Automated synthesis of biodiversity knowledge requires better tools and standardised research output

ECOGRAPHY

Forum

Automated synthesis of biodiversity knowledge requires better tools and standardised research output

Richard Cornford, Joseph Millard, Manuela González-Suárez, Robin Freeman and Thomas Frederick Johnson

R. Cornford (https://orcid.org/0000-0002-9963-3603) ✉ (richard.cornford16@imperial.ac.uk), Dept of Life Sciences, Imperial College London, London, UK. – RC and R. Freeman, Inst. of Zoology, Zoological Society of London, London, UK. RC, Dept of Life Sciences, Natural History Museum, London, UK. – J. Millard (https://orcid.org/0000-0002-3025-3565), Dept of Genetics, Evolution and Environment, Univ. College London, London, UK and Leverhulme Centre for Demographic Science, Univ. of Oxford, Oxford, UK. – M. González-Suárez (https://orcid.org/0000-0001-5069-8900), School of Biological Sciences, Univ. of Reading, Reading, UK. – T. F. Johnson (https://orcid.org/0000-0002-6363-1825), Dept of Animal and Plant Sciences, Univ. of Sheffield, Sheffield, UK.

Ecography
2022: e06068
doi: 10.1111/ecog.06068
Subject Editor: Barry Brook
Editor-in-Chief:
Jens-Christian C. Svenning
Accepted 17 January 2022

As the impact of anthropogenic activity on the environment has grown, research into biodiversity change and associated threats has also accelerated. Synthesising this vast literature is important for understanding the drivers of biodiversity change and identifying those actions that will mitigate further ecological losses. However, keeping pace with an ever-increasing publication rate presents a substantial challenge to efficient syntheses, an issue which could be partly addressed by increasing levels of automation in the synthesis pipeline.

Here, we evaluate the potential for automated tools to extract ecologically important information from the abstracts of articles compiled in the Living Planet Database. Specifically, we focused on extracting key information on taxonomy (studied species names), geographic location and estimated population trend, assessing the accuracy of automated versus manual information extraction, the potential for automated tools to introduce biases into syntheses, and evaluating if synthesising abstracts was enough to capture the key information from the full article.

Taxonomic and geographic extraction tools performed reasonably well, although information on studied species was sometimes limited in the abstract (compared to the main text) preventing fast extraction. In contrast, extraction of trends was less successful, highlighting the challenges involved in automating information extraction from abstracts, such as deficiencies in the algorithms, linguistic complexity associated with ecological findings, and limited information when compared to the main text.

In light of these results, we cautiously advocate for a wider use of automated taxonomic and geographic parsing tools for ecological synthesis. Additionally, to further the use of automated synthesis within ecology, we recommend a dual approach: development of improved computational tools to reduce biases; and enhanced protocols for abstracts (and associated metadata) to ensure key information is included in a format that facilitates machine-readability.

Keywords: data extraction, ecology, literature synthesis, machine learning, population trends, text mining

NORDIC SOCIETY OIKOS

www.ecography.org

Introduction

Anthropogenic activity is negatively impacting the natural world (IPBES 2019); vertebrate populations are declining (WWF 2020) and species are being lost at rates reminiscent of mass extinction events (Ceballos et al. 2015). This biodiversity loss threatens ecosystem function (Rockström et al. 2009, Leclère et al. 2020) – which humans rely on (Díaz et al. 2018) – placing people and their livelihoods at risk. Much of our knowledge regarding environmental change impacts draws on global syntheses such as the PREDICTS (Hudson et al. 2017) and BioTIME (Dornelas et al. 2018) datasets, in addition to intergovernmental reports (IPCC 2014, IPBES 2019). The rapid growth in environmental literature over the last 30 years (Anderson et al. 2021) has been essential for facilitating these syntheses. However, as the literature continues to grow, syntheses become ever more challenging and time consuming (Ananiadou et al. 2009, Cohen et al. 2012).

'Big data' approaches, and associated computational tools, provide a means to wrangle the extensive ecological literature into usable information (Westgate et al. 2018). Much of the recent development in synthesis methods has been in expediting and automating the searching for (Grames et al. 2019), and screening of (Wallace et al. 2012, Shackelford et al. 2020, Cornford et al. 2021), papers to address research questions. Within the medical literature, some approaches have even managed to automate the entire systematic review procedure (Marshall and Wallace 2019, Gates et al. 2020, Marshall et al. 2020, Yang et al. 2020, Brassey et al. 2021). In the environmental sciences, automated topic models have provided insight into research trends (Hintzen et al. 2020) and the identification of knowledge gaps (Westgate et al. 2015), with text-classifiers allowing for the automated analysis of social media content to understand public opinions of nature (Johnson et al. 2021a). Complementing these broader, summarisation approaches, direct extraction of ecologically valuable information from literature (e.g. species names and geographic locations) is a growing field, with recent examples including Akella et al. (2012), Millard et al. (2020) and Kulkarni and Di Minin (2021).

Harnessing big data approaches to automatically synthesise data found within individual publications could support the environmental sciences in capturing the abundance of primary literature for compilation projects (e.g. Hudson et al. 2017) and evidence reviews (e.g. the Conservation Evidence project; <www.conservationevidence.com>) within a fully reproducible pipeline. However, validating the outputs of automated approaches is crucial to ensure the tools are accurate and do not introduce unwanted biases (Westgate et al. 2018). Benchmarking also helps users compare between alternative approaches, and track performance gains as techniques improve. Unfortunately, collating data for such validation often requires extensive manual effort, making evaluation a challenge. An exception is the Living Planet Database (LPD; <https://livingplanetindex.org/data_portal>), a collection of vertebrate population time series, each tagged with a species name and monitoring location. The LPD is useful to test automated approaches for biodiversity assessment and ecological research for two core reasons. First, the LPD underpins the Living Planet Index, an aggregated index of changing vertebrate populations with important policy implications (WWF 2020). Second, the LPD is largely based upon research in the primary literature, meaning many records in the LPD can be traced back to a publication and, central to this work, an associated abstract. Here we use the LPD as a reference point to test the performance of automated synthesis approaches in an ecological context. Specifically, we evaluate the performance of automated approaches for three important tasks relating to the synthesis of biodiversity trends and provide recommendations on how to address detected limitations. While our analyses focus on ecological and biodiversity change data, the approach and identified issues are relevant widely to all environmental sciences. The three tasks are as follows:

1. *Taxonomic entity extraction* i.e. finding which species were studied. Recent papers have used automated extraction approaches to identify species in text (Gerner et al. 2010, Akella et al. 2012) and general taxonomic patterns in ecological research (Millard et al. 2020). However, formal assessments of extraction accuracy and vulnerability to bias are still relatively scarce and warrant investigation.
2. *Geographic location extraction* i.e. finding where the study was conducted (in this work, we group locations based on country borders, but other geopolitical/biogeographic boundaries could be specified). Whilst development in taxonomic extraction and application has accelerated in recent years (Gerner et al. 2010, Akella et al. 2012), geographic extraction has a far greater history and wealth of available methods (Buscaldi and Rosso 2008, Kitamoto and Sagara 2012, Speriosu and Baldridge 2013, Ding et al. 2018, Magge et al. 2018, Kokla and Guilbert 2020, Wang et al. 2020). Automated geographic extraction could be valuable for extracting countries in coarse spatial resolution synthesis projects, or in the pre-screening phase of fine resolution syntheses. However, automated geographic extraction has been rarely used in ecology and conservation, with very few examples of successful application (Fisher et al. 2011, Millard et al. 2020). As a result, whilst many methods have been developed (Kokla and Guilbert 2020), there is a general need to validate geographic extraction in the field of ecology.
3. *Population trend extraction* i.e. summarising estimated population trends for studied species and locations. Developing methods that can synthesise ecological findings and data could help manage the ever-growing literature. Population trends, describing change in abundance over time, are amongst the most valuable types of data to compile as they meet the criteria of an essential biodiversity variable, and can thus directly support conservation management and policy (Pereira et al. 2013, Jetz et al. 2019).

In addressing these tasks we explicitly consider two potential 'leaks', which could limit the accuracy of automatically generated output. First, automated synthesis tools used to

extract information from abstracts may be ineffective or biased, e.g. favouring the extraction of certain species or locations, and failing to detect population trends. Second, even if automated tools perform well in extracting information from abstracts, abstracts may not accurately represent full studies, e.g. the population trend in the abstract is over-emphasised or only example species are listed for a multi-species study.

To explore these tasks and leaks, we compiled 1556 English language abstracts from the LPD and assessed how well the outputs from the automated extraction aligned with that reported in the LPD (for species names and geographic locations). For 300 randomly sampled abstracts we also manually extracted species, locations and population trends, producing a dataset of publications with information extracted using three methods: 1) LPD estimates, manually extracted from full texts (full-text data); 2) information manually curated (by the authors) from abstracts (manually assessed abstracts); and 3) data automatically extracted from abstracts (automated). We compared alignment between these three extraction types to determine whether leaks were due to ineffective automated tools (automated estimates differ from both types of manual extraction) or abstracts lacking information present in the full text (manual extractions differ).

Methods

Taxonomic entity extraction

To automatically extract species names from abstracts, we used a two-step approach (detailed in the Supporting information and Millard et al. 2020). First, we used taxize::scrapenames (Chamberlain and Szocs 2013, Chamberlain et al. 2018) to extract potential taxonomic names from abstracts. We then applied string-matching to retain only Latin binomials also present in the 2017 Catalogue of Life (Roskov et al. 2017), ignoring non-vertebrates.

We used three comparisons to evaluate the performance of automated taxonomic extraction: a) automated versus full-text data in the LPD; b) automated versus manually assessed abstracts; and c) manually assessed abstracts versus full-text data in the LPD. For each comparison we calculated recall (percentage of species in latter present in the former, per publication), and bias (proportion of species within each order in the former divided by the proportion of species within each order in the latter). We investigated if this bias had phylogenetic signal, whereby certain clades would be under- or over-represented, by measuring Pagel's λ across orders.

Geographic location extraction

We used the CLIFF-CLAVIN geoparser (D'Ignazio et al. 2014) to extract focal geographic locations (countries and coordinates) from abstracts. As country strings can differ between the LPD records and those resolved by CLIFF-CLAVIN, we used the geographic coordinates from both to identify associated countries (see Supporting information for details). We measured the effectiveness of the automated geographic extraction in a comparable way to the taxonomic extraction, using recall based on country names, and bias as the proportional difference in country frequency between data extraction approaches.

Population trend extraction

We trained machine learning classifiers to predict aggregate, paper-level population trends using a paper's title and abstract (full details in the Supporting information). We assigned paper-level trend categories (increase, stable, decline or varied) based on the proportion and direction of significant population-level trends, which were themselves estimated from a $\log_{10}$-linear model of population time-series.

Both random forests and neural networks (constructed in Python; van Rossum 1995) were used to predict trends, representing two well-known and high-performing text classification techniques. The performance of these machine learning approaches is improved by larger amounts of training data and/or better data quality (Liu et al. 2019), which can be generated using data augmentation (Wei and Zou 2019). We explored the impact of text augmentation (e.g. randomly replacing words with a synonym) on the accuracy of our trend predictions, using the Python library EDA (easy data augmentation; Wei and Zou 2019).

Initial analyses (Supporting information, using the 1256 texts remaining after setting aside the 300 manually assessed abstracts), indicated that random forest classifiers incorporating data augmentation but ignoring texts containing 'varied' population trends, performed best. We therefore tested a classifier of this specification on the 300 manually assessed abstracts, comparing the performance of our automated approach to both manual alternatives using accuracy and Cohen's kappa (Kuhn 2020).

Results

Taxonomic entity extraction

For all 1556 texts, the automatic taxonomic extraction recalled an average of 80.8% of species relative to the full text (SD = 35.0%). When considering only the 300 manually assessed texts, our automated approach achieved average recall of 83.7% (SD = 34.1%; Fig. 1a). Cases of low recall were primarily influenced by under-reporting of species within abstracts, as only 82.5% (SD = 35.2%; Fig. 1c) of species with population data in the full texts were recorded in the abstracts. In contrast, loss of information from the automated method was low, with an average of 93.6% recall (SD = 17.9%; Fig. 1b) when compared to data manually extracted from abstracts.

Our analysis also suggests taxonomic bias in both the automated tool and abstract content, when compared to the full text, with some orders substantially under- and over-detected (Fig. 1d). Despite these disparities, we did not detect any significant phylogenetic signal (Pagel's λ likelihood-test

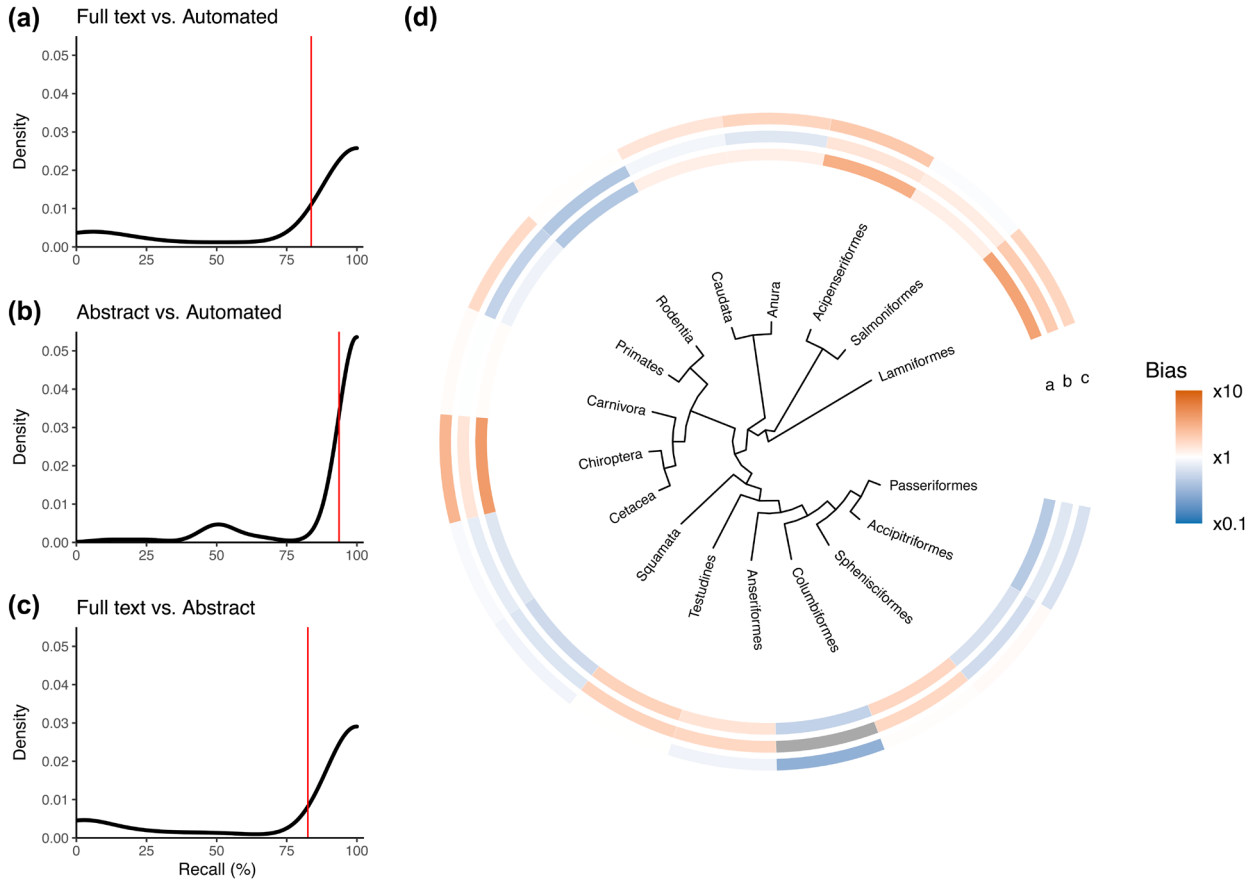

Figure 1. Recall and bias of automated taxonomic extraction using taxize and Catalogue of Life on our sample of 300 texts. (a–c) Distribution of recall (percentage of species successfully detected within each study) for the automated tool relative to the full-text data (a), automated relative to manually assessed abstract data (b) and the manually assessed abstracts relative to the full texts (c). The red line represents the mean recall within each of these comparisons. (d) Phylogenetic variability in detection bias of species within texts. Despite visible variation, we found no phylogenetic signal in detection bias (Pagel's λ likelihood-test p-value > 0.05). Each ring around the phylogenetic tree (a, b, c) relates to the comparisons between extraction approaches indicated in the left-hand column titles. Bias ranges from ×0.1 to ×10, where a value of 0.1 in ring a, for example, would indicate that a given order occurs 10 times less frequently in the automated extraction than in the data taken from the full text. The bias colouring is on the $\log_{10}$ scale. Grey indicates an absence of the order in the reference dataset.

p-value > 0.05) in detection bias for any of our comparisons, although sample sizes of 16 and 17 orders may be a limitation here (Fig. 1d and Supporting information).

Geographic location extraction

The automated geographic extraction generally performed worse than the taxonomic extraction, accurately identifying an average of 69.1% (SD = 45.5%) of countries relative to the full-text extraction when considering the 1556 records. For the 300 manually assessed texts, average recall rose to 77.9% (SD = 40.4%; Fig. 2.1a). However, unlike the taxonomic extraction, accuracy error was driven by the poor performance of the automated geographic extraction (mean recall = 82.1%, SD = 36.7%; Fig. 2.2a), as the manually assessed abstracts and full texts were well aligned (mean recall = 93.9%, SD = 22.6%; Fig. 2.3a).

The automated geographic extraction also showed bias, tending to over-assign records to countries with English as the first language (e.g. USA, UK and Australia; Fig. 2.1b and 2.2b) and under-assign records across South America and Southeast Asia. In contrast, comparing between full texts and manually assessed abstracts suggests more moderate over-/under-reporting of countries in abstracts (Fig. 2.3b). As an example, records labelled as France in the full texts were split between seven countries in automated extraction, nine countries when comparing the manually assessed abstracts to the automated extraction, and only two countries when comparing the full texts to the manually assessed abstracts (Fig. 2.1c, 2.2c and 2.3c).

Population trend extraction

Of the 300 manually assessed abstracts, 21 were classified as varied based on full-text data, and 180 as either varied or unclear by manual assessment. Here, we therefore present results based on the subset of these 300 texts where all classification approaches (automated, full-text data and manually assessed abstracts) produced categories of either increase, decline or stable (111 studies). These results allow for a fair

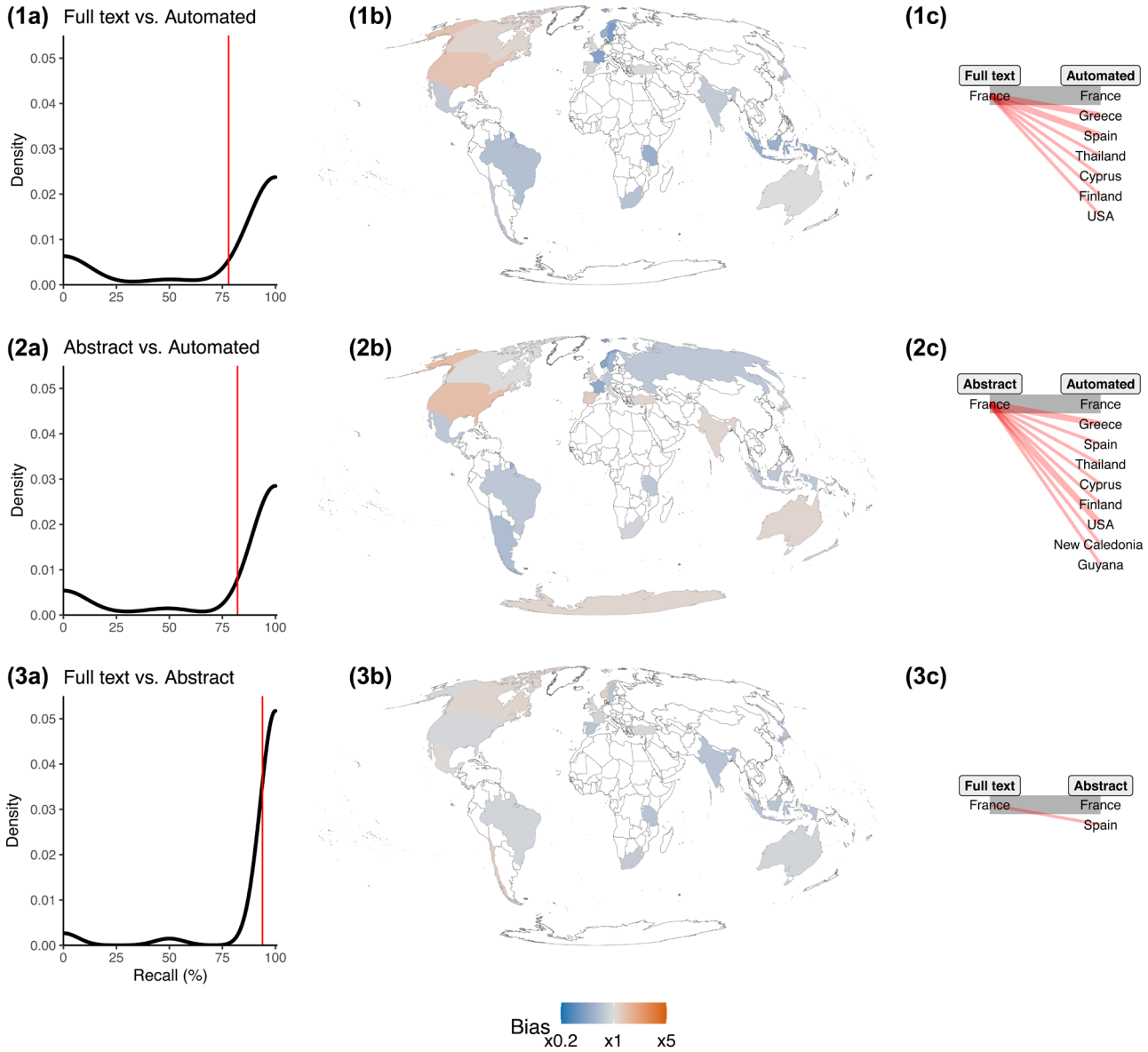

Figure 2. Recall and bias of automated geoparsing using CLIFF-CLAVIN on our sample of 300 texts. (1a, 2a, 3a) Distribution of recall (percentage of countries successfully detected) for the automated tool relative to the full-text data in the Living Planet Database (1a), the automated tool relative to the manually assessed abstracts (2a) and the manually assessed abstracts relative to the full-text data (3a). The red line represents the mean recall within each of these comparisons. (1b, 2b, 3b) Spatial variability in detection bias of countries within texts, comparing full-text information versus automated, manual abstract assessment versus automated, and full texts versus manually assessed abstracts. Bias ranges from ×0.2 to ×5, where a value of 0.2 in 1b, for example, would indicate that a given country occurs five times less frequently in the automated extraction than in the data taken from the full text. The bias colouring is on the $\log_{10}$ scale. White countries indicate no representation in the reference dataset. (1c, 2c, 3c) Assignment of records in the comparison groups relative to France in the reference group. The grey line indicates a match between the reference and comparison group, whilst red indicates a mismatch. Line thickness describes its proportional frequency.

comparison between approaches, but also likely over-estimate the performance of both our automated and manual approaches, as we focus on the simplest ecological (and textual) scenarios. Results based on all 300 manually assessed texts can be found in the Supporting information.

Automated population trend prediction performed worse than either taxonomic or geographic data extraction, with accuracy of 64.9% compared to the full text (kappa: 0.473, 'moderate'; Fig. 3a). Interestingly, the accuracy of manual abstract categorisation compared to estimates based on data from the full text was lower still, at 57.7% (kappa: 0.387, 'fair'; Fig. 3c). Agreement between the manually assessed abstracts and automated classifications was also low (accuracy: 58.6%, kappa: 0.369, 'fair'; Fig. 3b), suggesting the automated and manual approaches made different mistakes.

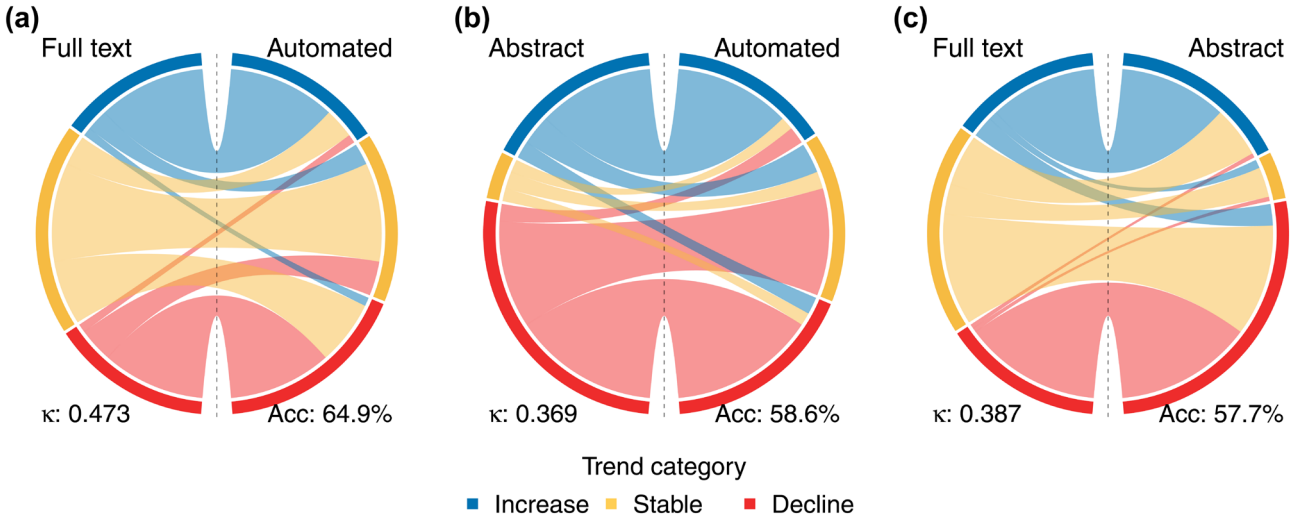

Figure 3. Accuracy of automated population trend extraction using random forest classifiers on 111 texts. The chord diagrams show the agreement between automated classifiers and the estimates based on full-text data (a), automated classifiers and manually assessed abstracts (b) and manually assessed abstracts and full-text data (c). Coloured semi-circles on the left of each panel show the distribution of trend categories in the reference data. Links indicate how these reference categories are distributed among the predicted categories (right-hand semi-circles), with link width proportional to the number of classifications represented. Accuracy (Acc) and Cohen's Kappa statistic (K) provide quantitative measures of performance. Accuracy ranges from 0% (all predictions incorrect) to 100% (all predictions correct). Kappa values of 0.21–0.40 indicate 'fair' agreement, and 0.41–0.60 'moderate' (Landis and Koch 1977).

Discussion

Here we evaluated three tools to explore key challenges in automated synthesis of ecological and biodiversity knowledge. Our study explored their limitations and tested the potential sources of errors or information 'leaks'. The first two tools for automated taxonomic and geographic extraction delivered moderately successful performance. These approaches are already being used (Millard et al. 2020), and, compared to manual extraction, offer much faster and more easily reproducible data collation. However, we found that automated extraction of species and locations can introduce biases (e.g. over-/under-representation of certain taxonomic orders), and so should be used cautiously. The performance of taxonomic and geographic extraction was affected by the representativeness of abstracts (relative to the main text), and the biases inherent in automated algorithms. For example, the automatic taxonomic tool performed well in extracting Latin binomials from abstracts, but abstracts poorly represented main texts in terms of taxonomic coverage. On the other hand, the automatic geographic extraction tool performed poorly in extracting locations from abstracts, but abstracts represented main texts well in terms of geographic coverage. The third tool we developed and tested was a population trend extractor which delivered relatively poor performance driven by a lack of clarity regarding trend descriptions in abstracts (a problem concerning how research is presented in the literature) and by the complexity associated with summarising multiple trends into one value (an issue related to limitations of automated tools).

The relatively good performance of automated taxonomic and geographic extraction is promising for current/future synthesis projects, through application as a text prioritisation tool (an example of a project already using these approaches is EntoGEM; <https://entogem.github.io/>). One current issue with global synthesis projects is their substantial taxonomic (McRae et al. 2017, Troudet et al. 2017) and spatial biases (Gonzalez et al. 2016, Tydecks et al. 2018). These biases hinder inference and erode our ability to predict over space (Yates et al. 2018) and phylogeny (Johnson et al. 2021b). Automatically analysing the content of collated studies early in the synthesis pipeline could reveal imbalances/gaps in geographic and taxonomic coverage, which if not addressed would undermine subsequent analyses and conclusions. Prioritising the collation of studies that fill data gaps has already been used in some synthesis projects (Jones et al. 2009, Hudson et al. 2017), traditionally relying on manual searches. Using automated taxonomic and geographic identification tools to also identify such publications could speed up the collation of representative ecological data, enabling more rapid and accurate syntheses, thereby better informing conservation decisions.

While taxonomic and geographic tools can be recommended, the poorer performance of the trend extraction tool limits our ability to automate the entire trend synthesis process. Although the accuracy demonstrated here may be sufficient for providing a coarse, preliminary overview of population trend distributions in scoping searches, we suspect obtaining more reliable estimates per-study is currently unfeasible for a variety of reasons. First, nature can be complex, making the estimation of population abundance trends difficult (Humbert et al. 2009), and potentially inaccurate (Fournier et al. 2019). Descriptions of such trends are therefore likely to be linguistically complex and could vary

depending on the trend estimation method used. Second, information is often not reported to facilitate synthesis, and abstracts can use polarising language, e.g. a population may be increasing, but this message could be confused if the text opens with negative or disaster-based language, or may describe only 'key results' that do not reflect the full content. Third, it is challenging to develop tools that can process complex texts or adequately capture information about multiple diverging trends. The first two issues reflect nature itself and academic writing and are embedded in much of the published ecological literature. It seems unlikely that the way in which researchers write will change given its importance in the framing of research and the complex nature of some biodiversity patterns. Although future developments in machine-learning tools may enable accurate automated trend extraction, we think a more ambitious, short-term change is needed in the form of standardization of abstract structures, language-use and inclusion of metadata. The importance of crafting titles, abstracts and keywords to ensure primary research is easily discoverable for use in syntheses has recently been highlighted (Hennessy et al. 2021), but automated synthesis would likely benefit from even more structure. Some journals, e.g. Global Ecology and Biogeography, structure abstracts into sections, where results and methods are isolated. These structures would limit the conflation of results with the disaster-based language often found in the introduction and discussion, thereby improving performance of metadata extraction.

Our study tackles some important challenges involved in automating ecological synthesis but there are limitations associated with the approaches we present. First, we focused solely on English language texts, and only explored tools designed for English. Although English is the main language of the scientific literature (Nuñez and Amano 2021), and LPD articles, we recognise that considering publications in languages other than English is important for ensuring biodiversity knowledge/inference is unbiased (Konno et al. 2020, Amano et al. 2021). We therefore encourage future work to develop/evaluate similar automated synthesis tools for texts in a variety of languages. Second, we have only evaluated the automated extraction of data from article abstracts. Text-mining approaches are known to improve when full-texts are used (Westergaard et al. 2018), with this work also finding that abstracts may not accurately represent the content of the full paper. However, we argue that as access to full-text articles is often restricted by paywalls, it is important that fast, accurate, automated syntheses can be performed using freely, and easily, available abstracts. Third, we assessed automated tools for extracting large-scale patterns of biodiversity change, i.e. qualitative population trends associated with countries and species. The collation of such information is vital for systematic maps/coarse resolution synthesis, but may struggle to capture the known nuances of biodiversity change, especially at local scales (Dornelas et al. 2019, Leung et al. 2020). Further work to enhance the granularity of data extraction and minimise identified biases is therefore needed before automated approaches are readily applied without caution. Finally, our analysis centres on a database of vertebrate population time-series. Previous research comparing automated and manual extraction of various animal pollinator species (mostly insects) from abstracts found recall of 79.5% (Millard et al. 2020), suggesting that the quality of automated taxon tagging may vary across phylogenetic groups. Further evaluation of automated species extraction across kingdoms and phyla is therefore required, especially as targeting data retrieval for less charismatic groups – i.e. not mammals and birds – is essential for furthering biodiversity knowledge (Guerra et al. 2020).

Conclusion

In this work we have explored the three broad tasks of extracting taxonomic names, geographic locations and population trends from article abstracts. We have shown that the species and country tagging tools perform sufficiently well for us to recommend their wider use, e.g. study prioritisation and coarse-scale literature summarisation, but caution is needed, as these automated approaches can introduce biases, e.g. under-representing certain countries. Our trend extraction approach delivered poorer performance, being constrained by poor alignment between abstracts and the main text, poor text classifier performance, and the complexity of the population trend data and its descriptions. To facilitate improved automated synthesis within ecology, we recommend both the improvement of computational tools, and better structuring of abstract text.

Acknowledgements – We would like to that thank the editor and two reviewers for their constructive comments which greatly improved the clarity of our paper.
Funding – RC was supported by the QMEE CDT, funded by NERC grant no. NE/R012229/1.

Author contributions

Richard Cornford: Conceptualization (equal); Formal analysis (equal); Methodology (equal); Writing – original draft (equal); Writing – review and editing (lead). **Joseph Millard**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal); Writing – original draft (equal); Writing – review and editing (supporting). **Manuela González-Suárez**: Conceptualization (supporting); Data curation (equal); Writing – review and editing (supporting). **Robin Freeman**: Conceptualization (supporting); Writing – review and editing (supporting). **Thomas Frederick Johnson**: Conceptualization (lead); Data curation (equal); Formal analysis (supporting); Methodology (equal); Project administration (lead); Writing – original draft (equal); Writing – review and editing (supporting).

Transparent Peer Review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.06068>.

Data availability statement

Code and data are available from the Zenodo Repository : (<https://doi.org/10.5281/zenodo.5866181>) and from Github (<https://github.com/rcornf/auto_synth_2021/tree/v1.0.0>).

Supporting information

The supporting information associated with this article is available from the online version.

References

Akella, L. M. et al. 2012. NetiNeti: discovery of scientific names from text using machine learning methods. – BMC Bioinform. 13: 1–10.

Amano, T. et al. 2021. Tapping into non-English-language science for the conservation of global biodiversity. – PLoS Biol. 19: e3001296.

Ananiadou, S. et al. 2009. Supporting systematic reviews using text mining. – Soc. Sci. Comput. Rev. 27: 509–523.

Anderson, S. C. et al. 2021. Trends in ecology and conservation over eight decades. – Front. Ecol. Environ. 19: 274–282.

Brassey, J. et al. 2021. Developing a fully automated evidence synthesis tool for identifying, assessing and collating the evidence. – BMJ Evid. Based Med. 26: 24–27.

Buscaldi, D. and Rosso, P. 2008. Map-based vs. knowledge-based toponym disambiguation. – Proceedings of the 5th workshop on geographic information retrieval, pp. 19–22.

Ceballos, G. et al. 2015. Accelerated modern human-induced species losses: entering the sixth mass extinction. – Sci. Adv. 1: e1400253.

Chamberlain, S. and Szocs, E. 2013. taxize – taxonomic search and retrieval in R. – F1000Research 2: 191.

Chamberlain, S. et al. 2018. taxize: taxonomic information from around the web [Computer software manual]. – R package ver. 0.9.5. <https://github.com/ropensci/taxize>.

Cohen, A. M. et al. 2012. Studying the potential impact of automated document classification on scheduling a systematic review update. – BMC Med. Inform. Decision Making 12: 33.

Cornford, R. et al. 2021. Fast, scalable and automated identification of articles for biodiversity and macroecological datasets. – Global Ecol. Biogeogr. 30: 339–347.

Díaz, S. et al. 2018. Assessing nature's contributions to people. – Science 359: 270–272.

D'Ignazio, C. et al. 2014. CLIFF-CLAVIN: determining geographic focus for news articles. – NewsKDD: data science for news publishing. KDD, NY, USA.

Ding, D. et al. 2018. Geographical feature extraction for entities in location-based social networks. – Proceedings of the 2018 world wide web conference, pp. 833–842.

Dornelas, M. et al. 2018. BioTIME: a database of biodiversity time series for the anthropocene. – Global Ecol. Biogeogr. 27: 760–786.

Dornelas, M. et al. 2019. A balance of winners and losers in the anthropocene. – Ecol. Lett. 22: 847–854.

Fisher, R. et al. 2011. Global mismatch between research effort and conservation needs of tropical coral reefs. – Conserv. Lett. 4: 64–72.

Fournier, A. M. et al. 2019. Site-selection bias and apparent population declines in long-term studies. – Conserv. Biol. 33: 1370–1379.

Gates, A. et al. 2020. Decoding semi-automated title-abstract screening: findings from a convenience sample of reviews. – Syst. Rev. 9: 1–12.

Gerner, M. et al. 2010. LINNAEUS: a species name identification system for biomedical literature. – BMC Bioinform. 11: 1–17.

Gonzalez, A. et al. 2016. Estimating local biodiversity change: a critique of papers claiming no net loss of local diversity. – Ecology 97: 1949–1960.

Grames, E. M. et al. 2019. An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks. – Methods Ecol. Evol. 10: 1645–1654.

Guerra, C. A. et al. 2020. Blind spots in global soil biodiversity and ecosystem function research. – Nat. Commun. 11: 1–13.

Hennessy, E. A. et al. 2021. Ensuring prevention science research is synthesis-ready for immediate and lasting scientific impact. – Prev. Sci., doi: 10.1007/s11121-021-01279-8.

Hintzen, R. E. et al. 2020. Relationship between conservation biology and ecology shown through machine reading of 32 000 articles. – Conserv. Biol. 34: 721–732.

Hudson, L. N. et al. 2017. The database of the PREDICTS (projecting responses of ecological diversity in changing terrestrial systems) project. – Ecol. Evol. 7: 145–188.

Humbert, J.-Y. et al. 2009. A better way to estimate population trends. – Oikos 118: 1940–1946.

IPBES 2019. – In: Díaz, S. et al. (eds), Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES Secretariat.

IPCC 2014. Climate change 2014: synthesis report. – In: Core Writing Team et al. (eds), Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change. IPCC.

Jetz, W. et al. 2019. Essential biodiversity variables for mapping and monitoring species populations. – Nat. Ecol. Evol. 3: 539.

Johnson, T. F. et al. 2021a. classecol: classifiers to understand public opinions of nature. – Methods Ecol. Evol. 12: 1329–1334.

Johnson, T. F. et al. 2021b. Handling missing values in trait data. – Global Ecol. Biogeogr. 30: 51–62.

Jones, K. E. et al. 2009. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals: ecological archives e090-184. – Ecology 90: 2648–2648.

Kitamoto, A. and Sagara, T. 2012. Toponym-based geotagging for observing precipitation from social and scientific data streams. – Proceedings of the ACM multimedia 2012 workshop on geotagging and its applications in multimedia, pp. 23–26.

Kokla, M. and Guilbert, E. 2020. A review of geospatial semantic information modeling and elicitation approaches. – ISPRS Int. J. Geo-Inform. 9: 146.

Konno, K. et al. 2020. Ignoring non-english-language studies may bias ecological meta-analyses. – Ecol. Evol. 10: 6373–6384.

Kuhn, M. 2020. caret: classification and regression training [computer software manual]. – R package ver. 6.0-86. <https://CRAN.R-project.org/package=caret>.

Kulkarni, R. and Di Minin, E. 2021. Automated retrieval of information on threatened species from online sources using machine learning. – Methods Ecol. Evol. 12: 1226–1239.

Landis, J. R. and Koch, G. G. 1977. The measurement of observer agreement for categorical data. – Biometrics 33: 159–174.

Leclère, D. et al. 2020. Bending the curve of terrestrial biodiversity needs an integrated strategy. – Nature 585: 551–556.

Leung, B. et al. 2020. Clustered versus catastrophic global vertebrate declines. – Nature 588: 267–271.
Liu, Y. et al. 2019. RoBERTa: a robustly optimized BERT pretraining approach. – arXiv preprint arXiv:1907.11692.
Magge, A. et al. 2018. Deep neural networks and distant supervision for geographic location mention extraction. – Bioinformatics 34: i565–i573.
Marshall, I. J. and Wallace, B. C. 2019. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. – Syst. Rev. 8: 1–10.
Marshall, I. J. et al. 2020. Semi-automated evidence synthesis in health psychology: current methods and future prospects. – Health Psychol. Rev. 14: 145–158.
McRae, L. et al. 2017. The diversity-weighted living planet index: controlling for taxonomic bias in a global biodiversity indicator. – PLoS One 12: e0169156.
Millard, J. W. et al. 2020. Text-analysis reveals taxonomic and geographic disparities in animal pollination literature. – Ecography 43: 44–59.
Nuñez, M. A. and Amano, T. 2021. Monolingual searches can limit and bias results in global literature reviews. – Nat. Ecol. Evol. 5: 264–264.
Pereira, H. M. et al. 2013. Essential biodiversity variables. – Science 339: 277–278.
Rockström, J. et al. 2009. A safe operating space for humanity. – Nature 461: 472–475.
Roskov, Y. et al. 2017. Species 2000 and ITIS Catalogue of Life, 2017 annual checklist. – <www.catalogueoflife.org/annual-checklist/2017>.
Shackelford, G. E. et al. 2020. Accumulating evidence using crowdsourcing and machine learning: a living bibliography about existential risk and global catastrophic risk. – Futures 116 102508.
Speriosu, M. and Baldridge, J. 2013. Text-driven toponym resolution using indirect supervision. – Proceedings of the 51st annual meeting of the association for computational linguistics. Vol. 1 (long papers), pp. 1466–1476.
Troudet, J. et al. 2017. Taxonomic bias in biodiversity data and societal preferences. – Sci. Rep. 7: 9132.
Tydecks, L. et al. 2018. Spatial and topical imbalances in biodiversity research. – PLoS One 13: e0199327.
van Rossum, G. 1995. Python reference manual. – CWI (Centre for Mathematics and Computer Science).
Wallace, B. C. et al. 2012. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. – Proceedings of the 2nd ACM SIGHIT international health informatics symposium, pp. 819–824.
Wang, J. et al. 2020. NeuroTPR: a neuro-net toponym recognition model for extracting locations from social media messages. – Trans. GIS 24: 719–735.
Wei, J. and Zou, K. 2019. EDA: easy data augmentation techniques for boosting performance on text classification tasks. Proceedings of the 2019 Conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), pp. 6382-6388.
Westergaard, D. et al. 2018. A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. – PLoS Comput. Biol. 14: e1005962.
Westgate, M. J. et al. 2015. Text analysis tools for identification of emerging topics and research gaps in conservation science. – Conserv. Biol. 29: 1606–1614.
Westgate, M. J. et al. 2018. Software support for environmental evidence synthesis. – Nat. Ecol. Evol. 2: 588–590.
WWF 2020. Living Planet Report – 2020. – In: Almond, R. E. A. et al. (eds), Bending the curve of biodiversity loss. WWF International.
Yang, C. et al. 2020. Perspective: towards automated tracking of content and evidence appraisal of nutrition research. – Adv. Nutr. 11: 1079–1088.
Yates, K. L. et al. 2018. Outstanding challenges in the transferability of ecological models. – Trends Ecol. Evol. 33: 790–802.