MJFAS
Malaysian Journal of Fundamental and Applied Sciences

REVIEW ARTICLE

Research Trends on Functional Data Analysis Using Scopus Database: A Bibliometric Analysis

Jamaludin Suhaila[a,b]*, Muhammad Fauzee Hamdan[b]

[a]UTM Centre for Industrial and Applied Mathematics (UTM-CIAM), Ibnu Sina Institute for Scientific and Industrial Research, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia; [b]Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

Abstract Functional data analysis (FDA) has received significant attention from researchers due to its flexibility and diverse applications in various fields. FDA provides a comprehensive framework for analysing and extracting information from complex and high-dimensional datasets, enabling researchers to obtain insights into the underlying processes, improve modelling, and make accurate predictions. Therefore, understanding the FDA topic and its features and tools, as well as identifying the collaborative networks, are crucial for the development of its research areas. The objective of the present bibliometric study is to analyse the global research trend in FDA areas based on publication outputs, authorships, co-authorships, affiliated countries, and the co-occurrence of author keywords, which will enable researchers to assess the existing knowledge environment, future trends, potential research gaps, and collaboration opportunities. The publications from the year 1989 to 2021 were retrieved from the Scopus database, resulting in 1712 articles in journals after screening. Results have shown that articles published in the Journal of the American Statistical Association received the highest citations. Nearly 43% of the published articles were contributed by the leading authors from the USA, followed by China (11.5%) and Spain (9.4%). According to the QS World University Ranking 2021, eight of the top 20 productive institutions were ranked among the top 100 best universities. The findings indicated that researchers had intensively developed and applied FDA tools and features, such as smoothing, principal component analysis, regression, and clustering, in various domains. In addition, the expansion of FDA tools could be seen based on the recent progress in author keywords. New keywords, including function-on-function regression, function-on-scalar regression, scalar-on-function regression, outlier detection, structural health monitoring, and COVID-19, have arisen recently. Due to public concern about emerging diseases, future FDA work is expected to rise, particularly in the health sciences and biomedical fields.

Keywords: Bibliometric analysis, Scopus database, Functional data analysis, VOSviewer, Author Keyword Co-occurrences.

***For correspondence:**
suhailasj@utm.my

Received: 23 Oct. 2022
Accepted: 27 July 2023

Introduction

Data analysis and modelling have become difficult, especially when dealing with complex and high-dimensional data. With the advancement of technology, a popular statistical tool for analysing data over a curve, surface, or continuum known as functional data analysis (FDA) is increasingly being employed in various scientific domains, including biomedical, public health, biology, environmental sciences, climatological, hydrological, and demographic research. Functional data can convert discrete observations at any discrete time interval into a continuous smoothing function that can be considered and analysed as a single entity [1, 2]. The FDA's flexibility, such as providing the additional information from smoothing functions and the lack of concerns about correlations between repeated measurements, makes the method extremely demanding [3, 4].

Although there has been a growing interest in FDA applications, limited research has been dedicated to measuring and analysing scientific publications on a global scale. Ullah and Finch [4] conducted a systematic review of FDA application studies between 1995 and 2010. They found relevant papers using 11 internet databases, including Academic Search Premier, ScienceDirect, Springer Link, Cambridge Journals, PubMed, Sage Journals Online, and Web of Science. They mainly concentrated on applying FDA to the public health and biological fields in their research. Wang *et al*. [5], on the other hand, examined FDA techniques such as Functional Principal Component Analysis (FPCA), functional linear regression, functional data clustering and classification, and other nonlinear functional data methods. Their research, however, was limited to FDA concepts. Aneiros *et al*. [6, 7] addressed recent breakthroughs in functional data analysis and high-dimensional statistics. Despite increasing FDA articles published on techniques, tools, and applications, quantitative analysis of the most cited articles, prolific authors, organisational affiliations, and national and international collaborations are still minimal.

Identifying relevant studies, screening them, extracting the data from them, and synthesising them are all time-consuming steps in conducting a systematic literature review. However, bibliometric analysis makes use of computing methodologies and tools in order to automate the process of data gathering and analysis. Using this strategy can save a substantial amount of time and effort, which is especially helpful when working with extensive datasets. Bibliometric analysis is a statistical assessment of published scientific articles, proceedings, and scientific reports based on an academic literature database. It is used to determine the impact of publications and global research trends in a specific area. The bibliometric analysis aids in identifying the top and the most cited authors, top countries and institutions in research fields, highly cited publications and leading journals, and the most frequently used keywords within a particular research topic. Hence, bibliometric analysis is employed to identify the development of FDA research areas. The purpose of this paper can be summarised as follows: i) to search for the top productive journals in FDA and the citation score; ii) to identify the most prolific authors, countries, and their organizational affiliations; iii) to highlight the research interest and examine recent research keywords in FDA. This paper will benefit researchers and policymakers in understanding the research trend in FDA and identify potential future research.

Materials and Methods

Scopus was used as a database to find relevant documents for this study. Scopus covered the largest abstracts and literature citations for a wide range of subjects [8,9]. Scopus searches can be divided into basic and advanced categories. The basic search can be done within the article's search, including the title, abstracts, keywords, authors and affiliations, funding information, ISSN, DOI, and ORCID. Authors can also utilise the advanced search to narrow the scope of their search by using field codes, Boolean operators, and proximity operators.

Search Strategy

The process of data mining was conducted between 1 February 2022 and 8 February 2022, using a Scopus database. The central theme in this study focused on all research articles containing the word "functional data analysis" either in the title, abstract, or author keywords. In Scopus, quotation marks were used as the loose phrase, while the asterisk was used as a wild card. The earliest publication date is 1989, and the most recent is 2021. The query strings used for this search were given as TITLE-ABS("functional data analysis") OR AUTHKEY("functional data analysis") AND (LIMIT-TO(DOCTYPE,"ar")) AND (LIMIT-TO (SRCTYPE,"j")) AND (EXCLUDE (PUBYEAR, 2022)). The author was only concerned with the article-type journal, which resulted in 1725 documents. The list of article journals was filtered to guarantee that the analysis did not include any review articles since the focus was on research articles. The search phrases were expanded to include words like recent, progress, review, scientometric, and bibliometric, yielding 74 articles. The process of screening the abstract and full text of the articles took place, and only 13 articles were identified as review articles. EID, a Scopus unique article identifier, was used in the search strings to exclude these review articles, and finally, only 1712 articles were documented.

Based on 1712 documents, the publications were stratified and systematically assessed according to publication year, country, authors, document type, organisational affiliations, and subject area. For specific journals, bibliometric indicators such as the total citations, Cite Score, Scientific Journal Rankings (SJR), and Source Normalized Impact per Paper (SNIP) can be obtained, which can help in deciding the journal rankings. The Scopus Author ID can be used to determine the total publications of the authors, the number of citations, and the authors' H-index. The H-index represents the number of publications for which an author has been cited by other authors. It represents the productivity and impact of a particular scientist or scholar. In addition, data on a single country publication (SCP) were retrieved by limiting the search result to a specific country using the field code AFFILCOUNTRY.

Figure 1 explains the flow of gathering data for article publication. In addition, a sub-theme was also created to explore the output trends in FDA methodologies. Finally, specific terms were added to the previous search strings depending on the methods used by the FDA. The details on the search strings used in Scopus are provided in Table 1.

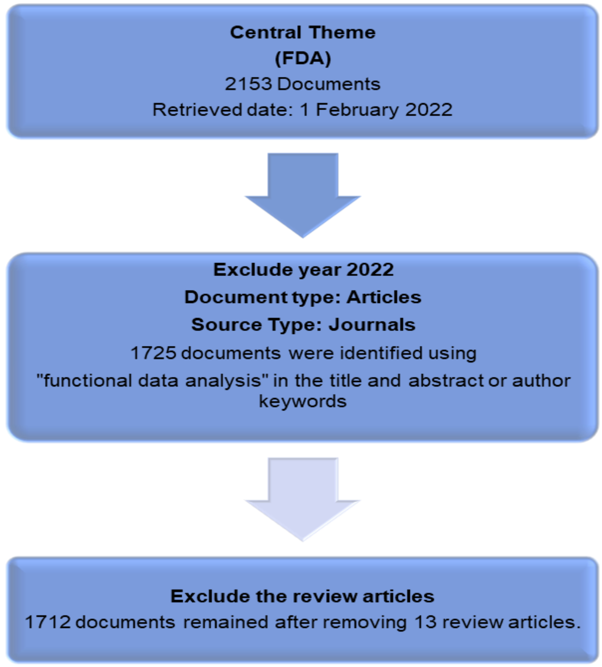

Figure 1. The flow process of gathering data publications for the central theme

Mapping through VOSViewer

VOSviewer is a software tool to construct and visualise bibliometric networks of scientific publications, journals, countries, authors, keywords, or terms. A bibliographic database file from Scopus was exported to VOSviewer to create network visualisation maps. In this study, our interest focused on the items that involved countries and author keywords. Items in these networks can be connected by co-authorship and co-occurrence. A link is defined as a connection or relation between two items, and each link has a strength represented by a positive numerical value. The higher the value, the stronger the link. The link strength (LS) in the case of co-occurrence analysis indicates the number of publications in which two keywords occur together, while the case of co-authorship represents the number of publications that two affiliated countries have co-authored. The total link strength (TLS) attribute indicates the total strength of the co-authorship links of a given country with other countries. A detailed explanation or interpretation of the VOSviewer can be found in the manual written by van Eck and Waltman [10].

Results and Discussion

Publications and Research Areas

A total of 1712 publications have been documented over the past 32 years. The first article was published in 1989 in the Journal of the American Statistical Association by Ramsay and Abrahamowicz [11]. After the first publication, there was no other publication record until 1994. The second FDA article was also written by Ramsay *et al*. [12]. Figure 2 shows that less than 20 publications were recorded until 2004. A strong research interest in the FDA started in 2009, with more than 50 publications. Since then, annual publications have increased steadily, and it is anticipated that they will continue to rise. In the year 2021, it recorded 240 published papers in the FDA. As of 2021, 14.3% (131 articles) of published articles are under the status of open access, which requires authors to pay a specific fee for the publication. With open access, researchers can download the published article freely.

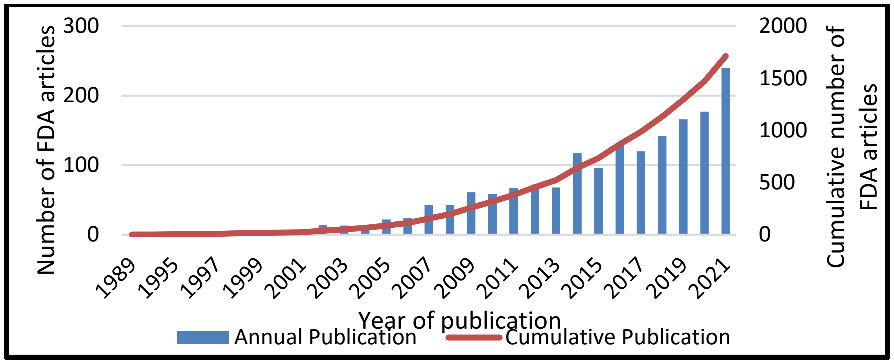

Figure 2. The annual and cumulative numbers of FDA articles indexed in Scopus from 1989 to 2021

Table 1. The search strategies and query strings used in Scopus.

Search	Step	String Search	Results
Any string related to "functional data analysis"	Insert the main keyword from 1989-2021	TITLE-ABS ("functional data analysis") OR AUTHKEY ("functional data analysis")	2153 documents
Any string related to "functional data analysis"	Limit to Article paper and journal exclude 2022	TITLE-ABS("functional data analysis") OR AUTHKEY("functional data analysis") AND (LIMIT-TO (DOCTYPE,"ar")) AND (LIMIT-TO (SRCTYPE,"j")) AND (EXCLUDE (PUBYEAR,2022))	1725 documents
Review articles	Find potentially review articles (74 papers) ●Screen the titles and abstracts to identify the real review articles ●Collect EID of the review articles (papers)	TITLE-ABS ("functional data analysis") OR AUTHKEY ("functional data analysis") AND (TITLE ("recent" OR pro gress OR review OR critical OR revisit O R advance OR development OR highlight OR perspective OR prospect OR trends OR bibliometric OR scientometric) OR (ABS (progress OR review OR bibliometric OR scientometric))) AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (EXCLUDE (PUBYEAR,2022))	74 documents
Exclude the review articles	Add additional phrase "AND NOT"	TITLE-ABS ("functional data analysis") OR AUTHKEY ("functional data analysis") AND NOT EID (2-s2.0-85085018836 OR 2-s2.0-85068329306 OR 2-s2.0-85058999433 OR 2-s2.0-85042619255 OR 2-s2.0-85052849404 OR 2-s2.0-84959387275 OR 2-s2.0-84949087732 OR 2-s2.0-84957612839 OR 2-s2.0-84928111503 OR 2-s2.0-84892512066 OR 2-s2.0-77952396887 OR 2-s2.0-66349121162 OR 2-s2.0-84884959104) AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (EXCLUDE (PUBYEAR , 2022))	1712 documents

FDA research areas have been intensively explored in several areas. The FDA's primary research areas are decision science and mathematics, as shown in Figure 3. The following subject areas cover the entirety of the publications: Mathematics (1047 articles), Decision Science (654 articles), Computer Science (273 articles), and other fields, including medicine, biology, and the environment, are all represented. The article was written in seven languages, with English being the most popular (1680:98%), followed by Chinese (21: 1.2%), Spain (7: 0.4%), and French (4: 0.2%). FDA has also published articles in German, Portuguese, and Russian, each featuring one article in each language.

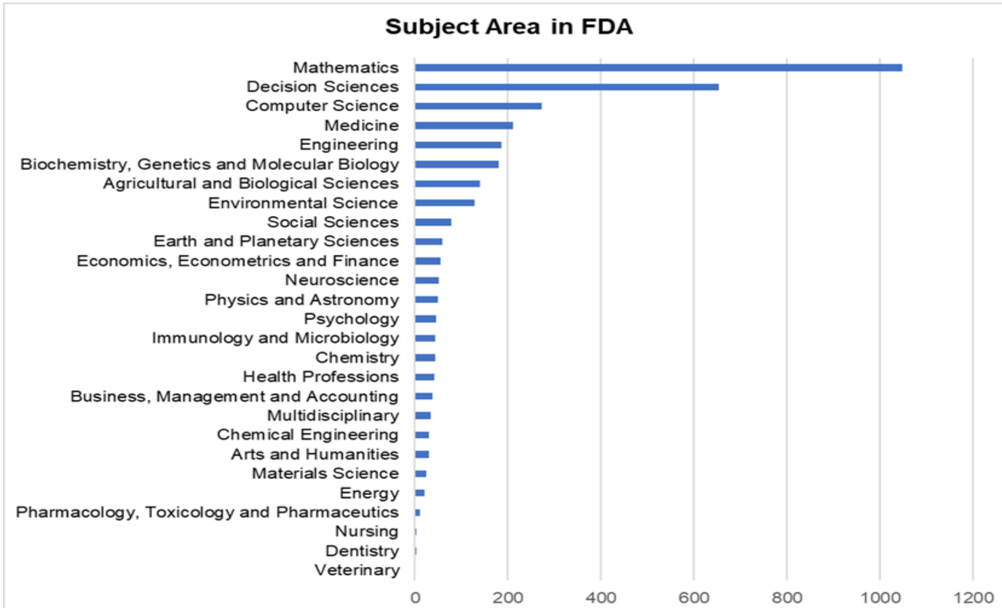

Figure 3. Documents by Subject Area in FDA

Top Productive Journals

Table 2 lists the most productive journals with more than 20 publications in FDA study areas. The Institute of Mathematical Statistics and Wiley-Blackwell have published three journals. Taylor & Francis and Elsevier each publish two journals. The Journal of the American Statistical Association was the most active journal, with 73 papers covering 4.3 percent of all publications, followed by Computational Statistics and Data Analysis (65, 3.8 %) and the Journal of Multivariate Analysis (52, 3.04%). With 2892 citations, the Journal of the American Statistical Association received the most. In addition, the article published by Yao *et al*. [13] in the journal in 2005 was the most cited article, with 748 citations. On the other hand, the Annals of Statistics obtained the second-highest number of citations, with 2362. With 2336 citations, Wiley-Blackwell of the Royal Statistical Society Series B Statistical Methodology was the third most cited journal. However, according to CiteScore (2020), this journal has the highest CiteScore, with a CiteScore of 8.

The CiteScore calculation is based on Scopus data. The CiteScore of a journal is the number of citations received in that year and the previous three years for documents published for four years divided by the total number of published documents in the journal during the same four-year period. According to the CiteScore (2020) report, only three journals had a CiteScore of 5 and above. The journal with the lowest CiteScore belonged to the Electronic Journal of Statistics, with a CiteScore of 1.9. This journal's ranking percentile is 60th, which is low compared to other journals.

Nevertheless, the value of CiteScore can sometimes assist researchers in making publication selection decisions for writers. CiteScore, on the other hand, cannot be used as a standalone metric because it is a metric without field normalisation and hence should not be compared across subject fields. The Journal Impact Factor (IF) is another frequently used metric to measure citations. The Journal Citation Reports publish the IF of the journal, which is determined from data obtained in the Web of Science database. CiteScore and IF are based on the same principles. Both rely on the number of citations compared to the number of articles published in a specific period. The only difference is the period considered in the calculations.

Table 2. The top productive journals in FDA research with their most cited article.

Journal	TP	TC	Cite Score 2020	Title of the most cited article	Times cited	Publisher
Journal of the American Statistical Association	73	2892	6.8	Functional data analysis for sparse longitudinal data [13]	748	Taylor and Francis
Computational Statistics and Data Analysis	65	1576	2.7	Robust forecasting of mortality and fertility rates: A functional data approach. [14]	309	Elsevier
Journal Of Multivariate Analysis	52	688	2.2	Functional canonical analysis for square integrable stochastic processes. [15]	77	Elsevier
Annals of Statistics	38	2362	5.8	Functional linear regression analysis for longitudinal data. [16]	392	Institute of Mathematical Statistics
Biometrics	34	982	3.0	Nonparametric mixed effects models for unequally sampled noisy curves. [17]	268	Wiley-Blackwell
Journal of the Royal Statistical Society Series B Statistical Methodology	31	2336	8	Parameter estimation for differential equations: A generalized smoothing approach. [18]	368	Wiley-Blackwell
Annals Of Applied Statistics	30	521	3.7	Modeling and forecasting electricity spot prices: A functional data perspective.[19]	59	Institute of Mathematical Statistics
Electronic Journal of Statistics	26	332	1.9	Longitudinal functional principal component analysis. [20]	95	Institute of Mathematical Statistics
Journal of Computational and Graphical Statistics	26	412	4.5	Functional Additive Mixed Model. [21]	105	Taylor & Francis
Statistics in medicine	26	332	3.4	Functional data analysis with application to periodically stimulated foetal heart rate data. II: Functional logistic regression. [22]	56	Wiley-Blackwell

TP: Total Publication; TC: Total Citations

Leading Authors

Table 3 shows the top ten prolific authors in the FDA research areas. Researchers from the United States of America (USA) dominated the list with four authors, followed by Canada (2 authors), Italy (2 authors), and one author from each of the following countries: France and Australia. The first publication in the Scopus database for the top 10 authors ranged from 1989 to the latest year of 2009, with most of them being co-authors and corresponding authors. Müller, H.G., an author from the University of California in the United States, has the most publications, with 70. In 2003, he was one of the co-authors of his first paper published in Scopus. However, there is a large gap in the publications between the first, second, and third-leading authors. A total of 29 publications were authored by the second-leading author, Vieu, P., who was affiliated with the Institut de Mathématiques, France. On the other hand, Ramsay, J.O., and Hall, P.G., the third and fourth leading authors, respectively, authored 23 and 21 publications.

Table 3. List of the ten most prolific authors in the FDA research area

	Authors	Scopus Author ID	Year of 1st publication	TP	H-index	TC	Current Affiliation	Country
1	Müller, H.G.	7404945300	2003 Co-author	70	26	2858	Department of Statistics, University of California, Davis, United States	United States
2	Vieu, P.	6603378253	2003 Co-author	29	15	832	Institut de Mathématiques de Toulouse, France	France
3	Ramsay, J.O.	7103327414	1989 Corresponding author	23	13	1194	Department of Psychology, Université McGill, Montreal, Canada	Canada
4	Hall, P.G.	57203346769	1998 Co-author	21	13	1347	School of Mathematics and Statistics, Australia University of Melbourne, Parkville, Australia	Australia
5	Kokoszka, P.S.	6603891972	2007 Co-author	21	12	409	Department of Statistics, Colorado State University, Fort Collins, United States	United States
6	Cao, J.	14622304000	2007 Corresponding author	20	9	497	Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, Canada	Canada
7	Sangalli, L.M.	22836567800	2009 Corresponding author	20	11	495	Department of Mathematics, Politecnico di Milano, Italy	Italy
8	Vantini, S.	23475624900	2009 Co author	20	10	368	Department of Mathematics, Politecnico di Milano, Italy	Italy
9	Morris, J.S.	7405897451	2001 Corresponding author	19	13	730	University of Pennsylvania Perelman School of Medicine, Philadelphia, United States	United States
10	Wang, J.L.	35239020400	2003 Co author	19	14	1860	Department of Statistics, University of California, Davis, United States	United States

TP: Total Publications; TC: Total Citations

The total number of citations obtained by the authors helps in the measurement of the H-index. The H-index is another metric that is used to measure both the citation impact and productivity of publications. It can be used as an indicator to measure the achievement of the authors. Müller HG has been awarded the highest H-index (26) in FDA publications, with 2858 citations by 2021. Vieu P, the second leading author, obtained an H-index of 15, followed by Hall PG (H-index: 13), Ramsay JO (H-index: 13), Wang JL (H-index: 14), and Morris JS with an H-index (13). The affiliations of the top 10 authors in FDA research areas come mainly from Statistics and Mathematics. However, a few authors, such as Ramsay JO, are affiliated with the Department of Psychology Université McGill Canada, while Morris JS is from the School of Medicine, Philadelphia, United States. FDA research appears to be more diverse than just mathematics and statistics.

Several authors in Table 3 worked collaboratively and published some articles together. For example, Müller, H.G., and Wang, J.L., wrote the first article in 2003, entitled 'Functional quasi-likelihood regression models with smooth random effects', published in the Journal of the Royal Statistical Society. Series B: Statistical Methodology [23]. Müller, H.G., and Wang, J.L., are affiliated with the same university, the University of California, Davis, United States. Other international collaborative works can also be seen between Müller H.G., Wang, J.L., and Hall, P. Their article entitled 'Properties of principal component methods for functional and longitudinal data analysis' was published in the Annals of Statistics in 2006. [24] Hall, P., is affiliated with the School of Mathematics and Statistics, University of Melbourne, Australia. Besides that, Ramsay J.O., the first author who published the FDA article, has collaborated with other leading authors listed in Table 3, such as Cao, J., and Sangalli, L.M. All the publications mentioned proved that international collaboration in FDA research areas exists. In a recent article, Ramsay, J.O. (Canada) and Sangalli, L.M. (Italy) worked collaboratively to model spatial anisotropy using functional regression with partial differential regularisation [25]. There is a high possibility that co-authors could sometimes be their colleagues or postgraduate students.

Leading Countries and Institutions

Figure 4 shows the top 20 active countries in the FDA publications. The USA contributed approximately 42.2% of the global FDA publications, as shown in Table 4. The authors from the United States contributed 722 papers, with 100 coming from the University of California, Davis. China is the second leading country with 197 publications, approximately four times less than the USA, followed by Spain, Canada, the United Kingdom (UK), Italy, and France. Even though Canada ranked 4th in the top 20 countries, the total publication (TPi) from Université McGill was found to be higher than Beijing University of Technology. Beijing University of Technology has only 13 publications, less than the National University of Singapore and Academia Sinica Taiwan, which placed 16th and 17th, respectively.

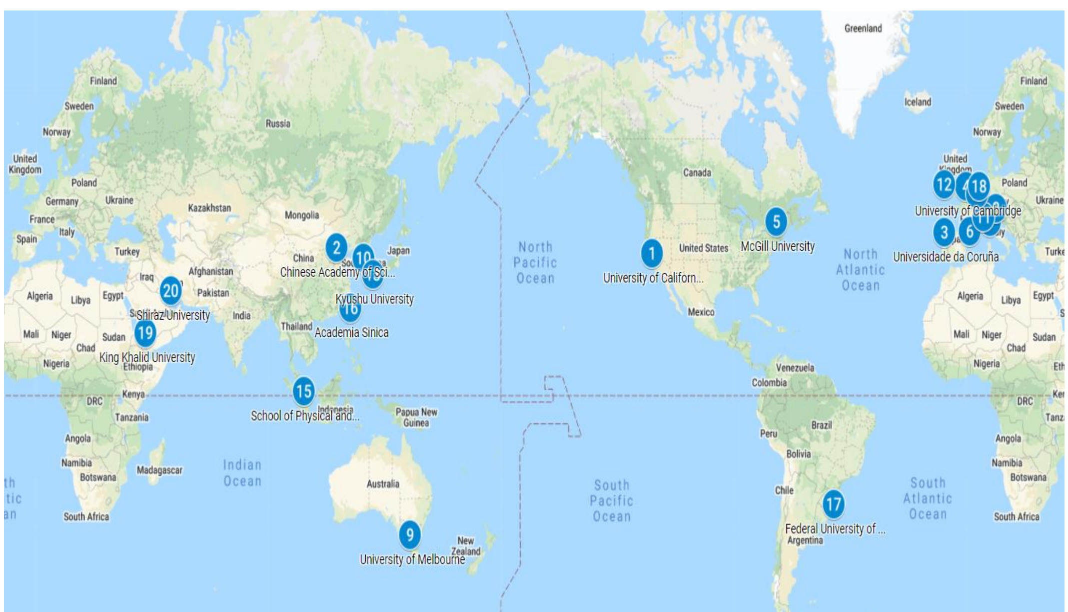

Figure 4. The top 20 most productive countries and academic institutions in FDA publications

Among the top 20 countries, the USA (62.7%), Spain (60.0%), Poland (62.5%), and Japan (60%) had nearly 2/3 single country publications (SCP). The findings show that the authors in these countries preferred to collaborate locally. Perhaps it is easier to converse in the same language, especially for Japanese authors or professionals from the same country. On the other hand, Saudi Arabia had the fewest SCPs at 6.25%. Similarly, despite its small size, Singapore has the second lowest SCP at 7.4%, indicating that 92.6% of its authors are affiliated with other countries. However, looking into the details, authors from Singapore collaborated with other co-authors, with some coming from the top 10 authors, as listed in Table 3. Apart from Singapore, South Korea, the Netherlands, and Ireland have over 80% of international collaborations. Through international collaborations, authors from different countries can share their knowledge and expertise, broadening the network between countries and boosting the university's ranking. In addition, eight universities from Table 4 were ranked in the top 100 best universities based on the QS (Quacquarelli Symonds) World University Ranking 2021. The University of Cambridge (7th), National University of Singapore (11th), Université McGill, Canada (31st), Seoul National University, Korea (37th), University of Melbourne, Australia (41st), Universiteit van Amsterdam, Netherlands (61st), Ludwig-Maximilians-Universität München, Germany (63rd), and KU Leuven (84th), Belgium, were on the list. It shows that FDA research has received attention from top universities globally.

The next stage is to create a map based on bibliographic data using VOSviewer software. The authorship was classified into four categories according to the continents. The co-authors are from 70 countries, with 3165 authors. Figure 5 depicts the distribution of countries involved in FDA research. The larger the box size, the more publications the country has, while the colours represent the average year of the publications. According to Figure 5, the United States has the highest total publications, with an average year of publication of 2014.2, followed by China (average publication: 2018.1), Spain (average publication: 2015.8), Canada (average publication: 2014.9), the United Kingdom (average publication: 2015.7), and Italy (average publication: 2016.9). If the location distances in VOSviewer are small, the two countries are significantly associated, whereas the line thickness suggests a strong link between them. Furthermore, the link strength represents the total number of publications linked to both countries. On the other hand, the total link strength describes the strength of a country's co-authorship links with other countries.

The largest number of countries comes from Europe (27), Asia (23), America (11), Africa (7), and Oceania (2). The result of the co-authorship indicates that the United States was the most affiliated country with a total link strength of 361 times of co-authorship, which links to 42 countries. The list is followed by the United Kingdom (37 links, 165 co-authorships), China (30 links, 152 co-authorships), Canada (24 links, 136 co-authorships), France (24 links, 111 co-authorships), Australia (31 links, 96 co-authorships), Spain (33 links, 94 co-authorships), Italy (24 links, 85 co-authorships), and others. Hence, the results show that FDA topics have received attention throughout the world. The authors from the United States and Canada collaborated the most, with 55 co-authorship publications, followed by the United States and China with 53 co-authorship publications (LS: 53), and the United States and the United Kingdom (LS:35). China also has strong international partnerships with Canada (LS:22), France (LS:11), and Singapore (LS:12). Aside from strong collaboration with the United States, the authors from the United Kingdom also established international collaboration with Ireland (LS: 14), Italy (LS: 10), Spain (LS: 12), and Australia (LS: 9).

Figure 5 shows that each of the 70 countries has at least one co-authorship with another country. In the VOSviewer, it is also apparent that countries like Georgia and Greece appear far apart, indicating that they have the least collaboration with authors from other countries. According to the average publishing year of 2021, countries including Tunisia, Lebanon, Bahrain, Palestine, Cameroon, and the Philippines have recently participated in FDA research topics. The increased number of international collaborators may be influenced by several factors, such as memorandum of understanding (MOU) activities, international conferences/seminars, postgraduate supervision, visiting researchers, and international research funding.

Author Keywords

Using VOSviewer, we examine the co-occurrence of author keywords. First, to group similar keywords, we construct a thesaurus file. Then, we look at all comparable terms, group them, and replace them with a single term. The total number of keywords was 4342, with 154 matching the VOSviewer's minimum five occurrence requirement.

Table 4. The top 20 most productive countries and academic institutions in FDA publications

Rank	Country	Total Publication (TPc) (%)	SCP(%)	The most productive academic institutions	Total publication of a given academic institution (TPi)
1	United States	722 (42.2%)	62.7	University of California, Davis	100
2	China	197 (11.5%)	48.2	Beijing University of Technology	13
3	Spain	160 (9.4%)	60.0	Universidade da Coruña	27
4	Canada	147 (8.6%)	36.7	Université McGill	34
5	United Kingdom	140 (8.2%)	26.4	University of Cambridge	20
6	Italy	130 (7.6%)	55.4	Politecnico di Milano	50
7	France	124 (7.2%)	39.5	Institut de Mathématiques de Toulouse	38
8	Germany	93 (5.4%)	50.5	Ludwig-Maximilians-Universität München	19
9	Australia	78 (4.6%)	24.4	University of Melbourne	24
10	South Korea	36 (2.1%)	22.2	Seoul National University	14
11	Switzerland	32 (1.9%)	53.1	Ecole Polytechnique Fédérale de Lausanne	15
12	Belgium	32 (1.9%)	37.5	KU Leuven	14
13	Saudi Arabia	32 (1.9%)	6.25	King Khalid University	19
14	Ireland	31 (1.8%)	16.1	University of Limerick	14
15	Japan	30 (1.8%)	60.0	Kyushu University	5
16	Singapore	27 (1.6%)	7.4	National University of Singapore	17
17	Taiwan	26 (1.5%)	30.8	Academia Sinica Taiwan	16
18	Czech Republic	24 (1.4%)	37.5	Univerzita Palackého v Olomouci	8
19	Netherlands	24 (1.4%)	23.8	Universiteit van Amsterdam	4
20	Poland	24 (1.4%)	62.5	Uniwersytet im. Adama Mickiewicza w Poznaniu	13

TPc: Total publications of a given country; TPi: Total publications of a given academic institution; SCP: Single Country Publication

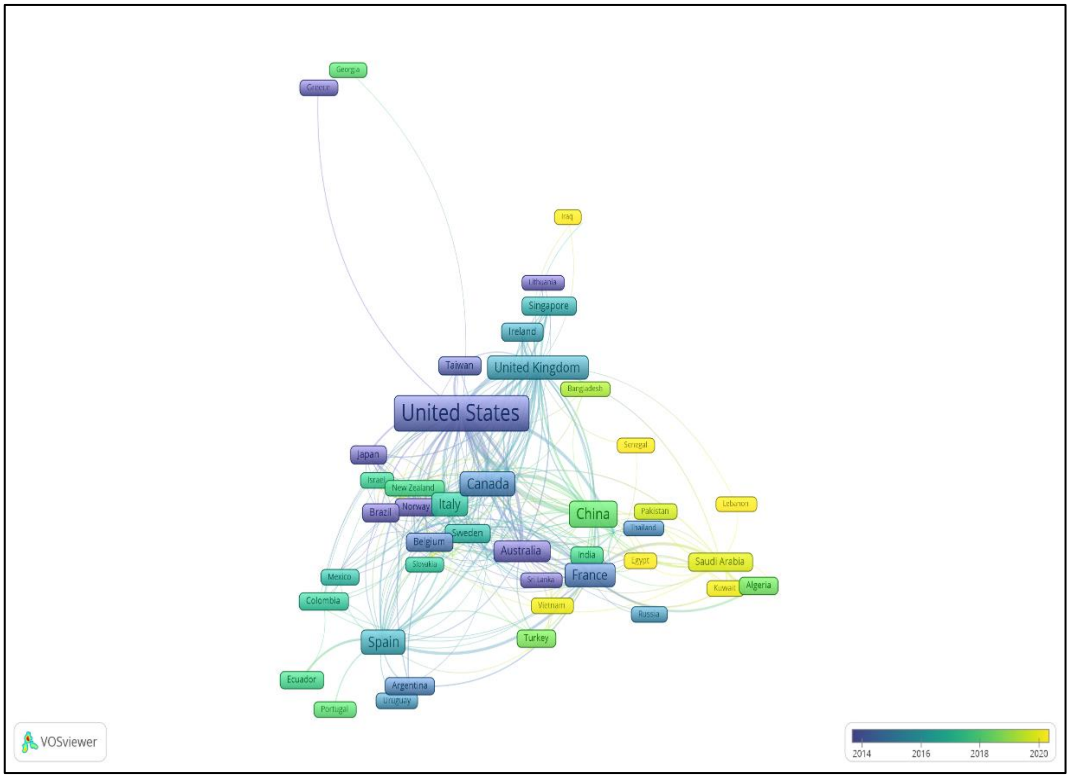

Figure 5. A screenshot of the bibliometric map created based on co-authorships with network visualization mode. The following URL can be used to open Figure 5 in VOSviewer: https://bit.ly/33NNazM

Terminology and Concepts

Referring to Figure 6, the author keyword 'functional data analysis' was the most popular in the articles (average publication: 2015.2), with 1242 occurrences and 153 links to other keywords. Other keywords such as 'functional principal component analysis (FPCA)' have 136 occurrences which link to the other 75 keywords and an average publication of 2016.2, 'principal component analysis (PCA)' (113 occurrences, 75 links, average publication: 2012.3), 'functional regression' (64 occurrences, 49 links, average publication: 2015.7), 'smoothing' (49 occurrences, 51 links, average publication: 2011.4), 'clustering' (44 occurrences, 33 links, average publication: 2014.0), 'dynamic time warping' (42 occurrences, 33 links, average publication: 2013.2) and 'longitudinal data' (43 occurrences, 44 links, average publication: 2012.2), These mentioned keywords are part of the statistical tools and features that were often used in FDA research areas. Functional data analysis also co-occurred with other keywords such as 'time series' (33 occurrences, 19 links), 'b-splines' (30 occurrences, 26 links), 'dimension reduction' (30 occurrences, 38 links), 'classification' (30 occurrences, 30 links), and 'smoothing spline' (30 occurrences, 33 links).

Tools and features of the FDA

The topic of the FDA has received considerable attention from various fields of research. Multivariate statistics, linear modelling, and time series analysis, which are part of conventional statistical methods, have been intensively employed using a functional concept. Some of the important features and tools of the FDA, based on the author's keywords, were also explained in this section.

Principal Component Analysis (PCA) and Functional Principal Component Analysis (FPCA)

The principal component analysis is a standard multivariate statistical method based on link strength related to functional data analysis. The PCA is generally used to reduce dimension. It reduces the number of correlated variables to a smaller number of uncorrelated variables while maintaining as much variation as possible [1]. The author keyword of PCA co-occurred with 'eigenvector' in six publications (LS:6), six publications with 'eigenfunction' (LS: 6), 'dimension reduction' (LS:7), 'singular value decomposition' (LS:5), 'consistency' (LS:6), and 'functional regression' (LS: 6).

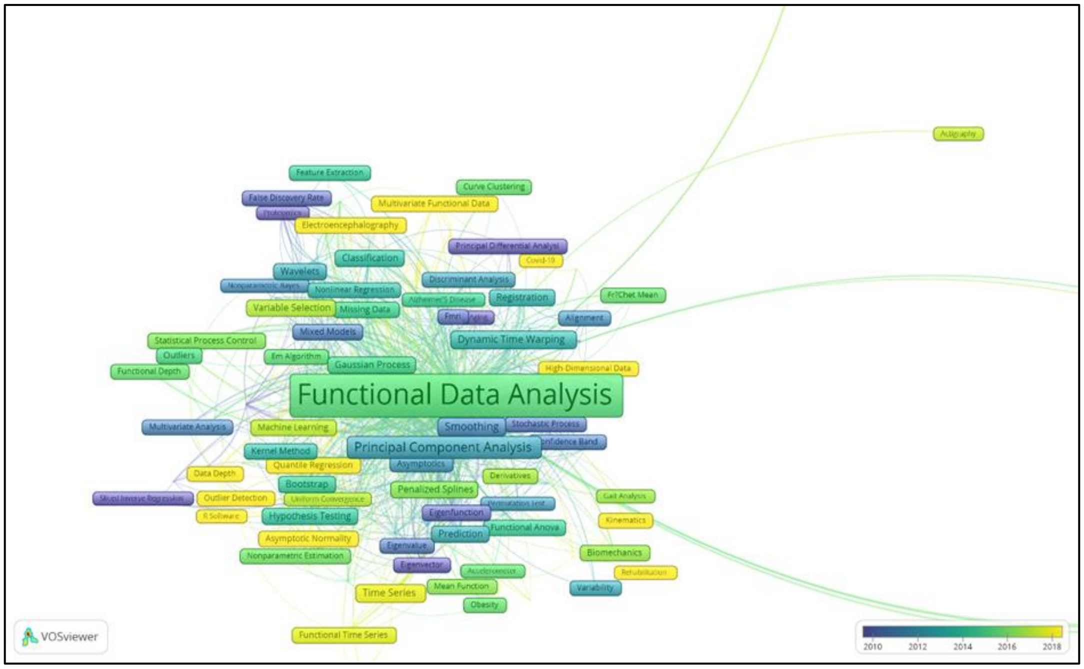

Figure 6. A screenshot of the bibliometric map created based on author keywords co-occurrence with overlay visualization mode. Minimum occurrences of a keyword are set to five. The following URL can be used to open Fig. 6 in VOSviewer: https://bit.ly/353Bc5Y

Ramsay and Abrahamowicz [11] were the first to publish articles attributable to principal component analysis. They used the method of principal component to describe the functional variation for the psychometric application. In another publication, Ramsay *et al*. [12] used a functional version of principal component analysis to study the variation in height acceleration. The method of principal component analysis has been successfully applied, such as in modelling the labelling index distribution [26], longitudinal data analysis [24, 27], growth modelling [28], mortality and fertility rates [14], patterns of traffic flow trajectories [29], and identifying species from bloodstains [30].

The principal component analysis method was later expanded to include the variations of smoothing curves in a functional concept known as functional principal component analysis. The keyword 'functional principal component analysis' appeared 136 times. It frequently co-occurred with 'functional regression' (LS:10), 'smoothing' (LS:7), 'longitudinal data' (LS: 6), 'functional linear model' (LS:5), 'asymptotic' (LS:4), and 'penalized spline' (LS: 5). Based on the Scopus database, the first article that used the keyword 'functional principal component analysis' was published by James [31]. In his study, he extended the method of a generalised linear model to handle the observation of predictor variables coming from a function or curve. The functional principal component analysis has been intensively used in several studies, such as interpreting lactate curves [32], analysing kinematic data [33], gene classification [34], exploring the major source of variations in glomerular filtration rate curves [35], modelling aircraft degradation [36], and estimating mean and covariance for functional snippets [37].

Smoothing

Discrete data can be converted into a function in two ways. The first is the interpolation approach, which assumes that the data are error-free, but if any observational errors need to be removed or minimised, a smoothing process is used [38]. Smoothing frequently occurs in conjunction with the terms 'principal component analysis' (LS:8), 'functional principal component analysis' (LS:7), 'eigenfunction' (LS:6), 'functional regression' (LS:6), and 'repeated measurement' (LS:5). The first step in functional data analysis is to create a set of basis functions that are used to smooth the functional data. B-spline smoothing, Cubic Spline smoothing, Fourier series smoothing, P-spline, Polynomial Smoothing, and Kernel smoothing are some of the most commonly used smoothing basis functions.

Silverman [39] introduced the smoothing technique for functional principal component analysis. He replaced the usual L_2-orthonormality constraint on the principal components with orthonormality to an inner product that takes account of the roughness of the functions. In addition, he implemented a Fourier transform for the periodic data. Ocaña *et al*. [40] also used the FDA feature of smoothing in functional principal component analysis to measure the distance. The term smoothing has been actively used in several publications related to FDA, such as smoothing functional varying coefficient models for longitudinal data [41], smoothing ground reaction force data using the B-spline basis function [42], smoothing two-dimensional functional datasets using spline basis functions [43], smoothing functional canonical correlation analysis of humidity and temperature data [44], and examining the choice of smoothing and differentiation for kinematic data [45].

Functional Regression

Functional regression is an extension of ordinary regression analysis where responses or covariates are functional data. The models can be categorised into four types depending on whether the responses or covariates are functional or scalar: (i) scalar responses with functional covariates, (ii) functional responses with scalar covariates, (iii) functional responses with functional covariates, and (iv) scalar or functional responses with functional and scalar covariates. Functional regression models can be linear, partially linear, or nonlinear. There are 64 occurrences in the article journal that use the keyword. It is often appeared with 'functional principal component analysis' (LS:10), 'smoothing' (LS:6), 'principal component analysis' (LS:6), 'small ball probability' (LS:3), 'prediction' (LS:2), and 'B-spline' (LS:2).

Chiou *et al*. [23] proposed a class of semiparametric functional regression models to describe the influence of vector-valued covariates on a sample of response curves. Each observable curve is thought to be the result of a random process with a mean function and random components. Manteiga and Vieu [46] explored the methodological aspects of regression in functional variables, curve classification, and functional data factorial analysis. Barber *et al*. [47] discussed the Least Absolute Shrinkage and Selection Operator (LASSO), a scalar regression function with applications to longitudinal genome-wide association studies (GWAS). Other published applications of functional regression include chemometrics [48], medical applications such as multiple sclerosis patients and cerebral aneurysms [49], traffic monitoring systems [50], seed germination coefficient comparison [51], improving the quality of optimal sampling schedules [52], and studying the effect of energy sector investment on energy security in the provinces [53].

Clustering

Clustering is a multivariate statistical approach for grouping things with a high degree of similarity to those in other groups. The clustering approach is used in the functional context to group smoothing or function curves that exhibit similar patterns to make inferences. Several authors have successfully investigated dissimilarities between functions in the case of functional data [54-58]. With applications to cluster analysis, Hitchcock *et al*. [54] investigated the influence of smoothing functional data on estimating the dissimilarities across objects in a dataset. Tokushige *et al*. [55] applied multivariate functional data analysis to the notion of crisp and fuzzy k-means clustering techniques. Song *et al*. [34] proposed a clustering technique to cluster time-dependent gene expression profiles. Giacofci *et al*. [59] presented curve clustering for the random effect in a high-dimensional context. Their method was tested on data from microarray comparative genomic hybridization (CGH).

The clustering technique was successfully used in sports analysis to classify patterns [60]. Liebl *et al*. [60] classified different footfall patterns in human runners. The proposed stochastic cluster analysis provides a robust and practical framework for grouping foot strikes. On the other hand, Misumi *et al*. [58] suggested a multivariate nonlinear mixed-effects model for clustering various longitudinal data. Data from typhoons that struck Asia between 2000 and 2017 was used in their study. In a recent publication, Léger and Mazzuco [61] employed FDA clustering to study smoothing mortality patterns from 32 countries based on the Human Mortality Database from 1960 to 2018. Usually, principal component analysis and clustering are closely related techniques. The factor scores extracted from the principal component analysis are normally used for clustering analysis, for example, in analysing bloodstains [30].

Development of the author keywords in the FDA

The link strength of the two keywords could also be used to analyse research interest. For example, 'functional data analysis' was used with 'principal component analysis' in 96 articles, followed by 88 articles with 'functional principal component analysis.' There are 42 articles containing both keywords: functional data analysis and functional regression, functional data analysis-smoothing (LS:39), longitudinal data-functional data analysis (LS:38), clustering-functional data analysis (LS:36), functional data analysis-dynamic time warping (LS:36), functional data analysis-dimension reduction (LS:28) and functional data analysis-classification (LS:26). It seems that research interest in functional data analysis and principal component analysis was the strongest. Other keywords such as 'association mapping', 'common variant', 'complex traits', 'rare variants', and 'complex diseases' were less related to FDA based on the longer distance between the keywords in VOSviewer.

We could analyse the evolution of the keywords in FDA if we used the overlay visualisation mode. With an average publishing year of 2003.7, the keyword 'curve estimation' was one of the first. It had seven other keywords associated with it. Faraway [62] produced the first article that used this keyword in performing regression analysis for functional responses. Faraway [63] once again mentioned this keyword while utilising the graphical method to investigate the mean structure for longitudinal data. This keyword has progressed with constant publishing in 2001, 2003, 2007, and 2015.

The keyword 'principal differential analysis' was also presented before 2010, with an average publication year of 2008.8. Ramsay [18] was the first to use it. In examining a sample of Chinese handwriting, he employed the differential equation method, and later, the same keyword was used to find a differential equation for lip motion [64]. The paper by Dalla Rossa *et al*. [65] demonstrated the advancement of this keyword. They explored using principal differential analysis to reduce the number of dimensions in functional data sets. Jang and Lim [66] have suggested a classification approach based on principal differential analysis. They combined the principal differential analysis results with logistic regression for binary classification.

Since COVID-19 has affected the global health systems of whole countries, many researchers have begun investigating the disease. 'Covid-19' is a recent author keyword with an average publication of 2020.8, with five occurrences. A link to six other keywords, including 'functional data analysis' (LS: 5), 'function-on-function regression' (LS: 1), 'b-splines' (LS: 2), 'phase variability' (LS:1), and 'functional principal component analysis' (LS:5). Acal *et al*. [67] authored the first article that mentioned 'Covid-19'. With the application to the COVID-19 pandemic in Spain, they compared the performance of modelling approaches based on the varimax rotation of FPCA. Later on, in 2021, a series of articles on COVID-19 was again published by Acal *et al*. [68,69]. Acal *et al*. [68] employed a functional Analysis of Variance (ANOVA) to detect the changes in air pollution during the COVID-19 pandemic. Two sets of data were used: before lockdown and during lockdown. The FDA techniques were used to investigate the impact of the lockdown on air quality. According to their findings, functional ANOVA has proven helpful in monitoring the changes in air quality over two time periods. In the most recent article, Acal *et al*. [69] employed a function-on-function regression model for the imputation of missing data on COVID-19 hospitalised and intensive care curves in several Spanish areas. Other authors, Kumar *et al*. [70], used functional data analysis to conceptualise and explore the comparative efficacy of various nonpharmaceutical interventions that countries could use to prevent or slow the spread of disease incidence and death. In addition, Scimone *et al*. [71] examined the spatio-temporal mortality patterns of COVID-19 cases in Italy. In their analysis, the mortality data were represented as densities of time of death, considered functional data.

Other author keywords with an average publication year of 2019 or higher include 'structural health monitoring' (average publication: 2019.8), 'function-on-scalar regression/scalar-on-function regression' (average publication: 2019.1), 'function-on-function regression' (average publication: 2019.33), 'statistics' (average publication: 2019.33), 'outlier detection' (average publication: 2019.36), 'quantile regression' (average publication: 2019). The author's keyword, 'structural health monitoring' was identified in five documents. The first was published in 2018, followed by a single publication in 2019 and 2020, respectively, and the last two in 2021. It also links to keywords such as 'bootstrap', 'dynamic time warping', and 'missing data'. In a recent paper, Chen *et al*. [72] addressed their concern about missing structural health monitoring data. They used a warping function regression method from the functional data analysis framework to reconstruct missing data distributions. The author keyword 'structural health monitoring' also appeared in Jiang *et al*. [73]. The study focused on the features of nonlinear phase variation between responses in modelling relationships for field strain data under thermal effects.

Based on the largest number of occurrences for the average publication in 2019 and above, the keyword 'quantile regression' has 13 occurrences with 13 links. The keyword is closely related to 'functional nonparametric statistics' (LS:2), 'asymptotic normality' (LS: 1), 'small ball probability' (LS: 2), and 'bayesian hierarchical model' (LS: 1). The first article that used 'quantile regression' was published in NeuroImage authored by Reiss *et al*. [74]. The keyword was used to describe the functional connectivity of an individual human brain. After a few years, the keyword was used again in Statistics and Probability Letters by Ding *et al*. [75]. They proposed a quantile estimation as an alternative to the least squares approach for a semi-functional linear model. Later on, Wang *et al*. [76] developed a wavelet-based regularised linear quantile regression framework for coefficient estimations, where the responses are scalars, and the predictors include both scalars and functions. The number of publications that highlight 'quantile regression' increased to seven in 2020, showing that the present keyword has received much attention in FDA research areas. Almanjahie *et al*. [77] employed the L1-norm approach in constructing the local linear estimator of the spatial regression quantile for functional regressors, while Xu and Du [78] proposed a functional approach of nonparametric quantile regression estimation with the responses missing at random. The recent publication by Laksaci *et al*. [79] focused on the efficiency of the nonparametric estimation of the conditional quantile when the response variable is a scalar given a functional covariate.

Based on the Scopus database, the author keyword 'function-on-function regression was first used by Meyer *et al*. [80] to describe the relationship between two functional variables. It was used for repeatedly sampled functional data since researchers often sample multiple curves per person, resulting in repeated functional measures. They used a Bayesian framework combined with function-on-function regression to analyse how the brain processes different types of images. Rügamer *et al*. [81] employed function-on-function regression models in analysing the functional relationship between electroencephalography and facial electromyography. Later on, Acal *et al.* [68] proposed a function-on-function regression model to estimate the missing values of the functional responses associated with hospitalised and intensive care curves related to a newly emerging disease, COVID-19. The method provided a prediction equation for the imputation of the missing values. The other keywords, 'function-on-scalar regression', and 'scalar-on-function regression' are parts of functional regression. The idea is to fit regression models for scalar and functional responses with the effects of scalar and functional covariates. Gellar *et al*. [82] introduced models with subject-specific functional predictor domains. The basic idea is to evaluate a bivariate functional parameter affected by the functional argument and the domain width of the functional predictor. The method was applied to data from intensive care units (ICUs). In addition, the mentioned keywords can also be observed in the application of stock return [83], positron emission tomography (PET) data [84,85], and modelling the effect of wildfires on vegetation index [86]. Serra-Burriel *et al*. [86] employ FDA regression models, 'function-on-scalar regression', in which wildfire impacts are treated as functional responses, with elapsed time after each wildfire acting as a scalar covariate and pre-wildfire information acting as a scalar covariate.

Outlier detection has recently been one of the most popular author keywords in FDA research. It appeared in 11 documents and is linked to 11 additional author keywords. Vilar *et al*. [87] used a robust functional principal component analysis to identify outliers in functional time series with applications to the electricity market. In a recent European Journal of Operational Research publication, Rennie *et al*. [88] used functional data analysis to identify and respond to outlier demand in revenue management. They discovered that functional outlier identification had a higher detection rate than other approaches. In general, it could be summarised that the development of new author keywords has shown the growth of research interest in the FDA.

Limitation of Study and Future Work

The current study has several limitations. First, the articles related to the FDA research areas have been retrieved only from the Scopus database. Only articles with a Scopus index will be considered in the analysis. Besides, the search strings in this analysis are focused only on 'functional data analysis' that appears within the title, abstract, or author keywords in the Scopus database. Finally, since the number of published articles is quite limited, multiple databases, such as Web of Science, Dimensions, and Microsoft Academic, are highly recommended to provide a complete review and analysis in future work. For future analysis, the search strings can be extended to tackle particular issues such as 'climate change', 'infectious disease', 'finance', and 'time series forecasting'. Hence, the findings will be more relevant to address the current issues.

Conclusions

The present study used bibliometric analysis to provide an overview of FDA research trends over the past three decades, based on 1712 publications retrieved from the Scopus database. A strong international collaboration between the US-China, US-Canada, and US-UK has contributed to a large number of published papers. Authors from other small countries have also initiated international collaborations with leading authors in FDA research areas. There have been rapid publications for the last 30 years, and the number of publications is expected to continue to rise. The FDA offers a flexible framework for modelling functional data. Several FDA features and tools, such as smoothing, principal component analysis, regression, and clustering, have been thoroughly explored by researchers and have been applied intensively in various domains. Some new tools, such as function-on function regression, scalar-on-function regression, and function-on scalar regression, have been presented and discussed in this study, providing potential hot topics for future studies. In addition, new author keywords such as 'Covid-19', 'structural health monitoring', and 'outlier detection' have recently appeared in published articles, indicating that FDA research is rapidly developing and still progressing. Based on the findings, it seems that lately, the biomedical and health sciences are the two research fields that many researchers have explored for functional data analysis applications.

In conclusion, bibliometric analysis allows researchers to become more alert and always updated with recent research developments. It assists in identifying emerging research areas based on keyword occurrences, prominent researchers, leading institutions, and countries. However, it is not a suitable replacement for the qualitative assessment that systematic literature reviews require. Both methods have their specialties and can be combined to provide more comprehensive literature reviews within a specific research domain.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

This work was funded by the Ministry of Higher Education (MOHE) under the Fundamental Research Grant Scheme (FRGS/1/2020/STG06/UTM/02/3). We are also grateful to Universiti Teknologi Malaysia for supporting this project under Research University Grant Q.J130000.3854.19J58.

References

[1] Suhaila, J. (2021). Functional data visualization and outlier detection on the anomaly of El Niño southern oscillation. *Climate*, *9*(118).
[2] Wang, D., Zhong, Z.; Bai, K., & He, L. (2019). Spatial and temporal variabilities of PM2.5 concentrations in China using functional data analysis. *Sustainability*, *11*(6), 1620.
[3] Ullah, S., & Finch C. F. (2013). Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology, 13*, 43.
[4] Suhaila, J., & Yusop, Z. (2017). Spatial and temporal variabilities of rainfall data using functional data analysis. *Theoretical and Applied Climatology, 129*, 229-242.
[5] Wang, J. L., Chiou, J. M., & Muller, H. G. (2016). Review of functional data analysis. *Annual Review of Statistics and its Application*, *3*, 257-295.
[6] Aneiros, G., Cao, R., Fraiman, R., Genest, C., & Vieu, P. (2019). Recent advances in functional data analysis and high-dimensional statistics. *Journal of Multivariate Analysis*,*170*, 3-9.
[7] Aneiros, G., Horová, I., Hušková, M., & Vieu, P. (2022). On functional data analysis and related topics. *Journal of Multivariate Analysis*, *189*.
[8] Md Khudzari, J., Kurian, J., Tartakovsky, B., & Vijaya Raghavan, G.S. (2018). Bibliometric analysis of global research trends on microbial fuel cells using Scopus database. *Biochemical Engineering Journal*, *136*, 51-60.
[9] Sweileh, W. M. (2020). Bibliometric analysis of peer-reviewed literature on climate change and human health with an emphasis on infectious disease. *Globalization and Health*, *16*, 44.
[10] van Eck, N. J. & Waltman, L. (2020). VOSviewer Manual.
[11] Ramsay, J. O., & Abrahamowicz, M. (1989). Binomial regression with monotone splines: A psychometric application. *Journal of the American Statistical Association*, *84*(*408*), 906-915.
[12] Ramsay, J.O., Altman, N., & Bock, R.D. (1994). Variation in height acceleration in the Fels growth data. *Canadian Journal of Statistics*, *22*(*1*), 89-102.
[13] Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association,100*(*470*), 577-590.
[14] Hyndman, R. J., & Shahid Ullah, Md. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis*, *51*(*10*), 4942-4956.
[15] He, G., Müller, H. G., & Wang, J. L. (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis*, *85*(*1*), 54-77.
[16] Yao, F., Muller, H. G., & Wang, J. L. (2005). Functional linear regression analysis for longitudinal data. *Annals of Statistics 33*(*6*), 2873-2903.
[17] Rice, J. A., & Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, *57*(*1*), 253-259.
[18] Ramsay, J. O. (2000). Functional Components of Variation in Handwriting. *Journal of the American Statistical Association*, *95*(*449*), 9-15.
[19] Liebl, D. (2013). Modeling and forecasting electricity spot prices: A functional data perspective. *Annals of Applied Statistics*, *7*(*3*), 1562-1592.
[20] Greven, S., Crainiceanu, C., Caffo, B., & Reich, D. (2010). Longitudinal functional principal component analysis. *Electronic Journal of Statistics*, *4*, 1022-1054.
[21] Scheipl, F., Staicu, A-M., & Greven, S. (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics*, *24*(*2*), 477-501.
[22] Ratcliffe, S. J., Heller, G. Z., & Leader, L. R. (2002). Functional data analysis with application to periodically

stimulated foetal heart rate data. II: Functional logistic regression. *Statistics in Medicine*, *21*(*8*), 1115-1127.

[23] Chiou, J. M., Müller, H. G., & Wang, J. L. (2003). Functional quasi-likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society Series B Statistical Methodology*, *65*(*2*), 405-423.

[24] Hall, P., Müller, H. G., & Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics*, *34*(*3*), 1493-1517.

[25] Bernardi, M. S., Carey, M., Ramsay, J. O., & Sangalli, L. M. (2018). Modeling spatial anisotropy via regression with partial differential regularization. *Journal of Multivariate Analysis*, *167*,15-30.

[26] Grambsch, P. M., Randall, B. L., Bostick, R. M., Potter, J. D., & Louis, T. A. (1995). Modeling the labeling index distribution: An application of functional data analysis. *Journal of the American Statistical Association*, *90*(*431*), 813-821.

[27] Dai, X., Lin, Z., & Müller, H. G. (2021). Modeling sparse longitudinal data on Riemannian manifolds. *Biometrics* *77*(*4*), 1328-1341.

[28] Hermanussen, M., & Meigen, C. (2007). Phase variation in child and adolescent growth. *International Journal of Biostatistics*, *3*(*1*).

[29] Li, P. L., & Chiou, J. M. (2021). Functional clustering and missing value imputation of traffic flow trajectories. *Transportmetrica B: Transport Dynamics*, *9*(*1*), 1-21.

[30] Huang, W., Gao, L., Guo, W., Cui, H., Li, Z., Xu, X., & Wang, G. (2021). Analysis into functional data of spectral images from bloodstains of human and two species of animal. *Forensic Science and Technology*, *46*(*6*), 551-558.

[31] James, G.M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society Series B Statistical Methodology*, *64*(*3*), 411-432.

[32] Newell, J., McMillan, K., Grant, S., & McCabe, G. (2006). Using functional data analysis to summarise and interpret lactate curves. *Computers in Biology and Medicine*, *36*(*3*), 262-275.

[33] Ryan, W., Harrison, A., & Hayes, K. (2006). Functional data analysis of knee joint kinematics in the vertical jump. *Sports Biomechanics*, *5*(*1*), 121-138.

[34] Song, J. J., Deng, W., Lee, H-J., & Kwon, D. (2008). Optimal classification for time-course gene expression data using functional data analysis. *Computational Biology and Chemistry*, *32*(*6*), 426-432.

[35] Dong, J. J., Wang, L., Gill, J., & Cao, J. (2018). Functional principal component analysis of glomerular filtration rate curves after kidney transplant. *Statistical Methods in Medical Research*, *27*(*12*), 3785-3796.

[36] Zhang, B., Zheng, K., Huang, Q., Feng, S., Zhou, S. & Zhang, Y. (2020). Aircraft engine prognostics based on informative sensor selection and adaptive degradation modeling with functional principal component analysis. *Sensors*, *20*(*3*), 920.

[37] Lin, Z.& Wang, J.L. (2022). Mean and Covariance Estimation for Functional Snippets. *Journal of American Statistical Association*, *117*(*537*).

[38] Suhaila, J., Jemain, A.A., Hamdan, M.F., & Wan, ZWZ. (2011). Comparing rainfall patterns between regions in Peninsular Malaysia via a functional data analysis technique. *Journal of Hydrology*, *411*,197-206.

[39] Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Annals of Statistics*, *24*(*1*), 1-24.

[40] Ocaña, F.A., Aguilera, A.M., & Valderrama, M.J. (1999). Functional principal components analysis by choice of norm. *Journal of Multivariate Analysis*, *71*(*2*), 262-276.

[41] Şentürk, D., & Müller, H.G. (2010). Functional varying coefficient models for longitudinal data. *Journal of American Statistical Association*, *105*(*491*),1256-1264.

[42] Din, W. R. W., Rambely, A. S., & Jemain, A. A. (2013). Smoothing of GRF data using functional data analysis technique. *International Journal of Applied Mathematics and Statistics*, *47*(*17*), 70-77.

[43] Ivanescu, A.E. (2018). Function-on-function regression for two-dimensional functional data. *Communication in Statistics – Simulation and Computation*, *47*(9), 2656-2669.

[44] Koymen Keser, I.,& Deveci Kocakoç, I. (2015). Smoothed functional canonical correlation analysis of humidity and temperature data. *Journal of Applied Statistics*, *42*(*10*), 2126-2140.

[45] Zin, M. A. M., Rambely, A. S., Ariff, N. M., & Ariffin, M. S. (2020). Smoothing and differentiation of kinematic data using functional data analysis approach: An application of automatic and subjective methods. *Applied Sciences*, *10*(*7*),2493.

[46] Manteiga, W. G., & Vieu, P. (2007). Statistics for functional data. computational statistics & data analysis, *51*(*10*), 4788-4792.

[47] Barber, R. F., Reimherr, M., & Schill, T. (2017). The function-on-scalar LASSO with applications to longitudinal GWAS. *Electronic Journal of Statistics*, *11*(*1*), 1351-1389.

[48] Saeys, W., De Ketelaere, B., & Darius, P. (2008). Potential applications of functional data analysis in chemometrics. *Journal of Chemometrics*, *22*(*5*), 335-344.

[49] Sørensen, H., Goldsmith, J., & Sangalli, L.M. (2013). An introduction with medical applications to functional data analysis. *Statistics in Medicine*, *32*(*30*), 5222-5240.

[50] Chen, K., & Müller, H-G. (2014). Modeling conditional distributions for functional responses, with application to traffic monitoring via GPS-enabled mobile phones. *Technometrics*, *56*(*3*),347-358.

[51] Talská, R., Machalová, J., Smýkal, P., & Hron, K. (2020). A comparison of seed germination coefficients using functional regression. *Application in Plant Sciences*, *8*(*8*): e11366.

[52] Rha, H., Kao, M. H., & Pan, R. (2021). Bagging-enhanced sampling schedule for functional quadratic regression. *Journal of Statistical Theory and Practice*, *15*, 91.

[53] Bamisile, O., Ojo, O., Yimen, N., Adun, H., Li, J., Obiora, S., & Huang, Q. (2021). Comprehensive functional data analysis of China's dynamic energy security index. *Energy Reports*, *7*, 6246-6259.

[54] Hitchcock, D. B.; Casella, G.; Booth, J. G. (2006). Improved estimation of dissimilarities by presmoothing functional data. *Journal of American Statistical Association*, *101*(*473*), 211-222.

[55] Tokushige, S., Yadohisa, H., & Inada, K. (2007). Crisp and fuzzy k-means clustering algorithms for multivariate functional data. *Computational Statistics*, *22*(*1*), 1-16.

[56] Dabo-Niang, S., Ferraty, F., & Vieu, P. (2007). On the using of modal curves for radar waveforms classification. *Computational Statistics & Data Analysis*, *51*(10), 4878-4890.
[57] Gattone, S. A., & Rocci, R. (2012). Clustering curves on a reduced subspace. *Journal of Computational and Graphical Statistics*, *21*(*2*), 361-379.
[58] Misumi, T., Matsui, H., & Konishi, S. 2019. Multivariate functional clustering and its application to typhoon data. *Behaviormetrika*, *46*(*1*), 163-175.
[59] Giacofci, M., Lambert-Lacroix, S., Marot, G., & Picard, F. (2013). Wavelet-based clustering for mixed-effects functional models in high dimension. *Biometrics*, *69*(*1*), 31-40.
[60] Liebl, D., Willwacher, S., Hamill, J., & Brüggemann, G-P. (2014). Ankle plantarflexion strength in rearfoot and forefoot runners: A novel cluster analytic approach. *Human Movement Science*, *35*, 104-120.
[61] Léger, A-E., & Mazzuco, S. (2021). What can we learn from the functional clustering of mortality data? An application to the human mortality database. *European Journal of Population*, *37*, 769-798.
[62] Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics*, *39*(*3*), 254-261.
[63] Faraway, J. J. (1999). A graphical method of exploring the mean structure in longitudinal data analysis? *Journal of Computational and Graphical Statistics*, *8*(*1*), 60-68.
[64] Lucero, J. C. (2002). Identifying a differential equation for lip motion. *Medical Engineering & Physics*, *24*(*7-8*), 521-528.
[65] Dalla Rosa, M., Sangalli, L.M., & Vantini, S. (2014). Principal differential analysis of the Aneurisk65 data set. *Advances in Data Analysis and Classification*, *8*(*3*), 287-302.
[66] Jang, E., & Lim, Y. (2021). Classification via principal differential analysis. *Communication for Statistical and Applications and Methods*, *28*(*2*), 135-150.
[67] Acal, C., Aguilera, A.M., & Escabias, M. (2020). New modeling approaches based on varimax rotation of functional principal components. *Mathematics*, *8*(*11*), 2085.
[68] Acal, C., Aguilera, A.M., Sarra, A., Evangelista, A., Battista, T.D., & Palermi, S. (2022). Functional ANOVA approaches for detecting changes in air pollution during the COVID-19 pandemic. *Stochastic Environmental Research and Risk Assessment, 36*(4),1083-1101.
[69] Acal, C., Escabias, M., Aguilera, A.M., & Valderrama, M.J. (2021). COVID-19 data imputation by multiple function-on-function principal component regression. *Mathematics*, *9*(*11*), 1237.
[70] Kumar, V., Sood, A., Gupta, S., & Sood, N. (2021). Prevention- versus promotion-focus regulatory efforts on the disease incidence and mortality of COVID-19: A multinational diffusion study using functional data analysis. *Journal of International Marketing*, *29*(*1*), 1-22.
[71] Scimone, R., Menafoglio, A., Sangalli ,L.M., & Secchi, P. (2021). A look at the spatio-temporal mortality patterns in Italy during the COVID-19 pandemic through the lens of mortality densities. *Spatial Statistics*, *49*(*1*).
[72] Chen, Z., Lei, X., Bao, Y., Deng, F., Zhang, Y., & Li, H. (2021). Uncertainty quantification for the distribution-to-warping function regression method used in distribution reconstruction of missing structural health monitoring data. *Structural Health Monitoring*, *20*(*6*), 3436-3452.
[73] Jiang, H., Wan, C., Yang, K., Ding, Y., & Xue, S. (2021). Modeling relationships for field strain data under thermal effects using functional data analysis. *Measurement*, *177*.
[74] Reiss, P. T., Mennes, M., Petkova, E., Huang, L., Hoptman, M. J., Biswal, B. B., Colcombe, S. J., Zuo, X-N., & Milham, M. P. (2011). Extracting information from functional connectivity maps via function-on-scalar regression. *NeuroImage*, *56*(*1*), 140-148.
[75] Ding, H., Lu, Z., Zhang, J., & Zhang, R. (2018). Semi-functional partial linear quantile regression. *Statistics & Probability Letters*, *142*, 92-101.
[76] Wang, Y., Kong, L., Jiang, B., Zhou, X., Yu, S., Zhang, L., & Heo, G. (2019). Wavelet-based LASSO in functional linear quantile regression. *Journal of Statistical Computation and Simulation*, *89*(*6*), 1111-1130.
[77] Almanjahie, I. M., Chikr Elmezouar, Z., Bachir, B. A., & Kaid, Z. (2020). Spatial local linear estimation of the L_1-conditional quantiles for functional regressors. *Communication in Statistics – Theory and Methods*, *49*(23), 5666-5685.
[78] Xu, D., & Du, J. (2020). Nonparametric quantile regression estimation for functional data with responses missing at random. *Metrika*, *83*(*8*), 977-990.
[79] Laksaci, A., Ould Saïd, E., & Rachdi, M. (2021). Uniform consistency in number of neighbors of the kNN estimator of the conditional quantile model. *Metrika*, *84*(*6*), 895-911.
[80] Meyer, M. J., Coull, B. A., Versace, F., Cinciripini, P., & Morris, J.S. (2015). Bayesian function-on-function regression for multilevel functional data. *Biometrics*, *71*(*3*), 563-574.
[81] Rügamer, D., Brockhaus, S., Gentsch, K., Scherer, K., & Greven, S. (2018). Boosting factor-specific functional historical models for the detection of synchronization in bioelectrical signals. *Journal of the Royal Statistical Society Series C Applied Statistics*, *67*(*3*), 621-642.
[82] Gellar, J. E., Colantuoni, E., Needham, D. M., & Crainiceanu, C. M. (2014). Variable-domain functional regression for modeling ICU data. *Journal of the American Statistical Association, 109*(*508*), 1425-1439.
[83] Brockhaus, S., Fuest, A., Mayr, A., & Greven, S. (2018). Signal regression models for location, scale and shape with an application to stock returns. *Journal of the Royal Statistical Society Series C Applied Statistics*, *67*(*3*), 665-686.
[84] Chen, Y., Goldsmith, J., & Ogden, R.T. (2019). Functional Data Analysis of Dynamic PET Data. *Journal of American Statistical Association,114*(*526*), 595-609.
[85] Shi, B., & Ogden, R.T. (2021). Inference in functional mixed regression models with applications to Positron Emission Tomography imaging data. *Statistics in Medicine*, *40*(*4*), 4640-4659.
[86] Serra-Burriel, F., Delicado, P., & Cucchietti, F. M. (2021). Wildfires vegetation recovery through satellite remote sensing and functional data analysis. *Mathematics*, *9*(*11*), 1305.
[87] Vilar, J. M., Raña, P., & Aneiros, G. (2016). Using robust FPCA to identify outliers in functional time series, with applications to the electricity market. *SORT*, *40*(*2*), 321-348.

[88] Rennie, N., Cleophas, C., Sykulski, A. M., & Dost, F. (2021). Identifying and responding to outlier demand in revenue management. *European Journal of Operational Research*, *293*(*3*),1015-1030.