Biol Invasions
https://doi.org/10.1007/s10530-023-03067-7

REVIEW

Synthesising 35 years of invasive non-native species research

Emily A. Stevenson · Peter Robertson · Emily Hickinbotham · Louise Mair · Nigel J. Willby · Aileen Mill · Olaf Booy · Kirsty Witts · Zarah Pattison

Received: 22 August 2022 / Accepted: 5 April 2023

Abstract The growing focus on the threat of invasive non-native species (INNS) in international biodiversity targets highlights a need for targeted research to support effective understanding, legislation, and management. However, the publishing landscape of invasion biology is complex and expanding rapidly, making consolidation of information increasingly challenging. To identify the major research themes in the INNS literature and to understand how these have changed over the last 35 years, we applied a topic modelling approach. We analysed approximately 10,000 peer-reviewed article abstracts to identify 50 key topics being discussed in the literature. We also quantified how publications on these topics changed over time and how commonly different topics interacted within articles as a measure of their connectedness. Topics covering *Population genetics, Policy, First records* and *Insect biocontrol* were the most frequent. Topics were grouped into broad themes, with the largest theme related to Ecosystems, followed by Monitoring, then Management and decision-making. Significant overrepresentation for particular geographical regions and taxa in the literature were apparent. Considering relative changes through time, the most prevalent topics in each decade reflected policy influences, and technological developments. When assessing the degree of connectedness- *Policy*, *Population Genetics* and *Management Strategies* showed low levels of co-occurrence with other topics. This is of particular concern for topics focussed on *Policy* and *Management Strategy* as it suggests a weakness at the science-policy interface around accessing and exchanging of evidence. If progress towards future global targets is to be made, we argue that more interdisciplinary research must be encouraged, in particular to better incorporate policy and management considerations into the wider research landscape.

Keywords Gap-analysis · Management · Alien · Policy · Topic modelling

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10530-023-03067-7.

E. A. Stevenson · P. Robertson · E. Hickinbotham · L. Mair · A. Mill · O. Booy · K. Witts · Z. Pattison
Modelling, Evidence and Policy Group, School of Natural and Environment Sciences, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK

N. J. Willby · Z. Pattison (✉)
Biological and Environmental Sciences, University of Stirling, Stirling FK9 4LA, UK
e-mail: zarah.pattison2@stir.ac.uk

O. Booy
Animal and Plant Health Agency, Non-Native Species Secretariat, Sand Hutton, York YO41 1JW, UK

Published online: 29 April 2023

Introduction

The frequency of biological invasions, driven by globalisation, is increasing and shows no sign of reaching saturation (Seebens et al. 2018; Essl et al. 2020). As invasions increase, there will be further disruption to ecosystems and negative consequences for native species (Ricciardi et al. 2013) and human livelihoods (Shackleton et al. 2019). Preventing a conservation crisis requires effective management of invasive non-native species (INNS). However, previous studies have found a disconnect between the research generated and its policy and management applications (Esler et al. 2010). This is reflected in concern over the limited achievement of international targets related to INNS. The assessment of global actions to meet the Convention on Biological Diversity's (CBDs) Aichi Biodiversity Targets highlight that the efforts to combat species invasions have been outpaced by globalisation, and in particular the impact of massively expanded trade (CBD 2020). It is therefore important to consider if and how research is supporting global conservation goals of preventing and managing the impacts of INNS.

The body of literature related to INNS is large and continues to proliferate (Richardson and Pyšek 2008; Esler et al. 2010). However, its accessibility can be hindered by differences in the approach and terminology adopted by each discipline within invasion science, be that theoretical, practical, policy or management focussed, for example (Pyšek et al. 2004; Blackburn et al. 2011; Enders et al. 2020; Robertson et al. 2020). Extracting key messages from such a complex and wide-ranging body of literature is therefore challenging, especially when attempting to navigate the full extent of the publishing landscape. Topic modelling provides a semi-automated statistical tool to assess the content of articles in a corpus (a large body of literature; Blei and Lafferty 2009). Information is collected from article abstracts, with the patterns of word co-occurrence then analysed to identify common ideas or topics (Griffiths and Steyvers 2004). This provides a more quantitative alternative to standard literature reviews, summarising trends without researcher bias and allowing information from different thematic, spatial, and temporal scales to be consolidated (Westgate et al. 2015). The approach has already been applied to conservation science and conservation planning research (Westgate et al. 2015; Mair et al. 2018).

We used topic modelling to assess how scientific research contributes to developments in the field of invasion science, using terms commonly utilised by separate disciplines to capture the diversity in the literature. For the purposes of this study, invasive non-native species are a species, subspecies, or lower taxon; including any part, gametes, seeds, eggs, or propagules of such species that might survive and subsequently reproduce, whose introduction and/or spread outside their natural past or present distribution threatens biological diversity (CBD 2002). Adopting a broad inclusive definition of INNS allowed us to determine potential gaps and understudied aspects within the literature.

Our aim was to identify the key themes and knowledge gaps in invasion science and how these have changed over the period 1985–2019. Our objectives were to (1) identify the topics most popular in invasion science research; (2) identify how topics have changed in prevalence over time; and (3) assess how different topics interact (co-occur) within articles, as a means of identifying more or less well-connected topics in the literature. We specifically focus on how well research topics on invasive species are connected to management and policy, which is key to achieving scientific impact. Highlighting such absences can help pinpoint knowledge gaps needing to be bridged, areas where a lack of connectedness between topics/disciplines could help resolve stubborn problems, or in prioritising future research agendas (Esler et al. 2010).

Methods

Literature search

Documents included in the literature search were classed as articles (peer reviewed publications) according to the search engine and refined to only include articles written in English. The article inclusion criteria for this study required that each article clearly focused on invasive non-native species, as opposed to non-native/alien or naturalised species that were not considered invasive. A study by Golebie et al. (2022) found the use of terminology to broadly align with the stage of invasion, in that "invasive" was most commonly used except when the research was conducted at early stages of invasion, when "non-native" was most commonly used. The results of this study may therefore be biased towards the biodiversity aspects of invasion biology, as other areas of research such as animal and plant health interact with INNS but may use different terminology, such as "pest" (see Roda et al. 2011 as an example).

We searched Web of Science and Scopus on April 17th, 2020, using the search terms: "invasive" AND ["non-native" OR "alien"] AND "species". These terms were considered to capture the most relevant publications specific to "invasive non-native species". A range of other terms are used in some literature (e.g., non-indigenous, feral) however, we found that additional search terms returned a larger volume of irrelevant articles (for a comparison of results obtained from an alternative search string, see Supplementary Information S1.1). We considered publications up to and including 2019. The citations and abstracts for each article were then downloaded into the reference management software EndNote X9 (Clarivate, 2020) for review. For a full repository of all included articles see Supplementary Information 2.

Data cleaning

Duplicate articles and those without abstracts were excluded, reducing the number of papers from 17,558 to 10,945. Titles and abstracts of the 10,945 articles were screened and then manually reviewed to ensure that they met the article inclusion criteria for this study. To be included, articles needed to focus specifically and primarily on INNS, so articles were excluded if they discussed irrelevant topics such as medicine, or only mentioned INNS in passing as an example or as part of a list of issues affecting biodiversity. This was done by two people (Stevenson and Witts), due to the large volume of articles returned. A subset of 600 articles were cross-checked to refine inclusion criteria and ensure consistency between reviewers, with disparities discussed in a larger group to solidify the inclusion criteria prior to reviewing the rest of the articles. Examples of article abstracts that were excluded, along with the reason for exclusion, are provided in Supplementary Information S1.2. This left 9882 articles for processing.

Following the methodology used by Westgate et al. (2015) abstracts were converted into a corpus and processed using the R package *tm* (Meyer et al. 2008; R Core Team, 2021). To optimise the output of the model, a range of commonly occurring or generic words were removed. This included all numbers, written as words or digits (Grün and Hornik 2011), and the search terms as these would be present in all abstracts. Words that did not provide information on the topic being discussed, such as "like'', "may" and "either", were removed using a combination of the list of "stop words" provided as part of the *tm* package (Meyer et al. 2008) and terms that we chose to remove following a preliminary run of the model, as they did not provide meaning to the results, such as "can" and "will" (Full list in Supplementary Information S1.3). Hyphens and forward slashes were changed to spaces, capital letters were converted to lowercase and all other punctuation was removed (Grün and Hornik 2011). The remaining words were then stemmed to their common root and all spelling converted to the American style to prevent duplicates. Finally, words that appeared in five or fewer articles were removed to speed up processing and because such infrequent words contribute little to topic generation (Griffiths and Steyvers 2004).

Topic modelling

Topic modelling identifies the main ideas being discussed in a corpus using sets of words that co-occur with unusual frequency, which are then grouped into topics (Griffiths and Steyvers 2004). The model generates the weight that each word contributes to a topic, allowing the focus of each topic to be determined from the most highly weighted words. Individual articles will often discuss multiple topics and the relative weighting of each topic will vary across articles. This allows the main topic and diversity of topics within an article to be identified.

The number of topics fitted to a particular corpus is determined a priori (Mair et al. 2018). We ascertained the most suitable number of topics by using ten-fold block cross-validation to run a series of models with varying numbers of topics (n = 10, 20, 30, 40, 50, 60, 100, 200). For each number of topics, models were fitted to 90% of the corpus. Model fit was tested by calculating perplexity on the with-held 10% of the corpus; lower perplexity indicates better model fit

(Grün and Hornik 2011). This process was repeated for each candidate number of topics. The block-cross validation showed that as the number of topics increased, model perplexity decreased, suggesting there were many topics discussed in the corpus (Supplementary Information S1.4).

It was necessary to balance the need to investigate the complexity of the corpus with our aim of simplifying the overarching patterns in the literature, for easier consolidation by all stakeholders (sensu Westgate et al. 2015; Mair et al. 2018). We therefore fitted 50 topics using a Latent Dirichlet Allocation (LDA) model with Gibbs sampling, using the R package *topicmodels* (Grün and Hornik 2011). Model perplexity also had declined substantially by 50 topics, indicating better model fit (Supplementary Information S1.4).

We used the 20 highest weighted words in each topic, with particular focus on the top five words (Supplementary Information S1.5), to name the topics generated by the model. Alongside this, a workshop was held with all authors who brought wide and varying perspectives on INNS, to agree on topic names based on the data and their expertise. Where words associated with a topic appeared vague, articles were extracted for those topics and reviewed to gain further insight into the weighted words describing a topic. Topics were named to simplify the information being presented. Thereafter, topics were grouped into broad themes, based on an analysis of topic similarity, to allow for easier identification of patterns in the data (Supplementary Information S1.6). Topic similarity was calculated using the weight with which each unique word was assigned to each topic, following the methods described in Westgate et al. (2015). We acknowledge that individual topics may reasonably be placed in more than one theme, however the combination of inspection of similarity and expert input provided us with a means to organise topics into themes in the most appropriate way to facilitate communication of results.

To assess topic frequency within the corpus, we assigned articles to their single highest weighted topic (for the distribution of weights of highest weighted topic per article versus the weights of all other topics per article see Supplementary Information S1.7). There are a range of approaches that can be used to assign articles topics, including applying a topic weight threshold, such that an article is assigned to one or more topics that have a weight greater than the selected threshold, and using a cumulative weight threshold, such that an article is assigned to topics with weights that cumulatively sum to a selected threshold. We investigated the effect of using a cumulative threshold to assign articles to topics on our results, and present these analyses in Supplementary Information S1.8.

Topic generality and specificity

Within the literature, some topics are "general". General topics reflect broad ideas that are discussed across many articles, often in association with other topics. Meanwhile, other topics are more "specific", and these topics tend to be the sole or primary focus of an article, contributing a large weight to the article. To determine whether individual topics tended to be more general or specific, the distribution of topic weights within articles was used. For each article, the topic with the highest weight was selected. We calculated the mean weight of a topic when it was selected, and the mean weight of a topic when it was not selected. These values were then plotted against each other, which allowed us to observe the generality or specificity of each topic (Westgate et al. 2015). To further investigate the effect of applying a cumulative weight threshold to article assignment to topics, we calculated topic generality and specificity across a range of cumulative weight thresholds, see Supplementary Information S1.9.

Topic prevalence

To assess how topic prevalence changed over time, we generated a matrix of the weights with which each article was assigned to each of the 50 topics. Article publication date was then used to calculate the mean topic weight per decade, to give the relative prevalence of each topic within each decade. This allowed interpretation of variations in relative topic prevalence over time. For this analysis, articles from the 1980s and 1990s were combined due to the low volume of articles in the corpus from the 1980s.

Co-occurrence of topics within articles

To assess the frequency with which any two topics co-occurred within any article in the corpus, we used the distribution of topic weights within articles. The matrix of the topic weight per article was $\log_{10}$ transformed and the Euclidean distance between each topic pairing was then calculated and scaled from zero to one. A distance of one showed that a pair of topics frequently co-occurred within the same article, while a distance of zero showed that a pair of topics rarely co-occurred (Westgate et al. 2015).

Results

The database search yielded 8411 articles in Web of Science and 9147 articles in Scopus, giving a combined total of 17,558 papers, on April 17th, 2020. The EndNote software identified 6613 duplicated articles, leaving a total of 10,945 unique papers. Ensuring inclusion criteria were met, 9882 papers were deemed suitable for inclusion. Following abstract cleaning the vocabulary of the corpus was reduced from 34,027 unique words to 7427.

The number of articles published in the field has increased dramatically over the last 35 years. The oldest articles in the corpus were published in 1985 (the search returned no articles before 1985). There were fewer than a dozen publications per year until 1998, after which the frequency of publications increased rapidly, surpassing 100 articles a year by 2004 and 500 articles a year by 2010. By 2019 the number of articles discussing INNS had doubled again, surpassing 1000 publications a year (Fig. 1).

Topics and themes

Fifty topics were identified to describe the content of the corpus (Table 1). Topics were assigned to seven broad themes, (i) Ecosystems: topics which discussed a specific region, biome, or focused on a particular species strongly associated with one ecosystem type (e.g. Topic 6: Crayfish); (ii) Monitoring: topics regarding all aspects of monitoring INNS, including detection, identification, and distributional mapping; (iii) Management and decision-making: topics discussing the management and socio-political aspects of invasion science, such as prevention, control, and policy; (iv) Interactions: topics discussing the interactions of INNS with native species, or the effects of those interactions; (v) Assessing change: topics focused on studying and analysing temporal and ecological change; (vi) Traits: topics that explored the characteristics of INNS; and (vii) Invasion

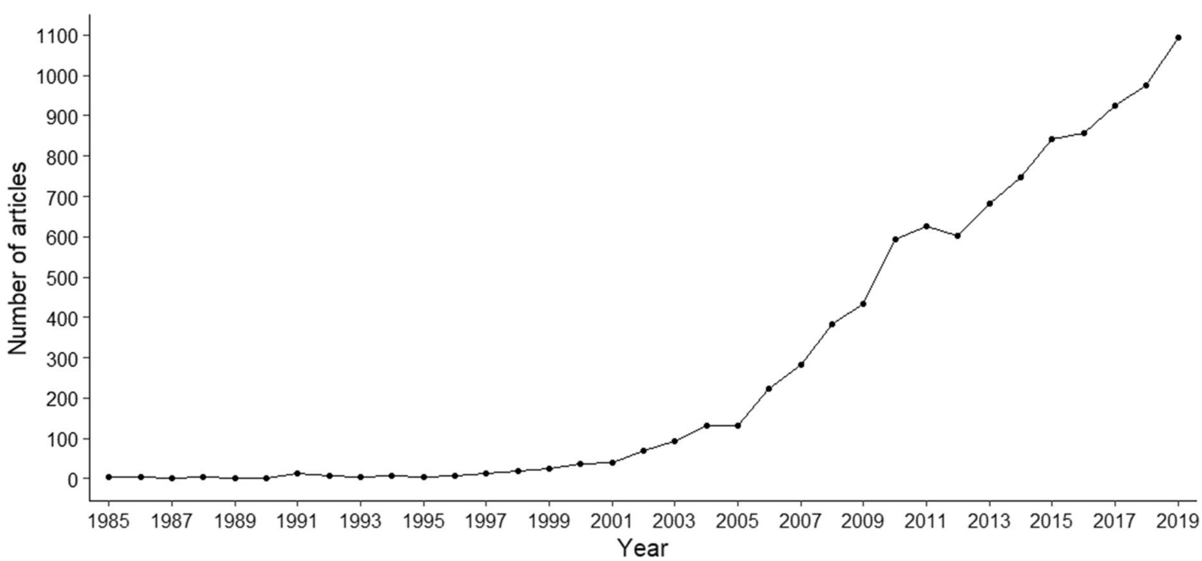

Fig. 1 The number of articles discussing invasive non-native species, published per year from 1985 (the earliest year in which articles in our corpus were published) to 2019

Table 1 Fifty topics generated by the topic model, with the five highest weighted words (reduced to the word stem in all cases) for each topic, topic name, and theme to which each topic was assigned. Each article was assigned to its highest weighted topic, generating the number of articles per topic

Topic no	Top 5 words (stemmed)	Topic name	Theme	No. articles
1	Mussel, marin, sea, coast, crab	Marine systems	Ecosystems	428
2	Fish, lake, freshwat, nativ, great	Freshwater fauna	Ecosystems	320
3	Forest, tree, stand, canopi, pine	Woodland structure	Ecosystems	310
4	Island, bird, nest, mammal, rat	Island threats and impacts	Ecosystems	307
5	River, riparian, stream, trout, flow	River systems	Ecosystems	219
6	Popul, crayfish, size, individual, red	Crayfish	Ecosystems	214
7	Cover, veget, grass, plot, grassland	Grasslands	Ecosystems	212
8	Water, aquat, system, studi, macrophyt	Aquatic plants	Ecosystems	167
9	South, africa, area, indigen, clear	South African invasions	Ecosystems	149
10	Wetland, salin, china, studi, coastal	Coastal systems	Ecosystems	145
11	State, unit, usa, north, include	Americas	Ecosystems	130
12	Popul, genet, hybrid, divers, sequenc	Population genetics	Monitoring	464
13	Record, first, report, mediterranean, new	First records	Monitoring	327
14	Model, distribut, predict, use, climat	Distribution modelling	Monitoring	320
15	Taxa, natur, number, flora, famili	Plant taxonomy	Monitoring	266
16	Use, data, detect, method, map	Detection	Monitoring	177
17	Scale, pattern, spatial, variabl, environment	Spatial patterns	Monitoring	171
18	Europ, european, north, america, region	Regional distribution	Monitoring	158
19	Site, area, along, survey, distribut	Distribution surveys	Monitoring	144
20	Area, conserv, biodivers, threat, protect	Protected areas	Monitoring	130
21	Research, manag, biolog, inform, public	Policy	Management and decision-making	421
22	Host, insect, parasit, biolog, pest	Insect biocontrol	Management and decision-making	327
23	Remov, treatment, restor, control, fire	Management methods	Management and decision-making	234
24	Manag, control, erad, strategi, effort	Management strategies	Management and decision-making	215
25	Econom, cost, use, valu, benefit	Economic impacts	Management and decision-making	192
26	Assess, risk, potenti, use, identifi	Risk Assessment	Management and decision-making	174
27	Predat, prey, food, behavior, feed	Food web interactions	Interactions	297
28	Soil, nutrient, invad, nitrogen, avail	Soil nutrients	Interactions	224
29	Pollin, flower, reproduct, plant, nativ	Pollination	Interactions	218
30	Growth, competit, biomass, experi, condit	Plant competition	Interactions	153
31	Higher, leaf, rate, litter, compar	Leaf litter	Interactions	142
32	Effect, interact, negat, affect, posit	Interactive effects	Interactions	93
33	Plant, herbivor, garden, natur, studi	Plant herbivory	Interactions	83
34	Temperatur, chang, climat, condit, degre	Climate change	Assessing change	202
35	Communiti, abund, divers, rich, composit	Community composition	Assessing change	174
36	Field, studi, found, spp, colon	Field studies	Assessing change	153
37	Impact, invad, ecosystem, ecolog, ant	Ecological impacts	Assessing change	105
38	Year, season, period, time, summer	Seasonality	Assessing change	89
39	Chang, increas, time, declin, long	Temporal change	Assessing change	86
40	Differ, studi, result, group, use	Analysis of results	Assessing change	69
41	Densiti, high, level, low, differ	Sampling	Assessing change	61
42	Nativ, suggest, compar, hypothesi, common	Comparative analysis	Assessing change	32
43	Seed, dispers, germin, fruit, bank	Seed dynamics	Traits	242
44	Trait, differ, histori, success, life	Life history	Traits	223

Table 1 (continued)

Topic no	Top 5 words (stemmed)	Topic name	Theme	No. articles
45	Adult, egg, surviv, develop, larva	Reproductive development	Traits	185
46	Activ, studi, concentr, effect, show	Chemical properties	Traits	165
47	Introduct, import, trade, spread, pathway	Pathways	Invasion mechanisms	272
48	Habitat, urban, disturb, land, area	Anthropogenic disturbance	Invasion mechanisms	127
49	Establish, success, factor, pressur, resist	Colonisation pressure	Invasion mechanisms	92
50	Rang, introduc, new, nich, expans	Non-native introductions	Invasion mechanisms	74

mechanisms: topics discussing INNS dispersal pathways and drivers of spread.

The largest theme was Ecosystems, which represented 26.3% of articles in the corpus and had the greatest number of topics (Fig. 2). The second largest theme was Monitoring which contained 21.8% of articles, followed by Management and decision-making, which contained 15.8%. The fourth largest theme, Interactions, contained 12.2% of articles across seven topics, while the theme Assessing change contained nine topics, but only 9.8% of articles. The smallest themes were each made up of four topics, with Traits containing 8.3% of articles, followed by Invasion mechanisms with 5.7%.

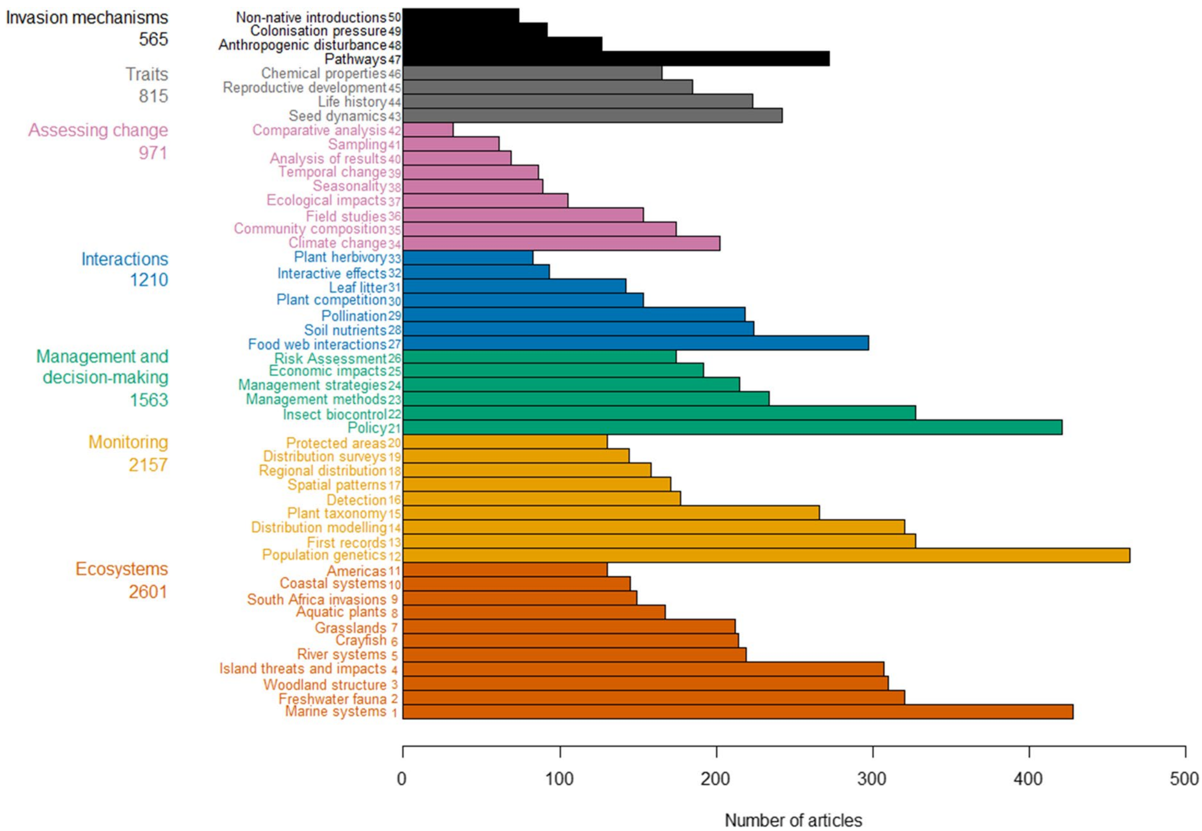

Fig. 2 The number of articles per topic. Articles were assigned to the topic with the highest weighting. The topic name and its respective topic number are given on the y-axis, alongside the theme each topic was assigned to, and the total number of articles within each theme

Some geographic regions and taxonomic groups had sufficient presence in the corpus to be represented as distinct topics. Two countries / regions had associated words weighted highly enough to be represented as a topic: *Americas* (Topic 11) and *South African invasions* (Topic 9) (Table 1). Europe and North America also featured as highly weighted words (Topic 18). Words associated with plants featured throughout the corpus and several topics were plant specific such as *Aquatic plants* (Topic 8); *Plant taxonomy* (Topic 15) and *Leaf litter* (Topic 31). For animals, the only topic specific to an organism was *Crayfish* (Topic 6). The top 20 words associated with each topic can be found in the Supplementary Information S1.5.

Topics associated with the largest number of articles, in decreasing order, were *Population genetics* (Topic 12), *Marine systems* (Topic 1), *Policy* (Topic 21), *First records* (Topic 13) and *Insect biocontrol* (Topic 22), which all came from the three largest themes (which were Ecosystems, Monitoring, and Management and decision-making). Topics associated with the smallest number of articles included *Non-native introductions* (Topic 50) and *Plant herbivory* (Topic 33).

Topic generality or specificity

The two topics with the largest number of articles, *Population genetics* and *Marine systems,* were highly specific (Topics 12 and 1 respectively, Fig. 3), meaning they were more likely to be the sole focus of an article. In contrast, the most specific topic in the corpus, *Pollination* (Topic 29), was associated with a relatively small number of articles (Fig. 2). Many topics within the theme Assessing change were more general (such as *Temporal change* and *Community composition*; Topics 39 and 35 respectively*)*. These topics were therefore broad and likely to be discussed in conjunction with other topics. Topics from the theme Invasion mechanisms were consistently general, whilst topics from the theme Traits showed high specificity, except for *Life history* (Topic 44). The theme Ecosystems predominantly contained specific topics, with *South African invasions* (Topic 9) being the most general. Within the Management and decision-making theme *Management strategies* (Topic 24) and *Insect biocontrol* (Topic 22) were the most and least general topics respectively, although *Policy* (Topic 21) and *Risk Assessment* (Topic 26) both had surprisingly high specificity.

Changes in topic prevalence over time

Within the corpus, the prevalence of topics changed across decades (Fig. 4; Supplementary Information S1.10). Many of the topics associated with the largest number of articles overall (Fig. 2) had lower prevalence in earlier decades. The relative prevalence of the topic with the largest number of articles overall, *Population genetics* (Topic 12, Fig. 2), increased dramatically between the 1980/90 s and 2010s. *Risk assessment* (Topic 26) also showed a dramatic increase in relative prevalence since the 1990s. *First records* (Topic 13) showed a similar pattern; *First records* exponentially increased in prevalence and by the 2010s ranked 2nd, rising from position 31 in the 1980s and 1990s, and position 29 in the 2000s

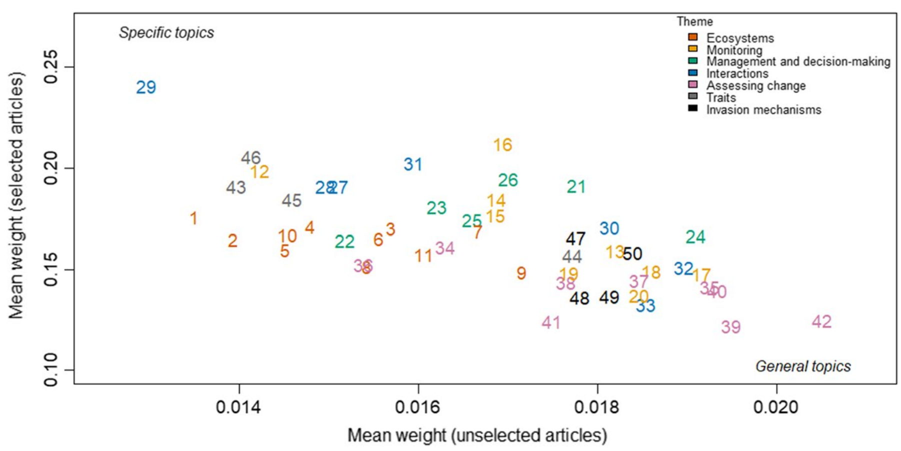

Fig. 3 Topic generality versus specificity. The topics in the bottom right corner were general (broad topics that tended to occur with other topics in an article), while topics in the top left corner were specific (more likely to be the only topic discussed in an article). Numbers represent Topics (see Table 1) and colours represent Theme

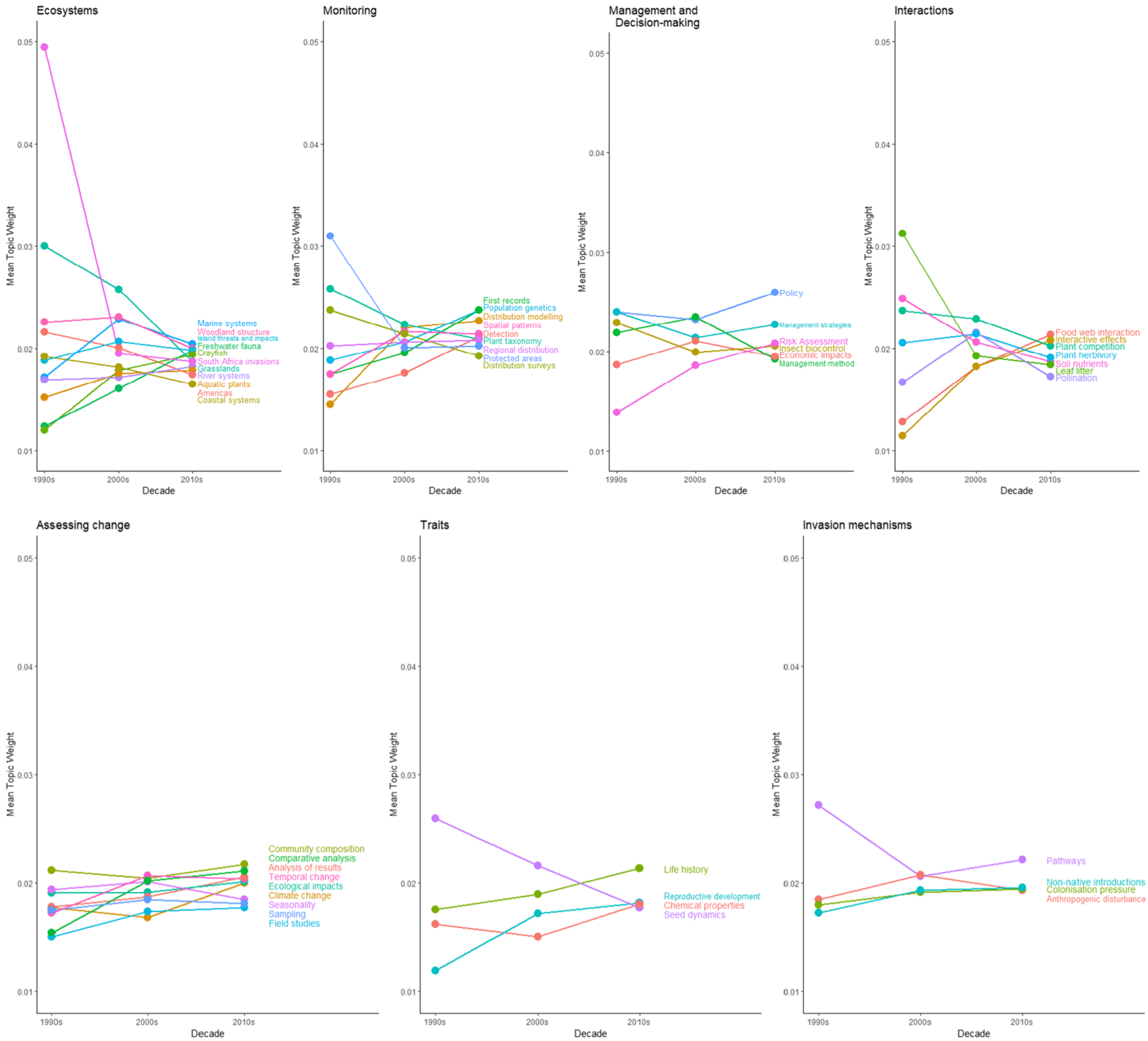

Fig. 4 Change in topic prevalence over time, calculated as mean topic weight per decade. Topics are grouped by theme

(Supplementary Information S1.10). *Policy* (Topic 21) was consistently prevalent across decades (1980s and 1990s position 9; 2000s position 3) and in the 2010s, was the most prevalent topic within articles of the corpus.

Some topics experienced both increases and decreases in relative prevalence; the prevalence of *Management methods* (Topic 23) and *Marine systems* (Topic 1) peaked in the 2000s, before dropping in the 2010s (Fig. 4). Several topics dropped in ranking in the 2000s but rose again in the 2010s, including *Island threats and impacts* (Topic 4), *Climate change* (Topic 34), *Economic impacts* (Topic 25) and *Detection* (Topic 16). *South African Invasions* (Topic 9) consistently decreased relative to other topics, from first position in the 1980s and 1990s, to 30th in the 2000s, to 36th in the 2010 (Supplementary Information S1.10). *Seasonality* (Topic 38) and *Plant herbivory* (Topic 33) also showed a decline in prevalence in the 2010s. While *Crayfish* (Topic 6) and *Freshwater fauna* (Topic 2) increased sharply in prevalence in recent decades, other freshwater related topics such as *Aquatic plants* (Topic 8) and *River systems* (Topic 5) remained consistently low ranked over time.

Co-occurrence of topics

Generally, topics were most likely to co-occur with topics from within the same theme; the most frequently co-occurring pairs of topics were *River systems* (Topic 5) and *Freshwater fauna* (Topic 2), followed closely by *River systems* (Topic 5) and *Aquatic plants* (Topic 8) (Fig. 5). The topic *Grasslands* (Topic 7) commonly occurred with *Management methods* (Topic 23), while *Leaf litter* (Topic 31) was very likely to co-occur with *Soil nutrients* (Topic 28) and *Chemical properties* (Topic 46).

The analysis identified gaps in topic co-occurrence, mostly between topics from the theme Management and decision making and other themes. *Policy* (Topic 21) had consistently low co-occurrence scores, and rarely co-occurred with any topic outside of the Management and decision-making theme. Topics such as *Marine systems* (Topic 1) had low interactions with *Economic impacts* (Topic 25) and *Management strategies* (Topic 24). Meanwhile, *Population genetics* (Topic 12) had very low co-occurrence values for *Ecological impacts, Interactive effects,* and *Management methods* (Topics 37, 32 and 23 respectively).

Beyond consideration of topic pairs, the average co-occurrence value for all pairs associated with a topic provides an overall measure of how well connected the topic is within the corpus (Fig. 6). The five topics with the highest average co-occurrence values were *Sampling* (Topic 41), *Coastal systems* (Topic 10), *Chemical properties* (Topic 46), *Field studies* (Topic 36) and *Colonisation pressure* (Topic 49). The five with the lowest average co-occurrence values were *Policy* (Topic 21), *Marine systems* (Topic 1), *Population genetics* (Topic 12), *Woodland structure* (Topic 3), and *Management strategies* (Topic 24).

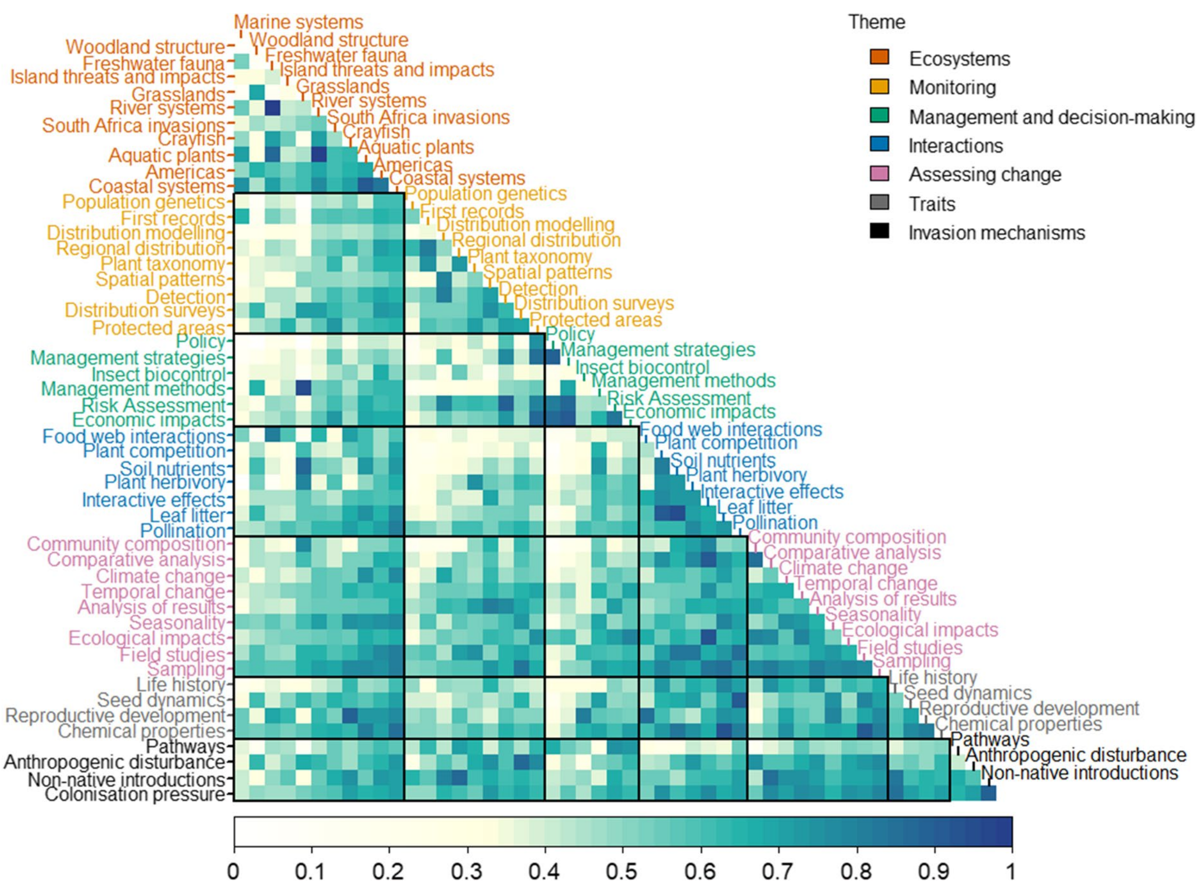

Fig. 5 Correlation matrix of topic co-occurrence within articles. Topics which co-occurred least often are scored zero, while the topics which co-occurred the most frequently scored one. The grids mark out theme groups for comparison

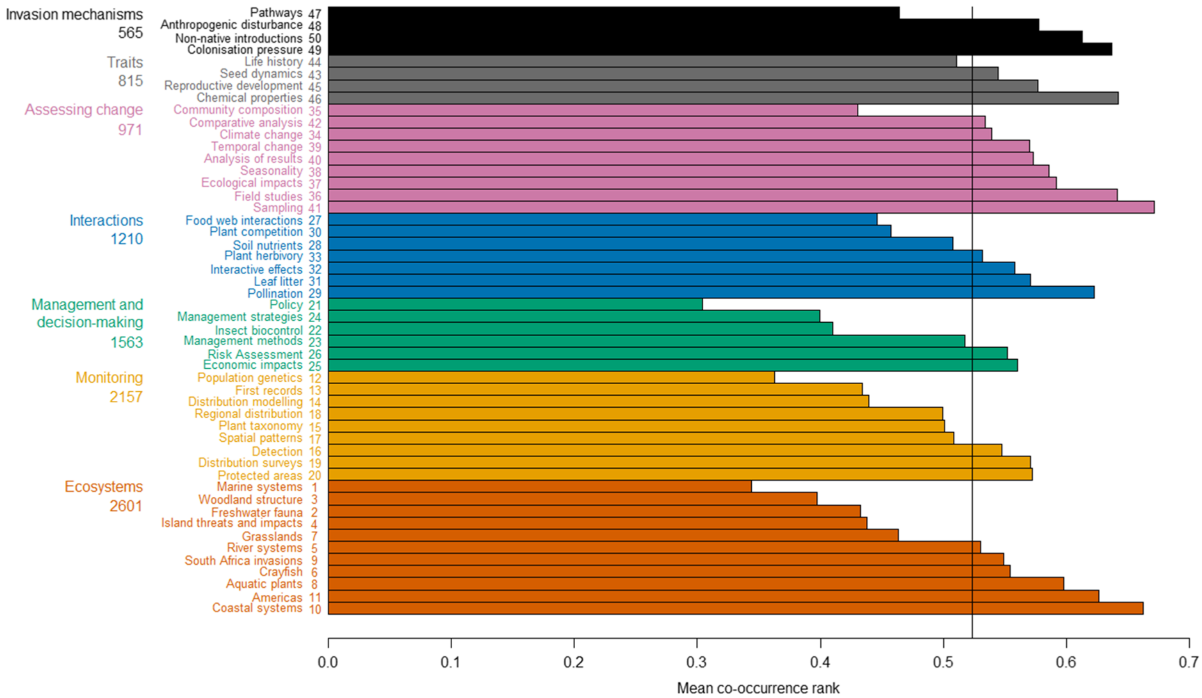

Fig. 6 Mean co-occurrence value for each topic across all articles. A lower co-occurrence rank indicates that the topic does not appear in articles with other topics as often as topics with a higher co-occurrence rank. The mean co-occurrence value across all topics is indicated by the solid vertical line. The topic name and its respective topic number are given on the y-axis, alongside the theme each topic was assigned to, and the total number of articles within each theme

Discussion

The publication landscape of invasion science is broad and covers a very wide range of topics, reflecting the multidisciplinary nature of the field. However, there was a focus on certain taxonomic groups, with several topics exclusively focused on plants, and articles on crayfish being popular enough to generate a unique topic. Topics related to freshwater systems were also represented in the literature, although the prevalence of *Aquatic plants* and *River systems* has remained consistently low over the last 35 years. The prevalence of topics changed dramatically over time, with articles on *Management methods* declining. Conversely, the advent of novel technologies has led to increases in the prevalence of articles on *Distribution modelling* and *Population genetics*, visibly impacting the publishing landscape. Assessment of topic co-occurrences revealed a distinct lack of interaction between articles discussing management or policy with other topics, showing a lack of integration of scientific research into policy and management on INNS.

Taxonomic and geographic patterns

There is an uneven global distribution of scientific publishing and readership in ecology (Nuñez et al. 2019), and we found that only two regions, *Americas* and *South African invasions,* were sufficiently represented in our corpus to emerge as distinct topics. There was wider representation within topics, with several countries mentioned by name including China, Canada, New Zealand, and Australia. However, the most dominant country was the USA, which was the only country to have a state, Florida, named within a topic (Supplementary Information Table S1.5).

While various taxa were listed in the top five highest weighted words for several topics, the only topic to represent a specific organism group was *Crayfish*. This topic exclusively focused on studies of this family, with "crayfish", "red" and "*clarkii*" appearing in the top 20 words, in reference to the red swamp crayfish, *Procambarus clarkii* (Supplementary Information Table S1.5). Analysis showed that the topic has grown in prevalence over the duration of the corpus, with the total number of articles in which it was the highest weighted topic being only fractionally smaller than the much broader topic *River systems.* Crayfish may provide an ideal model for studying many aspects of invasion science (Twardochleb et al. 2013; Manfrin et al. 2019), but a large focus on one taxon might also increase the risk that other potentially more damaging species or impacts are overlooked.

Taxonomic bias is a recognised issue in the literature, with plants commonly being the most studied taxa in invasion science (Pyšek et al. 2008), confirmed by our results given the number of topics focused exclusively on plants. This bias can lead to over- and under-representation of biota within the literature that prevents novel synthesis of information in invasion science (Matzek et al. 2015). The species most popular in invasion science have changed over time (Richardson and Pyšek 2008), and it is likely that other species will become more prevalent in the literature as the drivers of invasion by non-native species change (Seebens et al. 2018). To avoid lack of attention to novel/unrecognised threats, researchers need to be more aware of understudied species which can be determined through conversations with stakeholders and horizon scanning (Ricciardi et al 2017). Future research could quantify if the popularity of species or groups of organisms in the literature is reflective of the relative threat they pose.

Emerging trends in the literature

Overall, *Population genetics* and *First records* increased fastest in relative prevalence. The prevalence of articles discussing *First records*, as well as *Pathways* and *Distribution modelling*, is likely a result of the emphasis on pathways, spread, risk assessment and management in the Convention on Biological Diversity's Aichi Biodiversity Targets (part of the CBD's Strategic Plan for Biodiversity 2011–2020; CBD 2010) and the increasing availability of dedicated journals, and large collated databases based on first records (Seebens et al. 2018), in some cases boosted by growing enthusiasm for community science and recording platforms. This trend in prevalence is likely to continue as awareness of INNS and the rate of invasions increase, and new invaders and pathways emerge (Ricciardi et al. 2017).

Population genetics had the largest numbers of articles in the corpus and appears set to increase. The growth in prevalence of genetics is likely a result of novel technology becoming more widely accessible. Developments in environmental DNA sampling and sequencing allow for the identification or improved detection of INNS, which enables monitoring in systems that were previously difficult to survey (Thomas et al. 2020). A more controversial development, that may support the resurgence in prevalence of the topic *Insect biocontrol,* is the advent of genome modification which could have widespread applications for management of INNS (Thresher et al. 2014; Champer et al. 2016). Discussion of genetic research and biocontrol is predicted to increase in the literature as these technologies develop (Ricciardi et al. 2017), yet there is currently a lack of interaction between *Population genetics* and the wider literature. Greater interdisciplinary research is required to achieve uptake of these technologies and realise their potential benefits for INNS monitoring and management efforts.

Management methods was one of the most prevalent topics in the 2000s yet declined in relative prevalence over time. Prevalence in the 2000s may be associated with efforts to achieve the CBD's 2010 Biodiversity Target but, as the 2010 target was superseded by the Aichi Biodiversity Targets, this does not explain the subsequent drop in prevalence (UNEP 2002; CBD 2010). Decline in prevalence may be because managing INNS is perceived as a "wicked problem" (Woodford et al. 2016), although improved methods to prioritise management actions based on cost and feasibility can improve effectiveness (Booy et al. 2017). Given ambitious global targets to halt the loss of biodiversity caused by INNS, there is a clear need for more evidence designed to support effective management and problem solving (IUCN 2019).

The policy-management–science interface

Policy has been a consistently popular topic across decades, and the most prevalent topic of the 2010s.

However, we found that *Policy* was a highly specific topic, therefore most likely being the only focus of an article. *Policy* also consistently showed low co-occurrence with other topics, particularly topics within other themes, such as Ecosystems. The lack of co-occurrence between policy and Ecosystems topics was particularly noticeable for *Marine systems*, despite the results showing that *Marine systems* were discussed in many articles. This may be due to the greater challenges of identifying, monitoring and managing INNS in the marine environment (Giakoumi et al. 2019).

The overall co-occurrence between management topics and the wider literature was similarly low. While there was a topic dedicated to the economic impacts of INNS, there was no evidence that the terms relating to the cost of management had high weight in either *Management strategy* or *Management methods*. In the same vein, there was no evidence of 'success' being a highly weighted term in any topic of the management theme. This may be because the number of successful control or eradication campaigns is relatively small, especially outside of island ecosystems (Gardener et al. 2010; Glen et al. 2013; Robertson et al. 2017). It is also difficult to guarantee whether any eradication scheme has been 100% effective, hence the cautionary use of 'success'. This leads to ongoing management, which in turn makes calculation of costs challenging (Rout et al. 2009). These issues are recognised in the literature, leading to calls for clearer definitions of management objectives and success criteria, and establishment of cost estimation frameworks (Jardine and Sanchirico 2018; Robertson et al 2020). Considering the limited budgets many management schemes operate under, facilitating widespread discussion of management outcomes and the costs of management is crucial for success (Larson et al. 2011).

Invasion science casts light on a range of disciplines, ranging from evolutionary theory, island biogeography to management. However, management is currently the focus of international targets to reduce the impacts of INNS and is worthy of separate consideration. Management-based research in invasion science has been found to receive fewer citations than more fundamental research (Pyšek et al. 2006; Esler et al. 2010). This result reinforces previous studies that found research outputs in this and other disciplines are often poorly translated into policy or management action (Esler et al. 2010; McGeoch et al. 2010; Matzek et al. 2015). Scientific articles are often not the primary mechanism for disseminating outcomes of management and policy actions, and often these documents can be difficult to access, both for practitioners and researchers (Catalano et al. 2019; Kadykalo et al. 2021). While there is a need for both broad and focused research on INNS, improved communication and collaboration between disciplines, and between researchers and stakeholders, would be mutually beneficial and will increase the prospect of meeting future global targets for INNS management.

Knowledge gaps in invasion research

When considering the topics generated and the words that represented them, there were some notable absences. The social dimensions of INNS were not well represented within the corpus. Terms such as socio-economic, livelihood and well-being for example, were not present amongst highly weighted words. The role and impact of INNS in these and related contexts is a complex and important topic that requires further research (Shackleton et al. 2019). Alongside this is a need to better understand human perceptions towards INNS to better inform communication, management and policy (Shackleton et al. 2018). Terms referring to invasions in the Arctic were also not present amongst highly weighted words. While this region has been viewed as having few currently established INNS and a low invasion risk, loss of sea ice and increased shipping activities are expected to drive an increase in invasions (Chan et al. 2019). Recent horizon scanning has suggested this issue will soon feature widely in the literature (Ricciardi et al. 2017), but our findings suggest that a greater scale of invasions may be required first, before the scientific community responds.

While the application of topic modelling provides a quantitative approach to literature analysis, there was still bias in this work. For example, using abstracts written in English will cause an uneven geographical representation. Within the corpus, the abstracts contained a high number of synonyms for INNS, such as "alien", "exotic", "introduced" and "non-indigenous". While the aim was to encompass the majority of relevant literature with the search terms applied, papers solely using these alternative terms are likely to have been excluded from the corpus which might have contributed to the prevalence of particular disciplines and uneven geographical spread linked to usage of these terms. However, we found that using expanded search terms accounting for synonyms yielded a higher proportion of irrelevant articles, including from e.g., the medical sciences, and that the same topics and key trends were returned. We conclude that our narrower search string captures the diversity of the scientific literature, while excluding a greater proportion of irrelevant articles. Based on the results of this study, future investigations could be conducted using a similar methodology on a subsection of invasion science, such as Marine systems or Management methods. As these were such broad topics it would be beneficial to determine the more specific topics and trends within these areas.

Conclusion

This study provides an insight into the publishing landscape of invasion science by quantitatively assessing the research trends within the literature. The topics addressed varied widely, reflecting the multidisciplinary nature of the field, and new technological developments visibly impacted the publishing landscape. However, our results revealed a prevalence of specific taxonomic groups and geographical regions within the literature, and a paucity of interdisciplinary collaboration in the critical areas of policy and management. The risk is that research outcomes have low relevance or perceived transferability for practical management and policy, or that relevant research fails to reach the right audiences, and that some stubborn problems will persist for lack of more creative or joined-up thinking. To achieve targets for INNS management there will need to be improvements in how research is targeted, communicated and used. While emerging topics such as genetic identification and control offer new capabilities, we found a general lack of cross-sectoral interdisciplinarity, and we suggest that addressing this gap, particularly in relation to policy and management, would strengthen and steer the policy-management-science interface.

Acknowledgements We would like to thank MJ Westgate for supplying code and advice to EH to enable the creation of Supplementary S1.10. We would also like to thank B Rowland for R coding assistance enabling the generation of the cumulative topic weight threshold Figures of S1.8. We thank the reviewers for their valuable comments.

Author contributions PR, ZP and ES contributed to the study conception and design. Data collection was performed by ES and KW. Data analyses were conducted by ES and analytical support was provided by EH and LM. All authors workshopped the topic model outcomes to determine topic names and themes. The first draft of the manuscript was written by ES, all drafts were edited by ZP. All authors commented on versions of the manuscript. All authors read and approved the final manuscript.

Funding No funding or grants were received to conduct this study.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

Blackburn TM, Pyšek P, Bacher S et al (2011) A proposed unified framework for biological invasions. Trends Ecol Evol 26:333–339. https://doi.org/10.1016/j.tree.2011.03.023

Blei DM, Lafferty JD (2009) Topic models. In: Srivastava AN, Sahami M (eds) Text mining classification, clustering, and applications. CRC Press, pp 71–89

Booy O, Mill AC, Roy HE et al (2017) Risk management to prioritise the eradication of new and emerging invasive non-native species. Biol Invasions 19(8):2401–2417. https://doi.org/10.1007/s10530-017-1451-z

Catalano AS, Lyons-White J, Mills MM, Knight AT (2019) Learning from published project failures in conservation. Biol Conserv 238:108223. https://doi.org/10.1016/j.biocon.2019.108223

Champer J, Buchman A, Akbari OS (2016) Cheating evolution: engineering gene drives to manipulate the fate of wild populations. Nat Rev 17:146–159. https://doi.org/10.1038/nrg.2015.34

Chan FT, Stanislawczyk K, Sneekes AC et al (2019) Climate change opens new frontiers for marine species in the

Arctic: current trends and future invasion risks. Glob Change Biol 25:25–38. https://doi.org/10.1111/gcb.14469
Enders M, Havemann F, Ruland F et al (2020) A conceptual map of invasion biology: integrating hypotheses into a consensus network. Glob Ecol Biogeogr. https://doi.org/10.1111/geb.13082
Esler KJ, Prozesky H, Sharma GP, McGeoch M (2010) How wide is the "knowing-doing" gap in invasion biology? Biol Invasions 12:4065–4075. https://doi.org/10.1007/s10530-010-9812-x
Essl F, Lenzner B, Bacher S et al (2020) Drivers of future alien species impacts: an expert-based assessment. Glob Change Biol 26:4880–4893. https://doi.org/10.1111/gcb.15199
Gardener MR, Atkinson R, Renteria JL (2010) Eradications and people: lessons from the plant eradication program in Galapagos. Restor Ecol 18:20–29. https://doi.org/10.1111/j.1526-100X.2009.00614.x
Giakoumi S, Katsanevakis S, Albano PG et al (2019) Management priorities for marine invasive species. Sci Total Environ 688:976–982. https://doi.org/10.1016/j.scitotenv.2019.06.282
Glen AS, Atkinson R, Campbell KJ et al (2013) Eradicating multiple invasive species on inhabited islands: the next big step in island restoration? Biol Invasions 15:2589–2603. https://doi.org/10.1007/s10530-013-0495-y
Golebie EJ, van Riper CJ, Arlinghaus R et al (2022) Words matter: a systematic review of communication in non-native aquatic species literature. NeoBiota 74:1–28. https://doi.org/10.3897/neobiota.74.79942
Griffiths TL, Steyvers M (2004) Finding scientific topics. PNAS 101:5228–5235. https://doi.org/10.1073/pnas.0307752101
Grün B, Hornik K (2011) Topicmodels: an R package for fitting topic models. J Stat Softw 40:1–30
Jardine SL, Sanchirico JN (2018) Estimating the cost of invasive species control. J Environ Econ Manag 87:242–257. https://doi.org/10.1016/j.jeem.2017.07.004
Kadykalo AN, Buxton RT, Morrison P et al (2021) Bridging research and practice in conservation. Conserv Biol 35(6):1725–1737. https://doi.org/10.1111/cobi.13732
Larson DL, Phillips-Mao L, Quiram G et al (2011) A framework for sustainable invasive species management: environmental, social, and economic objectives. J Environ Manag 92:14–22. https://doi.org/10.1016/j.jenvman.2010.08.025
Mair L, Mill AC, Robertson PA et al (2018) The contribution of scientific research to conservation planning. Biol Conserv 223:82–96. https://doi.org/10.1016/j.biocon.2018.04.037
Manfrin C, Souty-Grosset C, Anastácio PM et al (2019) Detection and control of invasive freshwater crayfish: from traditional to innovative methods. Diversity 11:5. https://doi.org/10.3390/d11010005
Matzek V, Pujalet M, Cresci S (2015) What managers want from invasive species research versus what they get. Conserv Lett 8:33–40. https://doi.org/10.1111/conl.12119
McGeoch MA, Butchart SHM, Spear D et al (2010) Global indicators of biological invasion: species numbers, biodiversity impact and policy responses. Divers Distrib 16:95–108. https://doi.org/10.1111/j.1472-4642.2009.00633.x
Meyer D, Hornik K, Feinerer I (2008) Text mining infrastructure in R. J Stat Softw 25:1–54. https://doi.org/10.18637/jss.v025.i05
Nuñez MA, Barlow J, Cadotte M et al (2019) Assessing the uneven global distribution of readership, submissions and publications in applied ecology: obvious problems without obvious solutions. J Appl Ecol 56:4–9. https://doi.org/10.1111/1365-2664.13319
Pyšek P, Richardson DM, Rejmánek M et al (2004) Alien plants in checklists and floras: towards better communication between taxonomists and ecologists. Taxon 53:131–143. https://doi.org/10.2307/4135498
Pyšek P, Richardson DM, Jarošík V (2006) Who cites who in the invasion zoo: insights from an analysis of the most highly cited papers in invasion ecology. Preslia 78:437–468
Pyšek P, Richardson DM, Pergl J et al (2008) Geographical and taxonomic biases in invasion ecology. Trends Ecol Evol 23:237–244. https://doi.org/10.1016/j.tree.2008.02.002
Ricciardi A, Hoopes MF, Marchetti MP, Lockwood JL (2013) Progress toward understanding the ecological impacts of nonnative species. Ecol Monogr 83:263–282. https://doi.org/10.1890/13-0183.1
Ricciardi A, Blackburn TM, Carlton JT et al (2017) Invasion science: a horizon scan of emerging challenges and opportunities. Trends Ecol Evol 32:464–474. https://doi.org/10.1016/j.tree.2017.03.007
Richardson DM, Pyšek P (2008) Fifty years of invasion ecology—the legacy of Charles Elton. Divers Distrib 14:161–168. https://doi.org/10.1111/j.1472-4642.2007.00464.x
Robertson PA, Adriaens T, Lambin X et al (2017) The large-scale removal of mammalian invasive alien species in Northern Europe. Pest Manag Sci 73:273–279. https://doi.org/10.1002/ps.4224
Robertson PA, Mill A, Novoa A et al (2020) A proposed unified framework to describe the management of biological invasions. Biol Invasions 22(9):2633–2645. https://doi.org/10.1007/s10530-020-02298-2
Roda A, Kairo M, Damian T et al (2011) Red palm weevil (Rhynchophorus ferrugineus), an invasive pest recently found in the Caribbean that threatens the region. EPPO Bull 41:116–121. https://doi.org/10.1111/j.1365-2338.2011.02446.x
Rout TM, Thompson CJ, McCarthy MA (2009) Robust decisions for declaring eradication of invasive species. J Appl Ecol 46:782–786. https://doi.org/10.1111/j.1365-2664.2009.01678.x
Seebens H, Blackburn TM, Dyer EE et al (2018) Global rise in emerging alien species results from increased accessibility of new source pools. Proc Natl Acad Sci USA 115:E2264–E2273. https://doi.org/10.1073/pnas.1719429115
Shackleton RT, Richardson DM, Shackleton CM et al (2018) Explaining people's perceptions of invasive alien species: a conceptual framework. J Environ Manag 229:10–26. https://doi.org/10.1016/j.jenvman.2018.04.045
Shackleton RT, Shackleton CM, Kull CA (2019) The role of invasive alien species in shaping local livelihoods and human well-being: a review. J Environ Manag

229:145–157. https://doi.org/10.1016/j.jenvman.2018.05.007
Thomas AC, Tank S, Nguyen PL et al (2020) A system for rapid eDNA detection of aquatic invasive species. Environ DNA 2:261–270. https://doi.org/10.1002/edn3.25
Thresher RE, Hayes K, Bax NJ et al (2014) Genetic control of invasive fish: technological options and its role in integrated pest management. Biol Invasions 16:1201–1216. https://doi.org/10.1007/s10530-013-0477-0
Twardochleb LA, Olden JD, Larson ER (2013) A global meta-analysis of the ecological impacts of nonnative crayfish. Freshw Sci 32:1367–1382. https://doi.org/10.1899/12-203.1
Westgate MJ, Barton PS, Pierson JC, Lindenmayer DB (2015) Text analysis tools for identification of emerging topics and research gaps in conservation science. Conserv Biol 29:1606–1614. https://doi.org/10.1111/cobi.12605
Woodford DJ, Richardson DM, Macisaac HJ et al (2016) Confronting the wicked problem of managing biological invasions. NeoBiota 31:63–86. https://doi.org/10.3897/neobiota.31.10038
CBD (2010) Strategic Plan for Biodiversity 2011–2020, including Aichi Biodiversity Targets. https://www.cbd.int/sp/
CBD (2002) Decision VI/23* of the Conference of the Parties to the CBD, Annex, footnote to the Introduction. https://www.cbd.int/invasive/terms.shtml
CBD (2020) Global Biodiversity Outlook 5. Montreal. https://www.cbd.int/GBO5
IUCN (2019) Post-2020 target on invasive alien species. https://www.iucn.org/files/2019-post-2020-target-invasive-alien-species6112019pdf
R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/
United Nations Environmental Programme (2002) COP 6 Decision VI/26. Strategic Plan for the Convention on Biological Diversity. http://www.cbd.int/decisions/?dec=VI/26