatial raster timeseries of human development index, which we used to describe the Human development and Change in human development in each pixel in each year. Our future projections of human development are unlikely to be correct, especially as historic growth in human development is at risk from environmental disasters (UNDP, 2020), and our projections ignore the different SSP and RCP scenarios. However, given the substantial influence of human development (and its change) on the model, we decided including crude projections was more robust than ignoring both covariates in their entirety.

Similar to human development index, there are no future projections of protected area coverage. However, unlike human development index, there is no reasonable approach for estimating future protected area coverage. Instead of attempting to predict future protected area placement locations and times, we instead opted to simply consider that protected areas will remain static from 2020 until 2050.

All future covariate projections are produced in an identical way to the past projections (detailed in the Covariates section above). However, for these projections the model did not include the population area covariate and values are reported at the raster pixel level each year instead of averaged over the population area and population monitoring period. The model used to project annual rates of change in each species included the covariate trait Body mass alongside the genus and species intercepts, but we note all of these were assumed to remain constant over time and space for each species (see *Projection trends* below).

Two non-covariate features of our model, the region and country intercept, also needed to be represented as spatial rasters, but were assumed to be constant over time. We developed raster projections for the region and country (separately) describing how the model intercept changes depending on the where the pixel occurs. Regions and countries not captured in the data of our inference model were given an intercept value of zero. All of the rasters in included in the model were mean aggregated to the same one-degree resolution in a WGS84 projection.

Projecting trends

We developed a trend projection for the 85 extant species within the target families by running the following projection model for each species in each year between 1970 and 2050:

Annual rate of change (%) ~
Overall model intercept +
Genus intercept + Species intercept +
Region intercept + Country intercept +
β(Body mass) * Body mass +
β(Primary land loss) * Primary land loss +
β(Change in extreme heat) * Change in extreme heat +
β(Change in drought) * Change in drought +
β(Human development) * Human development +
β(Change in human development) * Change in human development +
β(Protected area coverage) * Protected area coverage +
β(Change in extreme heat * Body mass) * Change in extreme heat * Body mass +
β(Change in extreme heat * Change in drought) * Change in extreme heat * Change in drought +
β(Change in extreme heat * Body mass) * Change in extreme heat * Body mass

Each projection model includes the median overall intercept from the inference model, the genus- and species-specific intercepts, and the region and country intercepts. The species-specific intercepts were set a zero when species were absent from the data in the main model, but these species were influenced by the genus intercepts. The projection model also includes all main and interactive effect parameters with an effect at the 80% credible interval, where the slope coefficients are represented by beta (β) and contained by brackets; the corresponding data are adjacent but not contained by brackets. For our beta slope coefficients, we only use the median value and ignore coefficient uncertainty within the projections due to long computational run times. All datasets were z-transformed using the mean and standard deviation of the covariates in the inference model. All other parameters in the inference model (the full model containing all parameters) were excluded from the projection model. We used a hyperbolic sine transformation to rescale the annual rate of change into a percentage – annual rates of change were inverse hyperbolic sine transformed in the inference model. This projection model was repeated for each of the three SSP/RCP scenarios, so each species had three annual spatial raster timeseries extending from 1970 to 2050. For simplicity, projections were initially completed for all terrestrial pixels but we only considered values in the area currently occupied by each species which was defined by the current extant polygons of IUCN distribution maps (IUCN, 2020b).

Abundance intactness

Our trend projections describe the rate at which species abundances are predicted to change across their current distribution, but they ignore the current abundance status i.e. populations cannot grow if they are at carrying capacity. To capture this important characteristic, we needed a baseline estimate of the abundance (relative to carrying capacity) for each species across space. These abundance estimates are not currently available in the literature. However, we found a partial proxy for this estimate in the Newbold et al. (2016) abundance intactness index, a spatial raster describing the average abundance of biodiversity in 2005, relative to their abundance in an intact ecosystem. This index ranges from 0 (local extinction) to 100 (intact population) and extends across the entirety of terrestrial planet. This index covers a wide array of text, describing the intactness of biodiversity as a whole, instead of species-specific estimates of intactness; so in any given pixel, all species would be treated as having the same abundance intactness. This may not be a fair assumption if species have responded differently to land-use pressures, but in the inference model we found no evidence of land-use change effecting species differently. Even if species have responded differently, and so would have different abundance intactness estimates within the same pixel, this intactness estimate is only being used as a baseline to derive abundance trends, instead of directly informing them. As a result, we used the abundance intactness index as a baseline period for species in 2005 and used the trend projections for each species to project the abundance backwards (to 1970) and forwards (to 2050). For example, if the abundance in 2005 was 50, and the rate of change between 2005-2006 was 2%, the abundance in 2006 would be 51. We capped the abundance at 100, which we considered carrying capacity. We also limited the rate of growth in the trend projections so growth in raster pixels could not exceed 5% per year. Whilst populations can grow faster than 5% per year, such high rates would be unlikely for these large carnivoran species at a large spatial extent (the one-degree resolution we use). Finally, we averaged (mean) the species' abundance projections across the SSP/RCP scenarios so each species only had one timeseries. We opted to average over the SSP/RCP scenarios to simplify the findings, and notably, the results were similar regardless of the scenario.

For each species abundance projection, we calculated three metrics: annual rate of change (%) from 1970 to 2020, annual rate of change (%) from 2020 to 2050, and future range loss (%). The annual rate of change metrics were derived by finding the mean abundance across all species extant in each year, and then modelling this abundance (response) in a log-linear regression with year as the predictor. This was done separately for the two periods (1970-2020 and 2020-2050). We converted the slope from these regressions into a percentage rate of change. We estimated future range loss by finding the percentage of pixels where abundance drops to zero (extirpated) by 2050, relative to the total number of pixels in the species current extant. We also derive one additional metric, Historic range loss (%), to accompany these species descriptors, by finding the percentage change in species current IUCN extant (IUCN, 2020b) relative to species counterfactual-natural ranges (Faurby et al., 2018), which we use as a proxy for historical ranges.

Guild abundance intactness

After deriving the species metrics from the abundance projection timeseries', we averaged species abundances over space to produce an average estimate of guild abundance in each year from 1970 to 2050.

Guild completeness

Whilst guild abundance intactness describes the status of species within their current distribution, it ignores historic range contractions. As a result, we determined the current completeness of the guild, to estimate the change in species richness over space. For this, we compiled the distribution maps from the IUCN (IUCN, 2020b) for the 85 extant species, and overlayed these to develop a one degree resolution raster describing the current species richness across the planet. We then compiled a historic species richness estimate over space by overlaying the counterfactual natural ranges (Faurby et al., 2018) of the 85 extant species and two extinct species. We divided the current estimate by the historic estimate to derive the proportion of currently extant species relative to extant and locally extinct species.

Guild status

We multiplied each year of the guild abundance intactness index by the proportional guild completeness to describe the overall status of large predators between 1970 and 2050. This guild status index captures abundance and distribution changes to describe the status of predators over space and time.

Supplementary results

Sensitivity analysis

We assessed how trend weighting, including censored observations, and specifying a lag period on the covariates influenced model fit and inference, in part to assess if results were particularly sensitive to specific parameters, but also to help choose the parameters which optimised fit and spatio-taxonomic coverage. Using censored observations and a lag period of 5 years on covariates, model fit was greater when using the simulated error weights compared to the unweighted model and the model weighted by sample size (Table S1). Using simulated error weights and a lag period of 5 years on covariates, we then tested the impact of including censored observations which showed higher marginal and conditional R^2 when censored observations were included. Using only high quality timeseries (compared to including censored observation) resulted in a higher conditional R^2, but at the cost of excluding 19 countries and 2 species from the dataset. We considered the gain in model fit did not outweigh the added spatial and taxonomic bias. Finally, using simulated error weights and the full dataset (including censored observations), we tested how the lag period of covariates influenced fit. The 0-year lag had a slightly greater overall fit than the 5- and 10-year lags, but also had a slightly lower marginal R^2. These differences were so minimal that we opted for the lag most supported by the literature – 10 years, with suggesting peak population change occurs 8 years after environmental change (specifically forest loss) in mammals

In our final model, we used the simulated error weights, included censored observations, and used a 10-year covariate lag. While these decisions optimized fit and data coverage, the type of weightings, the data or time lags used had little impact on inference, as model coefficients were largely similar across all parameter types (Figure S1 – S3).

Table S1. Fit of nine models tested in the sensitivity analyses split across three categories: Weighting – influence of different trend observation weighting options; Censoring – impact of including different qualities of trend data; and Lag – fit under different lag periods for covariates (e.g. for a predator population monitored between 1995-2000, the Change in human density would be measured from 1995-2000, 1990-2000, and 1985-2000, respectively under the 0, 5, and 10-year lags. Fit measured as the marginal and conditional R^2. There are varying levels of data in each model, and we summarise the frequency of countries and species this data occurs in. For the weighting models, all quantitative and qualitative-censored trends were included, with a 5-year lag on the covariates. For the censoring models, all trend observations were weighted by the simulated error, with a 5-year lag on the covariates. For the lag model, all quantitative and qualitative-censored trends were included and weighted by the simulated error.

	Category	Marg. R^2	Cond. R^2	Countries (N)	Species (N)
Weighting	Unweighted	0.09	0.26	75	50
	Weighted by trend sample size	0.10	0.29	75	50
	Weighted by simulated error	0.11	0.29	75	50
Censoring	High quality timeseries trends	0.11	0.30	56	48
	All quantitative trends	0.10	0.28	69	49
	All quantitative and qualitative-censored trends	0.11	0.29	75	50
Lag	0 years	0.10	0.30	75	50
	5 years	0.11	0.29	75	50
	10 years	0.11	0.29	75	50

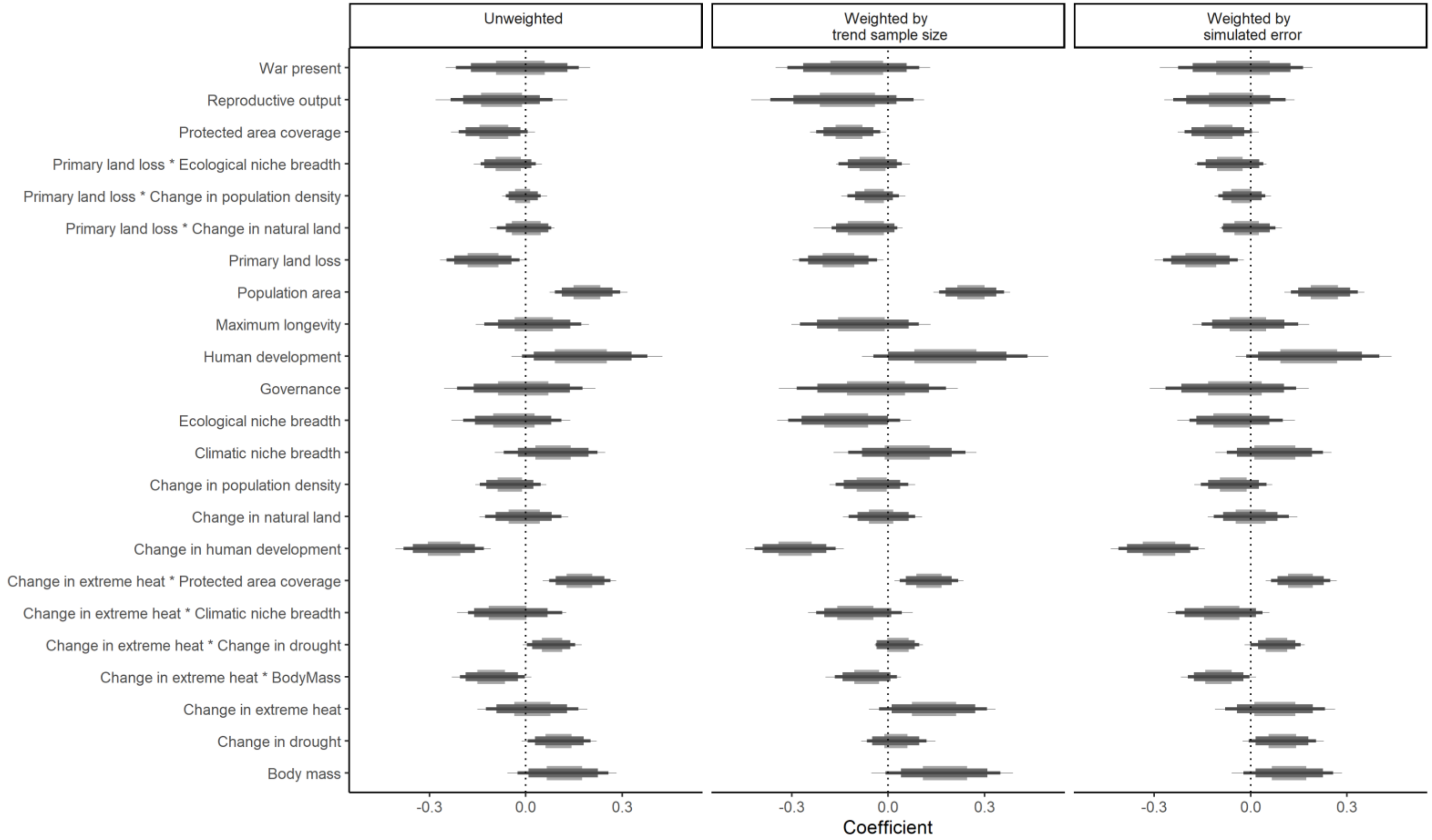


Figure S1. Standardized slope coefficients for the 23 main effects and interactions on the annual rate of change, comparing three models with different types of trend weighting: 1) trend values are unweighted; 2) trend values are weighted by the sample size (frequency of abundance observations used to derive trend); and 3) trend values are weighted by the simulated error. The four widths of the error bars represent different credible intervals: 50% (thickest), 80%, 95%, and 97.5% (thinnest)

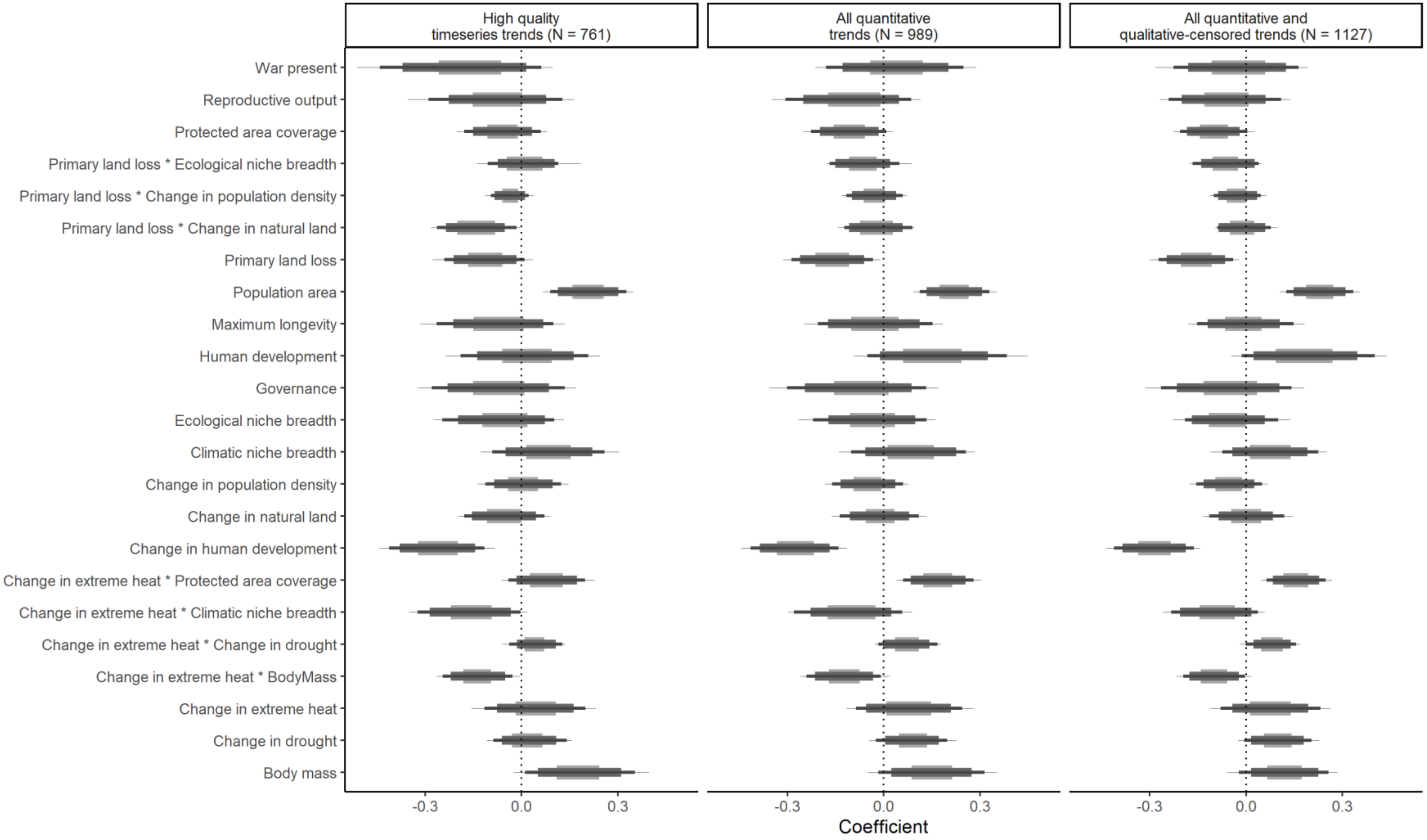


Figure S2. Standardized slope coefficients for the 23 main effects and interactions on the annual rate of change, comparing three models with different levels of inclusion for the trend data: 1) all timeseries trends with at least three abundance values are used; 2) all quantitative trend values are used; and 3) all trend values are used. The numbers in brackets alongside the facet titles describe the sample size of trends in the model. The four widths of the error bars represent different credible intervals: 50% (thickest), 80%, 95%, and 97.5% (thinnest)

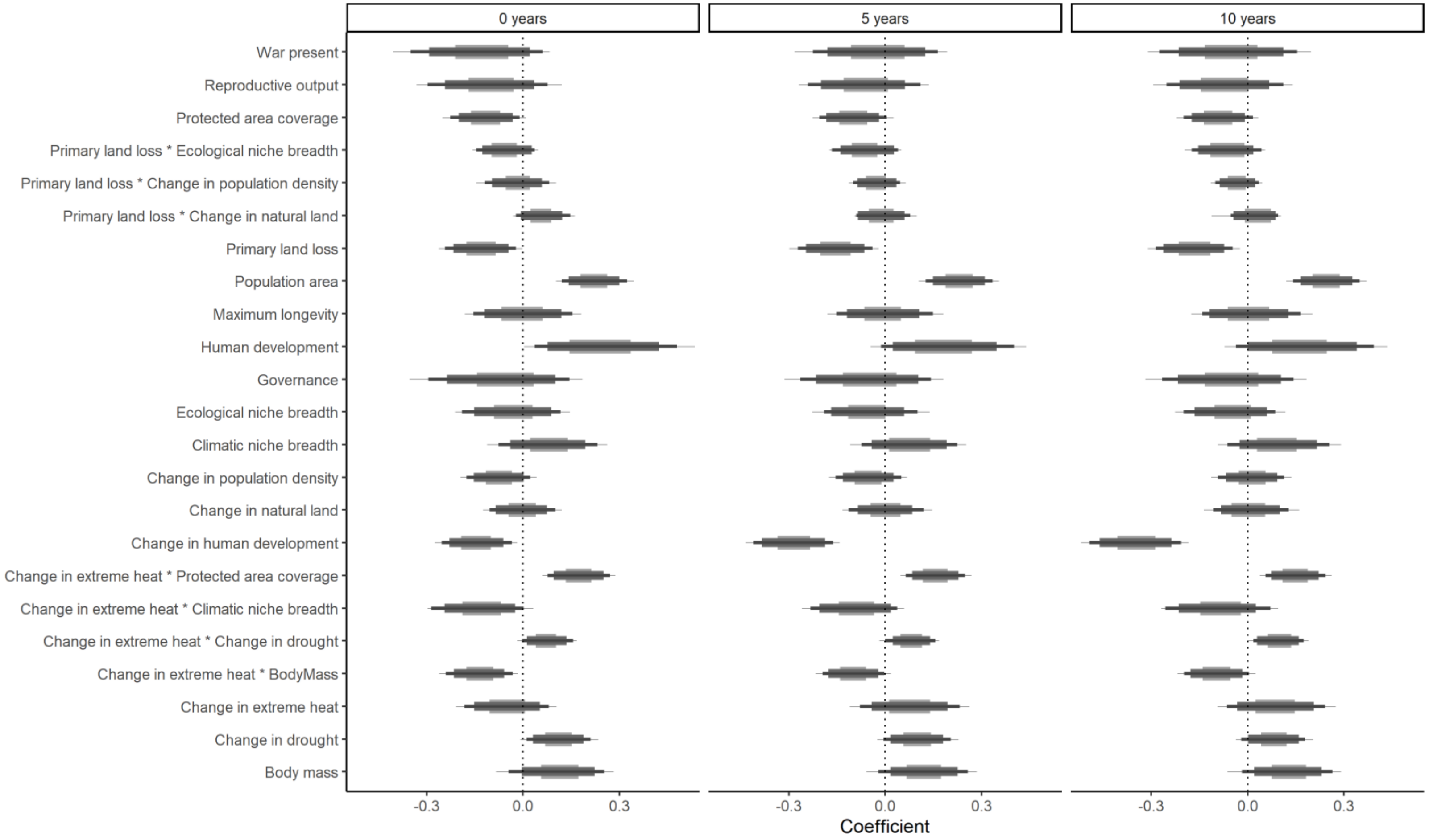


Figure S3. Standardized slope coefficients for the 23 main effects and interactions on the annual rate of change, comparing three models describing different lags of the model covariates: 1) 0 years; 2) 5 years; and 3) 10 years. The different lag periods only effect covariates that measure a change in the covariate over time. For example, for a predator population monitored between 1995-2000, the Change in human density would be measured from 1995-2000, 1990-2000, and 1985-2000, respectively under the 0, 5, and 10-year lags. The four widths of the error bars represent different credible intervals: 50% (thickest), 80%, 95%, and 97.5% (thinnest).

Model selection and effects

We recorded 1967 unique covariate parameter combinations across the 30,000 stored model iterations. However, 176 combinations accounted for the majority (80%) of the iterations (Figure S4).

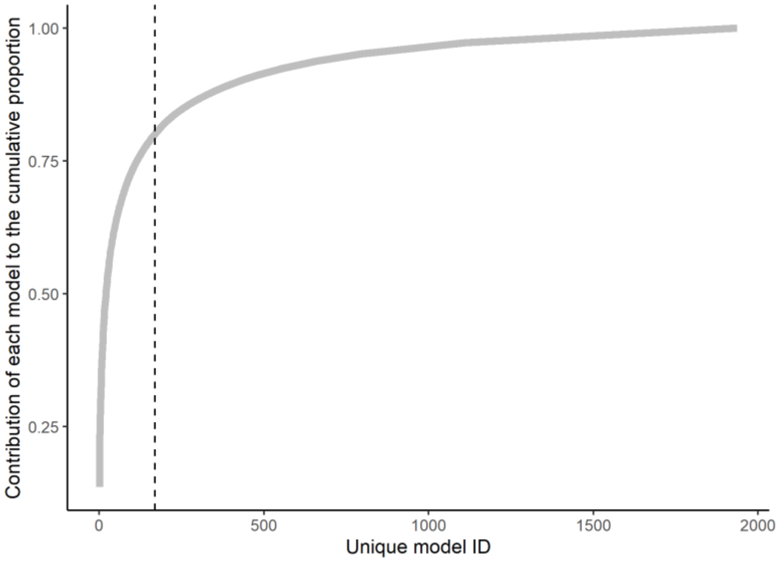


Figure S4. From the 30,000 stored iterations, we recorded 1967 unique model covariate/parameter combinations, most were rare and 80% of the iterations represented just 176 parameter combinations. We only report the posterior distributions and effects from this 80% subset.

Some parameters occurred far more frequently than others. Change in human development was the most common optional parameter, occurring in all the selected iterations. Whilst Change in extreme heat * Protected area coverage was the most common interactive parameter (Figure S5). Two parameters, the interaction of Primary land loss with Change in natural land and Ecological niche breadth, were present in none of the selected iterations.

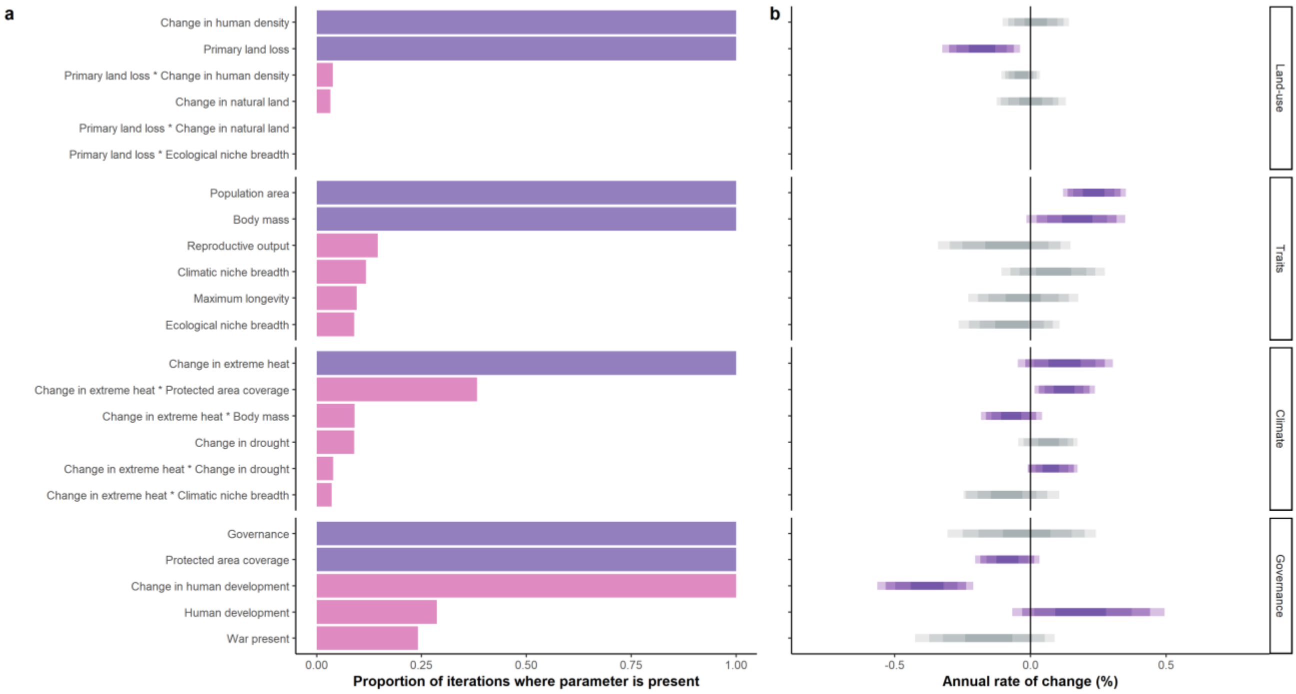


Figure S5. a) Proportion of iterations where parameter is present in the stored 80% of model runs. Core parameters (purple) are present in all models, whilst optional and interactive parameters (pink) only occur in models when they are selected from Bernoulli distributions. b) Standardised coefficients of parameters main effect and interactive model parameters. Parameters in purple have an effect at the 80% credible interval, whilst those in grey do not. The different credible interval thresholds are shown for each parameter, with darkest centre representing the 50% credible intervals, followed by the 80%, 90%, and 95% thresholds (the maximum and minimum point on each bar).

Model assumptions and checks

The inference model passed all standard linear mixed effect model assumptions, with residuals not showing signal of spatial or phylogenetic autocorrelation (Figure S6). However, the inference model failed to represent the more extreme observed annual rate of change (%) values, with the predicted rates of change largely ranging from -2 to 2, whilst the observed rates of change range -5 to 5; both on the inverse hyperbolic sine scale (Figure S7a). The qualitative-censored predictions largely agreed with the observed data, where censored-increasing values where primarily predicted to increase, and censored-stable values exhibited small increases and decreases (Figure S7b). The only category with reasonably poor alignment was censored-decreasing, where populations were predicted to be both increasing and decreasing. Our posterior predictive checks suggest the model produced broadly plausible values, with the independently simulated values occurring within the distribution of the observed trends, albeit failing to represent extreme values (Figure

S7c). For the qualitative-censored trends, the quasi-observed values matched the simulated values almost identically (Figure S7d), which is to be expected. The inference model had a median root mean square error of 9.2%, a median marginal R^2 of 0.2, and a median conditional R^2 of 0.5 (Figure S8).

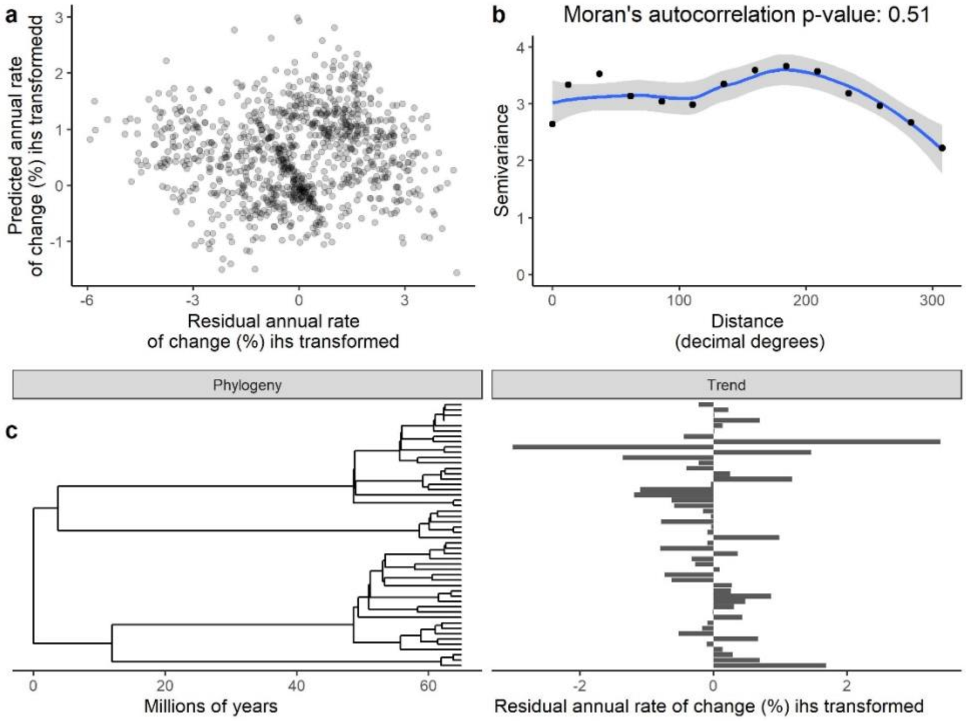


Figure S6. a) Median predicted annual rate of change (%) values from the inference model plotted against the median residual rates of change (%), both displayed with an inverse hyperbolic sine (ihs) transformation – the transformation used on the annual rate of change (%) within the inference model. b) Semivariance and Moran's autocorrelation of inference model's median residual annual rate of change (%) across distance/space (decimal degrees). c) The median residual annual rate of change (%) averaged (mean) across each species, plotted on the species' phylogeny to test for phylogenetic autocorrelation. The annual rate of change (%) is ihs transformed.

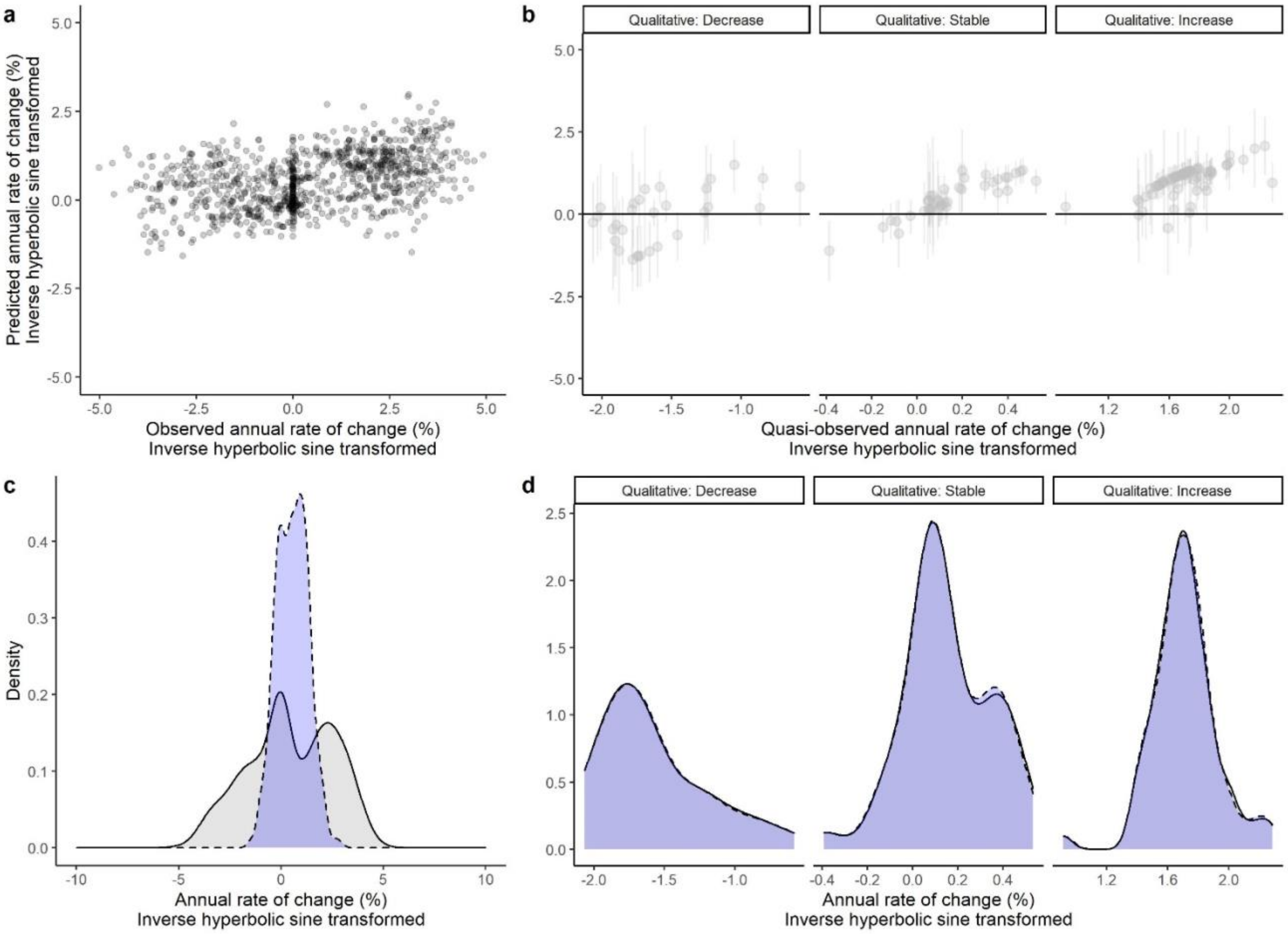


Figure S7. a) Median predicted annual rate of change (%) values from the inference model plotted against the observed rates of change (%), both displayed with an inverse hyperbolic sine (ihs) transformation – the transformation used on the annual rate of change (%) within the inference model. b) Median predicted annual rate of change (%) values (and 95% credible intervals) from the inference model plotted against each category of qualitative-censored values (median quasi-observed rates of change) - both displayed with an inverse hyperbolic sine (ihs) transformation. Values are quasi-observed as the true observed values are unknown. c) Distribution of observed annual rates of change (grey), compared to model simulated median annual rates of change (blue). d) Distribution of median quasi-observed annual rates of change (grey), compared to model simulated annual rates of change (blue).

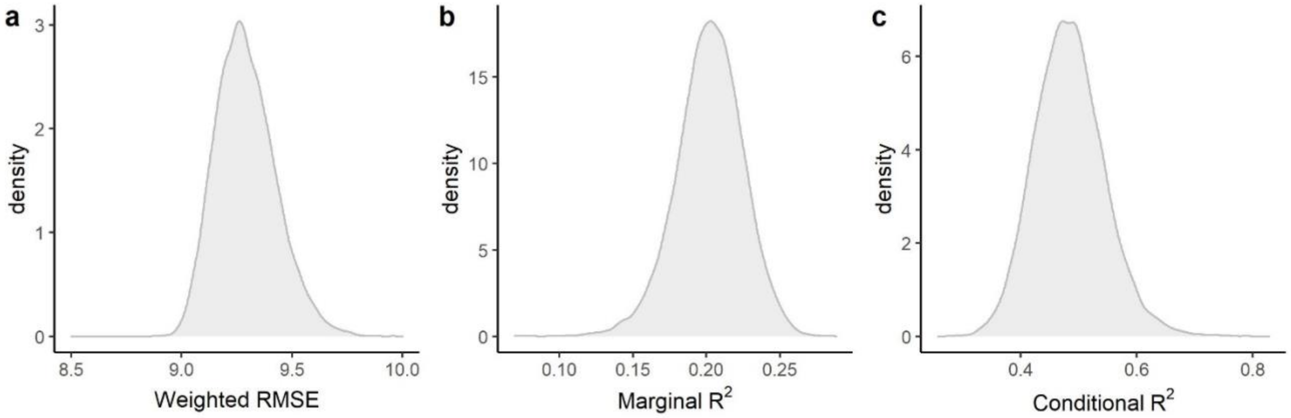


Figure S8. a) Distribution of the weighted root-mean-square error of the annual rate of change (%) in the inference model, comparing true to predicted values. b) Distribution of the inference model's marginal R^2. c) Distribution of the inference model's conditional R^2.


A. Cardwell

Chapter 5: `classecol` - classifiers to understand public opinions of nature

Authors: Thomas F. Johnson[1], Hebe Kent[1], Bethan M. Hill[1], Georgia Dunn[1], Leonie Dommett[1], Natasha Penwill[1], Tom Francis[1], Manuela González-Suárez[1]

Corresponding author: Thomas Frederick Johnson (Thomas.frederick.johnson@outlook.com)

1: Ecology and Evolutionary Biology, School of Biological Sciences, University of Reading, Reading, UK

Author contributions: TFJ developed the classification protocol, which was reviewed by MGS. TFJ, LD, GD, TF, BH, HK, and NP labelled the training datasets. TFJ developed and refined the classification models, and prepared the first manuscript draft. All authors critically reviewed the manuscript.

Citation: **Johnson, T.F.,** Kent, H., Dunn, G., Dommett, L., Penwill, N., Francis, T., and González-Suárez, M., Accepted. classecol: classifiers to understand public opinions of nature. **Methods in Ecology and Evolution**

Abstract

1) Human perceptions of nature, once the domain of the social sciences, are now an important part of environmental research. However, the data and tools to tackle this research are lacking or are difficult to apply.
2) Here, we present a collection of text classifier models to identify text relevant to the broad topics of hunting and nature, describing whether opinions are pro- or against-hunting, or show interest, concern, or dislike of nature. The methods also include a biographical classification – describing whether the author of the text is a person, nature expert, nature organisation, or 'Other'. The classifiers were developed using an extensive social media dataset and are designed to support qualitative analysis of big data (especially from Twitter).
3) The classifiers accurately identified biographies, text related to hunting and nature, and the stance towards hunting and nature (weighted F-scores: 0.79 - 0.99; 1 indicates perfect accuracy).
4) These classifiers, alongside an array of other text processing and analysis functions, are presented in the form of an R package classecol. classecol also acts as a proof of concept that nature related text classifiers can be developed with high accuracy.

Keywords: conservation, culturomics, cultural ecosystem services, human-nature relationship, human perceptions, text classifier, sentiment analysis.

Introduction

Ecology has become more transdisciplinary to better understand our environment. For example, ecosystem services reflect health, economic and cultural values (Kareiva et al., 2011), and journals and societies want to study human relationships with nature (Gaston et al., 2019; Society for Conservation Biology working groups, 2020). This transdisciplinary shift has brought the human dimension of nature into focus, but the study of human-nature relationships largely falls outside the traditional expertise of an ecologist or conservationist, who may be unfamiliar with the available methods and data.

Social media could help us understand human-nature relationships. Historically, surveys (or other qualitative approaches) have assessed perceptions, often providing detailed understanding of the person's thoughts. Social media does not offer such detail, but is cost-effective, less time-intensive, and offers enormous amounts of information (Fox et al., 2020). In 2020, social media has become widely used in most countries, with approximately half of the world's population (and increasing) being active users (Clement, 2020). Social media captures many data types (e.g. text, photos, videos, sound and interaction networks with other people) with spatial representation and temporal timeseries that could allow holistic analyses (Toivonen et al., 2019).

In recent years, the use, and diversity of uses, of social media analysis across the environmental sciences has rapidly increased (Ghermandi & Sinclair, 2019). Social media has been used to develop species distribution models (August et al., 2020), measure aesthetic and recreational ecosystem services (Van Zanten et al., 2016; Graham & Eigenbrod, 2019), track illegal wildlife trade (Di Minin et al., 2018), and determine the role of wildlife in nature-based tourism (Hausmann et al., 2017). The abundance and availability of data on these platforms – many now 15 years old, open the door for more research. Analyses of social media could revolutionise our understanding of the human-nature relationship and how it impacts the environment, but this requires new and improved tools (Toivonen et al., 2019).

There are many approaches to 'mine' opinions and gain insights from text data (Aggarwal & Zhai (2013). For example, sentiment analysis aims to understand the emotion of a text, often classifying the text's language-use as negative, neutral, or positive (Liu, 2020). This can be done with machine learning approaches, but a more readily accessible approach for interested ecologists and conservationists would be lexicon-based sentiment analysis. Lexicon-based approaches assign scores to words to calculate an average score for a text passage e.g. if more negative words are used, the text will be labelled as negative. Overall scores are effective for describing sentiment, but meaning may be unclear (Mohammad et al., 2017; Aldayel & Magdy, 2019). For example, lexicon-based sentiment analyses would return negative scores for these two messages ‘It is sad that Pangolin are vanishing’ and ‘Pangolins are bad’ (both use negative language), failing to recognize that only the second message indicates a dislike for pangolins. Furthermore, in some lexicons, species names can have negative scores (e.g. ‘shark’) which can bias results if we are interested in human-nature relationships (Lennox et al., 2020).

Stance analysis is an alternative approach (Srivastava & Sahami, 2009; Aggarwal & Zhai, 2013; Liu, 2020), more targeted towards assessing opinions about topics or specific questions. Stance analysis could help recognize the dislike of pangolins in the example above, but this method is often time-consuming to develop as it requires large training datasets alongside complex machine learning models. Furthermore, the generality of the stance analysis models can be low. For example, if a stance analysis model was built to detect fondness of pangolins, it may be of limited use for other species. So whilst stance analysis gets far closer (relative to lexicon-based sentiment analysis) to understanding a user’s opinion, for it to be useful, it would also need to be derived from a broad array of training data themes and answer general and pertinent questions.

`classecol` description

With the massive growth in social media analysis, and especially in studies using text data to look at people’s perceptions of and relationships with nature (Ghermandi & Sinclair, 2019), there is a great need for text analysis tools (Toivonen et al., 2019). To meet this demand, we present `classecol` a text cleaning, processing, and classification tool to support analysis of public opinions of nature in a big data setting. `classecol` avoids the interpretation issues of sentiment analysis and the specificity issues of stance analysis. `classecol` can identify relevant texts, describe their stance, and determine the type of user that produced the text. This provides a proof of concept to guide and encourage further text analysis development for ecology, and we hope other groups developing classifiers would consider uploading them to our package - becoming formal contributors (see package vignette). `classecol`‘s ten text classifiers, have been trained and tested on Twitter data, and fall within three topics:

1. Hunting – Are texts discussing the hunting of wildlife? If so, what’s the user’s opinion e.g. pro or against hunting?
2. Nature – Are texts relevant to nature? If so, what’s the user’s opinion e.g. expressing interest, concern, or dislike of nature?

3. Biographical (bio hereafter) – Is the author of the text a person? If so, is that individual a member of the general public or an individual discussing nature in a professional or academic capacity?

Developing classifiers

Prior to developing the ten classifiers in the `classecol` collection, we developed base classifiers for each of the three topics following eight steps: 1) Defined a protocol to describe the criteria text must meet to fall in a category (e.g. What text characteristics distinguish pro- and against-hunting?). 2) Ensured the human classifiers could accurately and consistently use the protocol. 3) Seven individuals classified 1,100 texts for each topic (tweets for hunting and nature, and user provided descriptions for bio) creating a training dataset of 7,700 texts per topic. 4) Built six text classification models for each topic including multinomial logistic regression, support vector machines, naïve Bayes, random forest, K nearest neighbour, and a four-layer neural network. A logistic regression was then used to merge the outputs from these models generating an ensemble text classifier. 5) Tested the performance of the ensemble model and identified cases of misclassification to refine the protocol and classification criteria. 6) Corrected misclassified training texts using the refined protocol. 7) Finalised the classification protocol. 8) Tested different text cleaning options (e.g. from raw text to very clean text – see Table S1) to identify that which maximised ensemble model precision and recall (both defined below). These eight steps are further detailed in Supplementary material: Developing classifiers.

In the final protocol there are three categories for the hunting topic and four for the nature and bio (one added during the reclassification steps) topics:

Hunting

1. Irrelevant – text does not discuss the hunting of animals.
2. Pro-hunting – text indicates support for hunting.
3. Against-hunting – text indicates opposition to hunting.

Nature

1. Irrelevant – text does not discuss nature or nature related activities.
2. Pro-nature (positive phrasing) – text endorses nature with positive language e.g. interest.
3. Pro-nature (negative phrasing) – text endorses nature with negative language e.g. concern.
4. Against-nature – text indicates opposition or frustration towards nature e.g. fear.

Bio

1. Expert – user has professional status, or qualifications to indicate expertise, in nature or a nature related field.
2. Person – user is an individual without nature expertise.
3. Nature org (added) – user is an organisation, company or group working in a nature related activity.
4. Other – user is none of the above.

Classifier accuracy

We report the F-score (Zhang & Zhang, 2009) accuracy of each category in each classifier, and an overall accuracy per classifier (average F-score weighted by the proportional abundance of each category). Accuracy was measured on an independent data sample i.e. not used to develop the classifiers. F=1 indicates perfect classification.

The hunting classifiers had high overall (0.87 – 0.97) and category accuracies (Figure 1), except for Irrelevant, where lower accuracy (0.64 - 0.72) was driven by low recall (0.54 – 0.61). Nearly half of the Irrelevant texts were assigned to the wrong category. In the nature classifiers, overall accuracies ranged from 0.82 – 0.92, with moderate to high accuracy across all categories except Pro-nature (negative phrasing) and Against-nature in the ‘full’ model. Against-nature had low model recall (0.67) and precision (0.4), probably because this category only represented 1.1% of all classifications. This low coverage could make the model unreliable, which may explain why Pro-nature (negative phrasing) also had low accuracy in the ‘full’ model, despite good accuracy in other models. Given this finding, we removed Against-nature from the stance and trimmed models and would recommend using the trimmed over the full model. Finally, in the bio models, overall accuracies ranged from 0.79 – 0.87, with moderate to high accuracy in all categories. All topics are characterised in Figures S6-8.

Using `classecol`

Prior to data collection and analyses, any research project involving public opinion should consider the legal and ethical requirements – see Data rights and ethics in the supplementary material.

The `classecol` functions fall into two groups: 1) general text cleaning and analysis and 2) text classification. The first group includes five functions of value for anyone interested in natural language processing. The `clean` function provides comprehensive text cleaning options, including the conversion of common emoticons, abbreviations, slang, and environmental related hashtags into readable text. `valence` detects the presence of terms that can alter, reverse, or amplify meaning. `contract` performs word stemming and lemmatisation to reduce term complexity (e.g. consulting becomes consult). `lang_eng` detects whether the text contains English terms or not. Finally, `senti_matrix` pulls together 11 popular sentiment analysis approaches into one function, to produce a matrix of average sentiment scores for each sentence. All of these functions can be used in conjunction, for example, to assess sentiment analysis of some text, you may use `lang_eng` to remove non-English terms, then `clean` and `contract` the text, before running the `senti_matrix` function.

Our second group of functions are the most important component of `classecol`. These text classifiers are processed through a Python backend, thus require downloading and installing Python (we recommend version 3.6). This can be done automatically in R through the `addeR::py_download` function (Johnson, 2021). The `load_classecol` function then automatically downloads the text classification models and Python module dependencies. `load_classecol` also links R to the Python backend and needs to be run every time a new R environment is loaded; the text classification models and Python modules will only need to be downloaded once. The `hun_class`, `nat_class`, and `bio_class` functions perform the text classifications in the hunting, nature and bio topics, respectively. Prior to using the classifiers, we recommend running `clean(level = “simple”)` for `hun_class` and `clean(level = “full”)` for `nat_class`, but no cleaning is required before using `bio_class`. `nat_class` also requires a matrix of valence and language indicators, as well as sentiment scores for each text record (see package vignette on *https://github.com/GitTFJ/classecol*).

The `hun_class`, `nat_class`, and `bio_class` functions each contain multiple text classifiers which could be valuable in different scenarios (Figure 1). For `hun_class`, the relevance model identifies whether text is relevant or irrelevant to hunting, stance classifies relevant texts as pro- or against-hunting, and full runs both relevance and stance. Similarly, for `nat_class`, relevance identifies whether text is relevant or irrelevant to nature, stance identifies whether relevant pro-nature texts are using positive or negative phrasing, and the trimmed model combines both. `nat_class` also has a full model which includes the low-accuracy Against-nature category, which should be used with caution. Finally, for `bio_class`, the person model identifies whether a user is a person or not, expert classifies persons as nature experts or general public, and full combines both, and adds the additional ‘Nature organisation’ category.

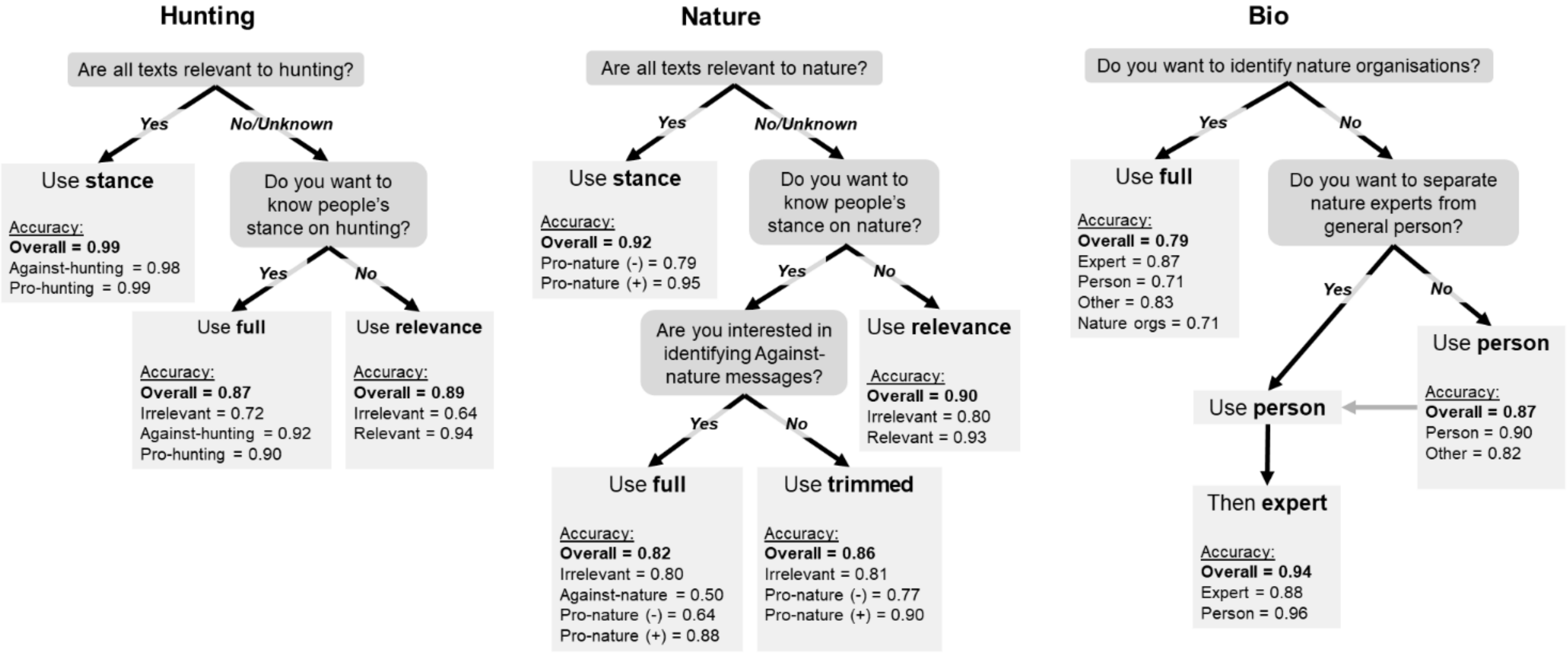


Figure 1. Flowchart to assist in selecting a suitable `classecol` classifier for each of the hunting, nature and bio topics. Flowchart questions are depicted in dark grey boxes with rounded edges, and classifier options are in the lighter shade of grey. The bold text in the classifier boxes describes the classifiers name and overall accuracy. Accuracy (measured as the classification F-score, a value of 1 is perfect classification accuracy) is also broken down into each classifier category.

Classifiers can be used hierarchically (e.g, use relevance followed by stance) rather than using the combined classifiers. This increased computational processing time but had little impact on accuracies, except in the bio model, where accuracy is improved by using the person classifier, followed by the expert classifier. Classifiers can also be stacked. For example, to explore the general public's stance towards hunting in the USA, we could remove non-English texts with `lang_eng`, identify members of the public with `bio_class(type = "full")`, and then determine hunting stance with `hun_class(type = "full")`. When running any of the text classifiers, we recommend manually classifying a sample of your data so classification accuracy can be determined.

`classecol`'s suite of text processing, analysis, and classifier functions can assist academics and policy-makers interested in exploring the human dimensions of nature in big data. This research theme, and in-turn `classecol`'s value, extends far beyond the fields of ecology and conservation, with social scientists, human geographers, and environmental scientists all working with human-nature relationship data. `classecol` proves that moderate to high accuracies can be achieved from text classifiers and we hope this will inspire future classifier development (methods and code are openly available). Admittedly, there are time-costs to consider as supervised classifiers like `classecol` require lengthy training datasets, which are laborious to compile, and as mentioned earlier, can lack generality. Whilst we have designed classecol across a broad array of training data themes, its generality (or accuracy) across different data-types is unknown. `classecol` should be used cautiously on non-Twitter data, and a sample of data must always be manually classified (by a human) so accuracy can be tested.

Despite hundreds of studies in environmental sciences using social media analysis, there is a scarcity of method comparison and testing which means the accuracy and representativeness of these text analysis tools remains largely unknown, and could be error-prone. For example, when we measure sentiment analysis scores for texts in our human-classified hunting and nature stance data, we may expect sentiment analysis to detect the opposing hunting stances, or the opposing language use in pro-nature tweets i.e. Against-hunting tweets would primarily have negative scores, and Pro-hunting tweets would have positive scores. However, the sentiment scores between the categories largely overlap in both the hunting and nature topics (Figure 2). Sentiment approaches were unable to distinguish the classifications and detect our stances (lexicon-based sentiment analysis can only describe the text's polarity, not infer meaning). To ensure social media data is used robustly in the environmental sciences, its pivotal that methods are tested and frameworks for analysis are developed.

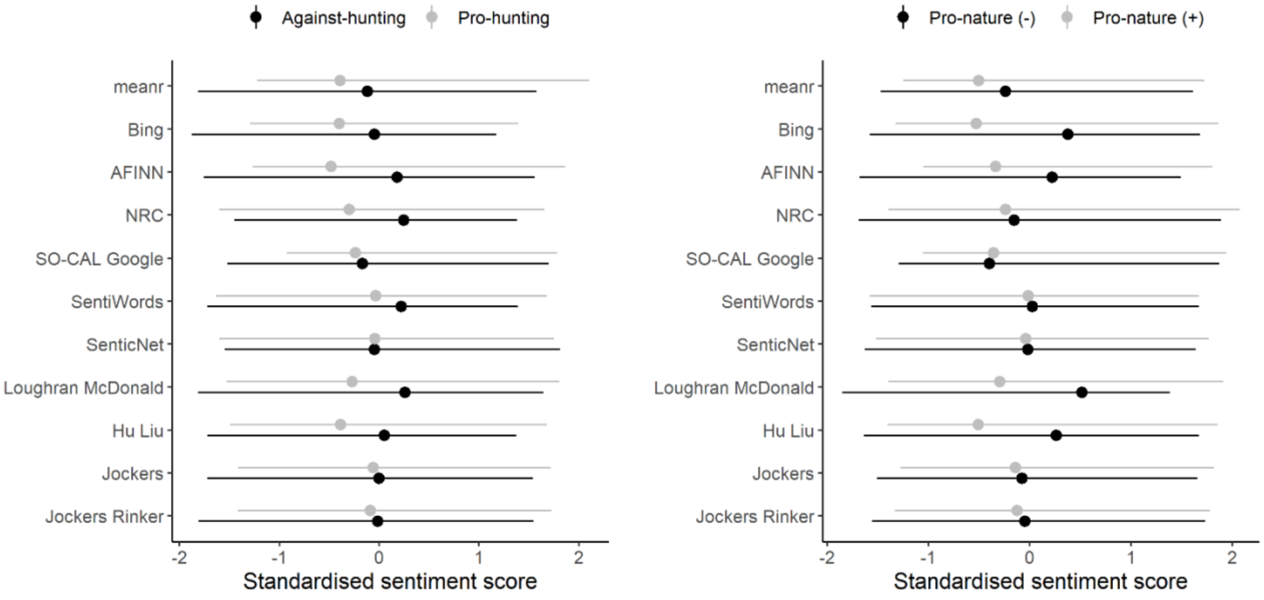


Figure 2. Assessment of 11 sentiment approaches ability to distinguish between the hunting and nature stances. The points represent the median sentiment score and error bars are the 95% quantiles [2.5%, 97.5%], displayed for each approach in each stance. If the sentiment analysis approaches were able to distinguish between the stances, we would expect to see little to no overlap in the black and grey points.

Big-data culturomics within the ecological and conservation sciences are already reliant on transdisciplinary work involving social science. Transdisciplinary research is key to harnessing the data's massive potential, but requires careful method development and testing. This scrutiny extends onto `classecol` for which next steps include further testing of the text classifiers especially on non-Twitter data. The full potential of `classecol`, to our knowledge the first publicly available text classifier of opinions on nature, is yet to be explored, but we hope this tool will be the first of many in a growing community.

Code and data availability

Twitter terms and conditions prevent sharing of the training data. Code to develop classification models at https://github.com/GitTFJ/classecol_dev and the classecol R package is located at https://github.com/GitTFJ/classecol

Acknowledgements

Thanks to RC and two reviewers for valuable feedback. MG for the package name. Funding to TFJ from NERC (Natural Environment Research Council) Centre for Doctoral Training studentship (J71566E), and to MGS from The Royal Society (IE160539).

Supplementary material

Developing classifiers

Prior to developing the 10 models in the `classecol` collection, we created base models in each topic, with the aim of identifying and addressing possible classification issues. These base models attempted to predict every category within the topic and were developed following eight steps (Figure S1): 1) Protocol – criteria for classifying tweets into categories for each topic; 2) Test classifiers – assess accuracy of human classifiers; 3) Classify – assign categories to text to create training datasets; 4) Models – build models with training data; 5) Test models – assess model accuracies and inadequacies; 6) Reclassify – correct misclassifications in the training data; 7) Final protocol – amend protocol to incorporate changes identified in steps 4 - 6. 8) Cleaning – improve model accuracy by cleaning text to pull out important features. Each of these eight steps are described in detail below.

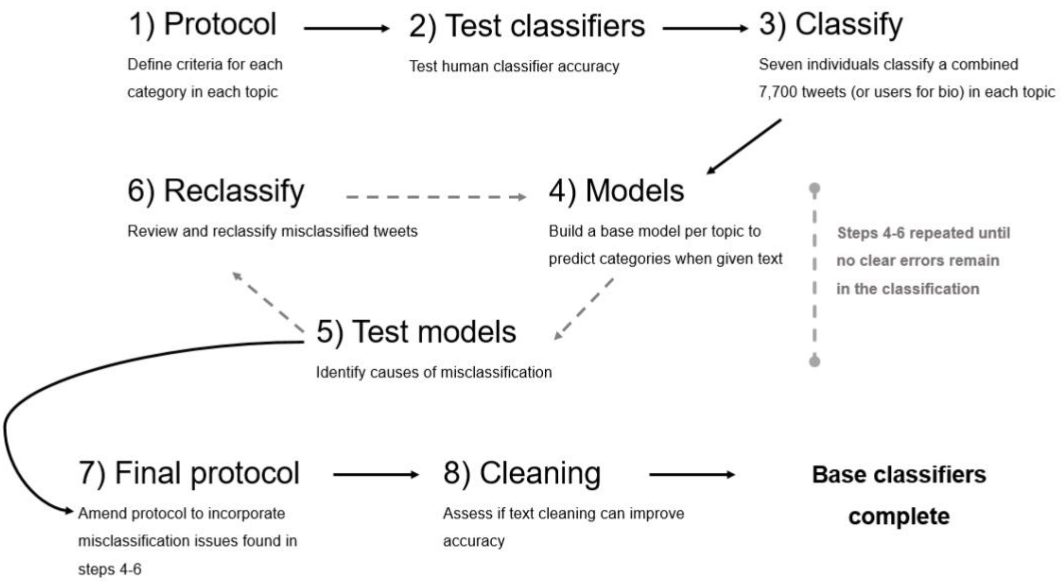


Figure S1. Eight steps to developing base classifiers. In steps 4-6, the dotted grey arrows indicate that these steps are iteratively repeated until no clear errors remained in the training datasets classifications. We define ‘clear errors’ as those which wrongly characterise the category according to the protocol.

Protocol

For each topic, we developed a series of classifications and specified the criteria the text must meet to be assigned to each classification. These classifications were determined by examining tweets under common hashtags (#hunting and #nature) and the associated user biographies to determine how the text could be split.

Importantly, this was just an initial criterion which was later amended after identifying inconsistencies – see 'Final protocol' below**.**

In the hunting classifier, tweets were assigned into one of the following classifications using only the tweet's text:

- Irrelevant – tweet discusses the hunting of something other than animals (e.g. #hunting for car keys), or contains insufficient information to determine the classification (e.g. text just says '#hunting'), or mentions hunting but the primary focus is on something else (e.g. bass fishing holiday #fishing #hunting).
- Against-hunting – tweet opposes the hunting of animals (e.g. poor creatures #bantrophyhunting).
- Pro-hunting – tweet describes involvement in the hunting of animals (e.g. #hunting for geese), or the support for hunting related activities (e.g. amazing lion head #trophy).

For the nature classifier, tweets were assigned into one of the following classifications using only the tweets text:

- Irrelevant – primary content of the tweet is not nature related (e.g. I love the new #Lionking).
- Against-nature – tweet describes nature as a bad thing (e.g. we need to control foxes, they are destroying my garden! #wildlife).
- Pro-nature (positive phrasing) – tweet describes nature as a good thing and primarily uses positive phrasing, showing an interest or curiosity in nature (e.g. Planet Earth is so cool, look at these forests #wildlife #nature).
- Pro-nature (negative phrasing) – tweet describes nature as a good thing and primarily uses negative phrasing, showing concern for nature (e.g. The destruction of our forests is absurd #nature). Note the use of the term primarily in the phrasing aspect of the Pro-nature categories. In some cases, nature may be described with both positive and negative phrasing, and in this case, the text should be categorised as the most frequent phrasing type (e.g. '*I love nature*, that's why I am so devastated by these dreadful wildfires #saveourspecies before it is too late' has both *positive* and negative phrasing, but the negative phrasing is most frequent, and so the text should be classified as Pro-nature with negative phrasing).

For the bio classifier, users were assigned one of the following classifications based on their screen name and biographical description:

- Expert – user labels themselves as having expertise or training in a nature related field (e.g. Graduate of nature-related degree), or describes a relevant job role (e.g. National park warden), or affiliation (e.g. chair of RSPB).

- Person – user appears as a personal account (e.g. Jim Bloggs 26, nurse and rock-climber).
- Other – covers all accounts falling outside the Expert and Person groups, including parody and alias accounts, as well as accounts with insufficient information.

Test classifiers

To ensure the protocol was well understood and human classifiers would judge tweets (or users for bio) consistently, each individual conducted a test where they had to classify a selection of text passages and describe why they made each choice. The test included questions on the hunting (six questions), nature (six questions), and bio topics (seven questions), and covered each of the classification categories. This test was conducted before individuals classified any text (pre-classification) and after (post-classification) to assess consistency in individual understanding of the protocol over time. After the pre-classification test, each individual was given a detailed report of their test results and a description of what the correct answers were and how those were obtained. From this test, we identified that records did not always fall within one category and sometimes touched on many. As a result, we updated the protocol so records were assigned to the most aligned category, or left unclassified if not well-aligned to any category.

Classify

For each topic, we randomly extracted Twitter records from three existing datasets managed by TFJ. The hunting and nature datasets contain millions of tweets which were queried through the Twitter API. The hunting dataset compiled all tweets mentioning any of the following terms: #Hunting, #TrophyHunting, #StopHunting, or #BanTrophyHunting. The nature dataset searched a more broad array of terms, including species names, celebrated events, natural and manmade disasters, and more specific environmental terms: #BiodiversityDay, #BigCats, #Biodiversity, #Birding, #Cheetah, #ClimateChange, #Conservation, #Ecology, #Elephant, #Leopard, #Lion, #Poaching, #PhD, #Nature, #Shark, #Sustainable, #SaveTheAnimals, #Tiger, #Wildlife, #WildlifeCrime, #WildlifeTrafficking, #WildlifeTrade, #Wildlife, #WWD, and #WorldWildlifeDay. The bio dataset contains a list of the name and description of millions of Twitter users. The hunting and nature classifications are solely based on tweet content, whilst the bio classification is based on the concatenation of the name and biographical description in each user. All tweets and user bios were downloaded between March 2018 and July 2019.

Records were classified by seven assessors (all authors except MGS). Each assessor was assigned 1,000 tweets per topic that were only scored by them, and

100 tweets per topic that were scored by everyone to check for agreement in classification and potential issues in the criteria. In total each assessor classified 1,100 per topic, resulting in three training datasets, each with 7,700 classified tweets and bios. The training datasets included 7,100 unique tweets, but because 100 tweets had been independently classified by each of the seven assessors, the total number of classifications was 7,700. All records were classified between July and October 2019.

Models

We developed a series of text classification models to predict the defined categories from the tweet or bio datasets. The same process was completed for each topic. Models had two levels, in the first level we developed multiple text classification models, in the second level we pooled the results across the multiple models. Before specifying any models, we randomly split the labelled training data into groups: a training set for the first level models (71%), an independent set to validate first level neural network fit (4%), a training set for the second level models (18%), and an independent set to test model accuracy (7%).

For the first level models, we took the first level training set (71%), conducted a basic clean of the text (see '*Cleaning*' below), and calculated the term-frequency inverse-document-frequency (tf-idf) for each tweet or bio, where each term represents a single word i.e. we only used unigrams. This transformation converts the text string into a matrix which described the relative frequency of each word within each text passage. The human-classified text label was then modelled against the tf-idf matrix under six model types: multinomial naïve Bayes, support vector machine with a stochastic gradient descent and hinge loss function, random forest, multinomial logistic regression, K nearest neighbour, and a four-layer neural network. All methods except the neural network were completed using scikit-learn Python package (Pedregosa et al., 2011), the neural network was developed in keras (Chollet, 2015) and tensorflow (Abadi et al., 2015). The default options were used for all models except for the neural network, where we limited the tf-idf matrix to the 1,000 most frequent features (or words). Limiting the number of features substantially reduced processing time, and trial runs showed it did not noticeably reduce model accuracy. The neural network had four layers: an input with 1,000 nodes, two hidden layers with 512 and 128 nodes, and an output layer with nodes equal to the number of classification categories. All nodes were linked through a fully connected (dense) network. In both hidden layers, we set the dropout at 0.2 to randomly block nodes and reduce the effect of overfitting. In the input and both hidden layers, we used a rectified linear unit activation function, and in the output layer we used softmax activation. Finally, we used a categorical cross-entropy loss function and adam optimizer within the neural network, and specified a batch size of 256 with 30 epochs. These epochs were later trimmed to between 3 and 4, to maximise accuracy and minimise loss in the 4% model validation set (Figure S2)

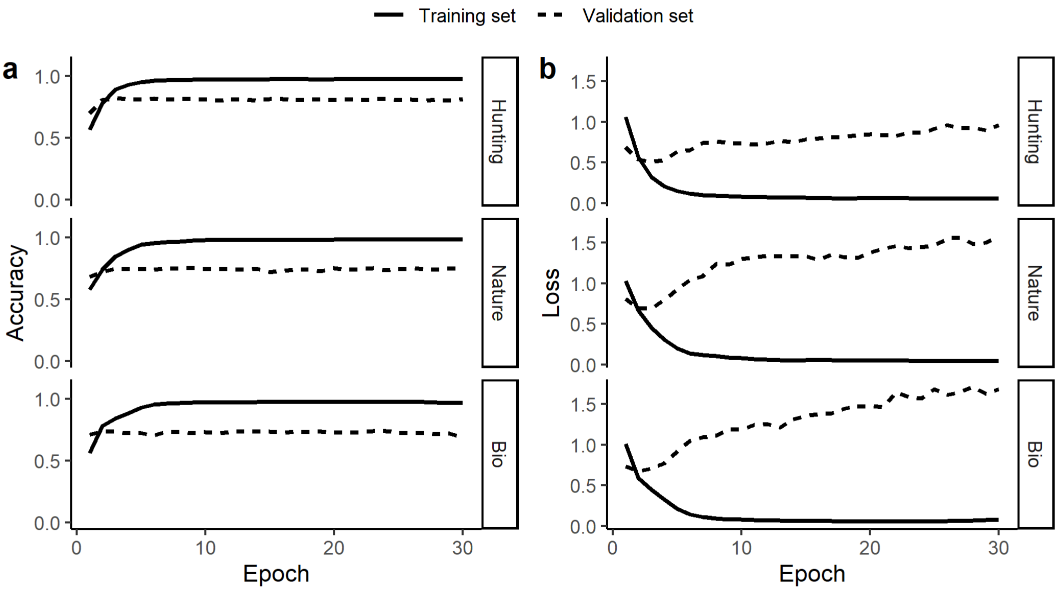


Figure S2. Change in base classifier accuracy and loss in the neural network text classifier. Accuracy and loss compared for the training set (71%) and an independent validation set (4%), across 30 epochs, with a batch size of 256.

There is evidence that using multiple machine learning models and pooling results through ensemble modelling can improve predictive accuracy (Thorne et al., 2018). As a result, for the second level models, we used a ‘blending’ ensemble approach where a new sample of data (18%) was run through the six models, producing a matrix of classification probabilities for each tweet or bio. This matrix was then run through an additional multinomial logistic regression, with the human-classified categories as a response, developing a model which can take a matrix of classification probabilities and predict the category. The classification accuracy of the six text models and the ensemble are tested on a final independent sample of the data (7%).

Test models

We examined a word cloud representation of each base model’s testing dataset (7%) confusion matrix to assess which terms may be wrongly influencing the category. We then searched for these terms in the full dataset to identify incidences where the assessor selected the wrong category, and incidences where assessors diverged in views on the category because the protocol was insufficient. We iteratively ran the models after each reclassification step (see ‘*Reclassify’* below) and reviewed potential problematic terms.

Reclassify

The purpose of the reclassify step was to remove classification errors and inconsistencies, as well as making sure each classification was representative of the protocols criteria e.g. the classification should align with the criteria. We addressed reclassifications in three ways: 1) searched frequent terms occurring in the misclassified categories (e.g. classified Irrelevant, but model predicted as Pro-hunting); 2) read through categories with a high a frequency of misclassifications to detect the root cause of any issues; 3) decided the protocol was insufficient and conducted a full review of all texts to amend issues. Finally, in cases where we felt there was insufficient data (e.g. the expert category in the bio classifier) to accurately characterise any classification, we compiled new records (where possible) that could bolster the category. We anticipated that correcting misclassifications may improve model accuracies, but this was not the purpose of the reclassifications. As a result, we only assessed the change in model accuracy (related to the reclassifications) after deciding we were happy the training dataset was aligned and representative of each classification and the overall protocol. We repeated the models, test models, and reclassify steps until we were content with the protocol.

Hunting reclassifications:

1) In the first review, preppers, snares and survival were assigned as Irrelevant in some cases and Pro-hunting in other cases. The protocol had no rules regarding these terms, which post-hoc were decided they should fall within the Pro-hunting category, not Irrelevant. As a result, all records containing 'prep' ,'snar', or 'surv' were reviewed, and reclassified where necessary. Outdoor also frequently appeared in the Irrelevant and Pro-hunting categories, so we reviewed all tweets mentioning 'outdoor', but found no large-scale misclassification issues and deemed the classification protocol sufficient. In the first review, 1.9% of tweets were re-classified.
2) In the second review, fishing related terms regularly occurred in the Irrelevant and Pro-hunting categories. Following the protocol, these should be classified as Irrelevant if the primary focus of tweet is fishing and #hunting is just used as a popular hashtag to improve the tweets reach. However, this was challenging to distinguish within the tweets. As a result, we reviewed every tweet containing 'fish', and if the tweet was primarily focussed on fishing, but still endorsed hunting, the tweet was re-assigned as Pro-hunting. Photography related terms also commonly occurred in the Irrelevant and Pro-hunting category, so we reviewed all tweets mentioning 'photo', but found no large-scale misclassification issues and deemed the classification protocol sufficient. In the second review, 11.2% of tweets were re-classified.

3) In the final review, tweets discussing deer were regularly classified as Irrelevant. We reviewed all tweets containing ‘deer’ and reclassified the 0.5% of tweets that were deemed to be misclassified under the current protocol. We found no additional-obvious causes of misclassification in the dataset.

Nature reclassifications:

1) In the first review, tweets discussing hunting activity were classified as Against-nature by some individuals. However, hunters may contest this classification and state they have an admiration and an interest in nature. As a result, we specified in the new protocol that hunting should not automatically be considered Against-nature and its classification should be dependent on the context e.g. ‘hunting this pesky fox’ would be Against-nature as nature is deemed pesky, however ‘hunting for natures majestic cat today #Lion’ would be deemed pro-nature (appreciation of nature and positive phrasing). We reviewed all Against-nature tweets and re-classified text following this new protocol. We also reviewed all tweets discussing ‘pest’ and ‘remov’ to assess if any tweets likely representing an Against-nature stance fell in a different category, but found no large-scale issues. In the first review, 1.4% of tweets were re-classified.
2) In the second review, we identified an issue in the original protocol, where the examples and rules described the classifier as a nature-topic, but when the tweets were classified, TFJ set the database label as wildlife, not nature e.g. Against-wildlife instead of Against-nature. We were concerned this label may have influenced the categories, so TFJ reviewed all tweets to ensure the classifier worked across the entire nature topic, not simply within the domain of wildlife. In the second review, 21.3% of tweets were re-classified.
3) In the final review, we did not find any large-scale issues but did identify a series of terms to be checked. We reviewed, and reclassified where necessary, all tweets containing any of the following: ‘conservation’, ‘sustain’, ‘water’, ‘pet’, ‘job’, ‘fashion’, ‘jewel’, ‘lion’, or ‘unicode’. In the final stage, 17.9% of tweets were re-classified. We found no additional-obvious causes of misclassification in the dataset.

Bio reclassifications:

1) In the first review, many terms were shared between the Expert and Other categories, as some ‘Other’ were leading organisations and groups in nature and the environment. As a result, we reviewed all users again and added a category ‘Nature org – nature organisation’ to capture this middle ground, re-classifying 17.8% of tweets. However, after splitting the data into these categories, we found there were too few experts to accurately predict the Expert category. As a result, an additional search was conducted by

downloading the biographies of all non-restricted accounts that the first author follows (N ~ 1,000). TFJs professional account contains a high frequency of Experts, but does not accurately represent all nature related disciplines, and so TFJ further sampled the accounts followed by 50 peers (accounts TFJ follows). This increased the number of users from 7,700 to 15,987. TFJ classified all new users and added them to the existing dataset.

2) In the second review, many of the terms we expected to be associated with the Expert or Nature organisation categories occurred in the general Other category, so we reviewed all users with the following terms in their name or description, and reclassified them where necessary: 'ecolog', 'enviro', 'climate', 'biolog', 'sustain', 'lectur', or 'prof'. In the second review, 2.3% of users were re-classified.
3) In the third review, we identified further terms that would likely only belong in Expert or Nature organisation, but fell in other categories, so we reviewed all users with the following terms in their name or description, and reclassified them where necessary: 'evol', 'conserv', 'biodiv', 'geog' and 'bird'. We also briefly scanned through the Other category to identify if people's names (e.g. John or Jane) were present, possibly indicating the user is a Person instead, and we reclassified these where necessary. Finally, there were a series of obvious errors in users where the first letter of their name began with a, b, or c. We are unclear what caused the error but reviewed each of the users to correct any misclassifications according to the protocol. In this third review, 4.7% of users were re-classified.
4) In the final review, we did not find any large-scale issues but did identify a series of terms to be checked. We reviewed, and reclassified where necessary, all tweets containing any of the following: 'author', 'teacher', 'father', 'mother', 'dad', 'mum', 'husband', 'wife', or 'gamer'. In this final review, 0.3% of tweets were re-classified. We found no additional-obvious causes of misclassification in the dataset.

Final protocol

Here, we represent the changes in the protocol, with removals crossed out and coloured red, and additions coloured blue.

Hunting

- Irrelevant – tweet discusses the hunting of something other than animals (e.g. #hunting for car keys), or contains insufficient information to determine the classification (e.g. text just says '#hunting')~~, or mentions hunting but the~~

~~primary focus is on something else (e.g. bass fishing holiday #fishing #hunting).~~

- Against-hunting – tweet opposes the hunting of animals (e.g. poor creatures #bantrophyhunting).
- Pro-hunting – tweet describes involvement in the hunting of animals (e.g. #hunting for geese), or the support for hunting related activities (e.g. amazing lion head #trophy), or the promotion of hunting (e.g. look at this cool knife and camo for just $30 #hunting), or the endorsement of hunting for survival (e.g. hunt to eat #survival), or support for hunting as an aside to fishing (e.g. bring on the #bass catch weekend of #camping #fishing #hunting).

Nature

- Irrelevant – primary content of the tweet is not nature related (e.g. I love the new #Lionking).
- Against-nature – tweet describes nature as a bad thing (e.g. we need to control foxes, they are destroying my garden! #wildlife). The action of hunting is not inherently Against-nature as hunters may appreciate nature e.g. ‘look at this majestic lion trophy #hunting’ would be classified as Pro-nature (positive phrasing).
- Pro-nature (positive phrasing) – tweet describes nature as a good thing and primarily uses positive phrasing, showing an interest or curiosity in nature (e.g. Planet Earth is so cool, look at these forests #wildlife #nature).
- Pro-nature (negative phrasing) – tweet describes nature as a good thing and primarily uses negative phrasing, showing concern for nature (e.g. The destruction of our forests is absurd #nature). Note the use of the term primarily in the phrasing aspect of the Pro-nature categories. In some cases, nature may be described with both positive and negative phrasing, and in this case, the text should be categorised as the most frequent phrasing type (e.g. ‘*I love nature*, that’s why I am so devastated by these dreadful wildfires #saveourspecies before it is too late’ has both *positive* and negative phrasing, but the negative phrasing is most frequent, and so the text should be classified as Pro-nature with negative phrasing).

Bio

- Expert – user labels themselves as having expertise or training in a nature related field (e.g. Graduate of nature-related degree), or describes a relevant job role (e.g. National park warden), or affiliation (e.g. chair of RSPB).
- Person – user appears as a personal account (e.g. Jim Bloggs 26, nurse and rock-climber).

- Other – covers all accounts falling outside the Expert, Person, and Nature organisation groups, including parody and alias accounts, as well as accounts with insufficient information.
- Nature organisation – organisations, companies, or groups with a core interest or expertise in nature.

Cleaning

Up to this point, all text had been cleaned using the Basic cleaning option (see Table S1); however, model accuracy is highly dependent on the quality of the feature extraction, and even small gains could be valuable. As a result, in the `classecol` package we have developed a series of cleaning options and assessed their impact on model accuracy. These options range from the very simple where text is modelled in its rawest form, to the complex - see full+ in Table S1. All of the cleaning options except full+ act on the first level text classification models, whereas full+ differs because it acts on both model levels by including its sentiment matrix in the second level ensembles. This sentiment matrix contains a matrix of 11 lexicon-based sentiment analysis approaches describing the polarity of text, a language indicator to specify if English terms are detected, and indicators to detect if any valence, negator, amplifier, or de-amplifier terms are used which can change the texts meaning (e.g. I love foxes… NOT) . We assess the impact of cleaning using only the final reclassification data, not the original classified data.

Table S1. Different cleaning options tested to assess their impact on model fit. These cleaning options are available through the `clean` function in `classecol`. Tick indicates the option present in the cleaning approach, whilst a cross indicates the option is absent.

	raw	basic	simple	full	full+
Text made lowercase	✗	✓	✓	✓	✓
Numbers removed	✗	✓	✓	✓	✓
Special characters removed	✗	✓	✓	✓	✓
URLs removed	✗	✓	✓	✓	✓
Text lemmatised *e.g. better = good*	✗	✓	✓	✓	✓
Text stemmed *e.g. walking = walk*	✗	✓	✓	✓	✓
Stopwords removed *e.g. the, and etc.*	✗	✓	✓	✓	✓
Hashtags split *e.g. '#BigCat' = 'big cat'*	✗	✗	✓	✓	✓
Twitter syntax removed *e.g. 'RTs'*	✗	✗	✓	✓	✓
Emoticons readable *e.g.* 😊 *= happy*	✗	✗	✓	✓	✓
Abbreviations expanded *e.g. lol = laugh out loud*	✗	✗	✗	✓	✓
Slang expanded *e.g. bro = brother*	✗	✗	✗	✓	✓
Grade expanded *e.g. A = excellent*	✗	✗	✗	✓	✓
Rating expanded *e.g. 10 = excellent*	✗	✗	✗	✓	✓
Sentiment matrix added	✗	✗	✗	✗	✓

Data rights and ethics

Before any project is designed and data are collected, it is important to consider the legal and ethical components of the work. From a legal perspective, tweets represent personal data and often fall under national and international (e.g. EU General data protection regulations) law that requires due care to ensure data are processed and stored correctly. Furthermore, even if the project meets legal standards, it is important to consider who owns the data. For example, tweets remain in the stewardship of Twitter and are owned by the users that posted the tweets. As a result, Twitter can recall data at any time and forbid its use, and Twitter forbids raw data sharing and publication. Finally, it is important to ensure the data use and research questions meet institutional and funding agency ethical requirements. However, these may be insufficient for this relatively novel data source, and efforts should be made by researchers to fully explore and implement ethical approaches for social media research, even in cases where the data are in the public domain, like Twitter – see Sula (2016); Ahmed, Bath, & Demartini (2017); Monkman, Kaiser, & Hyder (2018); and Toivonen et al., (2019). `classecol` was built and developed in alignment with the terms and conditions of the data provider (Twitter, 2020), the ethics guidelines of the lead author's research institution and funding agency, and data protection law (UK Government, 2018).

Assessing classifiers

Consistency in human classification

Within the pre-classification tests, classification accuracy ranged from 51% - 87% across the seven human assessors (Figure S3a). We found no significant difference between pre- and post-classification accuracies in a paired t-test (mean-difference = -5.55%, t = -1.62, p = 0.16). There was also no relationship (in a linear model) between accuracy and time taken to complete the test (coef = -0.23, t = -0.92, p = 0.38).

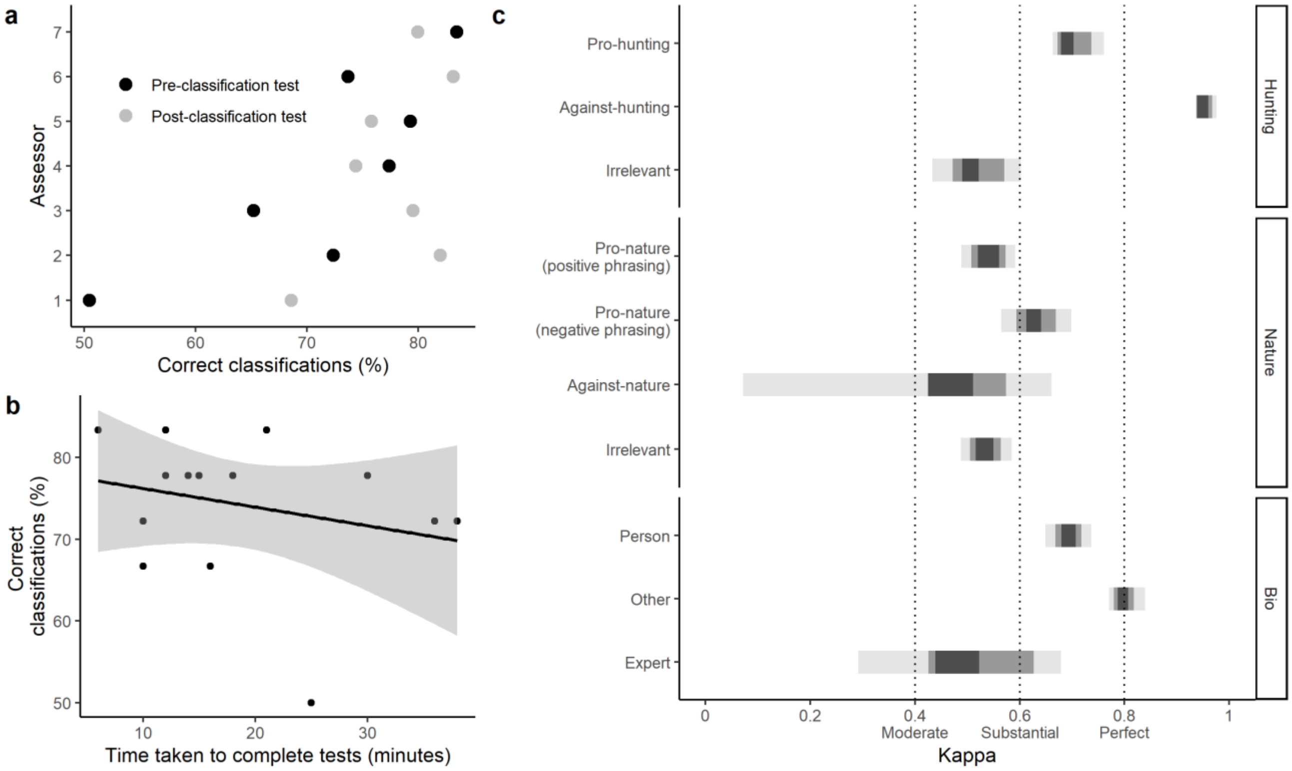


Figure S3. a) Pre- and post-classification test accuracy for the seven human assessors. Test accuracy was measured by getting each assessor to classify the same 19 (six in hunting, six in nature, and seven in the bio topic) passages of text, each compared to correct answers determined by TFJ and MGS. b) Change in classification test accuracies related to time taken to complete the test. c) Consistency in classifications across the 100 shared texts per topic, represented by Cohen's Kappa statistic, where values above 0.4 indicate moderate agreement, above 0.6 indicates substantial agreement, and above 0.8 indicates perfect agreement (Landis & Koch, 1977). The error for the kappa statistic was determined through 1000 bootstrap simulations, where the darker grey, grey, and lighter grey represent the 50%, 80%, and 95% quantiles, respectively. This consistency assessment is based on the original protocol and classification, not the labels reclassified by TFJ.

Within the 100 shared records per person per topic, we found substantial variation in the classification consistency (measured with Cohen's Kappa), with high consistency in the Against-hunting category and low consistency in the Against-nature and Expert categories (Figure S3c). To ensure this low consistency was not driven by poor classification from any one individual, we conducted a sensitivity analysis, assessing how consistency changes when each individual is removed from the calculation and we found that the results did not shift significantly when any one individual was removed (Figure S4).

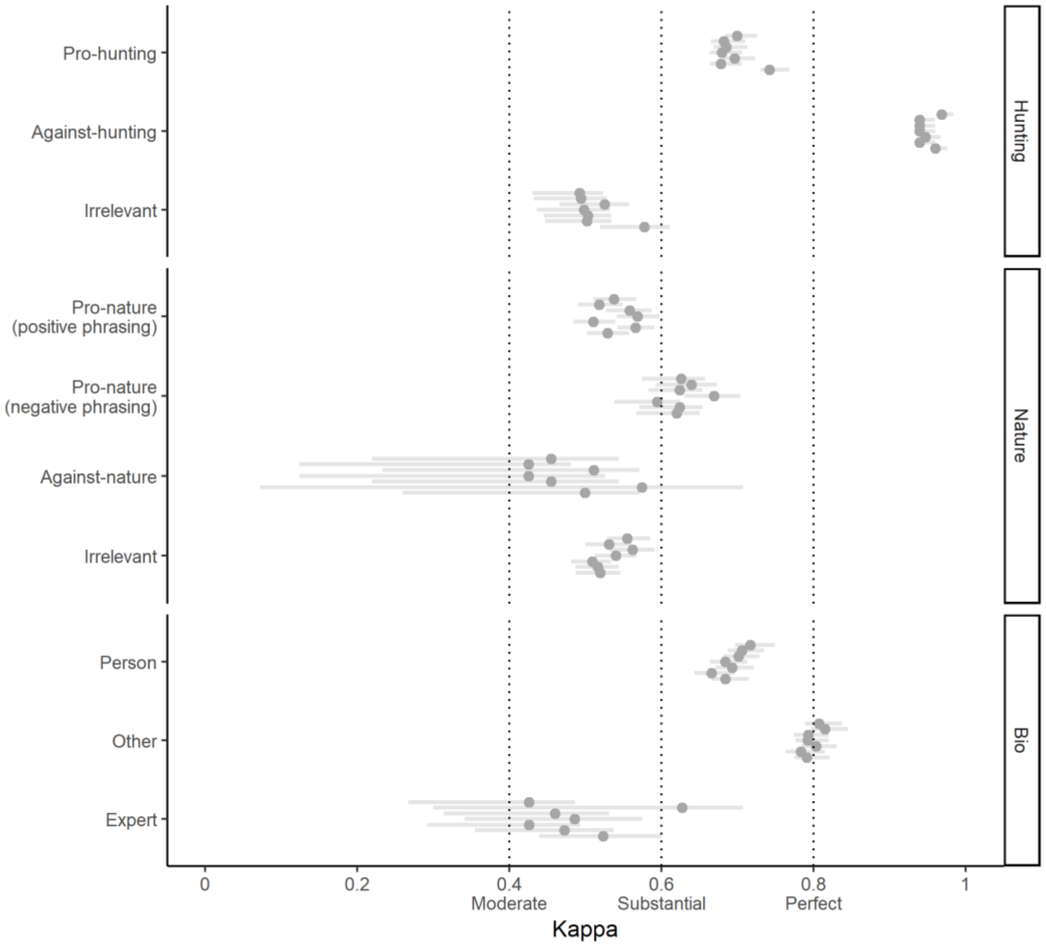


Figure S4. Sensitivity analysis in classification consistency across the 100 shared texts per topic, represented by Cohen's Kappa statistic, where values above 0.4 indicate moderate agreement, above 0.6 indicates substantial agreement, and above 0.8 indicates perfect agreement (Landis & Koch, 1977). Sensitivity analysis removed each human assessor iteratively to assess if any one individual had a detrimental effect of consistency e.g. if any one point had a substantially higher accuracy consistently across all categories, it would indicate one assessor performed inconsistently to the rest of the assessors. The point represents the median consistency and error bars display the 95% quantiles calculated through 1000 bootstrap simulations. This consistency assessment is based on the original protocol and classification, not the labels reclassified by TFJ.

Improving classifiers

The ensemble model outperformed each of the single text-classification models (Figure S5a). In the hunting and nature topics, each reclassification led to an increase in model accuracy, and the weighted F-score increased (respectively) by 0.05 and 0.09 between the original and final classifications (Figure S5b). However, in the bio topic, all reclassifications reduced the model accuracy, although, the 95% confidence intervals for the final reclassification overlap the original classification. The reclassifications had little impact on the bio topics accuracy but contributed to clarify the protocol. Each topic performed best under different levels of cleaning, with hunting accuracies greatest when using the Simple cleaning options, the bio accuracy was greatest with no cleaning (‘raw’), and cleaning had little effect on the nature topic (Figure S5c).

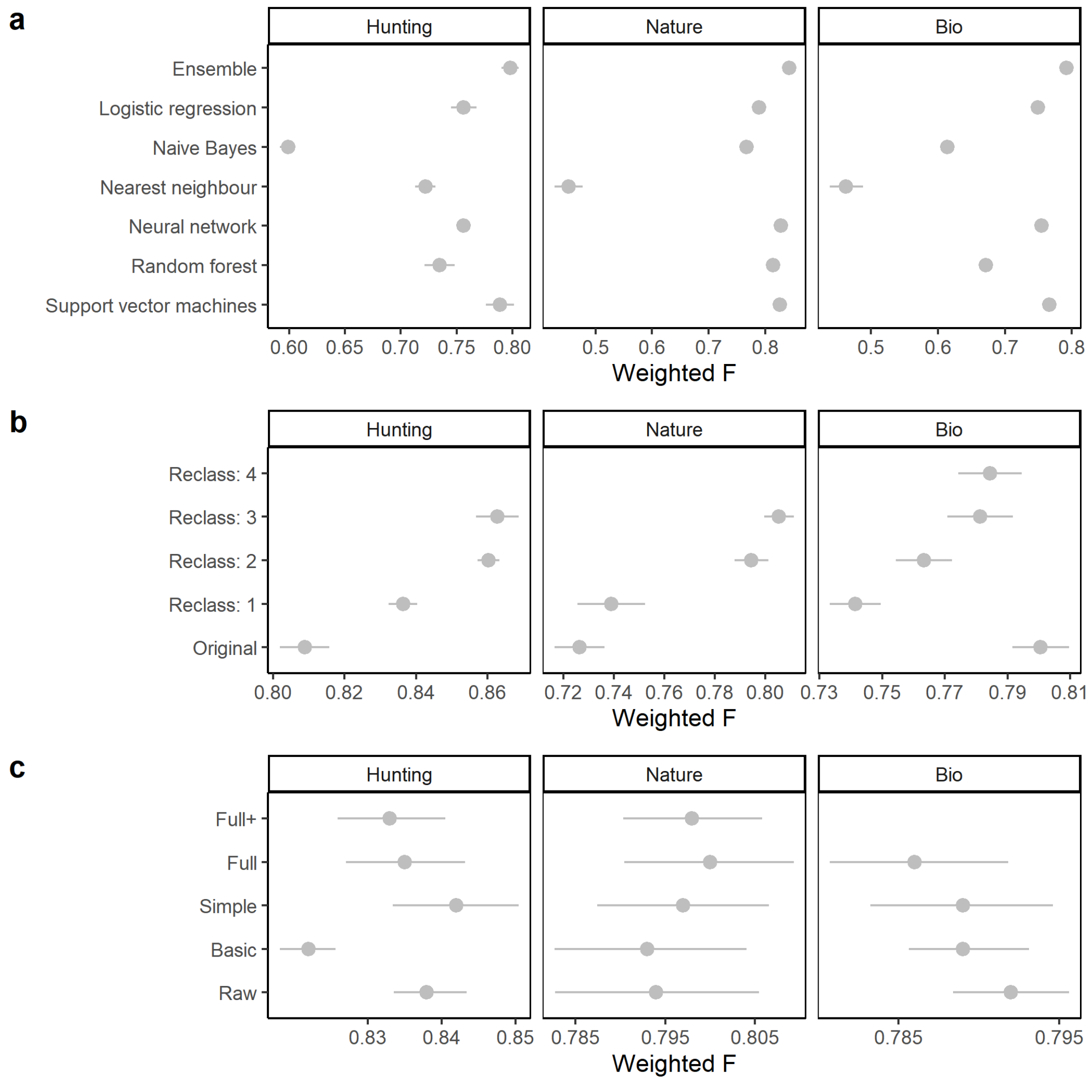


Figure S5. Model accuracy, represented by the weighted F statistic, under a) different model types, b) label reclassifications, and c) cleaning options. Notably, x-axis ranges differ between panels.

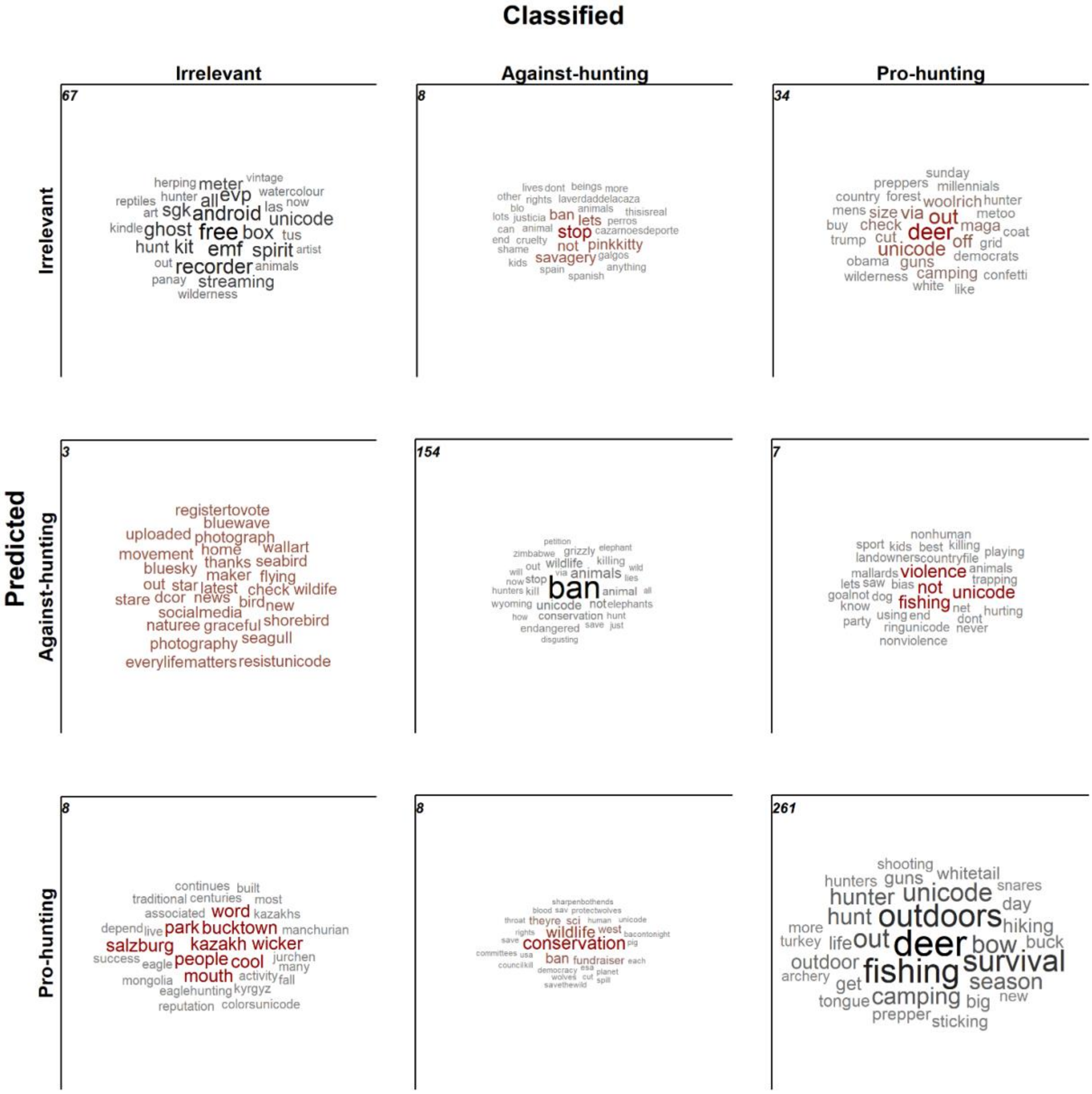


Figure S6. Confusion matrix indicating correct classifications (black) and misclassifications (red) in the hunting topic, with the frequency of misclassification indicated in the upper left of each panel, with a total testing dataset sample size of 550 tweets. In each panel, the thirty most frequently used words are depicted, and the size of the word determines this frequency. Panels with < 3 tweets in the test dataset are not represented by a word cloud (are left blank), as there are essentially zero cases of that misclassification combination.

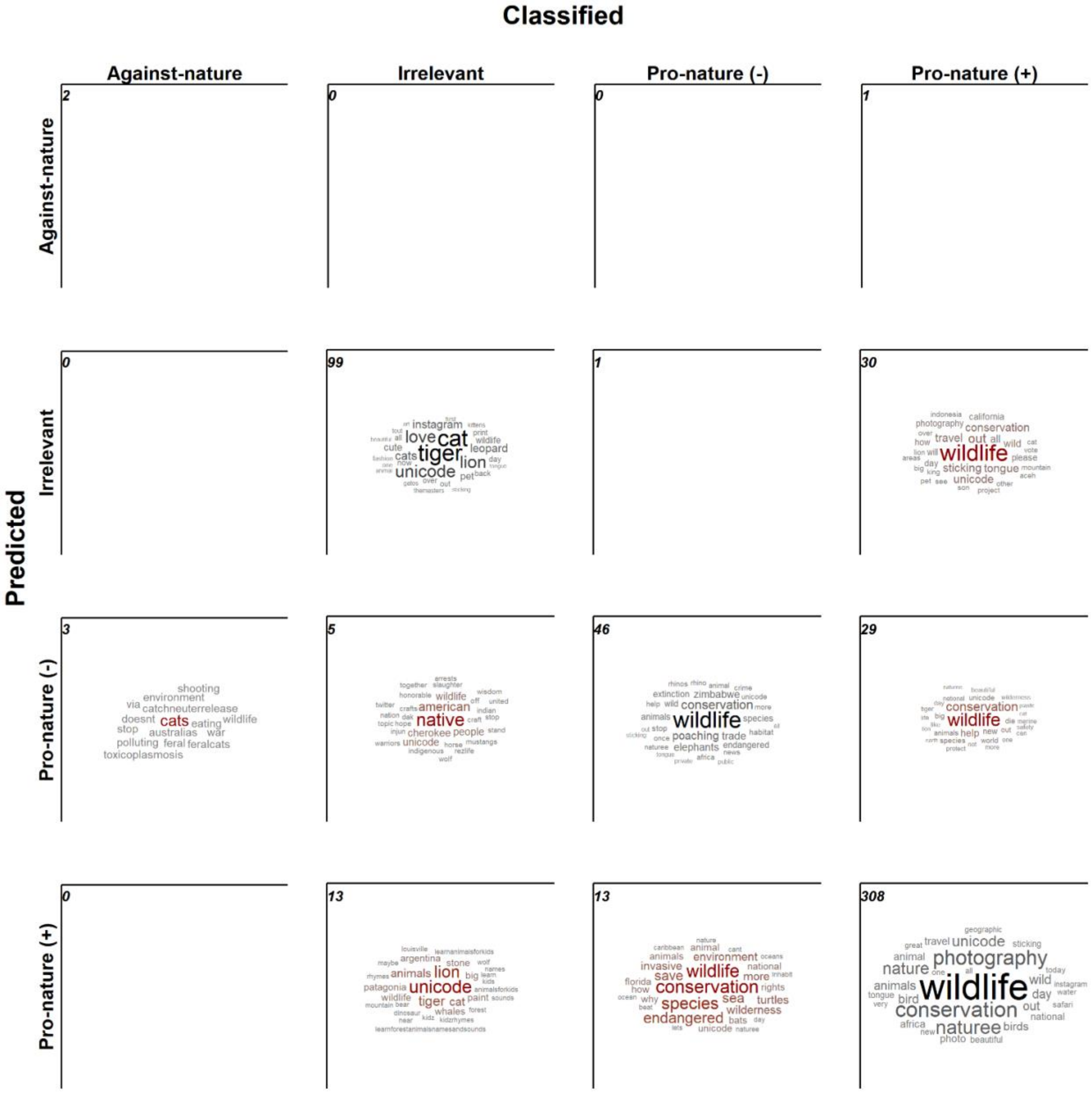


Figure S7. Confusion matrix indicating correct classifications (black) and misclassifications (red) in the nature topic, with the frequency of misclassification indicated in the upper left of each panel, with a total testing dataset sample size of 550 tweets. In each panel, the thirty most frequently used words are depicted, and the size of the word determines this frequency. Panels with < 3 tweets in the test dataset are not represented by a word cloud (are left blank), as there are essentially zero cases of that misclassification combination. The '+' and '-' in the Pro-nature categories indicate positive phrasing and negative phrasing, respectively.

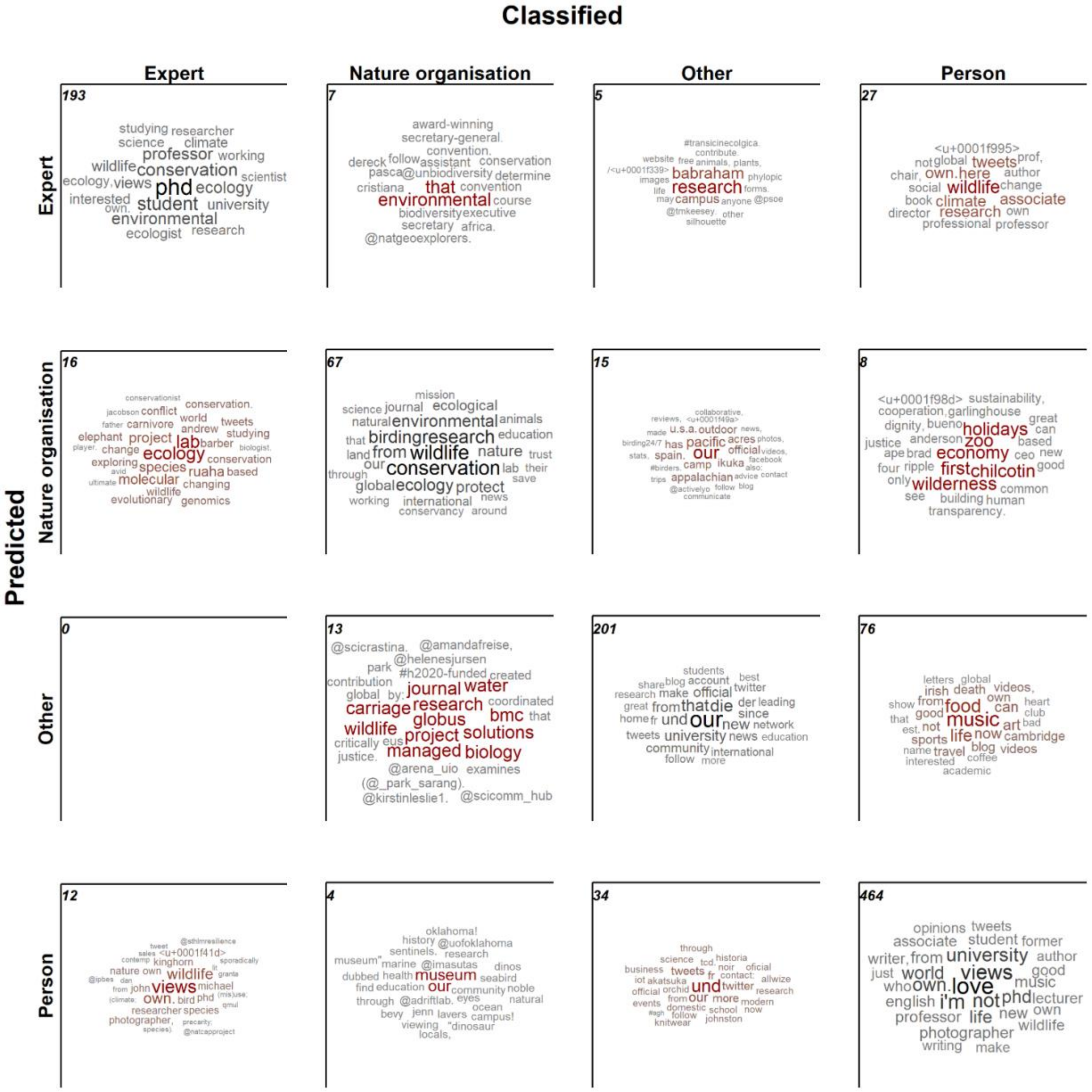


Figure S8. Confusion matrix indicating correct classifications (black) and misclassifications (red) in the bio topic, with the frequency of misclassification indicated in the upper left of each panel, with a total testing dataset sample size of 1142 users. In each panel, the thirty most frequently used words are depicted, and the size of the word determines this frequency. Panels with < 3 tweets in the test dataset are not represented by a word cloud (are left blank), as there are essentially zero cases of that misclassification combination.

`classecol` functions

Table S2. Description of functions available within the `classecol` R package.

Function	Description
`clean`	Cleans social media text converting up to 1000 multi-word nature-related hashtags and over 1500 abbreviations into readable text. This function also converts nearly 150 emoticons and 175 slang-words into readable text using terms sourced from the lexicon R package (Rinker, 2019a).
`valence`	Checks for the presence of negator (flips the meaning of the text e.g. I do not like it), amplifier (adds intensity to the text e.g. I really like it), de-amplifier (softens the text e.g. I hardly like it), and adversative-conjunction terms (overrules the previous sentiment e.g. I like it but it's not worth it), which could alter the meaning of the text. Terms and examples sourced from the lexicon (Rinker, 2019a) and sentiment (Rinker, 2019b) R packages.
`contract`	Performs stemming (trims word to their simplest form e.g. cars becomes car) and lemmatisation (identifies the core theme of a term, bringing synonymous terms into one word e.g. automobile becomes car) within R. Function is a wrapper for textstem (Rinker, 2018), qdap (Rinker, 2020), and lexicon (Rinker, 2019a) R packages. This function is not necessary when running any of the `hun_class`, `nat_class`, and `bio_class` models, as the stemming and lemmatisation are conducted within the substantially faster Python program.
`lang_eng`	Checks if the language is English or not, the `hun_class`, `nat_class`, and `bio_class` models are designed for English text. Function is a wrapper to the cld2 (Ooms & Sites, 2018) R package.
`senti_matrix`	Conducts sentiment analysis (assesses polarity of text) using 11 approaches and pulls all approaches into a matrix of sentiment. Approaches are drawn from the sentimentr (Rinker, 2019b), lexicon (Rinker, 2019a), syuzhet (Jockers, 2017), and meanr (Schmidt, 2019) R packages.
`Load_classecol`	Downloads all required `hun_class`, `nat_class`, and `bio_class` models, python dependencies, and links R to Python. This function needs to be run before any text classifier can be used..
`hun_class`	Classifies tweets. Can select any of the hunting models in Table 1 using the 'type' parameter.
`nat_class`	Classifies tweets. Can select any of the nature models in Table 1 using the 'type' parameter.
`bio_class`	Classifies tweets. Can select any of the bio models in Table 1 using the 'type' parameter.


A. Cardwell

Chapter 6: Discussion

In this thesis, I used interdisciplinary approaches covering evolutionary biology, biodiversity science, data science, and social science to explore influences of biodiversity change, specifically understanding declines and recoveries in large carnivores from the order *Carnivora*. I focussed on a combination of methodological development, improving available data, and global-scale modelling to explore this biodiversity change and describe the status of these species. However, our results are of relevance to more than just the carnivore community, and our developments could be applied to the wider biodiversity change research community.

Thesis overview

Traits have been essential in improving the mechanistic basis of ecological modelling (Webb et al., 2010). However, trait data are plagued with missing values. The impact of these missing values has been explored in recent years, with studies proposing the uptake of imputation, or gap-filling methods (Penone et al., 2014; Kim et al., 2018; Molina-Venegas et al., 2018). Yet, there are still a host of scenarios where these imputation methods have not been tested. In ‘Chapter 2: Handling missing values in trait data’, I evaluated the performance of approaches for handling missing values under a variety of these scenarios. Overall, I showed that imputation methods can be valuable for filling data, especially using Rphylopars imputation (Goolsby et al., 2017; Johnson et al., 2021b). However, I also found cases where imputation was not the best method and analysing only the available data (i.e. complete-case analysis) would be preferable. Furthermore, in some cases, especially under severe biases in the data, none of the approaches for handling missing trait data were effective. These findings are important because they stress that no single approach is a golden bullet for dealing with missing values, and instead, caution and care should be taken when handling trait values, or missing data more generally. Traits can provide a means of making biodiversity change modelling more mechanistic, but this is only valuable if traits and their missing values are treated with due care. I summarised our findings into recommendations and warnings (see Chapter 2 – Table 1), but more work is required – I identified a series of facets warranting further investigation.

Data compilation efforts have been important in scaling research up to a global extent (Hudson et al., 2017; Dornelas et al., 2018), creating more generalisable conclusions. This was largely unfeasible for large carnivores using the currently available public datasets, as they are plagued with extreme data biases. In ‘Chapter 3: CaPTrends - a global database of population trends in large terrestrial Carnivorans; I conducted an extensive search of the primary literature to increase the availability of large carnivore population trend data. Through this, I tripled the size of the already compiled data in the Living Planet Index (WWF, 2020a), and addressed taxonomic and spatial biases – to a degree. However, the real strengths of this data extend beyond addressing biases, in the CaPTrends database I have compiled an extensive array of metadata to accompany each trend, including information on how the population trend was determined, as well as the threats and management interventions experienced by the population. This information could be used to account for important features that limit the effectiveness of large-scale biodiversity change models e.g. is the population managed with hunting. Once published, CaPTrends will be useful to biodiversity change researchers and people working with large carnivores.

In Chapter 4, I used the data in Chapter 3 and developed a model to explore covariates that influence population change in large carnivores (or predators). I found that land-use change, the usual suspect (Newbold et al., 2015), was an important influence of population change, but less so than human development variables (which described the quality of life for the human population), which had a substantial impact on both population declines and recoveries. I also found evidence of climate and protected area effects, and that many influences of change varied depending on species intrinsic traits e.g. large species are more likely to decline under extreme heat. Using this model, I determined the status of large carnivores, identifying that some species have increased, whilst others have decreased. Overall, the carnivore guild is projected to change very little between 1970 and 2050, but there is substantial spatial variation – most notably, parts of Africa are projected to experience severe population declines and extirpations. This work emphasised the value capturing socio-ecological variables in biodiversity change models, but many of these variables are cryptic and currently unavailable. Furthermore, the finding that a diverse array of features influences population change is also important because much of the existing biodiversity change research has focussed on one or two small aspects in detail, ignoring the sheer complexity of the drivers influencing change. This finding suggests that biodiversity change modelling needs to expand beyond its current scope, which focusses primarily on the impacts of land-use and climate change, into a broader view of features driving change.

One of the issues with the analysis in Chapter 4 was that I failed to determine the mechanism by which human development influenced population trends. I predict that this mechanism, at least in part, explains people’s tolerance for living alongside carnivores, and this assumption would be supported by other work (Cimatti et al., 2021). However, there are currently no datasets on tolerance, or even perceptions of nature, at a global scale. This is likely because of challenges in inferring meaning across different cultures and languages. As a result, in Chapter 5, I made the first step to try and support this global-scale perception analysis within the environmental sciences. Specifically, I developed a machine learning text classifier designed to understand public opinions of nature from social media data. This classifier could be used to develop variables like ‘tolerance towards wildlife’, something that was previously undetectable at a global scale. But it also acts as proof of concept that I can extract mass opinions about nature from social media data. However, for this classifier, and other classifiers that will likely replace it over time, there are still many steps that must be taken if we are to develop this global perception data. Firstly, it is important to validate that this data truly represent people’s opinions, and not just their online persona. It is also important to expand this to more languages beyond English, and ensure social media covers a variety of demographics and lifestyles.

Improving biodiversity change research

In this thesis, I endeavoured to make our analysis of large carnivore trends (focussed on Chapter 4) as robust as possible, but I propose that more work is still needed if I are to fully untangle the complexity of biodiversity change. Firstly, given that missing trait values had a substantial impact on inference in Chapter 2, more work is needed to collect trait data from the field, and then compile this into trait datasets. In the literature, there is a relatively haphazard approach to dealing with these missing values, and even though uptake of imputation approaches has increased in recent years, there have been no assessments of whether these imputation approaches have been used correctly. To avoid erroneous inference, and perhaps a reproducibility crisis, a more formalised framework for handling and reporting missing values could be developed. Similar has been developed for systematic reviews, e.g. PRISMA and Cochrane guidelines (Stewart et al., 2015; Higgins & Green, 2019). Lessons could also be learnt from exploring other fields, especially medicine, where approaches for handling missing values are more regularly embraced (Van Buuren, 2012).

Another area to explore is developing more flexible modelling frameworks to address data biases. In Chapter 4, I extended beyond the relatively simple models used in other biodiversity change studies (Newbold et al., 2015; Spooner et al., 2018), in an attempt to decrease the data biases and capture uncertainties. For example, I captured both quantitative and qualitative population trends within our regression by treating the qualitative observations as if they were censored quantitative ones i.e. increasing records would have an annual rate of change of between 0% and 20%. This approach improved our model fit and increased our sample size in poorly represented areas e.g. records in South America increased 7-fold. These benefits were only achieved by developing a more flexible modelling approach, and this could be applied to a variety of settings. For example, this approach could be used to study changes in species distribution ranges where the magnitude of change is unknown, but the direction of range change is.

Large-scale biodiversity change models are becoming more common, but many have one fundamental issue – a moderately low predictive accuracy. Our population trend model in Chapter 4 is no exception, with a marginal R^2 of 0.2 and conditional R^2 of 0.5. This limited predictive capacity occurred despite accounting for a comprehensive array of potential trend influences. These models can offer insights into general patterns in the data, but the end goal should be to develop better biodiversity change models that can be used for high-quality local prediction. To reach this goal, its essential to find a means of capturing system (in our case the population trend) stochasticity. In large carnivores, an important development would be to capture the status of trends of the wider community, as a carnivore population could simply be declining as it is depicting a natural predator-prey abundance cycle. However, this alone would be insufficient. To fully address this stochasticity, a more holistic approach will likely be necessary, focussed on 1) improving population monitoring to reduce measurement error; 2) capturing measurement and observation errors more explicitly within biodiversity change models; 3) defining and clarifying the mechanisms behind biodiversity change; 4) developing new model frameworks for to capture stochasticity and mechanisms.

The status of large carnivores

Large carnivores are both adored for their charisma, but also feared for the threat they carry to lives and livelihoods. This dichotomy makes their population status particularly complex, as they are likely host to the extreme benefits of conservation management, but also impacted by the extreme costs of persecution, poaching, and hunting, to name but a few. This complexity fed forwards into our modelling, which is why I made such an effort to capture the multitude of influences on their trends. As a result, finding that traits, land-use, climate, and governance features all influenced trends is an important result, and future biodiversity change research should endeavour to account for these variety of influences. Further, I provide evidence that large carnivore populations have experienced substantial declines, and even some extinctions, but many are stable, and the future provides hope for these species. Specifically, whilst population are projected to decline in some regions and species, these declines will be balanced by recoveries in other areas, and these recoveries may even present opportunities for range expansions. This finding is encouraging not just for large carnivores, but also for wider biodiversity, as large carnivores can act as an indicator species.

References

Abadi, M., Agarwal, A., Barham, P., et al. (2015). *Large-Scale Machine Learning on Heterogeneous Systems*.

Aggarwal, C. C., & Zhai, C. X. (2013). *Mining text data*. *Mining Text Data*.

Ahmed, W., Bath, P. A., & Demartini, G. (2017). Using Twitter as a Data Source: An Overview of Ethical, Legal, and Methodological Challenges.

Akande, O., Li, F., & Reiter, J. (2017). An Empirical Comparison of Multiple Imputation Methods for Categorical Data. *The American Statistician*, *71*(2), 162–170.

Aldayel, A., & Magdy, W. (2019). Assessing Sentiment of the Expressed Stance on Social Media. In *Lecture Notes in Computer Science*.

Allen, W. L., Street, S. E., & Capellini, I. (2017). Fast life history traits promote invasion success in amphibians and reptiles. *Ecology Letters*, *20*(2), 222–230.

Altrichter, M., Boaglio, G., & Perovic, P. (2006). The decline of jaguars Panthera onca in the Argentine Chaco. *Oryx*, *40*(03), 302.

Amano, T., Székely, T., Sandel, B., et al. (2018). Successful conservation of global waterbird populations depends on effective governance. *Nature*.

Atkins, J. L., Long, R. A., Pansu, J., et al. (2019). Cascading impacts of large-carnivore extirpation in an African ecosystem. *Science*, *364*(6436), 173–177.

August, T. A., Pescott, O. L., Joly, A., et al. (2020). AI Naturalists Might Hold the Key to Unlocking Biodiversity Data in Social Media Imagery. *Patterns*.

Baker, J., Humphries, S., Ferguson-Gow, H., et al. (2020). Rapid decreases in relative testes mass among monogamous birds but not in other vertebrates. *Ecology Letters*, *23*(2), 283–292.

Bauer, H., Chapron, G., Nowell, K., et al. (2015). Lion (Panthera leo) populations are declining rapidly across Africa, except in intensively managed areas. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(48), 14894–14899.

Belaire, J. A., Westphal, L. M., Whelan, C. J., et al. (2015). Urban residents' perceptions of birds in the neighborhood: Biodiversity, cultural ecosystem services, and disservices. *Condor*.

Benítez-López, A., Alkemade, R., Schipper, A. M., et al. (2017). The impact of hunting on tropical mammal and bird populations. *Science*.

Boakes, E. H., McGowan, P. J. K., Fuller, R. A., et al. (2010). Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data. *PLoS Biology*, *8*(6), e1000385.

Braga-Pereira, F., Peres, C. A., Campos-Silva, J. V., et al. (2020). Warfare-induced mammal population declines in Southwestern Africa are mediated by species life history, habitat type and hunter preferences. *Scientific Reports*.

Broekhuis, F., Kaelo, M., Sakat, D. K., et al. (2020). Human-wildlife coexistence: Attitudes and behavioural intentions towards predators in the Maasai Mara, Kenya. *ORYX*.

Cardillo, M., Mace, G., Jones, K. E., et al. (2005). Multiple causes of high extinction risk in large mammal species. *Science*.

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., et al. (2015). Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances*.

Chapron, G., Kaczensky, P., Linnell, J. D. C., et al. (2014). Recovery of large carnivores in Europe’s modern human-dominated landscapes. *Science*, *346*(6216), 1517–1519.

Chollet, F. (2015). Keras: The Python Deep Learning library. *Keras.Io*.

Cimatti, M., Ranc, N., Benítez-López, A., et al. (2021). Large carnivore expansion in Europe is associated with human population density and land cover changes. *Diversity and Distributions*.

Clement, J. (2020). Number of social network users worldwide from 2017 to 2025.

Clucas, B., McHugh, K., & Caro, T. (2008). Flagship species on covers of US conservation and nature magazines. *Biodiversity and Conservation*.

Cohen, J., Cohen, P., Stephen, G., et al. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). London: Routledge.

Cornford, R., Deinet, S., De Palma, A., et al. (2021). Fast, scalable, and automated identification of articles for biodiversity and macroecological datasets. *Global Ecology and Biogeography*, *30*(1), 339–347.

Cucchi, M., P. Weedon, G., Amici, A., et al. (2020). WFDE5: Bias-adjusted ERA5 reanalysis data for impact studies. *Earth System Science Data*.

Dalerum, F., Shults, B., & Kunkel, K. (2008). Estimating sustainable harvest in wolverine populations using logistic regression. *Journal of Wildlife Management*, *72*(5), 1125–1132.

Daskalova, G. N., Myers-Smith, I. H., Bjorkman, A. D., et al. (2020). Landscape-scale forest loss as a catalyst of population and biodiversity change. *Science*.

Davis, K. T., Dobrowski, S. Z., Holden, Z. A., et al. (2019). Microclimatic buffering in forests of the future: the role of local water balance. *Ecography*.

De MagalhÃes, J. P., & Costa, J. (2009). A database of vertebrate longevity records and their relation to other life-history traits. *Journal of Evolutionary Biology*.

Di Lorenzo, M., Guidetti, P., Di Franco, A., et al. (2020). Assessing spillover from marine protected areas and its drivers: A meta-analytical approach. *Fish and Fisheries*.

Di Minin, E., Fink, C., Tenkanen, H., et al. (2018). Machine learning for tracking illegal wildlife trade on social media. *Nature Ecology and Evolution*.

Díaz, S., Pascual, U., Stenseke, M., et al. (2018). Assessing nature’s contributions to people. *Science*.

Díaz, S., Settele, J. S., Brondízio, E., et al. (2019). *IPBES Global Assessment Summary for Policymakers*.

Diniz, J. A. F., Villalobos, F., Bini, L. M., et al. (2015). The best of both worlds: Phylogenetic eigenvector regression and mapping. *Genetics and Molecular Biology*, *38*(3), 396–400.

Diniz-Filho, J. A. F., Bini, L. M., Rangel, T. F., et al. (2012). On the selection of phylogenetic eigenvectors for ecological analyses. *Ecography*, *35*(3), 239–249.

Dornelas, M., Antão, L. H., Moyes, F., et al. (2018). BioTIME: A database of biodiversity time series for the Anthropocene. *Global Ecology and Biogeography*.

Dornelas, M., Gotelli, N. J., Shimadzu, H., et al. (2019). A balance of winners and losers in the Anthropocene. *Ecology Letters*.

Duncan, C., Chauvenet, A. L. M., McRae, L. M., et al. (2012). Predicting the Future Impact of Droughts on Ungulate Populations in Arid and Semi-Arid Environments. *PLoS ONE*.

Etard, A., Morrill, S., & Newbold, T. (2020). Global gaps in trait data for terrestrial vertebrates. *Global Ecology and Biogeography*.

Eziz, A., Yan, Z., Tian, D., et al. (2017). Drought effect on plant biomass allocation: A meta-analysis. *Ecology and Evolution*.

Faurby, S., Davis, M., Pedersen, R., et al. (2018). PHYLACINE 1.2: The Phylogenetic Atlas of Mammal Macroecology. *Ecology*.

Fernández-Llamazares, Á., Western, D., Galvin, K. A., et al. (2020). Historical shifts in local attitudes towards wildlife by Maasai pastoralists of the Amboseli Ecosystem (Kenya): Insights from three conservation psychology theories. *Journal for Nature Conservation*.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*.

Florczyk, A. J., Corbane, C., Ehrlich, D., et al. (2019). *GHSL Data Package 2019*. *JRC Technical report*.

Fournier, A. M. V., White, E. R., & Heard, S. B. (2019). Site-selection bias and apparent population declines in long-term studies. *Conservation Biology*.

Fox, N., August, T., Mancini, F., et al. (2020). ‘photosearcher’ package in R: An accessible and reproducible method for harvesting large datasets from Flickr. *SoftwareX*.

Fritz, S. A., Bininda-Emonds, O. R. P., & Purvis, A. (2009). Geographical variation in predictors of mammalian extinction risk: Big is bad, but only in the tropics. *Ecology Letters*.

Fritz, S. A., & Purvis, A. (2010). Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits. *Conservation Biology*.

Gaston, K. J., Aimé, E., Chan, K. M. A., et al. (2019). People and Nature —A journal of relational thinking . *People and Nature*.

GBIF.org. (2018). GBIF occurence download.

Ghermandi, A., & Sinclair, M. (2019). Passive crowdsourcing of social media in environmental research: A systematic map. *Global Environmental Change*.

Gloor, S., Bontadina, F., Hegglin, D., et al. (2001). The rise of urban fox populations in Switzerland. *Zeitschrift Fur Saugetierkunde*.

González-Suárez, M. (2014). *Carnivora trait dataset*.

González-Suárez, M., Bacher, S., & Jeschke, J. M. (2015). Intraspecific Trait Variation Is Correlated with Establishment Success of Alien Mammals. *The American Naturalist*, *185*(6), 737 – 746.

Gonzalez-Suarez, M., Gomez, A., & Revilla, E. (2013). Which intrinsic traits predict vulnerability to extinction depends on the actual threatening processes. *Ecosphere*, *4*(6), 1 – 16.

González-Suárez, M., Lucas, P. M., & Revilla, E. (2012). Biases in comparative analyses of extinction risk: mind the gap. *Journal of Animal Ecology*, *81*(6), 1211–1222.

González-Suárez, M., & Revilla, E. (2013). Variability in life-history and ecological traits is a buffer against extinction in mammals. *Ecology Letters*, *16*(2), 242–251.

Goolsby, E. W., Bruggeman, J., & Ané, C. (2017). Rphylopars : fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods in Ecology and Evolution*, *8*(1), 22–27.

Graham, L. J., & Eigenbrod, F. (2019). Scale dependency in drivers of outdoor recreation in England. *People and Nature*.

Hanley, N., Breeze, T. D., Ellis, C., et al. (2015). Measuring the economic value of pollination services: Principles, evidence and knowledge gaps. *Ecosystem Services*.

Hardt, J., Herke, M., & Leonhart, R. (2012). Auxiliary variables in multiple imputation in regression with missing X: A warning against including too many in small sample research. *BMC Medical Research Methodology*.

Harrison, R. D., Sreekar, R., Brodie, J. F., et al. (2016). Impacts of hunting on tropical forests in Southeast Asia. *Conservation Biology*, *30*(5), 972–981.

Hausmann, A., Toivonen, T., Heikinheimo, V., et al. (2017). Social media reveal that charismatic species are not the main attractor of ecotourists to sub-Saharan protected areas. *Scientific Reports*.

Hautier, Y., Isbell, F., Borer, E. T., et al. (2018). Local loss and spatial homogenization of plant diversity reduce ecosystem multifunctionality. *Nature Ecology and Evolution*.

Higgins, J. P., & Green, S. (2019). *Cochrane Handbook for Systematic Reviews of Interventions*. *Cochrane Handbook for Systematic Reviews of Interventions*.

Hudson, L. N., Newbold, T., Contu, S., et al. (2017). The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. *Ecology and Evolution*.

Hurtt, G. C., Chini, L., Sahajpal, R., et al. (2020). Harmonization of global land use change and management for the period 850-2100 (LUH2) for CMIP6. *Geoscientific Model Development*.

IPBES. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*.

Isaac, N. J. B., & Carbone, C. (2010). Why are metabolic scaling exponents so controversial? Quantifying variance and testing hypotheses. *Ecology Letters*, *13*(6), 728–735.

Isaac, N. J. B., Jarzyna, M. A., Keil, P., et al. (2020). Data Integration for Large-Scale Models of Species Distributions. *Trends in Ecology and Evolution*.

IUCN. (2018). Threats Classification Scheme (Version 3.2.

IUCN. (2020a). Habitats Classification Scheme (Version 3.1).

IUCN. (2020b). Spatial Data Download.

Jennions, M. D., & Møller, A. P. (2002). Publication bias in ecology and evolution: An empirical assessment using the ‘trim and fill’ method. *Biological Reviews of the Cambridge Philosophical Society*.

Jockers, M. (2017). syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text. R CRAN repository.

Johnson, T. F. (2021). addeR. https://github.com/GitTFJ/addeR.

Johnson, T. F., Cruz, P., Isaac, N. J. B., et al. (2021a). CaPTrends: A global database of population trends in large terrestrial Carnivorans. *Unpublished*.

Johnson, T. F., Isaac, N. J. B., Paviolo, A., et al. (2021b). Handling missing values in trait data. *Global Ecology and Biogeography*.

Johnson, T. F., Kent, H., Hill, B. M., et al. (2021c). classecol: Classifiers to understand public opinions of nature. *Methods in Ecology and Evolution*.

Jones, K. E., Bielby, J., Cardillo, M., et al. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, *90*(9), 2648–2648.

Josse, J., Tierney, N., & Vialaneix, N. (2020). CRAN Task View: Missing Data.

Kareiva, P., Tallis, H., Rickets, T. H., et al. (2011). *Natural Capital* (1st editio). Oxford: Oxford University Press.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The worldwide governance indicators: Methodology and analytical issues. *Hague Journal on the Rule of Law*.

Kim, S. W., Blomberg, S. P., & Pandolfi, J. M. (2018). Transcending data gaps: a framework to reduce inferential errors in ecological analyses. *Ecology Letters*, *21*(8), 1200–1210.

Kuo, L., & Mallick, B. (1998). Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B*.

Lancaster, L. T., Morrison, G., & Fitt, R. N. (2017). Life history trade-offs, the intensity of competition, and coexistence in novel and evolving communities under climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1712), 20160046.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*.

Lange, S. (2019a). Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0). *Geoscientific Model Development*.

Lange, S. (2019b). WFDE5 over land merged with ERA5 over the ocean (W5E5). V. 1.0.

Laurance, W. F., Goosem, M., & Laurance, S. G. W. (2009). Impacts of roads and linear clearings on tropical forests. *Trends in Ecology and Evolution*.

Leclère, D., Obersteiner, M., Barrett, M., et al. (2020). Bending the curve of terrestrial biodiversity needs an integrated strategy. *Nature*.

Lehikoinen, P., Santangeli, A., Jaatinen, K., et al. (2019). Protected areas act as a buffer against detrimental effects of climate change—Evidence from large-scale, long-term abundance data. *Global Change Biology*.

Lennox, R. J., Veríssimo, D., Twardek, W. M., et al. (2020). Sentiment analysis as a measure of conservation culture in scientific literature. *Conservation Biology*.

Leung, B., Hargreaves, A. L., Greenberg, D. A., et al. (2020). Clustered versus catastrophic global vertebrate declines. *Nature*.

Little, R. J. a, & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. *Statistical analysis with missing data Second edition*. John Wiley & Sons.

Liu, B. (2020). *Sentiment analysis: Mining Opinions, Sentiments, and Emotions* (2nd ed.). Cambridge University Press.

Macdonald, E. A., Hinks, A., Weiss, D. J., et al. (2017). Identifying ambassador species for conservation marketing. *Global Ecology and Conservation*.

Millard, J., Gregory, R. D., Jones, K. E., et al. (2021). The Species Awareness Index (SAI): a Wikipedia-derived conservation culturomics metric for public biodiversity awarenes. *Unpublished*.

Mohammad, S. M., Sobhani, P., & Kiritchenko, S. (2017). Stance and sentiment in Tweets. *ACM Transactions on Internet Technology*.

Molenberghs, G., Fitzmaurice, G., Kenward, M., et al. (2015). *Handbook of missing data methodology*. *Handbook of Missing Data Methodology*. CRC Press.

Molina-Venegas, R., Moreno-Saiz, J. C., Castro Parga, I., et al. (2018). Assessing among-lineage variability in phylogenetic imputation of functional trait datasets. *Ecography*, *41*(10), 1740–1749.

Møllerand, A. P., & Jennions, M. D. (2001). Testing and adjusting for publication bias. *Trends in Ecology and Evolution*.

Monkman, G. G., Kaiser, M., & Hyder, K. (2018). The Ethics of Using Social Media in Fisheries Research. *Reviews in Fisheries Science and Aquaculture*.

Moons, K. G. M., Donders, R. A. R. T., Stijnen, T., et al. (2006). Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology*, *59*(10), 1092 – 1101.

Mukaka, M., White, S. A., Terlouw, D. J., et al. (2016). Is using multiple imputation better than complete case analysis for estimating a prevalence (risk) difference in randomized controlled trials when binary outcome observations are missing? *Trials*, *17*(1), 341.

Newbold, T., Hudson, L. N., Arnell, A. P., et al. (2016). Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science*.

Newbold, T., Hudson, L. N., Hill, S. L. L. L., et al. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, *520*(7545), 45–50.

Nielsen, M. R., Meilby, H., Smith-Hall, C., et al. (2018). The Importance of Wild Meat in the Global South. *Ecological Economics*.

Nyakatura, K., & Bininda-Emonds, O. R. P. (2012). Updating the evolutionary history of Carnivora (Mammalia): A new species-level supertree complete with divergence time estimates. *BMC Biology*.

Ooms, J., & Sites, D. (2018). cld2: Google's Compact Language Detector 2. R CRAN repository.

Pacifici, M., Visconti, P., Butchart, S. H. M. M., et al. (2017). Species' traits influenced their response to recent climate change. *Nature Climate Change*, *7*(3), 205–208.

Pecl, G. T., Araújo, M. B., Bell, J. D., et al. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*.

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825 – 2830.

Penone, C., Davidson, A. D., Shoemaker, K. T., et al. (2014). Imputation of missing data in life-history trait datasets: which approach performs the best? *Methods in Ecology and Evolution*, *5*(9), 961–970.

Pettersson, T. (2019). *UCDP/PRIO Armed Conflict Dataset Codebook Version 19.1* . *Journal of Peace Research*.

Plummer, M. (2016). rjags: Bayesian graphical models using MCMC. *R Package Version 3-13*.

Polaina, E., González-Suárez, M., & Revilla, E. (2019). The legacy of past human land use in current patterns of mammal distribution. *Ecography*.

Polaina, E., Revilla, E., & González-Suárez, M. (2016). Putting susceptibility on the map to improve conservation planning, an example with terrestrial mammals. *Diversity and Distributions*.

Poyatos, R., Sus, O., Badiella, L., et al. (2018). Gap-filling a spatially explicit plant trait database: Comparing imputation methods and different levels of environmental information. *Biogeosciences*, *15*(9), 2601.

Prokosch, J., Bernitz, Z., Bernitz, H., et al. (2019). Are animals shrinking due to climate change? Temperature-mediated selection on body mass in mountain wagtails. *Oecologia*.

Quartagno, M., Grund, S., & Carpenter, J. (2019). jomo: A Flexible Package for Two-level Joint Modelling Multiple Imputation. *R Journal*.

R Core Team. (2018). R: A language and environment for statistical computing.

R Development Core Team. (2020). R Development Core Team, R: a language and environment for statistical computing. *R: A Language and Environmental for Estatistical Computing*.

Riggio, J., Caro, T., Dollar, L., et al. (2016). Lion populations may be declining in Africa but not as Bauer et al. suggest. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(2), E107–8.

Rinker, T. (2018). textstem: Tools for Stemming and Lemmatizing Text. R CRAN repository.

Rinker, T. (2019a). lexicon: Lexicons for Text Analysis. R CRAN repository.

Rinker, T. (2019b). sentimentr: Calculate Text Polarity Sentiment. R CRAN repository.

Rinker, T. (2020). qdap: Bridging the Gap Between Qualitative Data and Quantitative Analysis. R CRAN repository.

Ríos-Saldaña, C. A., Delibes-Mateos, M., & Ferreira, C. C. (2018). Are fieldwork studies being relegated to second place in conservation science? *Global Ecology and Conservation*, *14*, e00389.

Ripple, W. J., Estes, J. A., Beschta, R. L., et al. (2014). Status and ecological effects of the world’s largest carnivores. *Science*.

Rockström, J., Steffen, W., Noone, K., et al. (2009). A safe operating space for humanity. *Nature*.

Rohatgi, A. (2015). WebPlotDigitizer.

Roth, T., Allan, E., Pearman, P. B., et al. (2018). Functional ecology and imperfect detection of species. *Methods in Ecology and Evolution*, *9*(4), 917–928.

Salmon, M. (2018). opencage: Interface to the OpenCage API.

Sandel, B., Guti??rrez, A. G., Reich, P. B., et al. (2015). Estimating the missing species bias in plant trait measurements. *Journal of Vegetation Science*, *26*(5), 828–838.

Santini, L., González-Suárez, M., Russo, D., et al. One strategy does not fit all: determinants of urban adaptation in mammals. (U. Brose, Ed.), 22 Ecology Letters § (2019). John Wiley & Sons, Ltd (10.1111).

Schmidt, D. (2019). meanr: Sentiment Analysis Scorer. R CRAN repository.

Schrodt, F., Kattge, J., Shan, H., et al. (2015). BHPMF - a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. *Global Ecology and Biogeography*, *24*(12), 1510–1521.

Sergio, F., Caro, T., Brown, D., et al. (2008a). Top predators as conservation tools: Ecological rationale, assumptions, and efficacy. *Annual Review of Ecology, Evolution, and Systematics*.

Sergio, F., Caro, T., Brown, D., et al. (2008b). Top predators as conservation tools: Ecological rationale, assumptions, and efficacy. *Annual Review of Ecology, Evolution, and Systematics*.

Shah, A. D., Bartlett, J. W., Carpenter, J., et al. (2014). Comparison of Random Forest and Parametric Imputation Models for Imputing Missing Data Using MICE: A CALIBER Study. *American Journal of Epidemiology*, *179*(6), 764–774.

Society for Conservation Biology working groups. (2020). Conservation Culturomics.

Soga, M., & Gaston, K. J. (2018). Shifting baseline syndrome: causes, consequences, and implications. *Frontiers in Ecology and the Environment*.

Soroye, P., Newbold, T., & Kerr, J. (2020). Climate change contributes to widespread declines among bumble bees across continents. *Science*.

Spooner, F. E. B., Pearson, R. G., & Freeman, R. (2018). Rapid warming is associated with population decline among terrestrial birds and mammals globally. *Global Change Biology*.

Srivastava, A. N., & Sahami, M. (2009). *Text mining Classification, clustering, and applications*. *CRC Press*.

Stekhoven, D. J., & Buhlmann, P. (2012). MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118.

Sterne, J. A. C., White, I. R., Carlin, J. B., et al. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ Clinical Research*, *338*, b2393.

Stewart, L. A., Clarke, M., Rovers, M., et al. (2015). Preferred reporting items for a systematic review and meta-analysis of individual participant data: The PRISMA-IPD statement. *JAMA - Journal of the American Medical Association*.

Suggitt, A. J., Wilson, R. J., Isaac, N. J. B., et al. (2018). Extinction risk from climate change is reduced by microclimatic buffering. *Nature Climate Change*.

Sula, C. A. (2016). Research Ethics in an Age of Big Data. *Bulletin of the Association for Information Science and Technology*.

Suraci, J. P., Clinchy, M., Dill, L. M., et al. (2016). Fear of large carnivores causes a trophic cascade. *Nature Communications*, *7*.

Swenson, N. G. (2014). Phylogenetic imputation of plant functional trait databases. *Ecography*, *37*(2), 105–110.

Thorne, J., Chen, M., Myrianthous, G., et al. (2018). Fake news stance detection using stacked ensemble of classifiers.

Thornthwaite, C. W. (1948). An Approach toward a Rational Classification of Climate. *Geographical Review*.

Toivonen, T., Heikinheimo, V., Fink, C., et al. (2019). Social media data for conservation science: A methodological overview. *Biological Conservation*.

Trisos, C. H., Merow, C., & Pigot, A. L. (2020). The projected timing of abrupt ecological disruption from climate change. *Nature*.

Troudet, J., Grandcolas, P., Blin, A., et al. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, *7*(1), 9132.

Twitter. (2020). Developer Agreement and Policy.

UK Government. Data Protection Act 2018 (2018).

UNDP. (2020). *Visualizing 25 Years of Human Development*.

UNDP. (2021). Human Development Index (HDI).

UNEP-WCMC, & IUCN. (2021). Protected Planet: The World Database on Protected Areas (WDPA).

Van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Chapman and Hall/CRC.

Van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., et al. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*.

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. *Journal Of Statistical Software*, *10*(2), 1 – 68.

van de Kerk, M., de Kroon, H., Conde, D. A., et al. (2013). Carnivora Population

Dynamics Are as Slow and as Fast as Those of Other Mammals: Implications for Their Conservation. *PLoS ONE*, *8*(8).

van Klink, R., Bowler, D. E., Gongalsky, K. B., et al. (2020). Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science*.

Van Zanten, B. T., Van Berkel, D. B., Meentemeyer, R. K., et al. (2016). Continental-scale quantification of landscape values using social media data. *Proceedings of the National Academy of Sciences of the United States of America*.

Venter, O., Sanderson, E. W., Magrach, A., et al. (2016). Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nature Communications*.

Vicente Serrano, S. M., Beguiria, S., & Lopez-Moreno, J. I. (2010). A multi-scalar drought index sensitive to global warming: the standardised precipitation evapotranspiration index. *Journal of Climate*.

Webb, C. T., Hoeting, J. A., Ames, G. M., et al. (2010). A structured and dynamic framework to advance traits-based theory and prediction in ecology. *Ecology Letters*, *13*(3), 267–283.

Wilson, D. E., & Reeder, D. M. (2005). *Mammal Species of the World*. *A Taxonomic and Geographic Reference*.

Wolf, C., & Ripple, W. J. (2017). Range contractions of the world’s large carnivores. *Royal Society Open Science*, *4*(7), 170052.

Woodroffe, R. (2000). Predators and people: Using human densities to interpret declines of large carnivores. *Animal Conservation*, *3*(2), 165–173.

WWF. (2016). *Living Planet Report 2016: Risk and resilience in a new era*. *WWF International*. Gland, Switzerland.

WWF. (2020a). Living Planet Index: Data Portal.

WWF. (2020b). *Living Planet Report 2020 - Bending the curve of biodiversity loss*. *Wwf*.

Zavaleta, E. S., Pasari, J. R., Hulvey, K. B., et al. (2010). Sustaining multiple ecosystem functions in grassland communities requires higher biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*.

Zhang, E., & Zhang, Y. (2009). F-Measure. In *Encyclopedia of Database Systems*.

The world's largest carnivores: understanding declines and recoveries is the PhD thesis of Thomas Frederick Johnson. This work was led by Thomas Frederick Johnson between 2017 and 2021, with lead supervision from Dr. Manuela González-Suárez, and co supervision from Drs. Nick Isaac and Agustin Paviolo

This work explores the features influencing population trends of the world's largest carnivores, and describes the population status of these charismatic species,. The thesis provides novel methodological and data advances, with the core aim to protect these endangered species, and foster coexistence with humanity.

University of Reading

CEH

CeIBA

NERC
SCIENCE OF THE
ENVIRONMENT