Received: 30 October 2017 | Accepted: 19 February 2018
DOI: 10.1111/faf.12280

ORIGINAL ARTICLE

WILEY FISH and FISHERIES

Narrow lenses for capturing the complexity of fisheries: A topic analysis of fisheries science from 1990 to 2016

Shaheen Syed[1,2] | Melania Borit[3] | Marco Spruit[1]

[1]Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

[2]Centre for Policy Modelling, Manchester Metropolitan University, Manchester, UK

[3]Norwegian College of Fishery Science, University of Tromsø – The Arctic University of Norway, Tromsø, Norway

Correspondence
Melania Borit, Norwegian College of Fishery Science, University of Tromsø – The Arctic University of Norway, Tromsø, Norway.
Email: melania.borit@uit.no

Funding information
EU-H2020-SAF21—Social Science Aspects of Fisheries for the 21st Century

Abstract

Despite increased fisheries science output and publication outlets, the global crisis in fisheries management is as present as ever. Since a narrow research focus may be a contributing factor to this failure, this study uncovers topics in fisheries research and their trends over time. This interdisciplinary research evaluates whether science is diversifying fisheries research topics in an attempt to capture the complexity of the fisheries system, or whether it is multiplying research on similar topics, attempting to achieve an in-depth, but possibly marginal, understanding of a few selected components of this system. By utilizing latent Dirichlet allocation as a generative probabilistic topic model, we analyse a unique dataset consisting of 46,582 full-text articles published in the last 26 years in 21 specialized scientific fisheries journals. Among the 25 topics uncovered by the model, only one (Fisheries management) refers to the human dimension of fisheries understood as socio-ecological complex adaptive systems. The most prevalent topics in our dataset directly relating to fisheries refer to Fisheries management, Stock assessment, and Fishing gear, with Fisheries management attracting the most interest. We propose directions for future research focus that most likely could contribute to providing useful advice for successful management of fisheries.

KEYWORDS
fisheries publications, fisheries research focus, fisheries research gaps, fisheries research topics, latent Dirichlet allocation, topic modelling

1 | INTRODUCTION

Following a similar trend to scientific research at large, fisheries science research output has significantly increased in the last three decades (Aksnes & Browman, 2016), in parallel with an increase in the number of fisheries scientific journals (Mather, Parrish, & Dettmers, 2008). This rapid expansion of the field is attributed to the growing concern about the state of global fish stocks and to the major role that science has been playing in fisheries management (Jarić, Cvijanović, Knežević-Jarić, & Lenhardt, 2012). However, despite the increased volume of fisheries science output and publication outlets, the global crisis in marine capture fisheries management is as present as ever, with unforeseen consequences ranging from fisheries-induced evolutionary changes among wild fish populations (Belgrano & Fowler, 2013) to conflicts between states over the implementation of best available science (Brooks et al., 2016). There are various hypotheses regarding causes and contributing factors for failures of fisheries management, including data uncertainty, model inadequacy, ecosystem structure, institutional efficacy, economic discord or research focus (Smith & Link, 2005). Among these, research focus is the least explored (Smith & Link, 2005). Using hybrid content analysis of a unique dataset consisting of 46,582 fisheries science full-text articles published in the last 26 years, we uncover focus topics in fisheries research and their trends.

Fisheries are a socio-ecological complex adaptive system (SECAS) in which macroscopic properties emerge from local actions that spread to higher scales due to agents' (fish and humans) collective behaviour; these properties then feed back, in a nonlinear way, influencing individuals' options and behaviours, but they typically only do so diffusely and over long timescales (Levin et al., 2013; Ostrom, 2009). A fishery can be defined as "the complex of people, their institutions, their harvest and their observations associated with and including a targeted stock or group of stocks (i.e. usually fish), and increasingly, the associated ecosystems that produce said stocks" (Link, 2010). Deconstructing the concept, the two main dimensions of a fishery are the human dimension (i.e. human agents, communities of these and their institutions) and the natural dimension (i.e. biotic, such as predator species and prey species, and abiotic, such as water temperature and nutrients) (Charles, 2001; Lennox et al., 2017; Österblom et al., 2013). The purpose of this study was to assess whether fisheries science output is reflecting this conceptual diversity of fisheries as SECAS, and if so, to what extent. Is science diversifying fisheries research topics in an attempt to capture the complexity of the fisheries system, or is it multiplying research on similar topics, trying to achieve an in-depth, but possibly marginal, understanding of a few selected components of this system? Based on the critical reflection that "the majority of fisheries scientists have a biologically oriented background, they can be a bit naïve regarding other factors when it comes to the prominence of economic or social considerations" (Link, 2010), the working hypothesis of this study is that the human dimension of fisheries might be underrepresented in the fisheries specialty literature.

The assessments of the development and trends in fisheries science have so far been mostly based on reviews (e.g. Johnson et al., 2013) or bibliometric evaluations (e.g. Aksnes & Browman, 2016) of scientific publications in the field. Limitations of these studies include the following: taking into account only a limited number of publications (e.g. Jarić et al., 2012), or a limited time period (e.g. 2000–09; Jarić et al., 2012); having a limited scope [e.g. artisanal coral reef fisheries research (Johnson et al., 2013); fish stock assessment research (Kumaresan, Ezhilrani, Vinitha, Sivaraman, & Jayaraman, 2014); shark by-catch research (Molina & Cooke, 2012)]; using proxies for full-text articles [e.g. titles (Jarić et al., 2012); abstracts (Aksnes & Browman, 2016)], or proxies for topics of research [e.g. one word per topic (Aksnes & Browman, 2016; Jarić et al., 2012)]. Most importantly, all these previous attempts to map the fisheries science field are top-down approaches, with topics of interest manually predefined by the analysts [e.g. species, region, habitat, study object (Jarić et al., 2012)], and the analysed data manually assigned to these topics. However, such approaches are limited due to the subjectivity inherent in human decisions, and the analysis of the same research field could yield opposite results (e.g. Rose, Janiger, Parsons, and Stachowitsch (2011) vs. Hill and Lackups (2010) evaluation of the field of cetacean research).

In contrast to previous approaches, we follow a completely novel strategy for the field of fisheries science, a bottom-up approach (Debortoli, Müller, Junglas, & vom Brocke, 2016) by utilizing topic modelling to uncover hidden research topics within fisheries science publications. Topic modelling algorithms are machine-learning methods to automatically uncover hidden or latent thematic structures from large collections of documents. Topic models can produce a set of interpretable topics that can be viewed as groups of co-occurring words that are associated with a single topic or theme (DiMaggio, Nag, & Blei, 2013). Such groups of co-occurring words (i.e. topics) are words that tend to come up together within the same linguistic context more frequently than one would expect by chance alone. These co-occurring words tend to purport similar meaning and refer to a similar subject. For example, in the context of fisheries science, an author might write a text to which she/he gave the keywords "community structure," "subtropical areas," "reference points" and "weight." This text might use more frequently the words "parameters," "estimation," "stock," "modeling," "male," "female," "sex" and "spawning." If we wanted to use topic modelling to uncover the latent topics of this hypothetical text, based on how often these most used words would appear together (i.e. co-occur), the automated topic model would group the first four words in one topic and the last four words in a different topic. These two topics would then be manually labelled by a domain expert most likely as Stock assessment modeling and Reproduction, respectively. Note that the subject of these two topics is not similar to the one that might be inferred from the keywords given to this hypothetical text. Thus, these topics are latent, and they are hidden in the pattern of co-occurring words. In essence, topic models are able to exploit the co-occurrence structure of texts and produce the topics as lists of words that frequently come up together, within and between documents; technically, such lists of words are probability distributions over words.

The topics emerge from the statistical properties of the documents and, thus, overcome the need for manual annotation of the collection of texts, although manual interpretation of the *subject* of a topic might still be needed, as it is yet considered the gold standard in the domain of topic modelling (Lau, Grieser, Newman, & Baldwin, 2011). As such, we allow the documents to speak for themselves and view the documents through the computational lens of the topic model, rather than relying on the manifest or reported content by their authors. Document collections that are too large to explore manually can now be analysed to study phenomena of the sort that can only be viewed through the macroscopic computational lens of the topic model (Mohr & Bogdanov, 2013). Topic modelling approaches have been helpful in elucidating the key ideas within a set of documents, such as articles published in the journal *PNAS* (Griffiths & Steyvers, 2004), political science texts (Grimmer & Stewart, 2013) or data-driven journalism (Rusch, Hofmarcher, Hatzinger, & Hornik, 2013). Moreover, it is considered that this approach could provide insight into the development of a scientific field and changes in research priorities (Neff & Corley, 2009) and could do so with greater speed and quantitative rigour than would otherwise be possible through traditional narrative reviews (Grimmer & Stewart, 2013). As such, this topic modelling method has been applied, for example, in the domain of transportation research (Sun & Yin, 2017), computer science (Hall, Jurafsky, & Manning, 2008; Wang & McCallum, 2006; Wang, Paisley, & Blei, 2011), fisheries modelling (Syed & Weber, 2018), conservation science (Westgate, Barton, Pierson, & Lindenmayer, 2015) and the fields of operations research and management science (Gatti, Brooks, & Nurre, 2015).

After identifying the hidden topics of fisheries science, we analyse the extent to which these topics cover the complexity of the fisheries domain. Afterwards, we examine topic similarity, topic co-occurrence, topic prevalence and topical trends over the last 26 years. We furthermore identify patterns in increasing and decreasing topic trends over specific periods of time (i.e. hot and cold topics in 1990–95, 1995–2000, 2000–05, 2005–10 and 2010–16) and describe the distribution of uncovered topics over journals.

2 | METHODS

2.1 | Latent Dirichlet allocation

This paper utilizes the topic model latent Dirichlet allocation (LDA) (Blei, 2012; Blei, Ng, & Jordan, 2003). LDA is a Bayesian probabilistic topic model and follows the assumption that documents exhibit multiple topics in mixing proportions, thus capturing the heterogeneity of, for example, research topics within scientific publications. In statistics, this is often referred to as a mixed-membership model (Erosheva, Fienberg, & Lafferty, 2004). Technically, a topic is a multinomial distribution of words in the vocabulary, where each word has a different probability within each topic; within a topic, more prominent words have a higher probability and groups of high probability words can be considered as co-occurring clusters or constellation of words that describe a certain underlying topic or theme. A document might be 60% about the topic Fisheries management and 40% about the topic Stock assessment. A topic "about" a subject (e.g. *fisheries management*) relates to the probability distribution of words that places high probability to words that would be used to describe the subject (DiMaggio et al., 2013). Note that the underlying topics and to what extent the document exhibits these topics are not known in advance. These details are the output of the LDA analysis and emerge automatically from the statistical properties of the documents and the assumptions behind LDA.

One way to think about LDA is to imagine a document in which one highlights words with coloured markers. Words that relate to one topic are coloured blue, words that relate to another topic are coloured red and so on. After all of the words have been coloured (excluding words such as "the," "a"), all the words with the same colour are the topics, and the article will blend the colours in different proportions. Different documents will have different blends of colours, and we could use the proportion of the various colours to situate this specific document in a document collection (e.g. documents addressing mainly the blue topic). Moreover, documents with the same blend of colours discuss the topics in similar proportion and are considered closely related from a topical perspective. Technically, documents with similar topic distributions are close in Kullback–Leibler divergence, a measure to calculate the distance between two probability distributions. LDA as a statistical model captures this intuition. We refer the interested reader to Blei (2012) for a concise introduction to LDA.

Latent Dirichlet allocation is best described by its generative process, that is the imaginary probabilistic recipe that generates the documents as well as the hidden structure. The hidden structure is the topics, modelled as distributions of words, and the topic proportions per document, where each document has some probability for each latent topic (i.e. mixing topic proportions). More formally, the generative process also assigns each word to a topic as to allow for documents to exhibit multiple topics, analogous to the coloured words example. More information on the generative process can be found in the supplementary material. Given the observed documents, the aim now is to infer the hidden structure to answer the question "what is the likely hidden topical structure that has generated these documents?", a process that can be seen as reverse-engineering the generative process. Technically, we want to infer the posterior distribution of the latent variables given the observed documents. An analogy to this process is described by the local farmers' market example; one might estimate what vegetables and what quantities are being sold at the local farmers' market by post hoc inspection of people's shopping basket.

Seeing more baskets refines the estimation of the products and their quantities and provides an estimate of the market's produce (Rhody, 2012). Mainly two types of inference techniques can be discerned: sampling-based algorithms (e.g. Newman, Asuncion, Smyth, & Welling, 2007; Porteous et al., 2008) and variational-based algorithms (e.g. Blei & Jordan, 2006; Teh, Newman, Welling, & Neaman, 2006; Wang et al., 2011). To simplify posterior inference, LDA uses a Dirichlet distribution as a conjugate prior for the multinomial distribution, hence the name LDA. The posterior distribution will reveal the probability distributions of words for each topic and the topic proportions per document. Note that the obtained structure is latent, and therefore, the probability distributions of words are not semantically labelled. However, when sorted, the words with the highest probability within a topic will relate to what one would call a topic or theme (DiMaggio et al., 2013; Mohr & Bogdanov, 2013). In this context, it is important to mention that the LDA model does not give a name to each identified latent topic (i.e. the model does not label the topics). The output of the model groups the co-occurring words under numbered topics (i.e. topic 1, topic 2, topic 3). Albeit a subjective endeavour, possibly affecting the statistical objectiveness of the LDA method, in order to increase the readability and interpretability of these topics, a human analyst can be used to interpret what is the common subject of the words within each topic and consequently give a name (i.e. a post hoc label) to this topic (DiMaggio et al., 2013). Research into automatic assignment of topic labels exists; however, manual annotation by a human expert is still considered the gold standard in labelling topics (Lau et al., 2011). For our study, instead of using the topic numbering provided by the LDA model (i.e. topic 1, topic 2, topic 3), in order to increase readability of the text and interpretability of results, we chose to give a specific label to each topic using the gold standard in this domain, that is manual annotation (see the section Labelling topics).

2.2 | Assumptions behind LDA

Latent Dirichlet allocation is a bag-of-words model in which documents are represented as unordered sequences of words. Such an assumption neglects word order and possibly important cues to the content of a document (Steyvers & Griffiths, 2007). Although an unrealistic assumption, it is reasonable when uncovering semantic structures of text (Blei, 2012; Blei & Lafferty, 2006). Consider a thought experiment where the words of a document are shuffled. After finding a high number of words like *spawning*, *eggs* and *growth*, one can imagine that the document deals with some aspects of reproduction. LDA further assumes document exchangeability, that is, the order in which documents are analysed is unimportant, yet all documents are analysed at the end of the LDA analysis. Consequently, LDA is unable to explicitly capture evolving topics from documents that cover large time spans (e.g. centuries). To do that, we would need to resort to a more complicated and computationally expensive dynamic topic model (Blei & Lafferty, 2006). Such an approach is currently not feasible given the large dataset used here, but would be interesting to explore in future work. Nevertheless, the assumption of document exchangeability captures the fact that current literature builds on top of previous literature, but is a limitation for topics that have radically changed in the way they are described (e.g. terminology) in literature. For example, the field of atomic physics was described by words relating to "matter" in the late 19th century, "electron" in middle of the 20th century and "quantum" in the late 20th century. Likewise, the field of neuroscience evolved from being described by words relating to "nerve," to "neuron," to "ca2" over the last 100 years (Blei & Lafferty, 2006). The dynamic topic model uses a sequence of time slices in which topics are conditioned on the previous topics, which is a limitation of the standard LDA model used in this study.

2.3 | Creating the data set

Taking into consideration the issue of having access to the electronic version of the text, we decided to include in our analysis only journal articles, as the majority of these are now available for download from online databases. Thus, we have excluded books, books chapters and reports, something that may have introduced bias in our results. Furthermore, due to computational and time constraints, we have limited the number of journals included in our analysis. Thus, the total volume of fisheries publications is underestimated in our analysis, something that limits the results of this study. The dataset was constructed following a set of inclusion criteria to obtain a diverse set of journals that reflect fisheries science while maintaining computational feasibility. First, we included all journals with the term "fisheries" or "fishery" in their title that are listed by the Fisheries Science Citation Index Extended (SCIE) 2016 provided by Thomson Reuters and having an impact factor of ≥1.0. Second, we included all journals from the Fisheries SCIE 2016 that do not include these words in their titles, but explicitly address fisheries in their aims and scopes, and having an impact factor of ≥1.0. Third, we included the top four journals with the highest 2016 impact factor with the term "marine" in their title, indexed by any list from SCIE or Social Science Citation Index, and explicitly addressing fisheries in their aims and scopes. All journals were subject to the University of Tromsø—The Arctic University of Norway subscription rights. A total of 21 journals satisfied these criteria (Table 1). Although journals that do not match these criteria also publish fisheries research, such journals were not considered to be specialized fisheries research outlets (e.g. the journal *Ecology and Society*).

Moreover, even though some of the most influential and highly cited fisheries papers are published in high-impact journals such as *Nature* and *Science*, they only marginally contribute to the total number of papers published in fisheries-related journals. Including in our analysis all publications from *Nature* and *Science* would result in a high number of fisheries-irrelevant topics (e.g. astrophysics), as these journals typically publish a broad range of topics. Using keyword searches to obtain only fisheries-related publications would be

TABLE 1 An overview of the dataset used when creating the latent Dirichlet allocation model to uncover latent topics from fisheries publications. The dataset consists of 46,582 full-text publications from 21 top-tier fisheries journals. Fisheries rank and impact factor are extracted from the 2016 Fisheries ISI Journal Citation Reports (JCR) provided by Thomson Reuters. Journals without a rank are "marine" journals not covered by the JCR fisheries index, but cover fisheries aspects within their aims and scopes. Y_{min} is the lowest publication year, Y_{max} is the highest publication year, N is the number of documents (publications) deemed fit for further analysis, $\bar{W}$ is the mean number of words, std. W is the estimated standard deviation of number of words, and $\bar{V}$ is the mean vocabulary size. Note that word and vocabulary statistics are obtained after the data cleaning process

Journal	Fisheries rank	Impact factor	Y_{min}	Y_{max}	N	$\bar{W}$	Std. W	$\bar{V}$
Fish and Fisheries	1	9.013	2000	2016	419	4,160	2,022	1,084
Reviews in Fish Biology and Fisheries	3	3.575	1991	2016	659	4,142	3,014	1,070
Fisheries	5	3.000	1997	2016	477	1,976	1,313	666
Aquaculture Environment Interactions	6	2.905	2010	2016	203	3,171	1,045	844
ICES Journal of Marine Science	7	2.760	1990	2016	3,903	2,431	1,039	689
Reviews in Fisheries Science & Aquaculture	9	2.545	1997	2016	375	4,442	4,429	1,060
Canadian Journal of Fisheries and Aquatic Sciences	10	2.466	1996	2016	4,423	3,205	1,120	828
Fisheries Research	11	2.185	1995	2016	3,610	2,491	1,083	678
Ecology of Freshwater Fish	13	2.054	1996	2016	932	2,470	971	708
Marine Resource Economics	14	1.911	2010	2016	159	3,609	1,279	835
Fisheries Oceanography	18	1.578	1997	2016	752	3,036	1,122	764
Journal of Fish Biology	21	1.519	1990	2016	7,075	2,112	1,550	651
Transactions of the American Fisheries Society	22	1.502	1997	2016	2,381	3,167	1,185	790
CCAMLR Science	24	1.429	1990	2016	314	1,722	1,123	505
Fisheries Management and Ecology	25	1.327	1994	2016	1,001	1,987	861	629
Knowledge and Management of Aquatic Ecosystems	26	1.217	1997	2016	590	1,860	958	622
North American Journal of Fisheries Management	27	1.201	1997	2016	2,517	2,705	1,206	680
Marine and Coastal Fisheries	28	1.177	2009	2016	274	3,538	1,143	875
Aquatic Conservation: Marine and Freshwater Ecosystems	n/a	3.130	1991	2016	1,328	2,872	1,250	841
Marine Ecology Progress Series	n/a	2.292	1990	2016	12,674	3,045	2,242	811
Marine Policy	n/a	2.235	1990	2016	2,516	3,145	1,263	889
				Total	46,582			

a top-down approach and, hence, would be biased by: (i) the search terms used and (ii) the way publications are indexed and, subsequently, retrieved.

We downloaded full-text research articles published in the 21 journals covering fisheries aspects for a time span of 26 years (1990–2016) to allow for enough variation in publication trends. Analysing full-text articles, compared to just abstract data, results in more detailed and higher quality topics (Syed & Spruit, 2017). Only research articles were considered, and other types of publications, such as errata, conference reports, forewords, announcements, dedications, letters, comments, and book reviews, were excluded. A total of 46,582 articles were deemed fit for further analysis. The year of publication was chosen to be the issue year in which the article appeared in print, regardless of the accepted year or (first) online publication. Information about the journal name, the time range for which articles were collected, the journal's impact factor, the total number of articles deemed fit for further analysis and word statistics are given in Table 1. Additionally, an overview of the number of publications per journal per year is shown in Figure 1. Not all journals provided articles for the complete time span of 26 years. For example, the journal *Fish and Fisheries* started in 2000 and, therefore, only articles from 2000 to 2016 were included in the study. Another example relates to the journal subscription rights, which did not allow data collection for all years. For example, the *Canadian Journal of Fisheries and Aquatic Sciences* started in 1901, but our subscription only allowed access from 1996.

All articles appeared in portable document format (PDF) and were first converted to their plain text representation. This resulted in a complete transformation from PDF to plain text for all elements of each article, including the header, title, author info, affiliation info, abstract, keywords, content, tables, bibliography and captions. Several articles, mainly from the early 1990s, were image-based PDFs that were unsuitable for direct conversion from PDF to plain text. In these cases, the Tesseract optical character recognition (OCR) software library was used to subsequently convert these articles into text-based PDFs and then to their plain text representation. To make sure that we only analysed the content text of each article, we used an advanced text pattern search method to remove boilerplate content such as journal information, article metadata, acknowledgments and bibliographies. Additionally, multilanguage abstracts or non-English articles (e.g. articles that appeared in French) were also removed.

Latent Dirichlet allocation is a bag-of-words model in which documents are represented as sequences of individual word features. As such, every document was tokenized. Tokenization is the process of obtaining individual words (also known as unigrams) from sentences. Unigrams lose important semantic cues that are encoded by compound words. To overcome this, bigrams were included by combining two consecutive unigrams that occurred ≥20 times within each document. As a result, compound words, such as "rainbow trout," are preserved. Additionally, we used named entity recognition (NER), a technique from natural language processing (NLP), to retrieve entities related to names, nationalities, companies, locations, objects etc., from the documents. Entities such as "the European Union," "the Norwegian Research Council" or "marine protected areas" are thus preserved and included in the analysis. The inclusion of bigrams and entities allows for a richer bag-of-words representation than a standard unigram representation.

Although all tokens within a document serve an important function, for topic modelling they are not all equally important. We proceed by filtering out numbers, punctuation marks and single-character words as they bear no topical meaning. Furthermore, we removed stop words (e.g. *the*, *is*, *a*, *which*), words that occurred only once (e.g. mainly typos and incorrectly hyphenated words) and words that occurred in ≥90% of the documents (e.g. *result*, *study*, *show*) as they serve no discriminative topical significance. Omitting frequently occurring words prevents such words from dominating all topics.

For grammatical reasons, different word forms or derivationally related words can have a similar meaning and, ideally, we would want such terms to be grouped. Stemming and lemmatization are two NLP techniques to reduce inflectional and derivational forms of words to a common base form. Stemming heuristically cuts off derivational affixes to achieve some normalization, albeit crude in most cases. Stemming loses the ability to relate stemmed words back to their

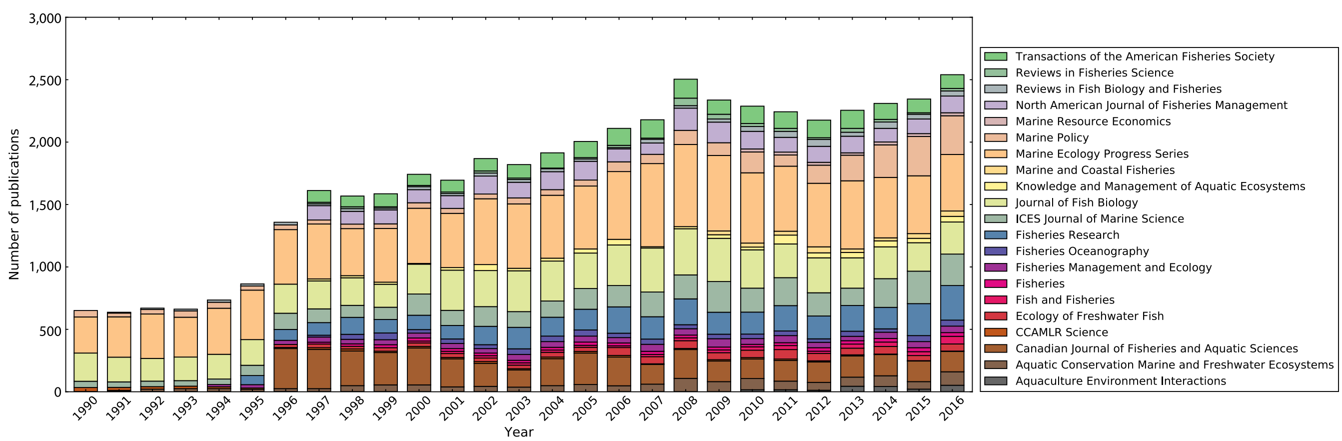

FIGURE 1 The number of publications (*y*-axis) per journal (colour-coded) for the years 1990–2016 (*x*-axis) that were used to create the latent Dirichlet allocation model. The total number of documents was 46,582

original part of speech, such as verbs or nouns, and decreases the interpretability of topics in later stages (Evangelopoulos, Zhang, & Prybutok, 2012). For example, the term "fishing" will be stemmed to "fish"; likewise, "modeling" will be stemmed to "model" and cannot be returned to its original part of speech (i.e. verb). Our analysis uses lemmatization, which is a more sophisticated normalization method that uses a vocabulary and morphological analysis to reduce words to their base form, called lemma. It is best described by its most basic example, normalizing the verbs "am," "are," "is" to "be," although such terms will be filtered out from our analysis. Likewise, lemmatization correctly normalizes "fisheries" and "fishery," and "policies" and "policy." Additionally, uppercase and lowercase words were grouped. The final corpus consisted of 46,582 full-text publications with around 130 million words and 170,000 unique words.

2.4 | Creating the LDA model

Latent Dirichlet allocation assumes that the number of topics to uncover is known in advance and is set by the *K*-parameter. As the optimal number of topics is not known in advance, we created 50 different LDA models by varying the *K*-parameter from 1 to 50. Measures to determine the optimal LDA model are described in the next section. The LDA models are created using the Python library Gensim (Rehurek & Sojka, 2010). Since LDA is a Bayesian probabilistic model, we can incorporate some prior knowledge into the model. Prior knowledge can be encoded by symmetrical or asymmetrical Dirichlet priors. A symmetrical prior distribution of topics within documents assumes that all topics have an equal probability of being assigned to a document. Such an assumption ignores that certain topics are more prominent in a document collection and, consequently, would logically have a higher probability to be assigned to a document. Conversely, specific topics are less common and, thus, not appropriately reflected with a symmetrical prior distribution. Logically speaking, an asymmetrical prior would capture this intuition and would, therefore, be the preferred choice (Syed & Spruit, 2018; Wallach, Mimno, & Mccallum, 2009). Additionally, we iteratively optimize the prior using the Newton–Rapson method (Huang, 2005) by learning it from the data. To infer the hidden variables (i.e. inferring the posterior distribution of the hidden variables given the observed documents), we use variational inference called "online LDA" (Hoffman, Blei, & Bach, 2010).

2.5 | Calculating model quality

Analogous to choosing the right number of clusters for techniques such as *k*-nearest neighbours, choosing the right number topics is an important task in topic modelling, including LDA, to appropriately capture the underlying topics in a dataset. A low number of topics will result in a few too broad topics, with high values capturing meaningless topics; such topics are merely the result of the statistical nature of LDA. Several approaches to determine the optimal number of topics have been proposed. One such approach is to fit various topic models to a training set of documents and calculate a model fit on a test set (held-out data) (Scott & Baldridge, 2013). The model that fits best on the test set would be considered a better model. However, topic models are used by humans to interpret and explore the documents, and there is no technical reason that the best-fitted model would aid best in this task (Blei, 2012). In fact, research has shown that such measures negatively correlate with human interpretation (Chang, Gerrish, Wang, & Blei, 2009).

Another approach is to assess the quality of topics with human topic ranking, which is considered the gold standard when assessing the interpretability of generated topics. Such ranking is often based on word or topic intrusion tests, in which an intruder word or topic needs to be recognized within a set of related or cohesive words or topics (Chang et al., 2009). However, this approach is time-consuming and expensive as for every created topic model (e.g. 1–50), and for every topic within that model, the interpretability of individual words and sets of words needs to be assessed. To circumvent this, a more quantitative approach is preferred while maintaining human interpretability. One way is to assess the quality of topics with coherence measures that are based on the distributional hypothesis (Harris, 1954), which states that words with similar meanings tend to co-occur within similar contexts. Such an approach, drawing on the philosophical premise that a set of statements or facts is said to be coherent if its statements or facts support each other, informs us about the understandability and interpretability of topics from a human perspective. This study utilized the C_V coherence measure (Röder, Both, & Hinneburg, 2015), which has shown the highest correlation with all available human topic ranking data, and is thus an appropriate quantitative approach (see supplementary material for an extensive and mathematical explanation of the C_V coherence measure). The C_V coherence score for all 50 LDA models was calculated and an elbow method, estimating the (inflection) point where adding more topics will not increase coherence, was used to obtain the optimal number of topics.

2.6 | Labelling topics

As previously described, the topical structure that permeates the document collection is latent, and the probability distributions of words (i.e. topics) are not semantically labelled (i.e. they are not given a name by the LDA model). When sorted, the top 10 or top 15 high probability words within each topic are used to describe the topic. However, algorithmic analyses of content remain limited in their capacity to understand latent meanings or the subtleties of human language (Lewis, Zamith, & Hermida, 2013) and manual labelling is still considered the gold standard in topic modelling (Lau et al., 2011). Thus, the labelling of each topic (i.e. giving a name to each topic; e.g. a topic with the five most probable words being "pig," "cow," "sheep," "goat," "horse" would be labelled as "domestic animals") was performed by a human analyst, that is a fisheries domain expert. When identifying the common subject of each topic (i.e. the name or the label of the topic), the analyst used the following procedure. First, the analyst closely inspected the 15 most probable words from each topic. Second, the analyst inspected the titles of the documents in the dataset that were included by the topic model in that respective topic. The interested reader can find a sample of publication titles that have high probability within a single topic in Table S1 of the supplementary material. Third, based on the previous two steps, the analyst labelled a topic (i.e. gave it a name; e.g., if the LDA model included in topic 1 the words “pig,” “cow,” “farm” and the titles of the documents included by the model in this topic have in common the subject of domestic animals, then the analyst gave topic 1 the label of “domestic animals”). Furthermore, to validate the labelling of the topics, we visualized the topics in a two-dimensional area by computing the distance between topics (Chuang, Ramage, Manning, & Heer, 2012) and applying multidimensional scaling (Sievert & Shirley, 2014). This two-dimensional topic representation displays the similarity between topics with respect to their word distribution over topics, that is the words and their corresponding probability within the topic. Clustering and overlapping nodes indicate similar word distributions, and the surface of the node indicates the relative topic prevalence in the complete set of documents. The topic prevalence indicates how widespread a topic is within all the documents, as all topic proportions add up to 100%. In a fourth step, the analyst used this visualization in order to validate the choice of the final label for each topic (e.g. topics using similar vocabulary usually refer to similar subjects; thus, for example, the topics labelled by the analyst “domestic animals” and “astrophysics” appearing close together in the two-dimensional topic representation would raise suspicions and the analyst would have to go through the labelling procedure again in order to find labels that make sense for the two vocabulary-close topics). The labels were further validated in a fifth step, as described in the section Validation of Results.

2.7 | Calculating topical trends over time

To gain insight into the topical temporal dynamics of the fisheries field, we aggregated the document topic proportions for each year and for every individual topic into a composite topic-year proportion (see supplementary material). Doing so provides a sense of how the prevalence of each topic within fisheries science publications has changed over the last 26 years. Additionally, to obtain insight into increasing and decreasing topical trends, we fit a one-dimensional least square polynomial for different time intervals. The polynomial coefficient is used as a proxy for the trend and defines the slope of the composite topic-year proportions for a range of years. Coefficients are multiplied by the number of years within each time interval to obtain the change measured in percentage points. Positive values indicate increasing or *hot* topics and negative values decreasing or *cold* topics. The time intervals allow for historical comparisons between 1990–95, 1995–2000, 2000–05, 2005–10 and 2010–16. Colour coding is used to resemble the hot (i.e. red) and cold (i.e. blue) topical trends.

2.8 | Calculating topic over journals

Following a similar approach as topical trends over time, we aggregate topic proportion per journal to gain insight into how topics are covered by the journals included in this study. Doing so enables us to identify broadly oriented or more focused oriented journals. Note that aggregation of topic proportions is handled per journal and covers only the years for which articles are downloaded (see Table 1). For example, for the journal *Fish and Fisheries*, journal topic distributions cover the time range 2000–16, whereas the journal *ICES Journal of Marine Science* covers the complete time range of 1990–2016.

2.9 | Relaxing LDA assumptions and future research directions

At the time of writing, the original LDA method proposed by Blei et al. (2003) had over 20,000 citations. The technique has received much attention from machine-learning researchers and other scholars and has been adopted and extended in a variety of ways. More concretely, relaxing the assumptions behind LDA can result in richer representations of the underlying semantic structures. The bag-of-words assumption has been relaxed by conditioning words on the previous words (i.e. Markovian structure) (Wallach, 2006); the document exchangeability assumption, relaxed by the previously mentioned dynamic topic model (Blei & Lafferty, 2006), and the Bayesian nonparametric model can be utilized to automatically uncover the number of topics (Whye Teh, Jordan, Beal, & Blei, 2004). Furthermore, LDA has been extended in various ways. Topics might correlate as a topic about “cars” is more likely to also be about “emission” than it is about “diseases.” The Dirichlet distribution is implicitly independent, and a more flexible distribution, such as the logistic normal, is a more appropriate distribution to capture covariance between topics. The correlated topic model aids in this task (Blei & Lafferty, 2007). Other examples extending LDA include the author-topic model (Rosen-Zvi, Griffiths, Steyvers, & Smyth, 2004), the relational topic model (Chang & Blei, 2010), the spherical topic model (Reisinger, Waters, Silverthorn, & Mooney, 2010), the sparse topic model (Wang & Blei, 2009) and the bursty topic model (Doyle & Elkan, 2009). Apart from its applicability to text, LDA can be applied to audio (Kim, Narayanan, & Sundaram, 2009), video (Mehran, Oyama, & Shah, 2009) and image (Fergus, Fei-Fei, Perona, & Zisserman, 2005) data. Those topic models that relax or extend the original LDA model bring additional computational complexity and their own sets of limitations and challenges; nevertheless, it would be interesting to explore these models in future research.

3 | RESULTS AND DISCUSSION

3.1 | Uncovering fisheries topics

The LDA model with the optimal coherence score contains 25 topics ($k = 25$). The ten most probable words (i.e. the words with the highest probabilities), together with the semantically attached label for each uncovered latent topic, are shown in Table 2. The manually assigned labels for the 25 topics are as follows: (1) Conservation, (2) Morphology, (3) Salmon, (4) Reproduction, (5) Non-fish species, (6) Coral reefs, (7) Biochemistry, (8) Freshwater, (9) Diet, (10) North

TABLE 2 Table showing the 25 uncovered topics from 46,582 fisheries science articles published in 21 fisheries-specialized journals in the period 1990–2016. Each topic displays the ten most probable words (i.e. words with the highest probability). The topics are manually labelled with a logical topic description that best captures the semantics of the top words

Topic	Label	Top-10 words	Theme
1	Conservation	Marine, Ecosystem, Change, System, Environmental, Impact, Environment, Process, Ecological, Research	Management
2	Morphology	Specimen, Mm, Body, Dorsal, Morphological, Length, Shape, Head, Form, Morphology	Aquatic organism biology
3	Salmon	Salmon, Chinook, Chinook Salmon, Pacific, River, Columbia, Fish, Year, Stock, Juvenile	Specific aquatic organisms
4	Reproduction	Female, Male, Sex, Size, Reproductive, Shark, Spawn, Mature, Maturity, Oocyte	Aquatic organism biology
5	Non-fish species	Sediment, Crab, Site, Mussel, Seagrass, Density, Treatment, Effect, Plant, Shell	Specific aquatic organisms
6	Corals	Reef, Coral, Site, Habitat, Area, Community, Abundance, Colony, Island, Depth	Aquatic habitats
7	Biochemistry	Concentration, Cell, Tissue, Acid, Protein, Lipid, Level, Sample, Activity, Exposure	Aquatic organism biology
8	Freshwater	Lake, Fish, Bass, Reservoir, Angler, Largemouth, Walleye, Population, Year, Perch	Aquatic habitats
9	Diet	Prey, Diet, Food, Predator, Size, Feed, Fish, Trophic, Value, Consumption	Aquatic organism biology
10	North Atlantic	Sea, Cod, Area, North, Lamprey, Fish, Atlantic, Herring, Parasite, Baltic	Geographical areas
11	Southern Hemisphere	Sea, Area, Water, Region, Island, Shelf, Whale, Temperature, South, Depth	Geographical areas
12	Development	Egg, Larval, Larvae, Spawn, Stage, Temperature, Larva, Early, Day, Hatch	Aquatic organism biology
13	Genetics	Population, Genetic, Sample, Analysis, Individual, Gene, Allele, Dna, Loci, River	Aquatic organism biology
14	Assemblages	Habitat, River, Site, Stream, Water, Flow, Area, Channel, Fish, Reach	Aquatic organism biology
15	Growth experiments	Fish, Temperature, Treatment, Experiment, Effect, Water, Rate, Control, Tank, Test	Aquatic organism biology
16	Stock assessment	Year, Population, Stock, Mortality, Rate, Recruitment, Estimate, Model, Biomass, Change	Modeling
17	Growth	Growth, Otolith, Length, Fish, Sturgeon, Sample, Mm, Size, Growth Rate, Rate	Aquatic organism biology
18	Tracking and movement	Fish, Tag, River, Release, Dam, Movement, Hatchery, Migration, Survival, Rate	Aquatic organism biology
19	Fishing gear	Catch, Fishing, Fishery, Fish, Gear, Net, Trawl, Size, Hook, Vessel	Fish technology
20	Primary production	Concentration, Water, Rate, Nutrient, Phytoplankton, Production, Sediment, Carbon, Chl, Sample	Specific aquatic organisms
21	Models	Model, Estimate, Value, Variable, Parameter, Analysis, Effect, Distribution, Base, Sample	Modeling
22	Salmonids	Trout, Fish, Stream, Rainbow, Population, Brook, Cutthroat, Creek, Brown, Salmonid	Specific aquatic organisms
23	Acoustics and swimming	Depth, Fish, Velocity, Water, Speed, Distance, Vertical, Surface, Sound, Night	Aquatic organism biology
24	Estuaries	Fish, Estuary, Water, Bay, Salinity, Estuarine, Area, Freshwater, Habitat, River	Aquatic habitats
25	Fisheries management	Fishery, Management, Fishing, State, Resource, Economic, Vessel, Policy, Area, Fish	Management

Atlantic, (11) Southern Hemisphere, (12) Development, (13) Genetics, (14) Assemblages, (15) Growth experiments, (16) Stock assessment, (17) Growth, (18) Tracking and movement, (19) Fishing gear, (20) Primary production, (21) Models, (22) Salmonids, (23) Acoustics and swimming, (24) Estuaries and (25) Fisheries management. These 25 topics can be grouped into overarching themes: aquatic organism biology (n = 11), specific aquatic organisms (n = 4); aquatic habitats (n = 3), geographical areas (n = 2), modelling (n = 2), management (n = 2) and

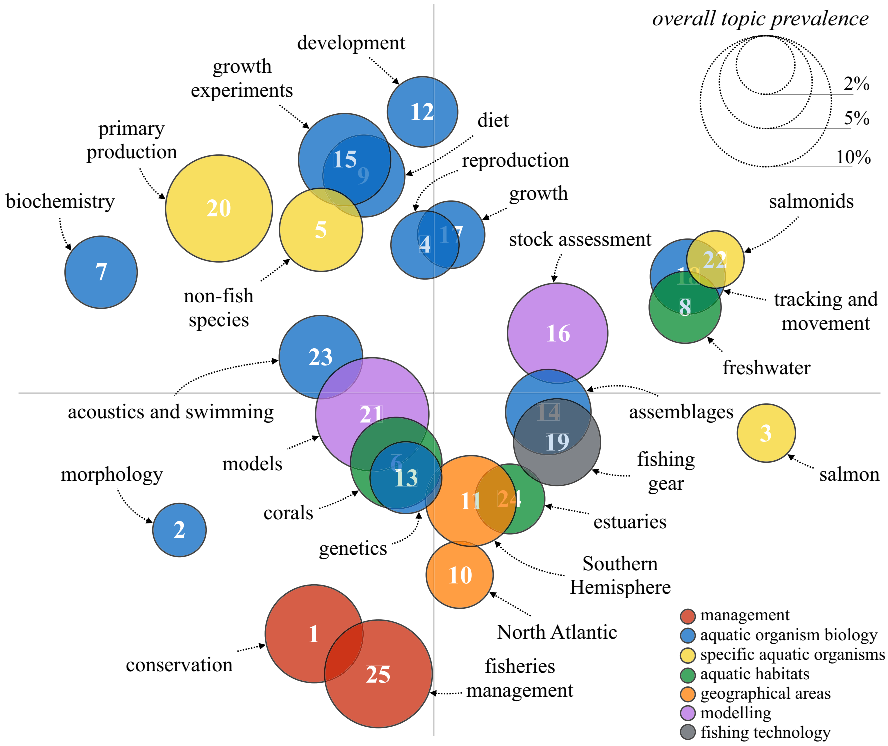

FIGURE 2 Intertopic distance map that shows a two-dimensional representation (via multidimensional scaling) of the 25 uncovered fisheries topics. The distance between the nodes represents the topic similarity with respect to the distributions of words (i.e. latent Dirichlet allocation's output). The surface of the nodes indicates the topic prevalence within the corpus, with bigger nodes representing topics being more prominent within the document collection (all nodes add up to 100%)

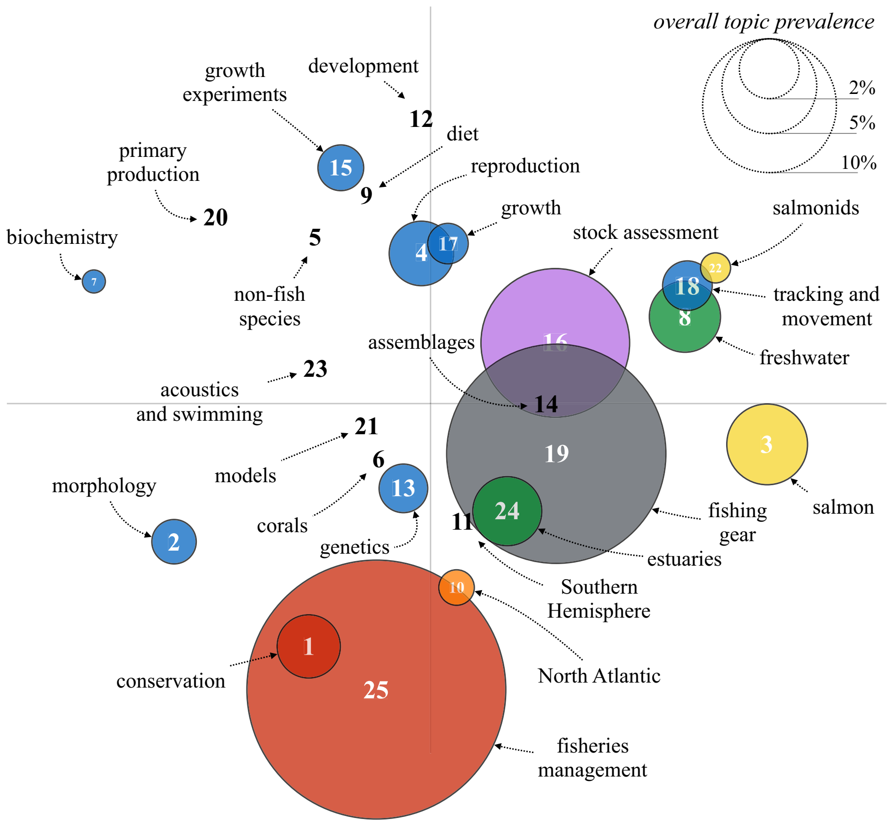

FIGURE 3 Intertopic distance map showing topics conditioned on the word "fishery" (including "fisheries"). The figure is similar to Figure 2 but expresses the differences in probability assigned to the word "fishery." Bigger nodes place higher probability to the word "fishery" and can be considered more directly related to fisheries science

fishing technology (n = 1). A visual representation of the topics, their proportions within the complete corpus, and their grouping in overarching themes can be found in Figure 2.

Conditioning the topics on the word "fishery"/"fisheries" (i.e. taking into consideration the probability assigned to this word), these 25 topics can be divided into four groups, the first two of which we considered to be directly relating to fisheries: using the word often (n = 3), moderately (n = 5), infrequently (n = 8) or almost not at all (n = 9) (Figure 3). The topics using this word often are, in descending order: (25) Fisheries management, (19) Fishing gear, and

(16) Stock assessment. Almost one-fifth of all the topics are using the word “fishery”/“fisheries” moderately: (1) Conservation, (3) Salmon, (4) Reproduction, (8) Freshwater, (24) Estuaries. One-third of all the topics are using the word “fishery”/”fisheries” infrequently: (2) Morphology, (7) Biochemistry, (10) North Atlantic, (13) Genetics, (15) Growth experiments, (17) Growth, (18) Tracking and moving, (22) Salmonids. Another one-third of all of the topics does not use this word almost at all: (5) Non-fish species, (6) Corals, (9) Diet, (11) Southern Hemisphere, (12) Development, (14) Assemblages, (20) Primary production, (21) Models, (23) Acoustics and swimming.

Considering all the 25 topics, only one [i.e. (25) Fisheries management] refers explicitly to the human dimension component of the fishery system, something that confirms our working hypothesis that the human dimension of fisheries is under-represented in the fisheries specialty literature. To evaluate whether the human dimension of fisheries as SECASs is further refined within the Fisheries management topic, following the same methodology as described above, we created a new LDA model that zooms in on this topic, thereby creating subtopics from the broader Fisheries management topic. The new model uncovered 12 subtopics from the topic Fisheries management (Table 3), out of which eight assign higher probability to the term “fishery”/”fisheries” (i.e. use this word often or moderately) and, thus, were considered directly relating to fisheries: three on various management approaches (i.e. Co-management, Precautionary approach and Quota systems); three on economics (i.e. Markets, Bioeconomics and Blue economy); and two on type of fishery (i.e. Small-scale fisheries and Recreational fisheries).

Out of the total of 25 topics uncovered by our analysis, three were considered generic [i.e. (10) North Atlantic, (11) Southern Hemisphere, (21) Models)]. Out of the remaining 22 topics, 20 cover the natural dimension of fisheries, reasonably mirroring the curriculum of fisheries biology and fisheries ecology higher education courses (e.g. Jennings, Kaiser, & Reynolds, 2009; King, 2007), but not addressing topics such as climate change. However, considering the focus of the two remaining topics, that is (1) Conservation and (25) Fisheries management, it is apparent that the research focus in fisheries during the last 26 years has not entirely captured the complexity of the fisheries domain, especially of the human dimension component, something also observed for specialized research areas such as, for example, by-catch reduction technology (Campbell & Cornwell, 2008; Molina & Cooke, 2012). These results seem to be confirmed by the bibliometric analyses published in Aksnes and Browman (2016) and Jarić et al. (2012), where no human dimension-related words were identified among the most frequent words used in fisheries publication titles and abstracts. This situation might not be surprising given the institutional context in which fisheries research is performed. For example, within the International Council

TABLE 3 Table showing the 12 uncovered subtopics from the documents (*n* = 3,390) dealing with the topic Fisheries management. The subtopics provide a “zoomed-in” view of the topical decomposition from the subset of documents on fisheries management. Each topic displays the ten most probable words and the semantic label that best describes the underlying latent topic

Topic	Label	Top-10 words	Theme
1	Spatial planning	Marine, Policy, Stakeholder, Process, Coastal, Development, Sea, Plan, Environmental, Regional	Non-fisheries
2	Markets	Price, Market, Fish, Product, Production, Model, Value, Seafood, Estimate, Sector	Economics
3	Bioeconomics	Cost, Model, Stock, Fishery, Effort, Value, Scenario, Harvest, Fish, Rate	Economics
4	Conservation/MPA	Marine, Mpa, Conservation, Protect, Ocean, Ecosystem, Mpas, Protection, Sea, Habitat	Non-fisheries
5	Small-Scale fisheries	Fishing, Fisher, Fish, Study, Catch, Fishery, Boat, Fisherman, Local, Community	Type of fishery
6	Blue economy	Marine, Fish, Aquaculture, Fishery, Shark, Water, Coastal, Development, Production, Specie	Economics
7	Pollution	Ship, Vessel, Oil, Port, Shipping, Pollution, Risk, International, Trade, Country	Non-fisheries
8	Legislation	Sea, Law, International, Convention, Agreement, Country, China, Water, Coastal, Maritime	Non-fisheries
9	Co-management	Fishery, Community, Social, Fishing, System, Right, Group, Fisher, Government, Local	Management approaches
10	Quota systems	Fishery, Vessel, Fishing, Catch, Quota, Fish, Fleet, System, Total, Stock	Management approaches
11	Precautionary approach	Fishery, Stock, Specie, Catch, Fishing, Datum, Assessment, Fish, Whale, Ecosystem	Management approaches
12	Recreational fisheries	Angler, Recreational, Fish, Fishing, Survey, Respondent, Catch, Fishery, Estimate, Value	Type of fishery

for the Exploration of the Sea (ICES), which is one of the most important fisheries-related intergovernmental organization, despite having along the years various groups working more or less directly working with different aspects of this human dimension, now only one out of more than 45 expert groups is working explicitly with this dimension of fisheries. This group, the Strategic Initiative Human Dimension, became operational in 2015.

Persisting in having this heavily unbalanced focus between the two dimensions of fisheries systems (i.e. the human dimension [i.e. human agents, communities of these and their institutions] and the natural dimension [i.e. biotic, such as predator species and prey species, and abiotic, such as water temperature and nutrients]) will not help in understanding the behaviour of fisheries stakeholders (from fishers to consumers), leading to unintended, and too often undesirable, management outcomes (Fulton, Smith, Smith, & van Putten, 2011), and thus unsustainable fisheries. Responding to the challenges posed by sustainable fisheries necessitates the development of stronger networks within the family of human dimension sciences and across disciplinary boundaries with the natural dimension sciences (Symes & Hoefnagel, 2010). Without providing an exhaustive list and in random order, the human dimension in fisheries could be included in fisheries science by addressing topics such as institutional aspects (enforcement and compliance, policy interactions, etc.), social aspects (gender, religion/beliefs, welfare, social cohesion, social networks, education and learning, human agency, health, safety and security at sea, food security, perception, attitudes, social norms, compliance, mental models of various actors involved in fisheries, etc.), economic aspects (poverty, innovation, distribution of benefits, spiritual, inspirational and aesthetic services of fisheries, etc.), political aspects (power structures, transparency, etc.) and cultural aspects (traditional/local ecological knowledge, history, cultural dimensions, culinary choices, heritage, blue humanities, fisheries literacy, etc.) (Charles, 2001; De Young, Charles, & Hjort, 2008; ICES, 2016; Österblom et al., 2013; Sowman, 2011; Spalding, Biedenweg, Hettinger, & Nelson, 2017; Stone-Jovicich, 2015).

Continuing our analysis, three topic clusters can be identified in Figure 2, indicating a similar probability distribution over words (i.e. topics that are, to some extent, related to the words they use to describe the theme): a growth cluster [the topics Growth experiments (15), Diet (9), Non-fish species (5), Primary production; Development (12), Reproduction (4) and Growth (17)]; an institutions cluster [the topics Fisheries management (25) and Conservation (1)]; and a salmonids cluster [the topics Freshwater (8), Tracking and movement (18), and Salmonids (22); one would expect to find here also the topic Salmon (3), but, interestingly from a linguistics point of view, this topic seems rather isolated]. The most isolated topics are Morphology (2) and Biochemistry (7), indicating most probably the use of a very specific topic distribution over words.

The most frequent aquatic organisms mentioned in our corpus are salmonids (e.g. salmon, trout) and other freshwater organisms (e.g. perch); shark (within the topic Reproduction); crab, mussel, and oyster (within the topic Non-fish species); cod, lamprey, and herring (within the topic North Atlantic); whale (within topic Southern Hemisphere); sturgeon (within the topic Growth); tuna (within the topic Fishing gear); and shrimp (within topic Estuaries). Commercially important species, such as anchoveta, pollock, and tilapia, were not included among the most frequent words of any of the 25 topics. These findings are relatively consistent with Aksnes and Browman (2016) and Jarić et al. (2012), who reported that the most frequently studied group of species was the Salmonidae, followed by the Atlantic cod, and that there is no correlation between the production of various species and the number of publications about these species.

Aquaculture has not been identified as a topic of its own in our dataset, and the word *aquaculture* was not included in the top 10 most frequent words of any of the 25 topics. However, the word *aquaculture* was included in the top 10 most frequent words for the subtopic Blue economy under the topic Fisheries management, possibly indicating the interest in this relatively new industry in a context that focuses on increasing economic activities in the marine and maritime sector.

With regard to the typology of fishers by Charles (2001), only the recreational type is specifically mentioned among the most frequent words in our corpus, with the word *angler* included in the topic (8) Freshwater. This might be because the research focusing on the other types (e.g. subsistence, artisanal) does not employ a very specific vocabulary, or that there might be a lag in research on, for example, small-scale and artisanal fishery (Purcell & Pomeroy, 2015). However, if we look only at topic (25) Fisheries management, the recreational type and small-scale type have each its subtopic, indicating that, from a management perspective, these two types of fisheries have been relatively extensively explored by fisheries scientists.

Out of the 25 topics uncovered by our LDA model, two refer to large geographical areas: the North Atlantic (10) and Southern Hemisphere (11). The words *Norwegian* (within the topic North Atlantic) and *Florida* [within the topic Estuaries (25)] are the only specific geographical references among the top 10 most probable words. These very few specific geographical references might indicate that most of the fisheries research is focused on a few areas around the globe, leaving large zones underexplored, as also indicated in Molina and Cooke (2012), or that research about other regions is published in other languages than English.

3.2 | Topic proportions within documents

For every document, the LDA model infers the topical decomposition, indicating which topics are found in that document and in what proportions. The assumptions behind LDA cause documents to exhibit mainly a small number of main topics, with other topics very close to zero (note that all topics per document sum up to 1). This structure assumes that documents are often about some topics, rather than being about all topics equally. Following this line of reasoning, we analysed the remaining topic proportions for documents exhibiting one of the topics as the dominant topic, this being defined as the document's topic proportion that exceeds all other topic proportions. Such an analysis provides insight into the document composition that comes from the mixing topic proportions, that is what topics co-occur together within fisheries research articles.

	(21) Models	(16) Stock assessment	(15) Growth experiments	(1) Conservation	(9) Diet	(14) Assemblages	(11) Southern Hemisphere	(6) Corals	(19) Fishing gear	(23) Acoustics and swimming	(24) Estuaries	(10) North Atlantic	(12) Development	(18) Tracking and movement	(4) Reproduction	(5) Non-fish species	(25) Fisheries management	(17) Growth	(7) Biochemistry	(8) Freshwater	(20) Primary production	(2) Morphology	(22) Salmonids	(13) Genetics	(3) Salmon
(25) Fisheries management (59.0%)	4.8	3.7	0.7	12.7	0.4	0.6	1.4	2.0	5.3	0.7	1.2	2.0	0.2	0.5	0.5	0.6		0.2	0.4	1.2	0.2	0.3	0.2	0.3	1.0
(20) Primary production (55.9%)	3.9	1.7	3.0	2.4	3.5	2.2	3.4	2.7	0.2	2.7	1.3	1.3	1.3	0.2	0.2	5.8	0.3	0.6	4.1	1.2		0.9	0.2	0.8	0.3
(13) Genetics (50.4%)	3.1	1.9	1.3	3.3	0.4	2.1	2.4	2.7	1.3	0.5	4.3	3.2	1.3	2.2	2.4	1.5	1.0	1.3	2.1	2.3	0.6	3.2	2.9		2.4
(19) Fishing Gear (45.4%)	9.1	4.3	3.0	2.3	1.1	0.5	4.0	3.0		4.6	2.9	3.4	0.5	1.3	2.0	1.2	5.9	1.2	0.3	1.2	0.2	0.9	0.3	0.2	0.9
(23) Acoustics and swimming (44.8%)	8.1	1.9	5.4	2.2	2.8	1.8	4.9	2.5	4.2		1.8	2.3	2.4	2.8	1.1	1.7	0.8	0.8	0.5	1.6	2.2	1.5	0.6	0.2	1.2
(6) Corals (44.4%)	4.9	3.6	2.9	4.7	2.6	1.6	2.2		2.9	2.9	2.8	1.7	1.9	1.0	1.8	6.8	2.0	1.3	1.4	0.5	2.7	1.5	0.6	1.0	0.3
(14) Assemblages (44.4%)	7.6	2.0	2.5	6.0	1.9		0.5	1.8	0.7	2.4	3.5	0.8	1.6	2.8	0.7	2.7	1.0	1.2	0.7	4.1	2.4	1.2	5.5	0.6	1.6
(7) Biochemistry (43.9%)	1.9	0.8	11.7	2.2	2.6	0.8	1.1	1.8	0.5	1.0	1.8	1.9	3.1	1.0	3.5	4.0	0.5	1.3		1.1	4.3	4.1	1.2	3.2	0.7
(15) Growth Experiments (43.6%)	2.9	1.5		2.2	4.1	1.7	0.7	1.4	1.7	4.6	2.0	1.9	2.8	3.1	2.6	1.7	0.7	1.9	7.2	2.6	1.6	2.1	3.2	1.2	1.2
(17) Growth (43.3%)	5.3	4.3	3.8	1.7	1.2	1.9	2.9	2.1	2.9	1.7	4.0	2.6	3.8	3.0	3.4	0.7	0.8		1.9	2.3	1.0	2.0	1.0	1.3	1.2
(21) Models (43.0%)		10.0	1.2	3.1	1.3	2.6	3.2	2.4	6.3	3.5	1.2	2.1	1.4	2.3	1.2	1.1	2.7	2.6	0.4	2.0	1.2	0.4	1.0	0.6	3.2
(5) Non-Fish Species (42.8%)	3.2	3.2	4.0	3.6	3.8	2.9	1.5	9.3	0.7	2.9	2.2	1.7	2.0	0.5	1.2		0.6	1.3	2.5	0.6	7.2	1.1	0.6	0.7	0.2
(11) Southern Hemisphere (42.7%)	6.3	5.5	1.5	2.8	5.4	0.7		2.4	4.4	5.6	2.1	2.6	4.1	1.2	2.6	0.8	1.3	1.3	0.9	0.2	2.8	0.6	0.1	0.6	1.6
(18) Tracking and movement (42.4%)	4.9	3.5	6.2	2.3	0.9	4.4	1.6	1.2	2.1	4.9	1.7	3.3	0.9		2.1	0.5	1.4	2.4	1.2	2.7	0.2	1.1	3.3	1.1	3.7
(8) Freshwater (41.9%)	5.4	6.4	3.3	2.6	6.3	5.4	0.7	0.9	2.7	2.1	1.4	0.9	2.0	3.1	1.6	1.0	2.0	2.6	1.0		1.4	1.4	2.3	1.1	0.7
(1) Conservation (40.5%)	4.1	3.7	2.3		1.4	4.1	2.4	3.9	1.4	1.4	2.9	2.4	0.6	1.6	0.8	2.2	13.6	0.4	1.2	1.8	1.8	1.3	1.4	1.8	1.0
(16) Stock Assessment (40.3%)	12.3		1.1	3.6	2.7	1.0	3.7	1.5	4.7	1.0	1.5	5.0	3.3	1.5	2.6	1.0	4.2	2.5	0.2	1.8	0.7	0.3	0.7	0.6	2.3
(2) Morphology (39.7%)	1.4	0.4	3.6	2.6	2.2	1.7	1.9	3.2	1.4	2.8	7.8	2.6	3.2	0.8	4.1	1.5	0.7	4.9	5.0	2.0	1.2		0.8	4.0	0.3
(3) Salmon (38.5%)	7.2	6.9	1.7	3.1	3.9	3.8	4.9	0.7	2.3	1.1	0.7	1.3	1.4	7.1	1.6	0.4	1.8	2.6	0.9	1.2	0.9	0.3	1.8	3.9	
(4) Reproduction (38.3%)	3.5	2.9	6.4	2.9	1.6	1.1	2.6	2.8	2.8	1.9	3.6	2.1	5.9	1.8		1.2	1.1	4.0	3.6	1.4	0.4	3.7	1.4	2.2	0.8
(22) Salmonids (38.0%)	4.7	3.6	6.4	2.5	3.6	11.5	0.4	0.4	0.7	2.6	0.7	1.4	1.8	6.4	2.1	0.7	0.8	2.1	0.8	3.5	0.5	0.9		2.2	1.5
(9) Diet (37.1%)	5.0	3.1	4.5	2.2		1.8	4.2	3.2	2.7	2.1	4.7	3.2	2.9	0.7	1.9	2.9	0.4	1.8	2.5	4.4	3.9	1.9	1.6	0.5	1.1
(24) Estuaries (37.0%)	3.3	3.1	1.6	5.3	4.3	6.9	2.5	4.2	4.0	2.1		1.8	2.6	1.4	2.8	3.0	2.1	2.3	0.8	2.6	1.3	2.7	0.6	1.6	0.2
(12) Development (34.3%)	5.3	6.1	6.3	1.4	3.1	1.4	4.9	2.7	1.3	4.4	2.1	3.7		0.7	4.3	2.9	0.3	3.1	3.2	1.1	3.2	1.8	0.7	0.9	0.7
(10) North Atlantic (34.1%)	4.6	7.0	2.9	3.1	3.8	1.1	4.2	2.8	4.4	2.3	2.6		4.0	4.0	2.0	1.7	1.2	2.1	2.0	1.3	1.3	2.8	1.4	1.9	1.4

FIGURE 4 Heat map displaying the dominant topic (left) and the remaining average topic proportions (top) for all 46,582 documents from the corpus. This indicates the extent to which documents about one main topic relate to the other uncovered topics (i.e. the degree of topic co-occurrence), decreasing from left to right. For example, documents that primarily focus on Fisheries management also focus on Conservation (12.7%) or Fishing gear (5.3%)

Figure 4 shows the average remaining topic proportions with respect to the dominant topic displayed as a heat map. The dominant topic, shown on the left (i.e. rows), has the average dominant topic proportion within parentheses. A higher number indicates that more of the document content deals with the dominant topic, while the other topics make up the smaller remaining portion of the document. Conversely, lower numbers reflect a dominant topic making up a smaller portion of the document, leaving more room for other topics to be part of that document. For example, documents dealing with the topic Fishing gear are on average allocating 45% of their content to their own topic, leaving 55% of the remaining content to other topics (for example, 9% to the topic Models and between 4% and 6% to each of the topics Fisheries management, Acoustics and swimming, Stock assessment, and Southern Hemisphere). Furthermore, the remaining average topic proportions, shown at the top (i.e. columns), are sorted from high to low and reflect which topics more frequently, or to a higher extent, co-occur with other topics. For example, given any document with a dominant topic, the remaining topic proportion deals more often with the topics Models, Stock assessment, Fish reproduction or Growth experiments (i.e. these three topics have the highest co-occurrence), and less often with the topics Salmonids, Genetics or Salmon (i.e. these three topics have the lowest co-occurrence).

Documents dealing with the most prevalent topic in the corpus directly relating to fisheries, that is Fisheries management, allocate to this topic almost 60% of their content, about 13% to Conservation, and between 3% and 5% to topics such as (in descending order): Fishing gear, Models and Stock assessment. Documents dealing with the most prevalent topic in the rest of the corpus, that is Primary production, allocate to this topic 55% of their content, and between 3% and 6% to topics such as (in descending order): Non-fish species, Biochemistry, Models, Diet, Southern Hemisphere, and Growth experiments.

Co-occurrence of such topics might be something natural. However, it would be interesting to consider whether other mixtures of topics would bring novel, and possibly also innovative, insight into fisheries science. Such an insight is highly needed in order to achieve sustainable fisheries exploitation and implement the fisheries-related actions of the international ocean governance objectives (European Commission, 2016).

3.3 | Topical trends over time and topic prevalence

To gain insight into the temporal changes of topics, we display the topical trend values and topic prevalence in Figure 5. The left-hand side displays the fitted increase (hot topics) or decrease (cold topics) in percentage points for different time intervals and represents the change in composite topic-year proportions within a certain time frame. Additionally, we display the average composite topic-year proportions for every topic on the right-hand side, referred to as topic prevalence. Individual trend lines for the 25 broad fisheries topics, as well as the topical trends and prevalence for the Fisheries management subtopics, can be found in the supplementary material—Figures S1–S3.

With regard to the groups of topics described in Table 2, during the entire period of 26 years, taken together, the 11 topics referring to aquatic organisms biology themes made up 40% of the corpus; the groups of topics on the themes of specific species, modelling, management and habitats, between 11% and 14% each; the two topics on the theme of geographical regions, 7% together; and the fishing gear topic, almost 5%. The topics Models, Primary production and Fisheries management are the most prevalent in the corpus during the entire period of 26 years, with approximately 7%, 6.5% and 6%, respectively, from the total number of documents. The topics Salmon, Salmonids and Morphology are the least prevalent, with only around 2% of the corpus accounting for each. The topic Models has been among the top three most prevalent topics since 1995. Primary production was the most prevalent topic in the first three time periods we analysed (from 1990 to 2005). Topics directly relating to fisheries, such as Fisheries management and Stock assessment, were among the top 3 of most prevalent topics in the period 2005–10 (both topics), and 2010–16 (only Fisheries management). The topic Conservation joined this top during the last 6 years.

Taken together, topics that could be grouped as relating to freshwater fisheries (i.e. Freshwater, Tracking and movement, Salmonids, Salmon, Estuaries, Assemblages) account for one-sixth of the corpus. These results, corroborated with the declining interest in these topics in the last 16 years (see Figure 5), confirm the findings of Jarić et al. (2012) that indicate a general decline in the frequency of studies focused on freshwater habitats.

The top four hottest topics of the last 26 years (overall column) are (in descending order): Fisheries management, Conservation, Models and Fishing gear (with Stock assessment, the third topic directly relating to fisheries, in 7th place). The interest in models has also been confirmed by Jarić et al. (2012), and this, in addition to the prevalence of the modelling topics, which was described above, provides empirical evidence for the fact that modelling is one of the most important research methods in fisheries science (Angelini & Moloney, 2007). The topic Fisheries management, the third most prevalent topic in the corpus, remained among the top three hottest topics during the last 11 years, while it was the coldest topic in the first half on the 1990s.

The configuration of top three hottest and top three coldest topics has fluctuated during the five time periods we have analysed. However, the topic Primary production has constantly been ranked among the coldest topics (with the topic Biochemistry also being in this top in the period 1990–2005). Overall the 26-year period, it is interesting to note that Primary production was the most prevalent topic, but at the same time had the steepest decline in interest. The topics Models and Conservation have been a constant occurrence in the top three hottest topics since the turn of the Century, whereas the topic Fisheries management only joined this top category in 2005. Another topic directly relating to fisheries, Fishing gear, was part of this top in the period 2000–05.

The increased prevalence in the corpus of the topic Fisheries gear and the constant interest shown in this topic, together with the constant prevalence in the corpus of the topic Stock assessment and the constant interest shown in this topic, indicate that fishing gear and stock assessment have been the central elements of fisheries science in the last 26 years. The increasing prevalence and interest in the topic Fisheries management might indicate the strengthening of the connection between fisheries science and management processes, in the light of the growing concern about the status of fish stocks worldwide. Stock assessments provide a scientific and quantitative basis to the process of developing and implementing a management plan (Hoggarth, Mees, & O'Neill, 2005) and, as such, are indispensable to management processes (Hoggarth et al., 2006). The interest in the topics Growth and Reproduction seems to be most stable among all the 25 topics when looking at the entire period of 26 years, even though the prevalence of these topics is rather small (around 3%). The constant interest in the latter can be explained by the importance of fisheries reproduction for fisheries assessment and management (Jakobsen, Fogarty, Megrey, & Moksness, 2016).

3.4 | Topical trends over journals

Although many journals included in our analysis overlap to some extent in their content, it is possible to identify journals that seem specialized in specific topics (Figure 6). For example, almost one-fifth of the publications that appeared in the journal *Fish and Fisheries* relate to the topic Fisheries management, whereas another approximatively one-fifth is related to the topic Conservation. Among the topics directly relating to fisheries, the journals *Marine Policy* and *Marine Resource Economics* are highly specialized in the

	Topic trends (percentage points)						Topic prevalence (percent)					
	1990–1995	1995–2000	2000–2005	2005–2010	2010–2016	overall	1990–1995	1995–2000	2000–2005	2005–2010	2010–2016	overall
Fisheries management (25)	−2.04	0.57	0.02	3.01	5.05	5.20	5.00	3.88	4.17	5.58	9.10	6.13
Conservation (1)	0.19	1.08	1.15	1.63	3.34	5.10	3.12	3.19	3.81	4.99	7.05	5.06
Models (21)	0.23	1.04	0.95	1.30	0.46	4.81	4.17	6.37	6.68	7.37	8.46	7.20
Fishing gear (19)	1.03	−0.74	0.95	0.32	0.31	3.67	2.38	3.73	4.16	4.88	5.38	4.51
Tracking and movement (18)	−0.07	1.83	0.70	−0.12	−0.10	2.87	1.35	2.73	3.47	3.56	3.82	3.34
Assemblages (14)	0.41	2.69	0.20	0.03	−0.42	2.57	1.69	4.01	4.46	4.48	4.17	4.07
Stock assessment (16)	−0.04	1.81	−0.64	−0.08	−0.18	1.94	3.82	5.59	5.73	5.46	5.79	5.45
Salmon (3)	−0.07	1.31	0.48	−0.01	−0.24	1.59	0.51	1.42	1.69	1.74	1.96	1.64
Freshwater (8)	−0.36	2.68	0.33	−0.93	−0.40	1.50	1.10	3.62	3.85	3.35	2.91	3.16
Genetics (13)	0.12	0.03	0.23	0.51	−1.17	1.13	2.73	3.12	3.05	3.91	3.56	3.39
Southern Hemisphere (11)	0.81	−0.11	−0.09	0.21	−0.76	0.63	4.21	4.38	4.68	4.72	4.60	4.54
Estuaries (24)	−0.47	−0.57	0.27	0.30	0.11	0.62	3.03	2.46	2.69	3.19	3.35	3.03
Corals (6)	1.17	−1.59	0.71	0.13	−0.63	0.31	5.19	4.10	4.89	5.43	4.83	4.89
Salmonids (22)	−0.09	1.46	−0.10	−0.52	−0.28	0.30	1.14	2.20	2.35	1.90	1.65	1.90
Growth (17)	−0.40	0.59	0.01	−0.47	−0.23	0.25	2.61	3.05	3.27	3.34	2.81	3.06
Reproduction (4)	0.29	−0.30	0.24	−0.18	−0.71	−0.12	3.07	2.80	3.20	3.15	2.85	2.99
Acoustics and swimming (23)	0.39	−0.13	0.22	−0.19	−0.18	−0.60	4.11	4.78	4.70	4.49	3.70	4.31
North Atlantic (10)	−0.13	−0.93	0.06	−0.51	−0.10	−0.95	3.45	3.06	2.80	2.80	2.64	2.84
Diet (9)	0.71	−0.26	−0.75	−0.61	−0.50	−1.46	4.72	4.34	4.05	3.77	3.49	3.90
Growth experiments (15)	0.49	−0.24	0.02	−0.30	−0.36	−1.72	6.28	6.06	6.22	5.36	4.94	5.62
Morphology (2)	−0.45	−0.89	0.11	−0.35	−0.63	−1.77	3.30	2.13	2.08	2.12	1.64	2.05
Development (12)	−0.05	−0.23	−0.82	−0.18	−0.44	−1.90	3.91	3.57	3.17	2.88	2.36	2.96
Non-fish species (5)	−0.44	−1.44	−0.30	−0.82	−0.36	−3.79	6.37	4.65	3.98	3.84	3.06	3.96
Biochemistry (7)	−0.77	−2.29	−0.86	−0.74	−0.75	−6.30	7.42	4.93	3.91	3.04	2.15	3.57
Primary production (20)	−0.46	−5.37	−3.10	−1.43	−0.83	−13.88	15.30	9.82	6.92	4.67	3.72	6.45

FIGURE 5 Topical trends and prevalence for all 25 uncovered fisheries topics displayed as a heat map. The topical trends (left) show the increasing/hot (red) and decreasing/cold (blue) topics for different time intervals. The value represents the fitted (via linear regression) change in percentage points within each time interval. The topic prevalence (right) shows the average cumulative topic proportion in percentage for different time intervals for each of the 25 fisheries topics. It shows how present a topic is within a certain time interval given all the scientific output within that time interval. Individual trend lines per year can be found in the supplementary material

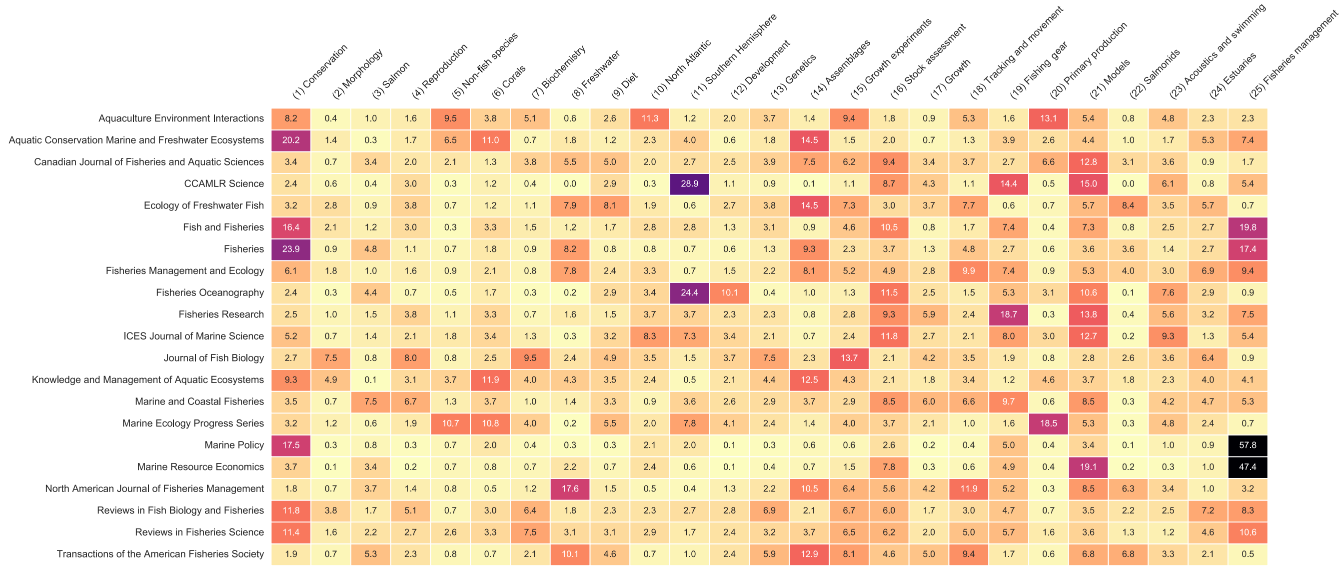

FIGURE 6 Topical distribution over journals displayed as a heat map. For each included journal (left), the coverage of topics (top) in percentages is displayed (percentage values in cells, row total = 100%). The heat map displays which journals publish which topics and in what proportions. See Table 1 for journal year coverage, as this differs between journals (e.g. 1990–2016, 1997–2016, 2000–16)

topic of Fisheries management, and the journals *Fisheries research* and *CCAMLR Science*, in the topic Fishing gear. It appears that no journal is highly specialized in Stock assessment, with this topic being addressed by almost all the journals. The top three journals publishing this topic are *ICES Journal of Marine Science*, *Fisheries Oceanography*, and *Fish and Fisheries*, with 10%–11% of the publication space of each of these journals covering this topic.

3.5 | Validation of results

We validated the output of the LDA model (including its labelling) by comparing the hot/cold LDA topics in the period 2000–09 with the hot/cold words used in publication titles in the same period identified by the bibliometric study of Jarić et al. (2012). Some of the words that Jarić et al. (2012) identified as having the greatest increase in frequency can be directly linked to some of the hot topics identified by our analysis, for example *by-catch*, *longline*, *seabird* in the bibliometric study relate to the topic Fishing gear (19) identified by our study; *genetic* relates to the topic Genetics (13). Likewise, some of the words with the largest decrease in frequency can be directly linked to some of the cold topics identified by the LDA analysis over the same period, for example *Atlantic* in the bibliometric study relates to the topic North Atlantic identified by our study; *growth*, to the topics Growth (17) and Growth experiments (15); *recruitment*, to the topic Stock assessment (14); *feeding*, to the topic Diet (9).

4 | CONCLUSION AND RECOMMENDATIONS

From the analysis of more than 46,500 full-text articles published in 21 top fisheries journals it is apparent that, during the last 26 years, the research focus of fisheries science has been predominately on the natural dimension of the fisheries system, with 22 out of 25 topics referring to this dimension. While the natural dimension of fisheries was split into various aspects covering topics from specific species to fish catch technology, the human dimension was explicitly expressed only through one, albeit the hottest topic in the data set and the second most prevalent: Fisheries management. Although there is undoubtedly some scientific production addressing various aspects of the human dimension of fisheries, it could be that the narrative used to describe the human dimension is not explicit enough to be captured by word co-occurrence, or that the human dimension is not prevalent enough to be recognized as a general topic or specific subtopic by the LDA model. Additionally, it might be that the scientific production on the human dimension is published in journals other than those specialized in fisheries or other types of outlets, such as books and book chapters. We could advance various hypotheses as to why this might be the case [e.g. most fishery scientists are biological/natural sciences trained and oriented (Link, 2010); those who are non-natural sciences trained and oriented tend to publish in outlets that foster recognition from within their own scientific communities, such as books, rather than journals], but this exercise would be outside the scope of this study. Instead, we want to emphasize two important recommendations: (i) diversification of the scientific focus so that it covers more of the complexity of fisheries, especially the human dimension (funding bodies play a crucial role here by setting the research agenda) and (ii) publication of fisheries research in outlets more likely to reach the intended audience (i.e. top interdisciplinary journals or specialized fisheries journals, if the objective of publishing is to contribute to fisheries sustainability by reaching fisheries policy- and decision-makers). A lack of interdisciplinary synthesis in research is one of the major factors in fisheries collapses (Smith & Link, 2005). Thus, more integrative research and research focused on the under-represented topics might provide insight into the fine mechanisms of fisheries as a SECAS, and, thus, a critical input for developing successful fisheries management approaches.

ACKNOWLEDGEMENTS

This research was funded by the project SAF21—Social Science Aspects of Fisheries for the 21st Century (project financed under the EU Horizon 2020 Marie Skłodowska-Curie [MSC] ITN-ETN Program; project number: 642080).

ORCID

Shaheen Syed http://orcid.org/0000-0001-5462-874X
Melania Borit http://orcid.org/0000-0002-1305-8581

REFERENCES

Aksnes, D. W., & Browman, H. I. (2016). An overview of global research effort in fisheries science. *ICES Journal of Marine Science: Journal Du Conseil*, *73*(4), 1004–1011. https://doi.org/10.1093/icesjms/fsv248

Angelini, R., & Moloney, C. L. (2007). Fisheries, ecology and modelling: An historical perspective. *Pan-American Journal of Aquatic Sciences*, *2*(2), 75–85.

Belgrano, A., & Fowler, C. W. (2013). How fisheries affect evolution. *Science*, *342*(6163), 1176–1177. https://doi.org/10.1126/science.1245490

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., & Jordan, M. I. (2006). Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, *1*(1), 121–143. https://doi.org/10.1214/06-BA104

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning – ICML '06* (pp. 113–120). New York, NY: ACM Press. https://doi.org/10.1145/1143844.1143859

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, *1*(1), 17–35. https://doi.org/10.1214/07-AOAS114

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Brooks, C. M., Crowder, L. B., Curran, L. M., Dunbar, R. B., Ainley, D. G., Dodds, K. J., … Sumaila, U. R. (2016). Science-based management in decline in the Southern Ocean. *Science*, *354*(6309), 185–187. https://doi.org/10.1126/science.aah4119

Campbell, L., & Cornwell, M. (2008). Human dimensions of bycatch reduction technology: Current assumptions and directions for future research. *Endangered Species Research*, *5*, 325–334. https://doi.org/10.3354/esr00172

Chang, J., & Blei, D. M. (2010). Hierarchical relational models for document networks. *The Annals of Applied Statistics*, *4*(1), 124–150. https://doi.org/10.1214/09-AOAS309

Chang, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *NIPS'09 Proceedings of the 22nd International Conference on Neural Information Processing Systems* (pp. 288–296). Vancouver, BC: Curran Associates Inc.

Charles, A. (2001). *Sustainable fishery systems*. Oxford, UK: Blackwell Science Ltd. https://doi.org/10.1002/9780470698785

Chuang, J., Ramage, D., Manning, C., & Heer, J. (2012). Interpretation and trust. In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems – CHI'12* (pp. 443–452). Austin, TX: ACM Press. https://doi.org/10.1145/2207676.2207738

De Young, C., Charles, A., & Hjort, A. (2008). *Human dimensions of the ecosystem approach to fisheries: an overview of context, concepts, tools and methods*. FAO Fisheries Technical Paper 489. Rome, Italy.

Debortoli, S., Müller, O., Junglas, I., & vom Brocke, J. (2016). Text mining for information systems researchers: An annotated topic modeling tutorial. *Communications of the Association for Information Systems*, *39*, 110–135. https://doi.org/10.17705/1cais.03907 https://doi.org/10.17705/1cais

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, *41*(6), 570–606. https://doi.org/10.1016/j.poetic.2013.08.004

Doyle, G., & Elkan, C. (2009). Accounting for burstiness in topic models. In *Proceedings of the 26th Annual International Conference on Machine Learning – ICML'09* (pp. 281–288). Montreal, QC: ACM Press. https://doi.org/10.1145/1553374.1553410

Erosheva, E., Fienberg, S., & Lafferty, J. (2004). Mixed-membership models of scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(Suppl 1), 5220–5227. https://doi.org/10.1073/pnas.0307760101

European Commission (2016). *International ocean governance: An agenda for the future of our oceans (SWD(2016) 352 final)*. Brussels, Belgium. Retrieved from https://ec.europa.eu/maritimeaffairs/sites/maritimeaffairs/files/swd-2016-352_en.pdf

Evangelopoulos, N., Zhang, X., & Prybutok, V. R. (2012). Latent semantic analysis: Five methodological recommendations. *European Journal of Information Systems*, *21*(1), 70–86. https://doi.org/10.1057/ejis.2010.61

Fergus, R., Fei-Fei, L., Perona, P., & Zisserman, A. (2005). Learning object categories from Google's image search. In *Tenth IEEE International Conference on Computer Vision (ICCV'05)* (pp. 1816–1823). Beijing, China: IEEE. https://doi.org/10.1109/iccv.2005.142

Fulton, E. A., Smith, A. D. M., Smith, D. C., & van Putten, I. E. (2011). Human behaviour: The key source of uncertainty in fisheries management. *Fish and Fisheries*, *12*(1), 2–17. https://doi.org/10.1111/j.1467-2979.2010.00371.x

Gatti, C. J., Brooks, J. D., & Nurre, S. G. (2015). A historical analysis of the field of OR/MS using topic models. *arXiv.org, stat.ML*. Retrieved from http://arxiv.org/abs/1510.05154

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(Supplement 1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Hall, D., Jurafsky, D., & Manning, C. D. (2008). Studying the history of ideas using topic models. In *EMNLP'08 Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 363–371). Honolulu, Hawaii: Association for Computational Linguistics. https://doi.org/10.3115/1613715

Harris, Z. S. (1954). Distributional structure. *WORD*, *10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hill, H., & Lackups, M. (2010). Journal publication trends regarding cetaceans found in both wild and captive environments: What do we study and where do we publish? *International Journal of Comparative Psychology*, *23*(3), 414–534.

Hoffman, M. D., Blei, D. M., & Bach, F. (2010). Online learning for latent Dirichlet allocation. In *NIPS'10 Proceedings of the 23rd International Conference on Neural Information Processing Systems – Volume 1* (pp. 856–864). Vancouver, BC: Curran Associates Inc.

Hoggarth, D. D., Abeyasekera, S., Arthur, R. I., Beddington, J. R., Burn, R. W., Halls, A. S., … Welcomme, R. L. (2006). *Stock assessment for* *fishery management : A framework guide to the stock assessment tools of the Fisheries Management and Science Programme*. FAO Fisheries Technical Paper. Rome, Italy: FAO.

Hoggarth, D., Mees, C., & O'Neill, C. (2005). *A guide to fisheries stock assessment using the FMSP tools*. London, UK: Marine Resource Assessment Group.

Huang, J. (2005). *Maximum likelihood estimation of Dirichlet distribution parameters. Carnegie Mellon University Technical Report*. Pittsburgh, PA.

ICES (2016). *Report of the SIHD survey of the current state of 'human dimension' in some ICES groups*. Retrieved from http://www.ices.dk/sites/pub/PublicationReports/Committeereport/SCICOM/2016/SIHD2016.pdf

Jakobsen, T., Fogarty, M. J., Megrey, B. A., & Moksness, E. (Eds.). (2016). *Fish reproductive biology*. Oxford, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118752739

Jarić, I., Cvijanović, G., Knežević-Jarić, J., & Lenhardt, M. (2012). Trends in fisheries science from 2000 to 2009: A bibliometric study. *Reviews in Fisheries Science*, *20*(2), 70–79. https://doi.org/10.1080/10641262.2012.659775

Jennings, S., Kaiser, M. J. J., & Reynolds, J. D. D. (2009). *Marine fisheries ecology*. Oxford, UK: Blackwell Science Ltd.

Johnson, A. E., Cinner, J. E., Hardt, M. J., Jacquet, J., McClanahan, T. R., & Sanchirico, J. N. (2013). Trends, current understanding and future research priorities for artisanal coral reef fisheries research. *Fish and Fisheries*, *14*(3), 281–292. https://doi.org/10.1111/j.1467-2979.2012.00468.x

Kim, S., Narayanan, S., & Sundaram, S. (2009). Acoustic topic model for audio information retrieval. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (pp. 37–40). New Paltz, NY: IEEE. https://doi.org/10.1109/aspaa.2009.5346483

King, M. (2007). *Fisheries biology, assessment and management*. Oxford, UK: Blackwell Publishing Ltd. https://doi.org/10.1002/9781118688038

Kumaresan, R., Ezhilrani, R., Vinitha, K., Sivaraman, P., & Jayaraman, R. (2014). Research trends in fish stock assessment during 1999–2013: A scientometrics study. *International Journal of Library and Information Science*, *3*(2), 24–36.

Lau, J. H., Grieser, K., Newman, D., & Baldwin, T. (2011). Automatic labelling of topic models. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (pp. 1536–1545). Portland, OR: Association for Computational Linguistics.

Lennox, R. J., Alós, J., Arlinghaus, R., Horodysky, A., Klefoth, T., Monk, C. T., & Cooke, S. J. (2017). What makes fish vulnerable to capture by hooks? A conceptual framework and a review of key determinants. *Fish and Fisheries*, *18*, 986–1010. https://doi.org/10.1111/faf.12219

Levin, S., Xepapadeas, T., Crépin, A.-S., Norberg, J., de Zeeuw, A., Folke, C., … Walker, B. (2013). Social-ecological systems as complex adaptive systems: Modeling and policy implications. *Environment and Development Economics*, *18*(2), 111–132. https://doi.org/10.1017/S1355770X12000460

Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media*, *57*(1), 34–52. https://doi.org/10.1080/08838151.2012.761702

Link, J. (2010). *Ecosystem-based fisheries management: Confronting tradeoffs*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511667091

Mather, M. E., Parrish, D. L., & Dettmers, J. M. (2008). Mapping the changing landscape of fish-related journals: Setting a course for successful communication of scientific information. *Fisheries*, *33*(9), 444–453. https://doi.org/10.1577/1548-8446-33.9.444

Mehran, R., Oyama, A., & Shah, M. (2009). Abnormal crowd behavior detection using social force model. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 935–942). Miami, FL: IEEE. https://doi.org/10.1109/cvpr.2009.5206641

Mohr, J. W., & Bogdanov, P. (2013). Introduction—Topic models: What they are and why they matter. *Poetics*, *41*(6), 545–569. https://doi.org/10.1016/j.poetic.2013.10.001

Molina, J. M., & Cooke, S. J. (2012). Trends in shark bycatch research: Current status and research needs. *Reviews in Fish Biology and Fisheries*, *22*(3), 719–737. https://doi.org/10.1007/s11160-012-9269-3

Neff, M. W., & Corley, E. A. (2009). 35 years and 160,000 articles: A bibliometric exploration of the evolution of ecology. *Scientometrics*, *80*(3), 657–682. https://doi.org/10.1007/s11192-008-2099-3

Newman, D., Asuncion, A., Smyth, P., & Welling, M. (2007). Distributed inference for latent Dirichlet allocation. In *Proceedings of the 20th International Conference on Neural Information Processing Systems* (pp. 1081–1088). Vancouver, BC: Curran Associates Inc.

Österblom, H., Merrie, A., Metian, M., Boonstra, W. J., Blenckner, T., Watson, J. R., … Folke, C. (2013). Modeling social–ecological scenarios in marine systems. *BioScience*, *63*(9), 735–744. https://doi.org/10.1093/bioscience/63.9.735

Ostrom, E. (2009). A general framework for analyzing sustainability of social-ecological systems. *Science*, *325*(5939), 419–422. https://doi.org/10.1126/science.1172133

Porteous, I., Newman, D., Ihler, A., Asuncion, A., Smyth, P., & Welling, M. (2008). Fast collapsed gibbs sampling for latent Dirichlet allocation. In *KDD'08 Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 569–577). Las Vegas, NV: ACM Press. https://doi.org/10.1145/1401890.1401960

Purcell, S. W., & Pomeroy, R. S. (2015). Driving small-scale fisheries in developing countries. *Frontiers in Marine Science*, *2*, 44. https://doi.org/10.3389/fmars.2015.00044

Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50). Valletta, Malta: European Language Resources Association (ELRA). https://doi.org/10.13140/2.1.2393.1847

Reisinger, J., Waters, A., Silverthorn, B., & Mooney, R. J. (2010). Spherical topic models. In *Proceedings of the 27th International Conference on Machine Learning* (pp. 903–910). Haifa, Israel: International Machine Learning Society (IMLS).

Rhody, L. (2012). Topic modeling and figurative language. *Journal of Digital Humanities*, *2*(1), 19–35.

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *WSDM'15 Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (pp. 399–408). Shanghai, China: ACM Press. https://doi.org/10.1145/2684822.2685324

Rose, N. A., Janiger, D., Parsons, E. C. M., & Stachowitsch, M. (2011). Shifting baselines in scientific publications: A case study using cetacean research. *Marine Policy*, *35*(4), 477–482. https://doi.org/10.1016/j.marpol.2010.11.002

Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004). The author-topic model for authors and documents. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence* (pp. 487–494). Banff, AB: AUAI Press Arlington.

Rusch, T., Hofmarcher, P., Hatzinger, R., & Hornik, K. (2013). Model trees with topic model preprocessing: An approach for data journalism illustrated with the WikiLeaks Afghanistan war logs. *The Annals of Applied Statistics*, *7*(2), 613–639. https://doi.org/10.1214/12-AOAS618

Scott, J. G., & Baldridge, J. (2013). A recursive estimate for the predictive likelihood in a topic model. *Journal of Machine Learning Research*, *31*, 527–535.

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (pp. 63–70). Baltimore, MD: Association for Computational Linguistics. https://doi.org/10.3115/v1/w14-31

Smith, T. D., & Link, J. S. (2005). Autopsy your dead…and living: A proposal for fisheries science, fisheries management and fisheries. *Fish and Fisheries*, *6*(1), 73–87. https://doi.org/10.1111/j.1467-2679.2005.00176.x

Sowman, M. (2011). New perspectives in small-scale fisheries management: Challenges and prospects for implementation in South Africa. *African Journal of Marine Science*, *33*(2), 297–311. https://doi.org/10.2989/1814232X.2011.602875

Spalding, A. K., Biedenweg, K., Hettinger, A., & Nelson, M. P. (2017). Demystifying the human dimension of ecological research. *Frontiers in Ecology and the Environment*, *15*(3), 119–119. https://doi.org/10.1002/fee.1476

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (Vol. 427, pp. 424–440). Mahwah, NJ: Lawrence Erlbaum Associates.

Stone-Jovicich, S. (2015). Probing the interfaces between the social sciences and social-ecological resilience: Insights from integrative and hybrid perspectives in the social sciences. *Ecology and Society*, *20*(2), art25. https://doi.org/10.5751/ES-07347-200225

Sun, L., & Yin, Y. (2017). Discovering themes and trends in transportation research using topic modeling. *Transportation Research Part C: Emerging Technologies*, *77*(June), 49–66. https://doi.org/10.1016/j.trc.2017.01.013

Syed, S., & Spruit, M. (2017). Full-text or abstract? Examining topic coherence scores using latent Dirichlet allocation. In *4th IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 165–174). Tokyo, Japan: IEEE. https://doi.org/10.1109/dsaa.2017.61

Syed, S., & Spruit, M. (2018). Selecting priors for latent Dirichlet allocation. In *The 12th IEEE International Conference on Semantic Computing* (pp. 194–202). Laguna Hills, CA: IEEE. https://doi.org/10.1109/icsc.2018.00035

Syed, S., & Weber, C. T. (2018). Using machine learning to uncover latent research topics in fishery models. *Reviews in Fisheries Science & Aquaculture*, 1–18. https://doi.org/10.1080/23308249.2017.1416331

Symes, D., & Hoefnagel, E. (2010). Fisheries policy, research and the social sciences in Europe: Challenges for the 21st century. *Marine Policy*, *34*(2), 268–275. https://doi.org/10.1016/j.marpol.2009.07.006

Teh, Y. W., Newman, D., Welling, M., & Neaman, D. (2006). A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. In *NIPS'06 Proceedings of the 19th International Conference on Neural Information Processing Systems* (pp. 1353–1360). Vancouver, BC and Cambridge, MA: MIT Press.

Wallach, H. M. (2006). Topic modeling: Beyond bag-of-words. In *Proceedings of the 23rd International Conference on Machine Learning – ICML'06* (pp. 977–984). Pittsburgh, PA: ACM Press. https://doi.org/10.1145/1143844.1143967

Wallach, H. M., Mimno, D., & Mccallum, A. (2009). Rethinking LDA : Why priors matter. In *NIPS'09 Proceedings of the 22nd International Conference on Neural Information Processing Systems* (pp. 1973–1981). Vancouver, BC: Curran Associates Inc.

Wang, C., & Blei, D. M. (2009). Decoupling sparsity and smoothness in the discrete hierarchical Dirichlet process. In *NIPS'09 Proceedings of the 22nd International Conference on Neural Information Processing Systems* (pp. 1982–1989). Vancouver, BC: Curran Associates Inc.

Wang, X., & McCallum, A. (2006). Topics over time: A non-Markov continuous-time model of topical trends. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD'06* (pp. 424–433). Philadelphia, PA: ACM Press. https://doi.org/10.1145/1150402.1150450

Wang, C., Paisley, J., & Blei, D. M. (2011). Online variational inference for the hierarchical Dirichlet process. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)* (Vol. 15, pp. 752–760). Fort Lauderdale, FL: PMLR.

Westgate, M. J., Barton, P. S., Pierson, J. C., & Lindenmayer, D. B. (2015). Text analysis tools for identification of emerging topics and research gaps in conservation science. *Conservation Biology*, *29*(6), 1606–1614. https://doi.org/10.1111/cobi.12605

Whye Teh, Y., Jordan, M. I., Beal, M. J., & Blei, D. M. (2004). Sharing clusters among related groups: Hierarchical Dirichlet processes. In *NIPS'04 Proceedings of the 17th International Conference on Neural Information Processing Systems* (pp. 1385–1392). Vancouver, BC and Cambridge, MA: MIT Press.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Syed S, Borit M, Spruit M. Narrow lenses for capturing the complexity of fisheries: A topic analysis of fisheries science from 1990 to 2016. *Fish Fish*. 2018;19:643–661. https://doi.org/10.1111/faf.12280