A deep dive into membrane distillation literature with data analysis, bibliometric methods, and machine learning approaches

Ersin Aytaç [1,2], Mohamed Khayet 1[1,3,*]

[1] Department of Structure of Matter, Thermal Physics and Electronics, Faculty of Physics, University Complutense of Madrid, Avda. Complutense s/n, 28040, Madrid, Spain.

[2] Zonguldak Bülent Ecevit University, Department of Environmental Engineering, 67100, Zonguldak, Turkey

[3] Madrid Institute for Advanced Studies of Water (IMDEA Water Institute), Calle Punto Net Nº 4, 28805, Alcalá de Henares, Madrid, Spain.

* Corresponding author:

khayetm@fis.ucm.es

Tel. +34-91-3945185

Abstract

Membrane distillation (MD) is a non-isothermal separation process applied mainly in desalination for the treatment of saline aqueous solutions including brines for distilled water production by different technological configurations. Various experimental and theoretical investigations have been carried out in many fields like membrane and module engineering, membrane fouling and scaling, crystallization, treatment of specific wastewaters, coupling with renewable energy sources, and other membrane separation technologies. However, no research has been conducted yet evaluating the MD literature with data analysis, bibliometric methods, and machine learning approaches for a better understanding of MD and quick identification of the necessary research lines to be carried out for its adequate and solid development toward industrial implementations. This study includes an in-depth review of published manuscripts on MD in refereed international journals. The data set was fetched from Scopus on 02.02.2022 with a wide spectrum of keywords. Interesting statistical and graphical information is presented in the data analysis section. By using different indexes of bibliometric analyzes, significant manuscripts, authors more active in MD research, and the corresponding institutions and countries that have contributed most to the progress of MD technology are presented together with the collaborations made between research groups. The most used membrane configurations and combined systems are revealed together with the materials most used in MD membrane formation and modification. Studies about membrane engineering were examined and guiding data were presented. The field of applications has been exposed and the use of renewable energy sources in MD has been revealed. With text mining approaches, the most used words, keywords analysis, and trending topics are visualized. Furthermore, the emotions of the authors toward MD perspectives were examined from the published abstracts of the manuscripts with sentiment analysis.

Keywords: Machine learning; Biblioshiny; membrane distillation; sentiment analysis; text mining; Upset graph; Venn diagram; word cloud.

1. Introduction

The non-isothermal separation technology Membrane distillation (MD) is gaining more popularity within the membrane science community during the last ten years because of its wide field of applications especially for the treatment of high saline aqueous solutions including brines [1]. The MD separation is carried out through a porous and hydrophobic membrane (i.e. in case of composite membrane at least one layer is hydrophobic) by means of temperature gradient, or which is the same a water vapor pressure difference between the feed and permeate sides of the membrane [2, 3]. It can be operated at low hydrostatic pressures (i.e. atmospheric pressure), under moderate temperatures (i.e. below the boiling point of feed solutions), allowing therefore the use of available waste heat and renewable energy sources and achieving very high separation efficiency (i.e. high non-volatile solute rejection factors) [4, 5].

Since its first publication in 1967 [6], MD process has been developed considerably expanding its application in different areas with different configurations. Generally, research MD studies has been focused on membrane and module engineering, effects of operating parameters on water production rate and separation performance, wastewater applications, simulation and theoretical modeling including computational fluid dynamics (CFD), Monte Carlo simulation, optimizations by response surface methodology (RSM) and artificial neural network (ANN) among others [7-13].

Nowadays one can find in the literature seven different MD configurations, four main variants (direct contact membrane distillation, DCMD; sweeping gas membrane distillation, SGMD; vacuum membrane distillation, VMD and air gap membrane distillation, AGMD) and three hybrid modes (thermostatic sweeping gas membrane distillation, TSGMD; liquid gap membrane distillation, LGMD; and material gap membrane distillation, MGMD) [14]. In fact, MD with its distinct modalities has proved its effectiveness not only in common applications such as desalination or removal of volatile organic compounds (VOCs), but also in lesser-known practices such as the concentration of traditional Chinese medicine and dairy solutions or treating endocrine-disrupting chemicals [15-19]. Some important MD details have thoroughly been gathered in the following review papers [20-24].

Bibliometric (scientometric) analysis, first mentioned by Otlet in his book *"Traité de Documentation"* (1934) [25] and used by Alan Pritchard in 1969 [26], is a way of evaluating various academic parameters of a given published literature under study [27-29]. It is a qualitative and numerical technique that seeks to identify the scholarly impact and characteristics of manuscripts within a specific research area, potentially providing valuable information to researchers involved in the development of this research area [30]. The methodologies followed in scientometric analyses are valuable in detecting, categorizing, defining growth patterns, and evaluating the key elements, and also help to understand the scientific production, the important topics, the most researched areas and mapping journals among others [31, 32]. It is worth quoting that bibliometric analysis were carried out by researchers in many fields such as membrane water treatment [33], capacitive deionization [34], forward osmosis [35] and disinfection by-products in drinking water [36].

Data analysis (DA) is commonly referred to a set of practices performed by humans in terms of understanding and defining raw data to produce new knowledge [37]. DA is emerging as a promising field for providing insights from large amounts of data in order to improve certain outcomes while lowering costs [38]. These kinds of methods are generally carried out on the basis of large data sets collected, derived from uniform or diverse sources [39]. Exploratory data analysis (EDA) is a step in the data analysis that involves visualizing data, identifying key characteristics, exposing hidden structures, identifying abnormalities and providing a better understanding through the use of visual methods and advanced techniques [40, 41].

Machine learning (ML), which is a field of artificial intelligence, seeks to predict an outcome by extracting patterns from big datasets, usually in the form of an algorithm [42, 43]. Due to the complexity of many systems in the field, ML has become increasingly significant as a data-driven technique [44]. Recently, it takes a massive role in time series analysis, image processing, cyber security, etc. [45-47]. In general, ML algorithms are divided into three classes: supervised learning, unsupervised learning, and reinforcement learning. The supervised learning algorithms, such as classification or regression, have the ability to train a classifier utilizing known inputs and output(s) to predict new data. The unsupervised learning algorithms, such as clustering and dimension (dimensionality) reduction (DR), are capable of discovering hidden patterns in given data. The reinforcement learning algorithms can lesson from previous experiences and identify the perfect actions in an unfamiliar environment in order to achieve the ideal state transition for reaching the goal. These approaches have found their way into numerous applications including healthcare, transportation, speech analytics, computer vision, market analysis, life sciences, etc. [48, 49]. In ML the process of extracting interesting and non-trivial patterns or knowledge from text documents is referred to as text mining (TM). This becomes crucial when the size of the document collection is big and manual content analysis is impossible to cluster, summarize, visualize, detect topics, extract concepts, etc. [50, 51]. Word cloud technique in TM is the graphical representation of texts, which is efficient when describing massive amounts of data. Word cloud is an amazing tool for visually interpreting data for gaining a fast information. In the interface of this approach, the most repeated words appear bigger than the lest repeated words[52]. Another approach in text mining is sentiment analysis also known as opinion mining or tendency analysis. It is an interesting and increasingly applied popular procedure of analyzing subjective texts with emotional subtexts [53]. It is a way to determine textual human comments as negative, positive or neutral emotions [54]. It allows researchers to get a sense of how the authors feel about certain topics [55]. Data classification, clustering, visualization and dimension reduction are very essential tasks in numerous scientific and engineering fields to identify the key data components or trends [56].

This study takes a glance to the published peer-reviewed MD papers from a broad perspective with bibliometric methods, exploratory data analysis and machine learning approaches. The main motivation of this study is to contribute to the future research strategies of researchers by presenting the MD literature. For this purpose, main analysis is introduced to the readers with graphics and statistics with the help of Orange Data Mining Tool, R and Python programming languages and Exploratory Tool.

2. Data and Methodology

2.1. Data

The used data was gathered from Scopus database. We chose Scopus database because it covers more than 20,000 journals including major publishers like Elsevier, Springer, Emerald and Taylor & Francis [57]. In addition, Scopus database provides the advantages of covering papers that fulfill a stringent set of indexing requirements, has more content than other databases and features, which make citation analysis and visual mapping easier [58-60].

In order to deepen the developed analysis, the reference counts of the papers, which was not in the downloaded data set, was added manually. The used data was downloaded in 02.02.2022 with the following searching criteria: TITLE-ABS-KEY (".........") AND (LIMIT-TO (SRCTYPE, "j")) AND (LIMIT-TO (PUBSTAGE, "final")) AND (LIMIT-TO (DOCTYPE, "ar")) AND (EXCLUDE (PUBYEAR, 2022)) AND (LIMIT-TO (LANGUAGE, "English")). The following keywords were written in the spaced place in TITLE-ABS-KEY "membrane distillation", "air gap membrane distillation", "direct contact membrane distillation", "vacuum membrane distillation", "vacuum enhanced membrane distillation”, "sweeping gas membrane distillation", "sweep gas membrane distillation", "membrane air stripping", "thermostatic sweeping gas membrane distillation", "permeate gap membrane distillation", "liquid gap membrane distillation", "water gap membrane distillation", "conductive gap membrane distillation", "material gap membrane distillation", "MD", "AGMD", "DCMD", "VMD", "VEMD", "SGMD", "MAS", "TSGMD", "PGMD", "LGMD", "WGMD", "CGMD", and "MGMD". The formerly used names for MD before the workshop held in Rome in 1986 like "transmembrane distillation", "thermopervaporation", "pervaporation", "membrane pervaporation", "capillary distillation" and also early articles that are not reflected in the searching results was added to the data set manually [61]. Subsequently, the collection was screened to remove the unrelated papers. The resulting dataset include 3227 papers. Depending on the searching criteria, MD configurations were classified as in Table 1 and the number of documents based on the configuration name was found to be 2052.

Table 1. Classification of MD configurations following the adopted searching criteria.

MD Configuration	Searching Keyword
DCMD	"DCMD" + "direct contact membrane distillation"
VMD	"VMD" + "vacuum membrane distillation" + "VEMD" + "vacuum enhanced membrane distillation"
AGMD	"AGMD" + "air gap membrane distillation"
SGMD	"SGMD" + "MAS" + "sweeping gas membrane distillation" + "sweep gas membrane distillation" + "membrane air stripping"
LGMD	"LGMD"+ "liquid gap membrane distillation" + “WGMD” + “water gap membrane distillation” + "PGMD" + "permeate gap membrane distillation"
MGMD	"MGMD" + "material gap membrane distillation" + "CGMD" + "conductive gap membrane distillation"
TSGMD	"TSGMD" + "thermostatic sweeping gas membrane distillation"

It must be pointed out that MD papers that are not indexed by Scopus or non-ISI indexed sources were not included in the data set. However, since Scopus is the largest database for ISI indexed publications, we assume that our analysis captures the main statistical trends. Screening and filtering explorations have been applied to ensure that only MD documents were selected. In order to maintain consistency in the developed text mining approaches, only articles published in English were collected. Although the dataset was downloaded in February 2022, articles published in 2022 were excluded. The reason for this type of filtration is to enable comparisons to be made in annual based analysis.

2.2. Methodology

The research was conducted based on three types of analysis, bibliometric, exploratory data and machine learning analysis. The analysis and graphics included in the study were obtained with a combination of R (Biblioshiny package), Exploratory, Orange and Python programs. In some cases, the data is processed by one program and visualized in another in order to create more descriptive figures. R–Biblioshiny is an open-source powerful package, recently developed (in 2017), dedicated to scientometric analysis [62]. With the help of this extension, mapping of the collection based on authors, sources, countries and articles were done and the results were analyzed. Exploratory Tool is a program that enables users to quickly discover data and explore insights with a simple interface [63]. Orange data mining tool was developed by Janez Demšar and Blaž Zupan for machine learning tasks [42] and Python is one of the most popular and highly used programming language [31, 64]. The analysis performed within this study gives a general information about the collection-based equations listed below.

The co-authors per document value is a measure that reflects the average number of co-authors per research computed as follows [65]:

$$\text{Co-Authors per Document} = \frac{\text{Authors Appearances}}{\text{Document}} \tag{1}$$

The collaboration index is the average number of authors per collaborative paper and it is a good metric to gain information about association in the domain. This index can be calculated as [66]:

$$\text{Collaboration Index} = \frac{\text{Authors of Multi−Authored Articles}}{\text{Multi−Authored Articles}} \tag{2}$$

Bradford's Law is an effective way to determine core journals in the data set. It states that relevant journals can be divided into three zones by an approximate number of articles (n) where each zone collects about one-third of all documents in the collection. Zone 1 (the core zone) shows the journals with the most articles in the domain, Zone 2 contains average amount of journals, and Zone 3 consists of a long tail of journals with few articles [67].

h – index (Hirsch index) is defined as the number of research papers (h) by the journal (or author), each of this has been cited at least h times in other articles. The g – index, proposed by Egghe in 2006 as an enhancement on the h – index, is a unique integer indicating that the top g articles received at least g^2 citations [65]. These indices are calculated based on the dataset by Bibliometric package [68].

The number of citations received by an article from all other papers indexed on a bibliographic database (in our study it is Scopus) is referred to as global citations. The number of citations that a publication has received from papers in the collection (dataset) is referred to as local citations [49].

The relationship between authors and the number of papers they publish is explained by Lotka's law, which is a good measure of scientific production. This law states that the percentage of researchers with n publications is proportional to the quotient $1/n^2$ [69]. Lotka's law is defined as [70]:

$$Y = \frac{C}{X^n} \quad (3)$$

where X represents the number of publications, Y is the frequency of authors with X publications, C and n are constants estimated by the software in the domain.

Articles Fractionized (AU_j) calculates each specific author's (j) contribution to a produced set of articles based on the assumption that all co-authors contribute equally to each document and can be calculated as follows [65].

$$AU_j = \sum_{h \epsilon AU_j} \frac{1}{number\ of\ co\text{-}authors\ (h)} \quad (4)$$

where h is a document included in AU_j of the author j in the collection.

Thematic maps are simple to use and allow to track the progression of themes across four quadrants. Each network cluster is stated by a bubble. The bubble name is the term with the highest occurrence value in the cluster. The size of the bubbles is related to the number of cluster word occurrences. The bubble position is determined by the Callon density and centrality of the cluster. The degree of inter-cluster interactions, or the extent to which a topic is connected to other topics, is measured by centrality. The density reflects the degree of intra-cluster cohesion, or how closely the terms in a given cluster are linked. The themes in the upper-right quadrant have a high centrality and density, indicating that they are both influential and well-developed. The lower-right quadrant depicts concepts that cross disciplines and have the ability to impact other topics while being weakly entrenched within. Topics that are emerging or fading are highlighted in the lower-left quadrant due of their low centrality and density. Last, the upper-left quadrant contains niche themes that are well-developed internally but unable to affect other themes [57, 68, 71].

In order to see the most repetitive words in the titles, abstracts and author keywords applied the word cloud approach and a pre-processing step was carried out. In the pre-processing step, textual data was transformed to lowercase, tokenized and stop words were removed. Tokenization is the procedure of dividing text strings into tiny pieces such as words and phrases [72]. After tokenization, word cloud method was used to summarize the most repeated words in titles, author keywords and abstracts.

Sentiment analysis was conducted using Oranges' Valence Aware Dictionary for Sentiment Reasoning (VADER) module, which divides sentiments into positive, negative and neutral polarities, as well as in the form of a compound, which is the normalized sum of all polarities in between -1 (most negative) and +1 (most positive) [73]. VADER is based on a dictionary that defines lexicons to emotion intensities, which are identified as sentiment scores. The sentiment score can be computed by summing the intensity of each word in the text [74, 75].

3. Results and Discussions

3.1. Main information of MD technology: Dataset, MD configurations, membrane engineering, hybrid MD separation and applications

Preliminary statistical values are important to draw a general framework about the data set. Basic information about the collection downloaded from the Scopus database can be seen in Table 2.

The data set includes 54 (1967-2021) years of research and development of MD technology from a total of 382 sources. The average citations per document value was calculated as 4.264. A total of 5305 authors contributed to MD research field, and the document per author value was determined as 0.608. These data show how popular MD is in the scientific community and it is a subject that has been studied by many researchers. The authors per document value (1.64) indicates that the articles have few authors. This phenomenon was also confirmed by the collaboration index, which is found to be 1.69, indicating that MD researchers are working together at a moderate level (MD researchers do not collaborate that much). Therefore, it is about time to start establishing scientific collaborations in MD research field. 56 authors wrote a total of 120 articles as a single author. The annual scientific papers published on MD technology can be seen in Fig. 1.

Table 2. Main information about the dataset.

Information	Result
Timespan	1967:2021
Sources (Journals)	382
Documents	3227
Average years from publication	7.13
Average citations per documents	34.37
Average citations per year per document	4.264
References	81416
Author's Keywords	5049
Authors	5304
Author Appearances	14632
Authors of single-authored documents	56
Authors of multi-authored documents	5249
Single-authored documents	120
Documents per Author	0.608
Authors per Document	1.64
Co-Authors per Documents	4.53
Collaboration Index	1.69

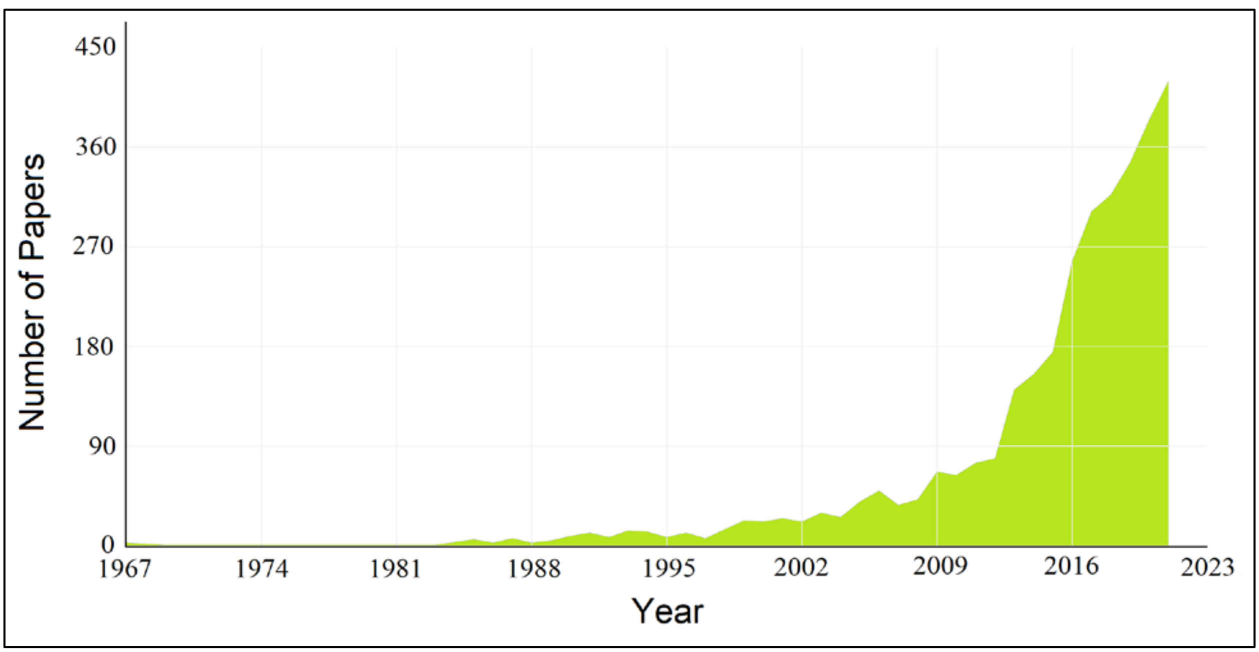


Figure 1. Yearly published peer-reviewed papers on MD in refereed journals.

The yearly scientific production graphic can provide important information about the change in popularity of the domain over time. After the first publication in 1967, MD was a field of study that did not attract attention until the early of 1980s. However, it is understood that MD being the center of attention by the researchers since 1983. It is clearly seen in Fig. 1 that the year MD started to stand out was 2012. Since this year, the number of MD studies has increased exponentially. In the next stage of the main information process, a scatter matrix was created for the published year, page count, cited by and reference count values of the papers in the data set and the results can be seen in Fig 2.

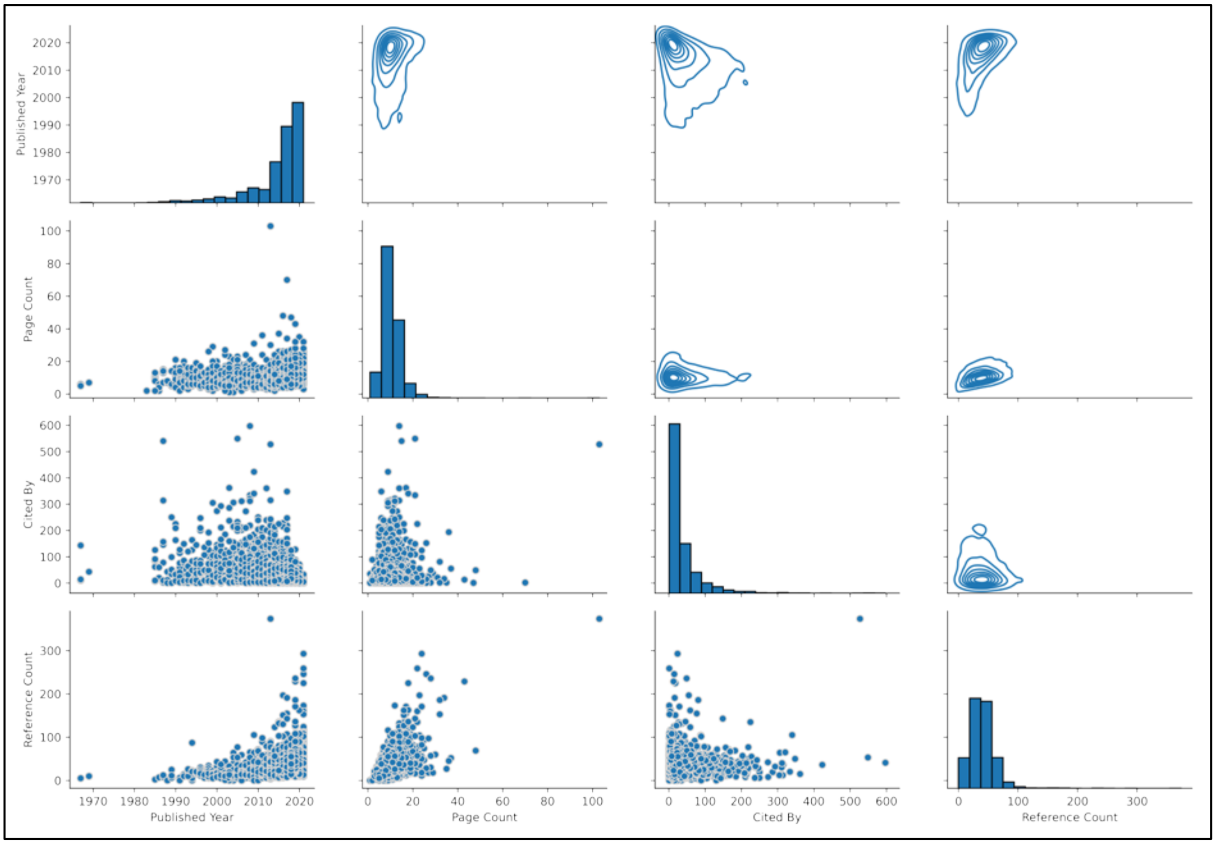


Figure 2. Scatter matrix of published year, reference count, page count and cited by values of papers in the dataset.

In Fig. 2 the diagonal subplots are the histogram plots of the interested data. The reported published year data is the same data in Fig. 1 thus the same interpretation can be applied. It has been revealed that the number of pages of MD papers are mostly between 6-11 pages. The MD papers in the data set were cited mostly in the range 1-30 (1941 publications) and commonly contain references between 18-56. Scatter plots were drawn to investigate the correlation between reference count, cited by, page count and published year variables. The results show that there is an increase of the number of pages and reference counts of the papers every year. On the contrary, there is a decrease of the number of citations received. It was determined that there was a linear relationship between the page count and reference count, as the pages of the papers increase the references increase too. However, there was no significant correlation between the page count and the cited by values. This means writing longer papers does not affect the citations received. Similarly, no correlation could be observed between cited by and reference count. Multivariate kernel density estimation plots were sketched to investigate the hotspots of binary data (i.e. the correlations above the diagonal plots) and to compare dataset across the same variable. Pairwise densities can be read easily in the upper triangle of Fig. 2. While the diagonal histogram plots show the intensities for a single data, each sub-plot in the upper triangle provides the opportunity to examine the dense points between attributes. As can be seen in the figure pairwise densities of sub-sketches all have single concentrated points, which reflects that attributes do not have irregularities.

As stated earlier, seven MD configurations have been proposed (DCMD, SGMD, VMD, AGMD, TSGMD, LGMD, MGMD). Some initial investigations on MD configurations have been applied with exploratory data analysis to visualize and try to discover patterns, and to highlight statistics as well. The classification method depending on the searching keywords has been already explained and summarized in Table 1. The distributions, the yearly scientific production and the Venn diagram of the different MD configurations are plotted in Fig. 3.

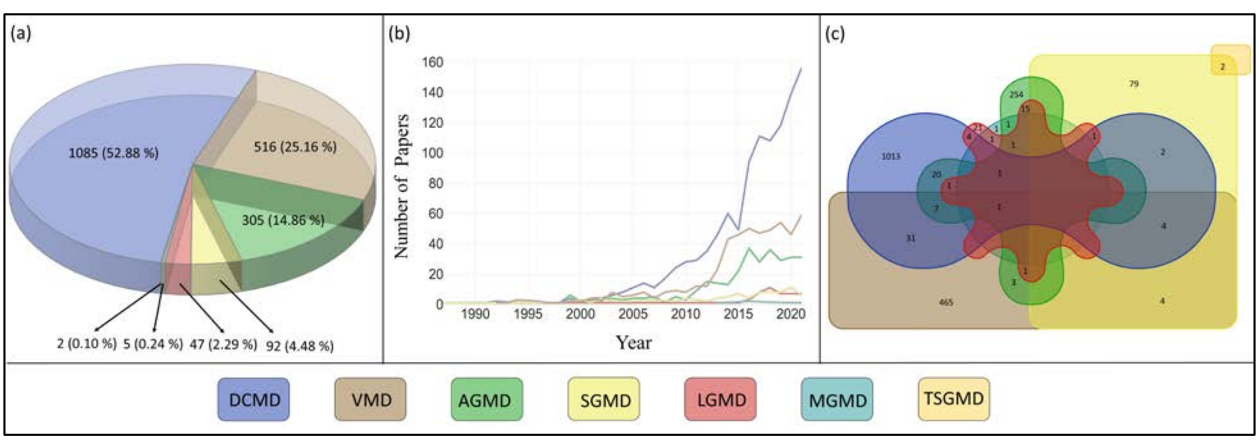


Figure 3. MD configurations **(a)** distribution by searching criteria **(b)** yearly scientific production and **(c)** Venn diagram.

As can be seen in Fig. 3(a) the most used configuration is DCMD. 1085 papers are related with DCMD (52.88%). This is because of the various advantages of this configuration such as the simplicity of the DCMD systems taking into consideration that condensation step is carried out at the permeate/membrane surface inside the membrane module [76]. The other two most used MD configurations are VMD (516 documents) and AGMD (305 documents). From Fig. 3(b) it can be observed that the annual distribution of DCMD published papers was always ahead of the other configurations, especially since 2005 when it has attracted more attention of researchers. Although the difference between DCMD and VMD published papers in 2015 seems to be very close, the subsequent increase of the DCMD number of published papers compared to those of VMD became significant. It is clear that the use of configurations such as LGMD, MGMD and TSGMD is limited in scientific studies.

Venn diagrams are useful for illustrating connections, similarities and contrasts between several data items or classes [77]. In this case, some papers mentioned more than one MD configuration. Therefore, a Venn diagram was created and plotted in Fig. 3(c) to illustrate the number of papers indicating different MD configurations. DCMD and VMD are the two most mentioned configurations in the same papers (31 papers). This statistic is an indication that these two competing configurations have been compared a lot by researchers to understand their pros and cons. DCMD and AGMD, which have been mentioned in 20 articles, come in the second place. Another remarkable point in Fig. 3(c) is that LGMD and AGMD are mentioned in 15 joint papers because both have been investigated in the same research studies taking into account the similarity of the used systems.

MD membrane engineering is one of the important active research line necessary for the development of MD technology, thus we have identified those papers focusing on membrane preparation and modification. Fig. 4 shows the annually published papers on MD membrane engineering while Fig. 5 indicates the membrane configurations for which the membranes were designed. In fact, the MD engineered membranes can be used in practically all MD configurations while the membrane pores are maintained dry (i.e. the liquid entry pressure of the membrane, *LEP*, should be as high as possible). However, the MD performance of some prepared or modified membranes is higher for a given MD configuration compared to the others and the risk of membrane pore wetting is much higher in VMD. For instance, the dual-layered hydrophobic/hydrophilic membranes are designed for DCMD and LGMD as the cold liquid permeate penetrates inside the big pores of the hydrophilic layer reducing the distance between the feed and permeate liquid/vapor interfaces and thus the water production rate is high.

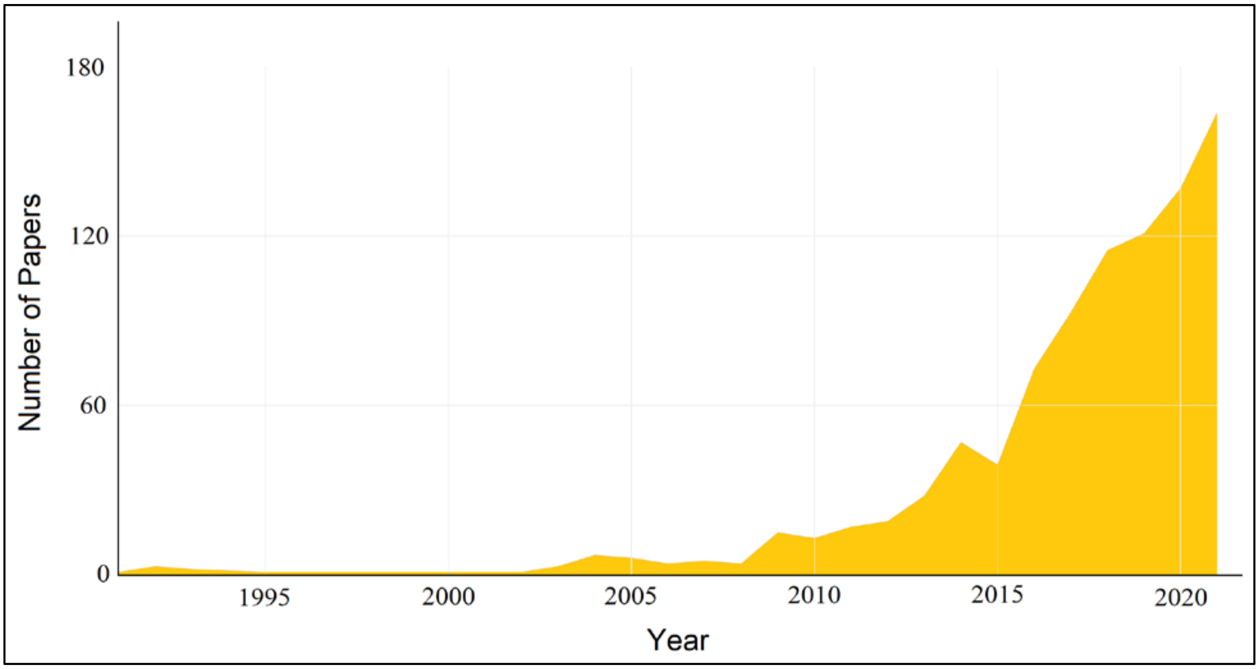


Figure 4. Yearly publications of MD membrane engineering.

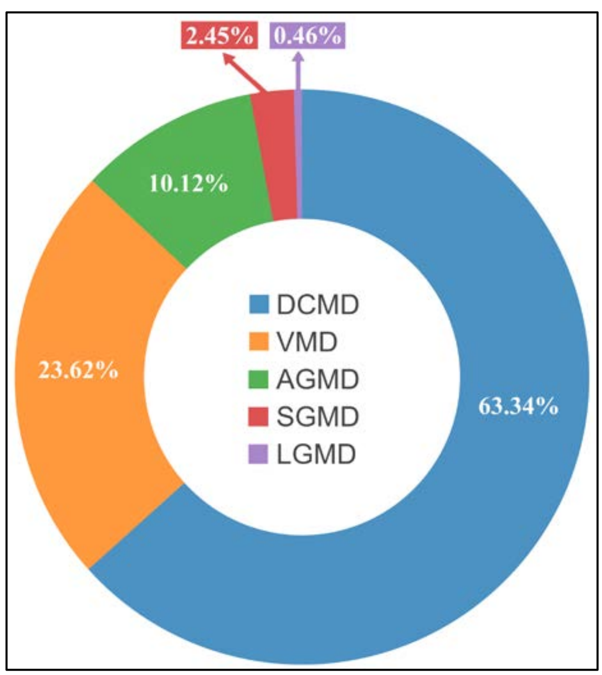


Figure 5. MD configurations for which membranes were engineered.

It can be seen in Fig. 4, the significant worldwide interest on MD membrane engineering since 2010 when the number of published papers starts to increase exponentially. This is a clear indication that this field of MD research is important and it will be maintained active for years. This is necessary for the industrial implementation of MD technology. It should also be noted that Fig. 4 shows certain parallelism with Fig. 1, which also shows the significant annual increase of MD studies since 2010. In addition, similar to Fig. 3, from Fig. 5 indicates that the most preferred MD configurations is DCMD, followed by VMD and AGMD. Among a total of 963 MD engineering papers, only 2.45% papers were published on SGMD and 0.46% papers of LGMD. It must be pointed out that membranes have not been prepared yet for the two configurations MGMD and TSGMD. The word cloud of the abstracts of the published papers on membrane engineering is presented in Fig. 6.

Word	Count
membrane	4926
membranes	2583
water	1636
pvdf	1355
surface	1350
distillation	1228
flux	1192
md	956
contact	955
high	796

Figure 6. Word cloud of abstracts of published papers on membrane engineering.

From Fig. 6 it can be seen that "*polyvinylidene fluoride (PVDF)*" is the most used polymer by MD membranologists. In addition, the term "*surface*" is the most frequent used word indicating that MD researchers are very interested in membrane surface modification. Moreover, the word "*flux*", is mentioned frequently showing that the main aim of the MD researchers is to develop membranes with improved MD permeate flux. The term "*contact*" appears repeatedly in abstracts showing that DCMD is the most preferred MD configuration and that researchers are interested in increasing the water contact angle of the designed membranes.

Various techniques have been employed for flat sheet, hollow fiber and nanofibrous membrane preparation and membrane surface modification (nonsolvent induced phase separation, NIPS;

thermally induced phase separation, TIPS; electrospinning; etc.). It must be pointed out that some proposed modified membranes were prepared by phase inversion using surface modifying macromolecules (SMMs) that migrate to the polymer/air interface during membrane formation [78]. Other modified membranes were prepared by radiation graft polymerization, plasma polymerization, grafting ceramic membranes, surface coating or dip coating, etc. Single, dual- and triple-layered membranes, either polymeric, ceramic or mixed matrix membranes, have been proposed for MD. A lot and considerable efforts have been made to develop new MD membranes with desirable structures and improved performance. Fig. 7 presents the classification made according to membrane material, preparation technique, membrane type and polymers used in membrane engineering papers.

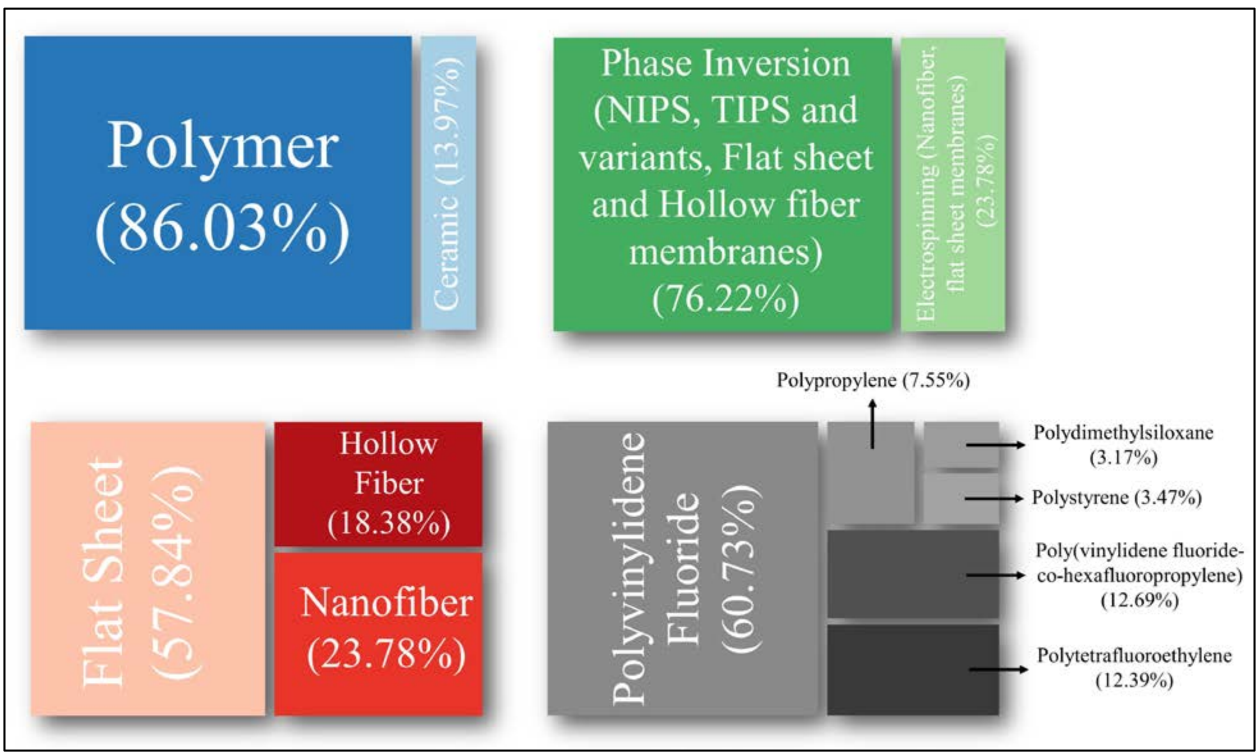


Figure 7. Details of MD membrane engineering: membrane material (blue), membrane preparation technique (green), membrane type (red) and polymer used (grey).

As can be seen in Fig. 7, polymeric membranes are the most dominant in membrane engineering contributing with 86 % of the prepared and modified membranes while the rest is ceramic based membranes. The most considered technique for membrane preparation is phase inversion (i.e.

76.2% either NIPS, TIPS or their variants for preparation of flat sheet or hollow fiber membranes). Although the electrospinning technique was proposed the first time in 2008 for the design of MD nanofibrous membrane [79] and the subsequent great interest in electrospun nanofibrous webs for MD because of various advantages such as their high void volume fraction and low heat transfer by conduction improving the thermal efficiency of the MD process [4, 80, 81], this technique is still less used than the phase inversion technique. Among the proposed membranes for MD technology, 57.84% are in flat sheet form, 23.78% are nanofibrous membranes and only 18.38% are hollow fibers. Taking into consideration the used technique, although nanofibrous membranes are also flat sheet membranes, we separated them from the flat sheet membrane group including either the modified or prepared membranes by the phase inversion method. As stated previously, PVDF is the most used hydrophobic polymer with a dominant percentage of 60.73% because it can be dissolved in a variety of solvents and porous membranes can be prepared easily by NIPS technique. Other polymers like polytetrafluoroethylene (PTFE) and poly(vinylidene fluoride-co-hexafluoropropylene) copolymer (PVDF-HFP) are considered less although both exhibit a higher hydrophobic character than PVDF. It is striking the few types of polymers that have been used so far in MD membrane engineering. A changing strategy is to promote the use of sustainable polymers and green solvents/diluents in order to implement the circular economy principles.

MD process is often combined with other water treatment processes in order to complete the separation procedure, improve the separation efficiency, mitigate MD membrane fouling, and reduce the water production cost and specific energy consumption among others. Fig. 8 indicates the combination of MD with other separation/treatment processes with an upset graph, which is a great tool to visualize intersections of multiple sets. As can be seen in Fig. 8, the detected processes combined with MD can be divided into 2 groups. The first group (highlighted with purple color) includes membrane-based separation processes such as microfiltration (MF), ultrafiltration (UF), nanofiltration (NF), reverse osmosis (RO), electrodialysis (ED), membrane crystallization (MCr), membrane bioreactor (MBR) and forward osmosis (FO). The second group (highlighted with blue color) contains other types of separation processes such as electrocoagulation (EC), coagulation/flocculation/precipitation (CFP), adsorption (Ads), oxidation (OX), aeration (Aer), freeze desalination (FD), mechanical vapor compression (MVC) and multi effect evaporation (MEE) among others. It must be pointed out that ED also includes electrodialysis reversal (EDR), in which the electrode polarity is inverted for a certain time to reduce both concentration polarization and scaling on the membrane surface [82]; and mechanical vapor recompression was considered within MVC. Aeration pretreatment was employed to mitigate membrane fouling by organic matter that may result in MD membrane wetting.

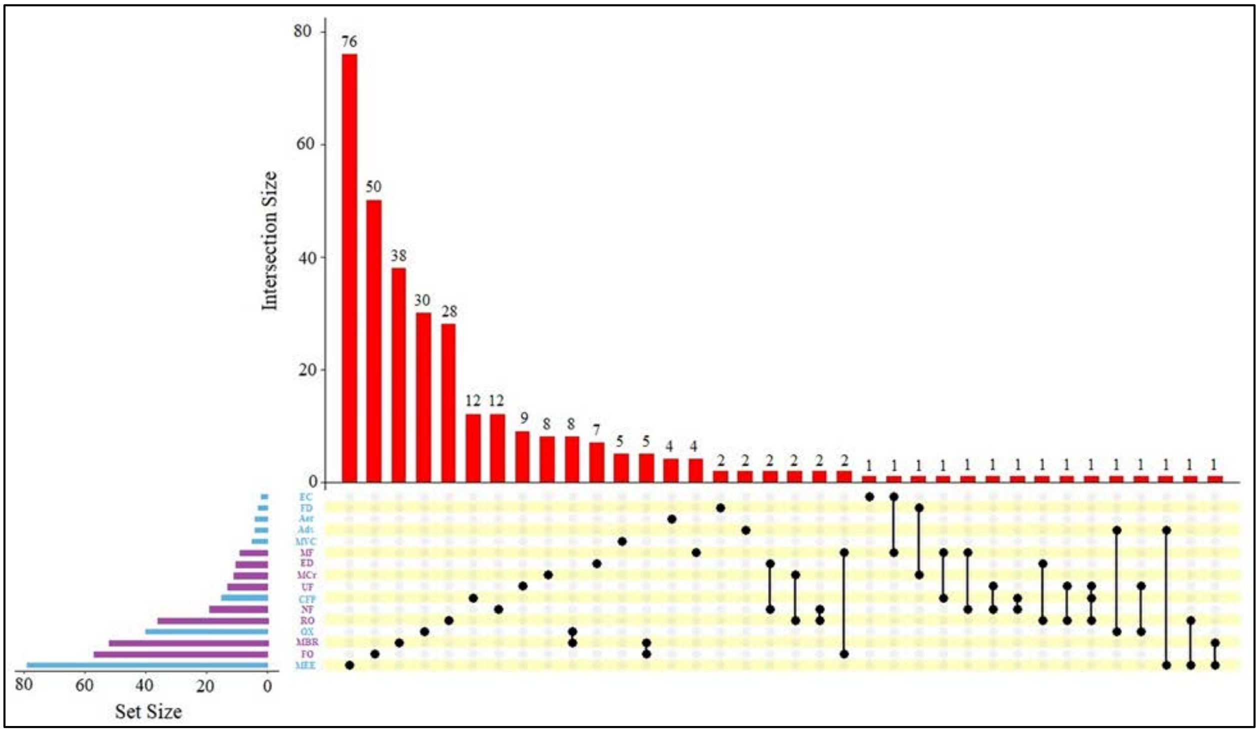


Figure 8. Combined separation processes with MD technology: (purple) [Membrane processes: microfiltration (MF), ultrafiltration (UF), nanofiltration (NF), reverse osmosis (RO), electrodialysis (ED), membrane crystallization (MCr), membrane bioreactor (MBR) and forward osmosis (FO)]; (blue) [other processes: electrocoagulation (EC), coagulation/flocculation/precipitation (CFP), adsorption (Ads), oxidation (OX : H_2O_2/Fenton/ultra-violet/photocatalysis/O_3/electrochemical oxidation/Zero valent Iron), aeration (Aer), freeze desalination (FD), mechanical vapor compression (MVC), multi effect evaporation (MEE : multi stage/multi effect distillation/multi effect evaporation)].

As can be seen in Fig. 8, the research on combined processes with MD technology has been studied by various researchers and MD has been combined not only with another single process (44.4% of the total number of MD combined systems) but also with two other treatment processes (52.8% of the total number of MD combined systems) and with three other processes (2.8% of the total number of MD combined systems). Among these MD combinations, the most preferred one is MEE with a contribution of 76 papers in order to reduce the specific energy consumption, followed by FO membrane technology with 50 papers in order to regenerate the used draw solutions by MD, and MBR with 38 papers in order to reduce MD membrane fouling and avoid wetting of membrane pores increasing therefore the quality of the produced water. It is clear that EC, FD, Aer, Ads, MF and MVC are less hybridized with MD and must gain more attention to explore further their benefits for MD technology, which has been used for the treatment of different water sources such as sea and brackish waters, industrial wastewaters (food, pharmaceutical, radioactive, municipal, etc.). The different fields of MD applications are plotted in Fig. 9.

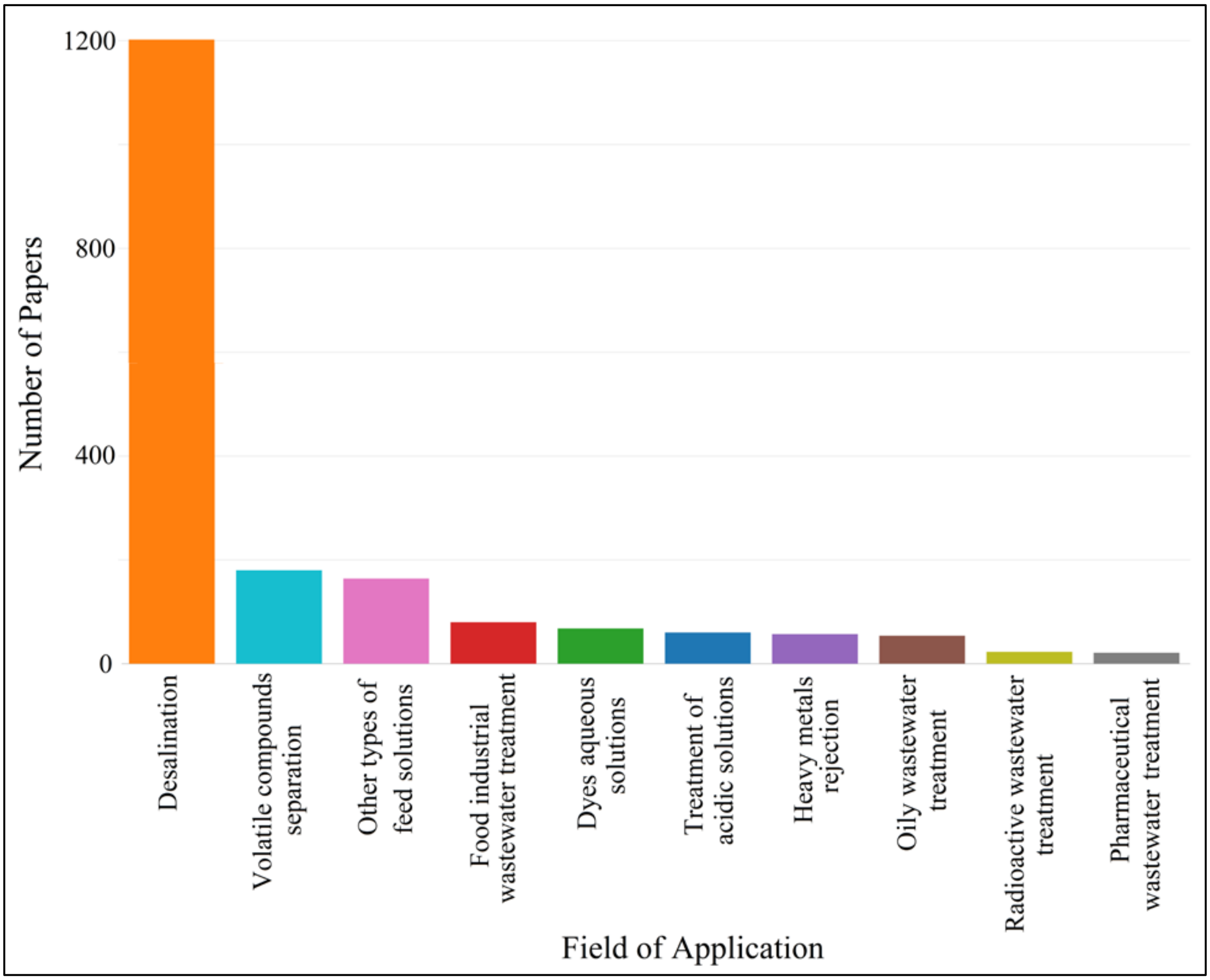


Figure 9. Fields of application of MD technology.

MD is preferred mostly for desalination (i.e. 37.25% of the total number of published MD papers). In desalination application real sea water, model feed solutions using various types of salts ($CaCl_2$, NaCl, $CaSO_4$, $Ca_3(PO_4)_2$, etc.), groundwater, geothermal water and brackish wastewater were used. MD is adequate of desalination because it can treat saline solutions up to their saturation including brines of reverse osmosis (RO) desalination plants producing not only distilled water but also ultrapure water for semiconductor, pharmaceutical, and hydrogen production industries. The second MD application is the treatment of water containing volatile compounds (with only ~6% contribution) being the most considered compounds ethanol and ammonia. The third largest group of MD application with ~5% contribution, named “Other types of feed solutions” corresponds to the treatment of different types of wastewaters such as municipal wastewater, landfill leachate, pesticide wastewater, plasma ultrafiltrate, etc. The treatment of other wastewaters originated from different types of industries (food, pharmaceutical, radioactive, dyes, acids, heavy metals, oils, indicated in Fig. 9) were not included within the group “Other types of feed solutions” because their own MD application is substantial compared to the others although it is clear that more MD research studies are still needed for these types of wastewaters in order to recover valuable components and simultaneously produce distilled water preserving the characteristics of MD membranes.

It must be mentioned that an adequate selection of the MD configuration dictates the efficiency of the MD separation and the specific energy consumption. In MD field desalination application, the distribution of the used configurations is illustrated in Fig. 10. Again, it was found that DCMD is the most used configuration in desalination (51.26%), which doubled that of VMD configuration (25.00%). The AGMD configuration comes in the third place with 17.47%, whereas the rest of MD configurations (SGMD, LGMD, MGMD) were not used much for desalination. The hybrid MD configuration TSGMD hasn´t been applied yet in desalination.

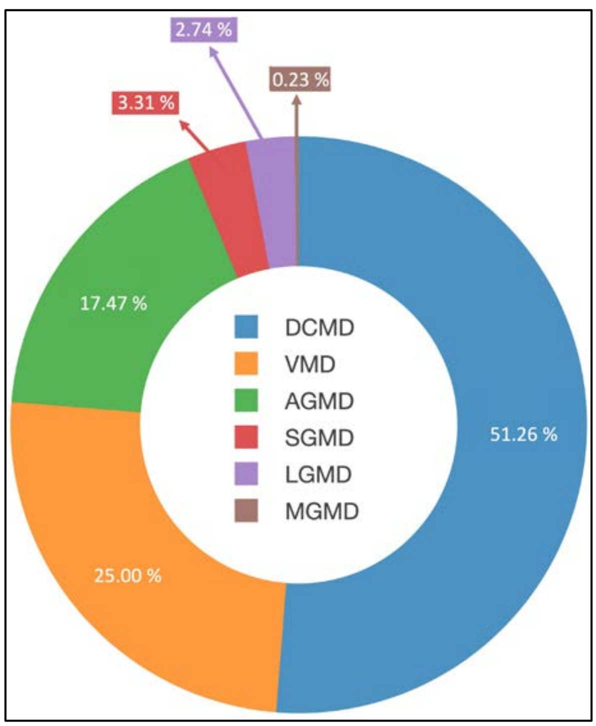


Figure 10. MD configurations used in desalination.

We have also investigated the application of the different MD configurations for the removal of volatile compounds and we found that VMD is the most preferred configuration with 46.22%. This may be attributed to the risk of membrane pore wetting from the permeate side that is absent for this configuration. In fact, applying DCMD or LGMD configurations for the removal of volatile compounds enhances the risk of wetting the pores of the membrane from the permeate side due to their high concentration, which reduces the surface tension of the permeate solution resulting in membrane pore wetting. The following used configurations for the removal of volatile compounds were found to be: DCMD (20.16%) > SGMD (17.65%) > AGMD (15.13%) > LGMD (0.84%). It draws attention that DCMD is the second most used configuration for the removal of volatile compounds although the risk of pore wetting from the permeate surface of the membrane is high. This occurs because the used membranes in this case may exhibit very high hydrophobic character and high *LEP* values to withstand the low surface tension of the resulting permeate aqueous solutions containing high concentrations of volatile compounds.

Similar to other separation processes, the application of renewable energy sources in MD technology, such as solar, wind and geothermal sources, is also of a great interest in order to reduce the specific energy consumption. Among the collected MD papers, the research studies using solar energy are 92%, whereas those considering geothermal sources, both solar and wind energies, and the three renewable energy sources together are only 5.3%, 2.00% and 0.7%, respectively. For solar energy applications the mostly preferred MD configurations are VMD, DCMD, AGMD, LGMD and SGMD, which contribute with 36.6%, 30.1%, 28.0%, 3.5% and 2.1%, respectively.

3.2. Sources publishing MD papers

The bibliometric methods are very useful approaches to identify the top sources of the collected papers in the dataset. Fig. 11 shows the top 10 most important journals publishing MD papers and the yearly cumulative MD publications of these journals.

The most remarkable point in Fig. 11(a) is that among all journals, two are in the forefront: Journal of Membrane Science and Desalination with 18.1% (584 papers) and 17.8% (574 papers) of MD publications, respectively. These are followed by the journals Desalination and Water Treatment (7.8%), and Separation and Purification Technology (5.9%). Figure 3(b) includes the cumulative annual MD published papers in the top 10 journals. It is clear that Journal of Membrane Science and Desalination have given quite similar importance to MD subject for many years. The number of MD publications in these two journals each year are very close. It can be seen that the number of MD papers published by these journals increased particularly since the beginning of the 1990s. In contrast, those published in the journal Desalination and Water Treatment appeared only since 2009, which coincides with the year of the journal foundation. In Fig. 12, the top 10 journals with the most local citations, based on our collected dataset, are plotted.

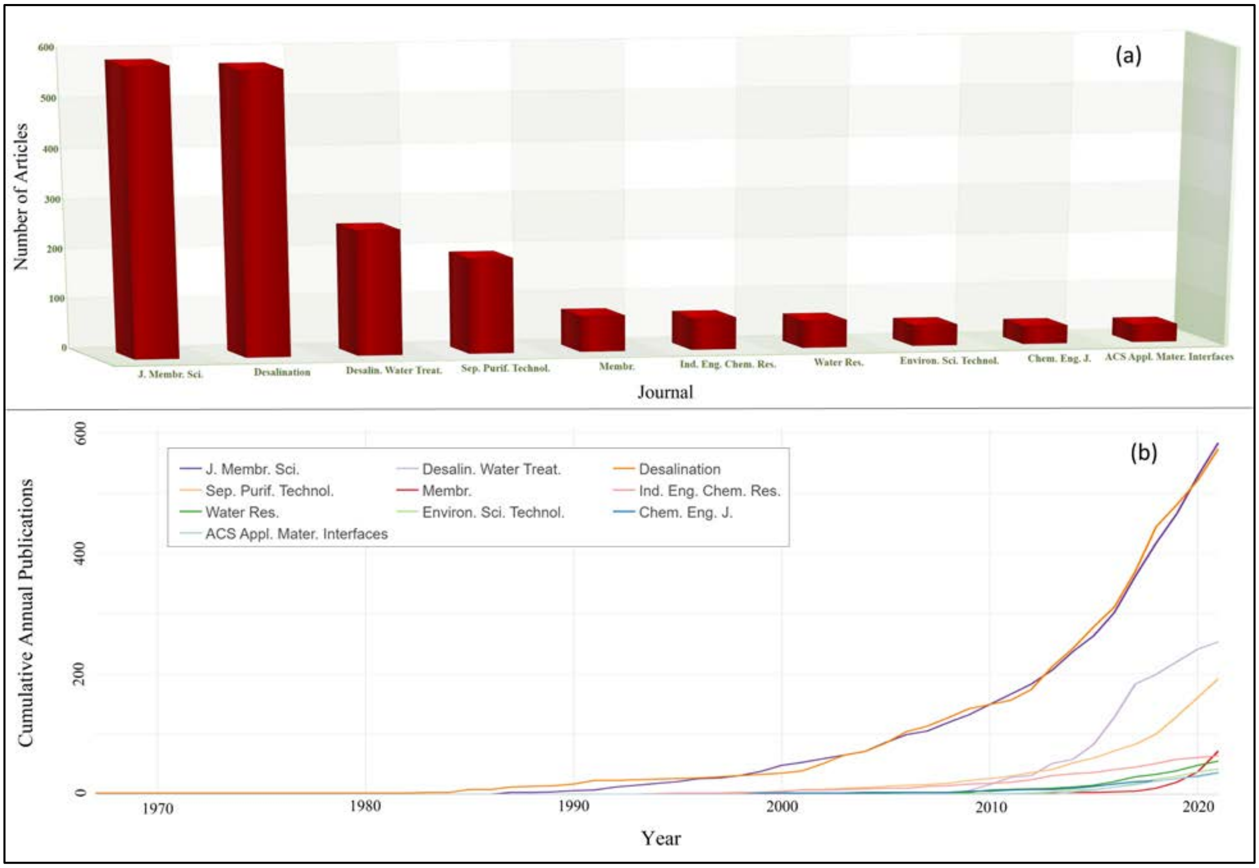


Figure 11. Top 10 (a) Most relevant sources publishing MD papers and (b) Their cumulative annual publications

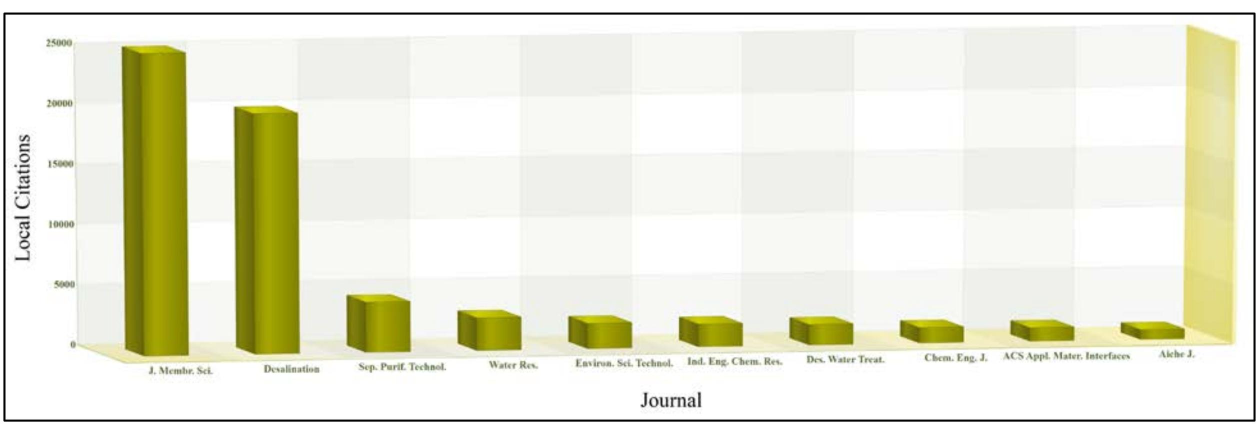


Figure 12. Top 10 most locally cited sources.

Although Journal of Membrane Science and Desalination are the most cited sources in Fig. 12, Journal of Membrane Science is one step ahead of Desalination in local citations. As of 02.02.2022 the date that the data set was fetched, Journal of Membrane Science received 23089 local citations while Desalination received 19915 local citations. It is to point out that although AIChe Journal is not within the top 10 most relevant sources (see Fig. 11), it is ranked 10th in local citations list (Fig. 12). This indicates that journals do not need to publish high number of papers but good quality MD studies (impact factor). This is also observed for the journal Desalination and Water Treatment, which is scored the third in terms of number of published MD papers in Fig. 11 but in the seventh place in terms of locally cited sources in Fig. 12. In this sense Bradford's Law is an effective method for identifying key journals in a data collection. Fig. 13 shows the Bradford's Law of the dataset.

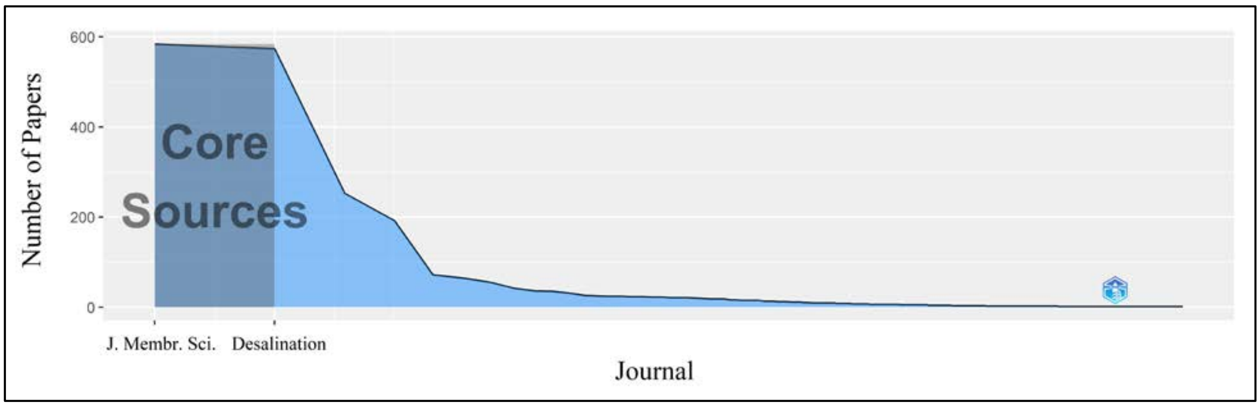


Figure 13. Bradford's law of MD technology.

This figure appears as a complementary illustration to Figs. 11 and 12. In our previous analysis, it is found that Journal of Membrane Science and Desalination are the leading journals in the MD field of research. These two journals cover approximately 36 % of the MD published papers and therefore appear as the core journals. The number of papers (quantity) published by a journal alone does not reveal the influence of that journal. In today's world, it is more important to show how effective a journal is in the scientific field by calculating metrics qualitatively. Two of the most important of these metrics are *h*-index and *g*-index. Fig. 14 shows the top 10 journals in the dataset based on *h*-index and their corresponding *g*-index.

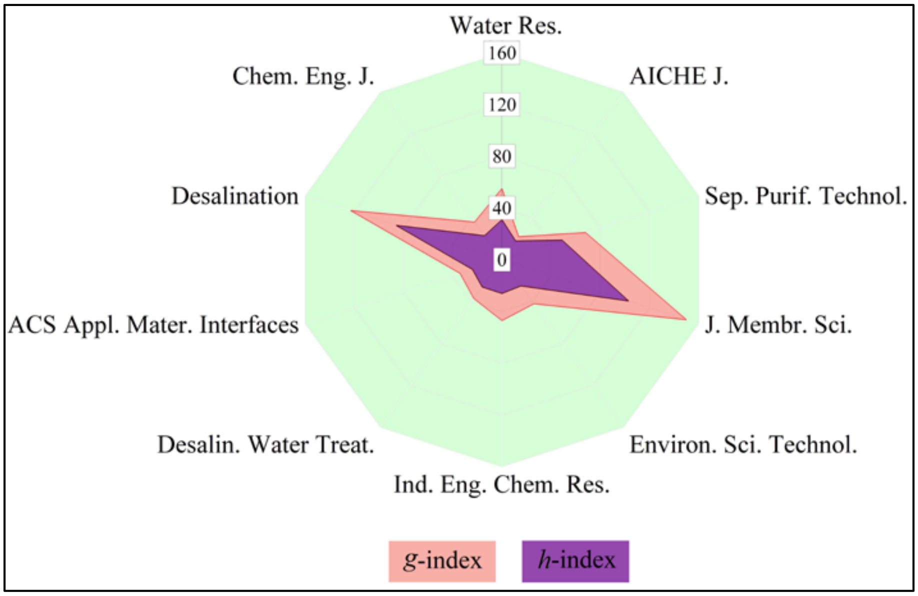


Figure 14. Impact of the top 10 sources publishing MD papers.

Fig. 14 proves that Journal of Membrane Science and Desalination are the two most important journals publishing MD papers, not only quantitatively but also qualitatively. The calculated *h*-index and *g*-index of the Journal of Membrane Science are 103 and 150, respectively. For Desalination these are lower, 86 and 123, respectively. The *h*-index and *g*-index of the other journals following Desalination are lower than 50 and 69, respectively.

3.3. MD researchers

The use of bibliometrics to find out significant authors' productivity is a good tool revealing the most important contributors to a given research area of interest. Fig. 15 shows the top 10 most productive and influential authors in the MD field.

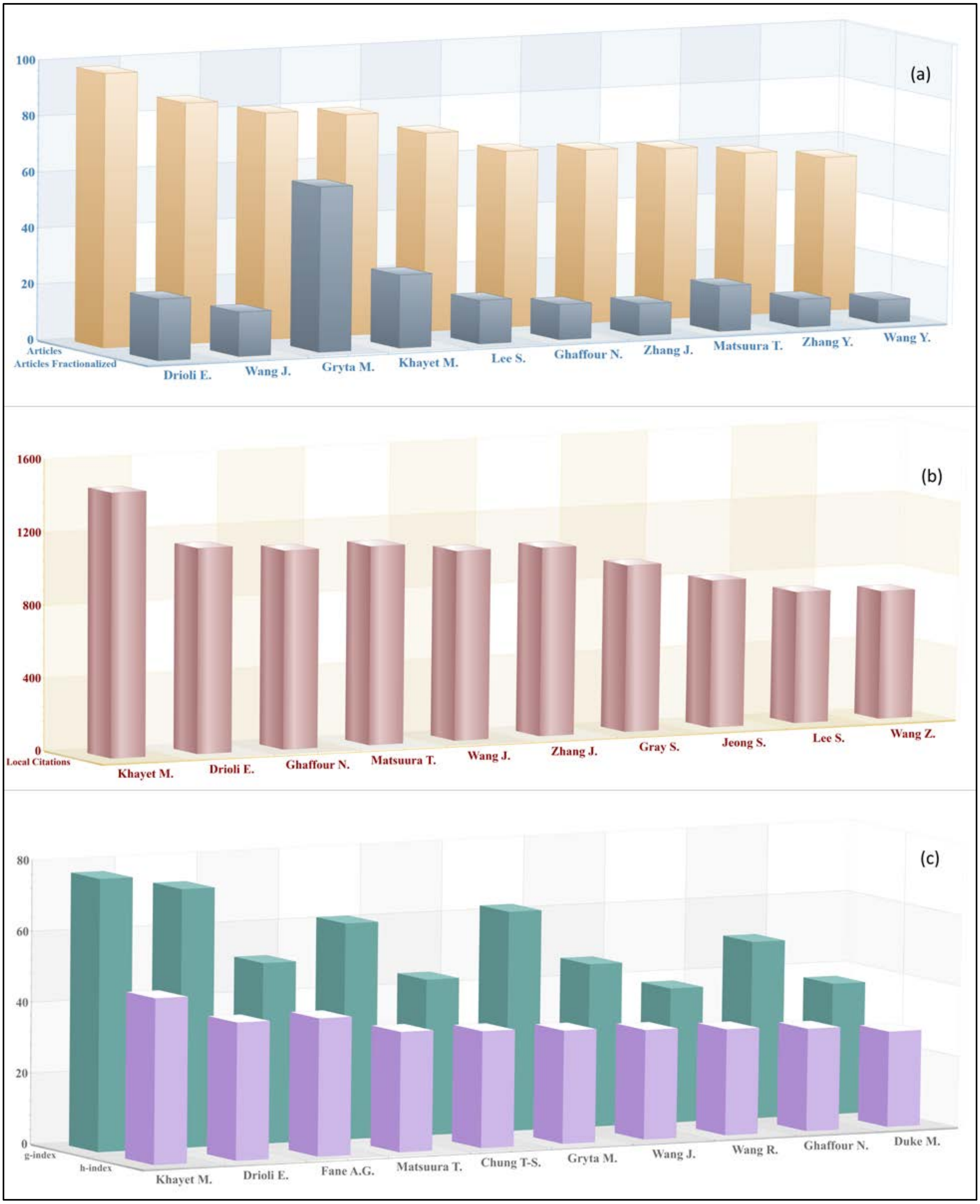


Figure 15. Top 10 authors **(a)** number of publications **(b)** local citations and **(c)** impact indexes.

The production over time of MD authors (Fig. 16) is one of the most important indicators showing the time dedicated by authors of published papers to MD research and their continuity.

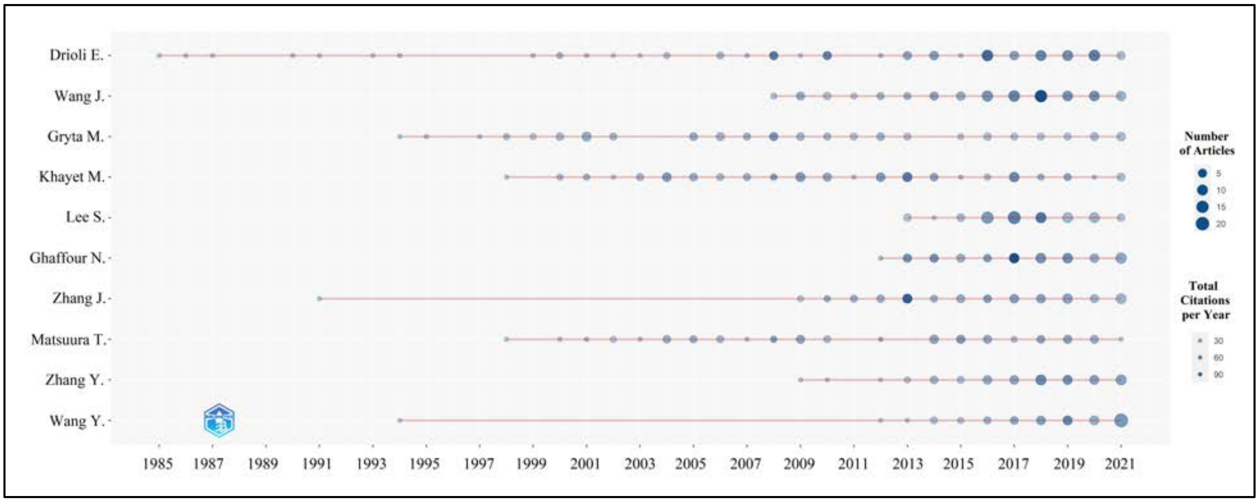


Figure 16. Production over time of the top 10 MD authors (The bubble size denotes the number of articles while the color intensity denotes total citations).

Drioli E. has been working on MD for the longest time (since 1985). Zhang J. is the second MD researcher working on MD (since 1991) but his second paper was published 2009. Gryta M. and Wang Y. published their first MD studies in 1994. Although Zhang J. and Wang Y. were among the first MD researchers, they did not continue their research on this topic until 2009 and 2012, respectively. For instance, after Zhang's first MD paper published in 1991, his next MD paper was published 18 years later. Similarly, Wang Y. published his first MD paper in 1994 and his second MD paper was published also 18 years later. The MD authors showing continuity in MD research are Khayet M. and Matsuura T. Both have been working on MD for a long time and have published papers almost every year. The MD author who published most papers per year is Wang Y. with 21 papers in 2021, while the author who reached the highest total citations per year is Wang J. with 114.8 citations in 2018. In this sense it is very informative to plot Lotka's Law illustration (Fig. 17) to examine the frequency of authors working on MD.

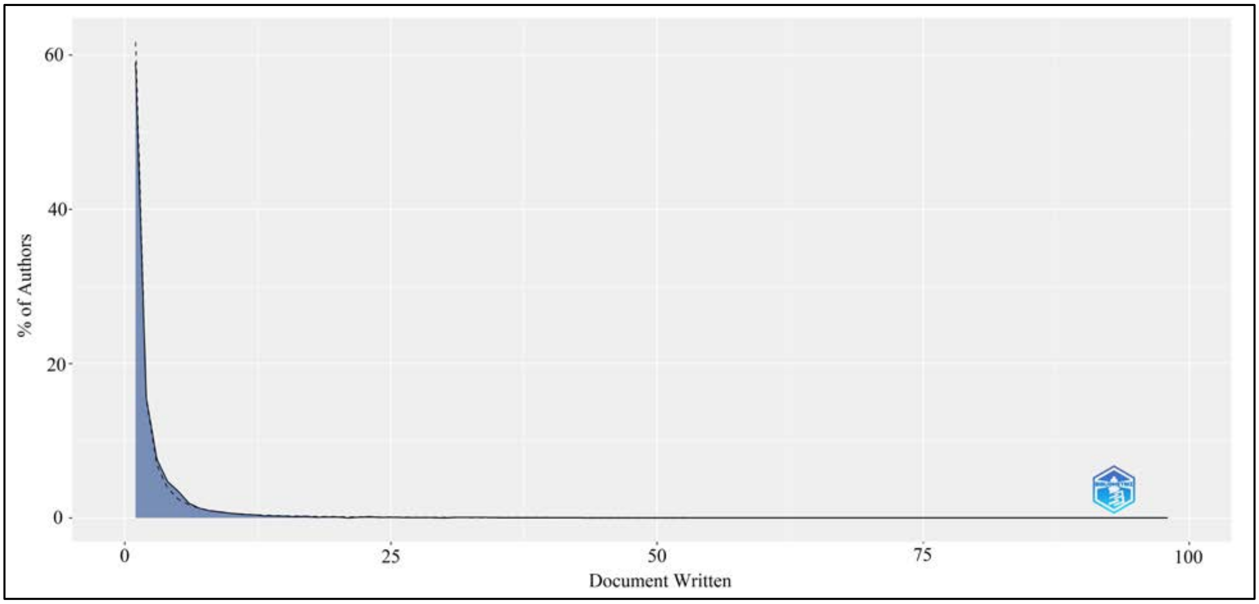


Figure 17. Lotka's Law showing the percentage of authors against the number of their published papers.

Fig.17 indicates that 3134 authors (approximately 60 % of total authors) contributed with only one published MD paper, whereas 9 % of the authors have contributed with 5 or more papers on MD technology. This situation reveals that MD is a subject that has been practiced by many researchers. Only 32 (0.6%) MD authors have more than 31 documents. The empty half of the figure shows that the MD scientists do not have continuity in their MD studies and their names are mentioned in very few papers.

As it is well known, collaboration networks between MD researchers, is an effective way for the transfer of knowledge, for the development of joint projects and for the socialization and collaboration network maps revealing partnerships between countries, researchers, universities or research centers. Fig. 18 summarizes the collaborations of researchers on MD technology.

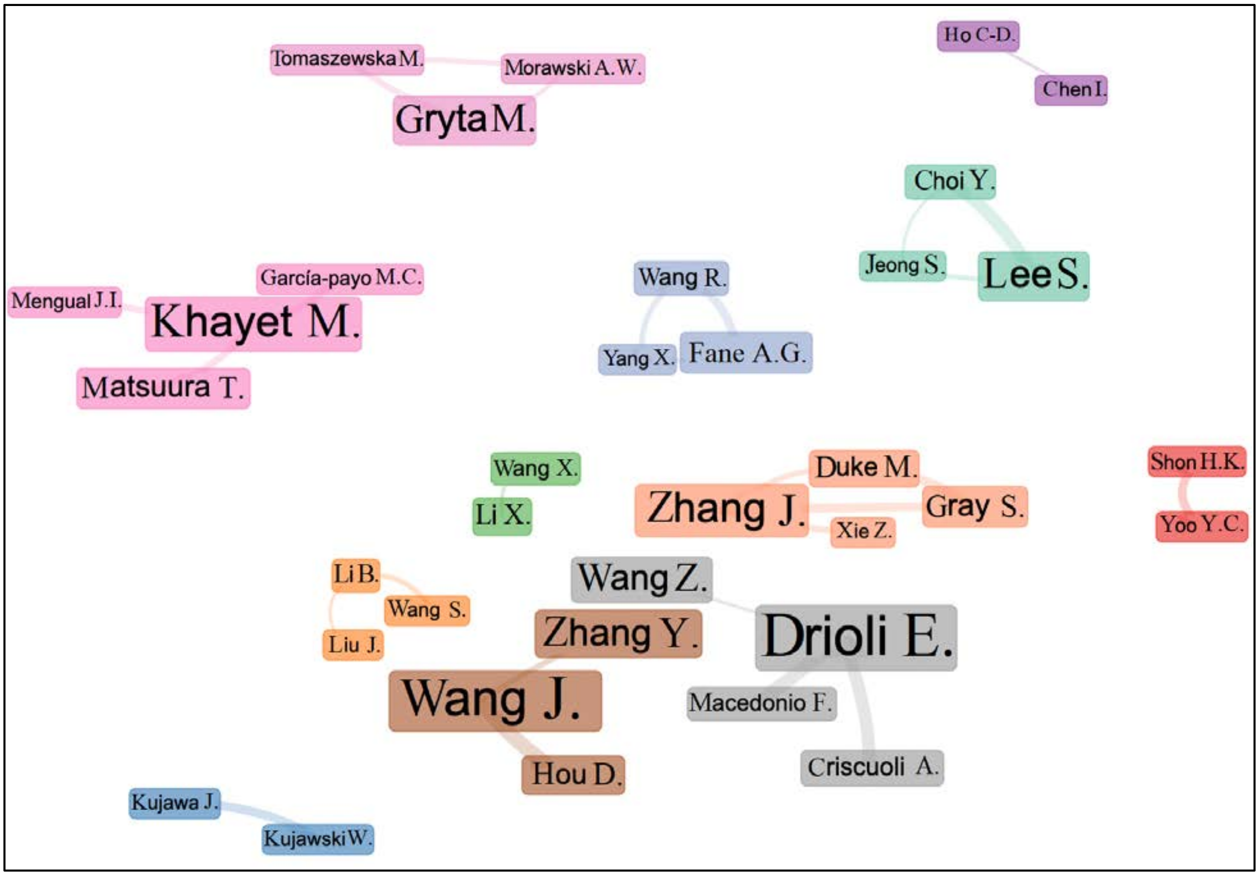


Figure 18. Collaboration networks of MD researchers (minimum number of edges =8).

In Fig. 18 the collaboration networks of MD authors are shown by the edges whose thickness implies the frequency of collaborations (i.e. a thicker edge denotes more MD collaborations between connected researchers). The size of each node shows the number of published papers by the author and different colors designates different workgroup clusters [83]. It is understood that there were 12 different research groups working intensively in MD field. The largest research groups are formed by 4 researchers' expert in MD whereas the group with the least number of researchers consists of 2 researchers. Drioli E., Wang J., Khayet M, Zhang J., Gryta M and Lee S. stands out as the most important center of gravities of MD collaborations.

3.4. Top cited MD papers

In general, document analysis evaluates the underlying intellectual structure of a knowledge area by determining the quality of referenced literature. The top 20 most global cited articles in

MD field (as per Scopus on 02.02.2022) and their corresponding local citations (within the dataset) are summarized in Fig. 19.

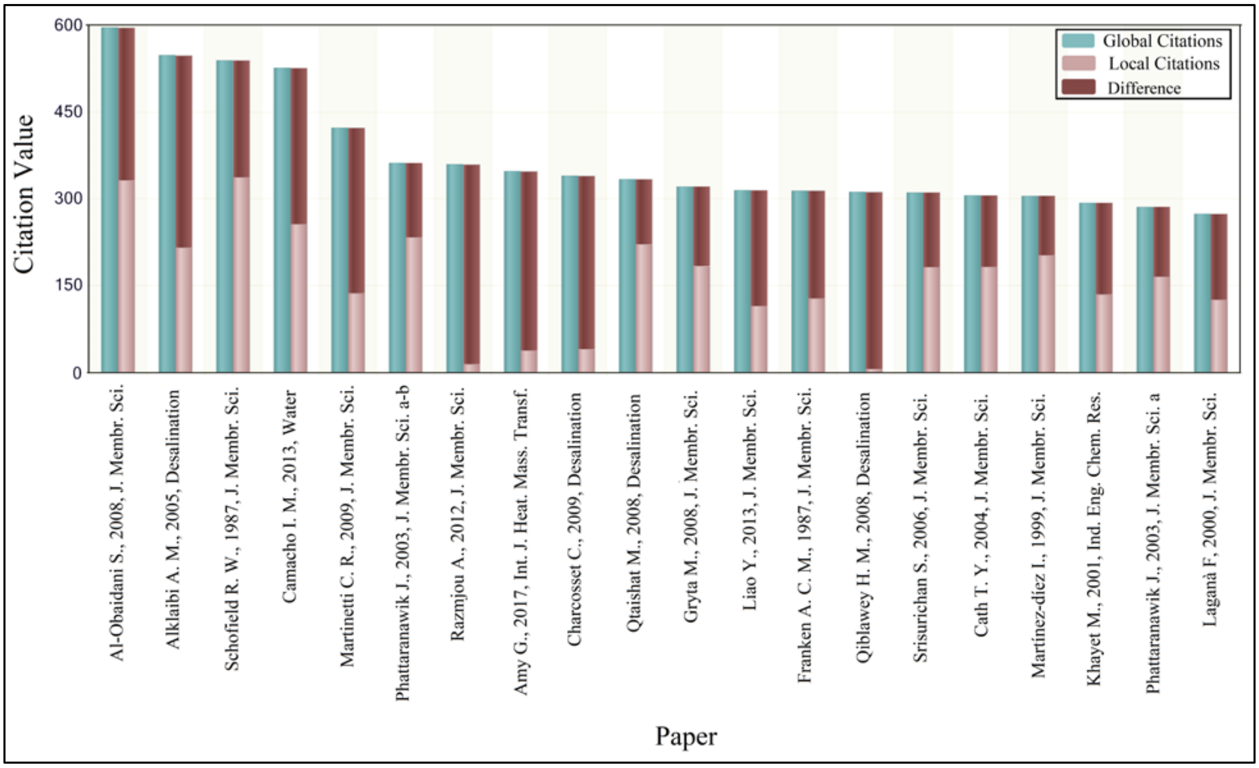


Figure 19. Top 20 global and local cited papers.

Depending on the Scopus citations statistic on 02.02.2022, the paper entitled "*Potential of membrane distillation in seawater desalination: Thermal efficiency, sensitivity study and cost estimation*" published in Journal of Membrane Science in 2008 by Al-Obaidani et al. [84] received the highest number of citations (597). This study was also the second most local cited paper in the collected dataset (332). The second most influential MD study globally is entitled "*Membrane-distillation desalination: Status and potential*", published in Desalination by Alklaibi and Lior in 2005 [85]. The paper that received the most local citations in our collected data set is "*Heat and mass transfer in membrane distillation*" published in Journal of Membrane Science in 1987 by Schofield et al. [86]. The fact that a paper's global and local citations are both high and close to each other is also an indication of the interesting level of the study carried out in the cited paper. The difference between the global and local citations is an indication of the importance of the developed study in other separation fields rather than in MD. For instance, it appears that the paper entitled "*Superhydrophobic modification of TiO_2 nanocomposite PVDF membranes for applications in membrane distillation*" published by Razmjou et al. in 2012 [87] attracts more attention in other fields of separations.

3.5. Countries active in MD technology

Another commonly used approach for determining the countries active on a certain research area is the country analysis. Figs. 20 and 21 show the most prolific countries in the collection in terms of MD research contributions (scientific production, total citations and average citation per published paper) and collaboration world map.

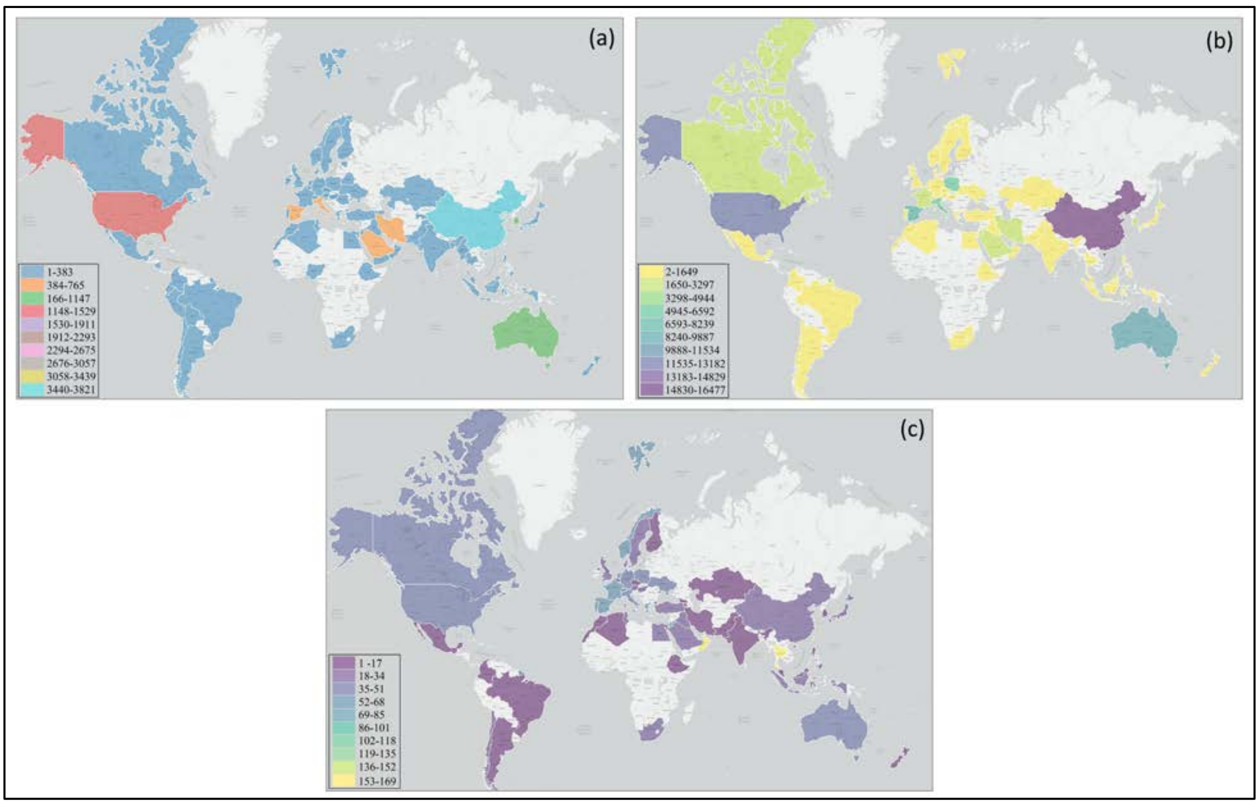


Figure 20. Country analysis: **(a)** MD Scientific production **(b)** total citations and **(c)** average citation per published paper.

Figure 21. MD collaborations world map (minimum number of edges = 5).

From Fig. 20(a) it can be seen that China contributed with most of the published papers (3821 papers) indicating that MD field attracts most attention by Chinese scientific researchers, followed by USA with 1458 published MD papers. Australia is the third active country in MD technology contributing with 1102 published papers. Fig. 20(b) presents the country-based total cited by values. In direct proportion with MD scientific production, the country with the most citations are China (16477 citations), followed by USA with 12137 citations and Australia with 9561 citations. Fig. 20(c) presents the average citation per published papers for countries. This indicates that China, which ranks first in scientific production and total citations metrics, has a low average citation per paper metric (24.74). The citations an article receives may depend on the significance and quality of the developed research, interest of the readers and relation to their investigated studies, and accessibility level of the paper. Apart from these, quantitative features such as page count, publication year, collaboration of authors, institutions, and countries, may also affect the citation metrics [88]. The two countries that exhibit the highest average citation per paper are Thailand and Oman, with 169.25 and 158.50, respectively. In fact, the number of authors affiliated from Thailand and Oman are 33 and 32, respectively. Although the number of MD researchers in these two countries is low, the published papers by researchers of these two countries are more influential. These data in Fig.20 indicate the importance of publication of high-quality papers rather than the quantity. Fig. 21 shows the cooperation between countries. To plot this figure, the minimum number of edges was selected as 5 (at least 5 connections between countries) to obtain a more readable figure. As can be seen in the figure, the strongest link is established between China and Australia (95 collaborations), while the second highest MD collaborations is made by China with USA (70 collaborations). The MD cooperation between Australia and Korea comes third with 43 joint studies.

3.6. Institutions active in MD technology

Today, it is difficult to produce scientific knowledge locally. Therefore, scientific collaborations between national and international universities or research centers are important in order to work more effectively, to progress faster in research projects and to disseminate knowledge globally. Fig. 22 indicates the detected collaborations between institutions/companies/universities in the field of MD.

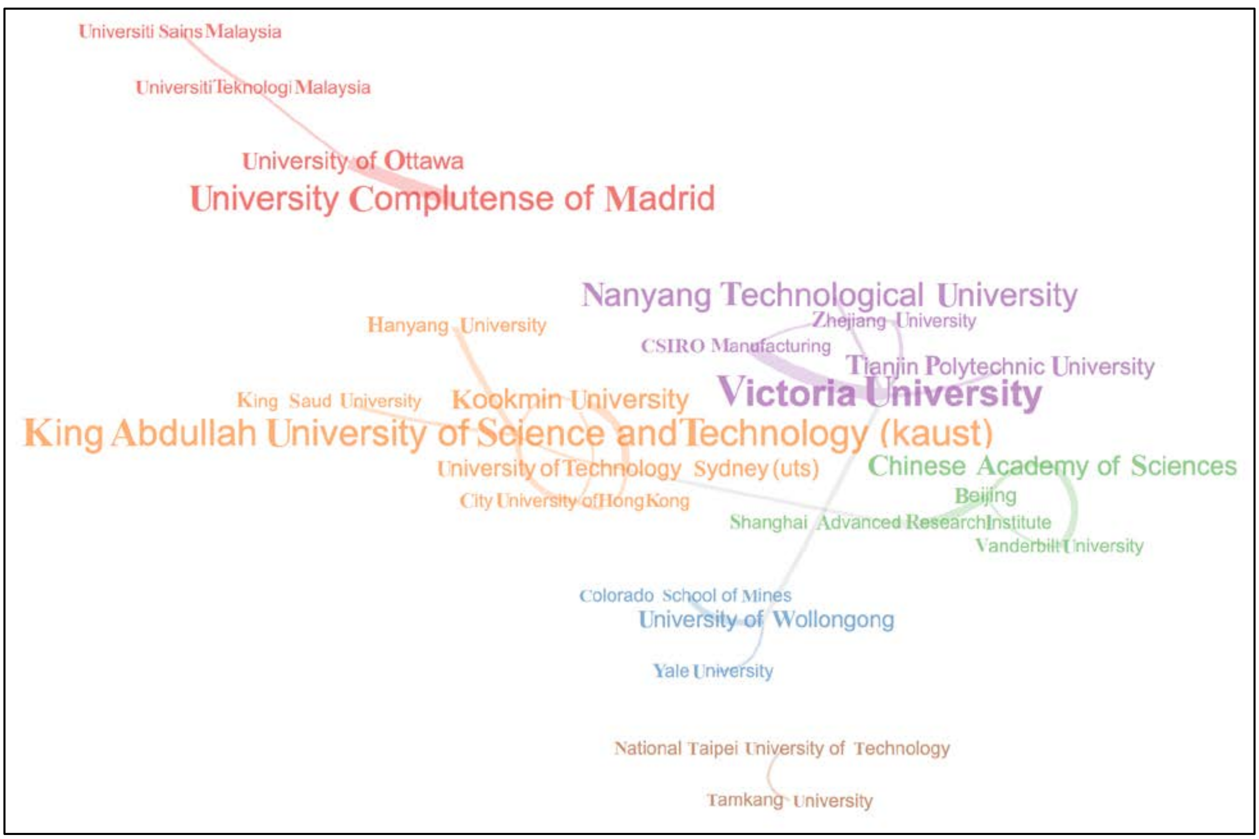


Figure 22. MD collaborations between institutions/universities/companies (minimum number of edges = 3).

As can be seen in Fig. 22, 6 clusters have been detected. The institutions located at the centers of the clusters are King Abdullah University of Science and Technology (Saudi Arabia), Victoria

University (Australia), University Complutense of Madrid (Spain), Chinese Academy of Sciences (China) and University of Wollongong (Australia). The two universities showing the most MD cooperation depending on the edge thickness are the University Complutense of Madrid - the University of Ottawa (Canada) and Victoria University (Australia) - CSIRO Manufacturing (Australia). As seen in the collaboration network, not only universities/institutions but also private companies are also very interested in MD. This shows that the theoretical and practical applications of MD are not limited to research centers, but also to industries especially for desalination application. The biggest cluster includes 6 affiliations while the second cluster has 5 affiliations.

3.7. Text mining of MD technology

Text mining or text data mining allows analysts to swiftly study large volumes of textual data to capture key concepts, meaningful trends and patterns, relationships and in general high-quality information that may not be detected easily by human. Fig. 23 shows the yearly cumulative growth of the top 10 keywords mentioned in MD papers.

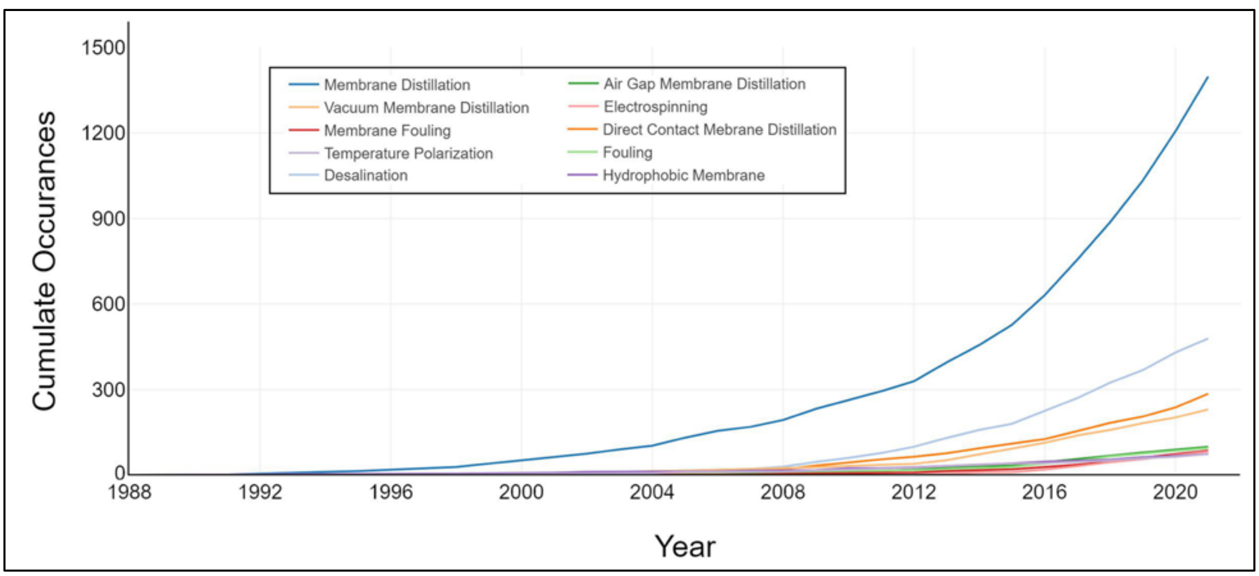


Figure 23. Cumulative growth of top 10 words based on author keywords.

In general, "*Membrane Distillation*" is the most preferred keyword by researchers. This is followed by "*Desalination*" as this application is the most considered by MD technology. The fact that "*Direct Contact Membrane Distillation*" is the third most used keyword proved that this configuration is also the most used among MD scientists as mentioned earlier. Fig. 23 strengthens the idea that MD researchers have been stuck around the same keywords for many years. Thus, it is advisable to start using other new keywords rather than the standard ones in order to attract more the attention of wider readers. In order to better understand the emerging MD topics a trend topic analysis was conducted based on authors` keywords and the obtained results are plotted in Fig. 24.

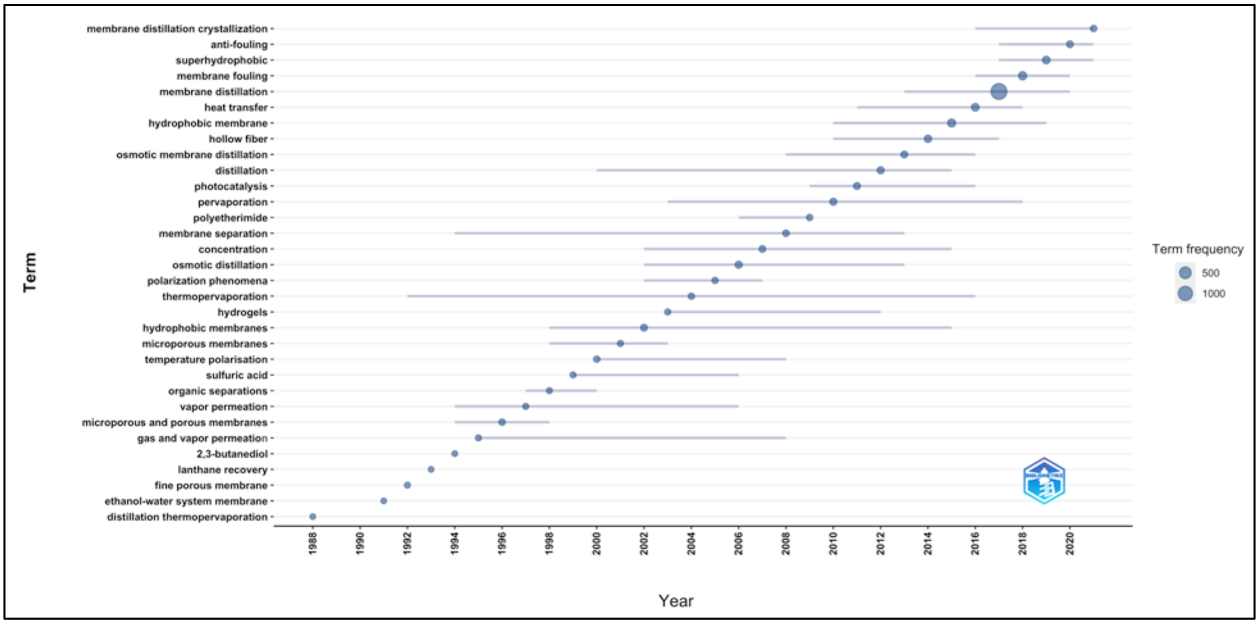


Figure 24. Trending topics in MD based on authors´ keywords. (word min. freq. = 1, number of words per year = 1)

In Fig. 24 the lines reflect the duration of the used word while the most common time that a word is used is emphasized by a bubble on the line. The size of the bubble shows the frequency of the used word so that the larger the bubble, the more frequently is used. As can be seen, various words have been mentioned as trending topics during some years between 1988 and 2021. The reason for the absence of trending topic words before 1988 is because either the published papers did not include author keywords and not enough papers were published before this year. While some keywords have been mentioned for a very short period of time (i.e. only for one year before 1995), other topics have been stated for various years showing their importance and impact on MD during these years. The most remarkable situation in Fig. 24 is the word "thermopervaporation", which was within the trending topics for the longest time (1992-2016). In the previous section 2.1. data, we stated that this term was used instead of MD separation process until the workshop held in Rome in 1986, when it was decided to use commonly MD replacing all other previously used terms including thermopervaporation. In fact, nowadays there is an essential difference between MD and thermopervaporation, which uses dense and selective membranes applying low hydrostatic pressure in the permeate side or vacuum and a high feed temperature [89]. We maintain this word in Fig. 24 because within the collected MD papers, comparisons of the two processes were carried out. As it may be expected, the word exhibiting the highest frequency is "membrane distillation" registered in 2017. The topics that have attracted most the attention of MD scientists during last four years are "membrane fouling" in 2018, "superhydrophobic" in 2019, "anti-fouling" in 2020 and "membrane distillation crystallization" in 2021. These show the focusing interests in membrane engineering designing superhydrophobic MD membranes with reduced fouling and the strong application of MD crystallization for the treatment of brines and valuable salts recovery.

Thematic maps also permit to detect certain themes of interest within a given research area showing their progression over time. Fig. 25 shows the MD thematic map based on authors´ keywords.

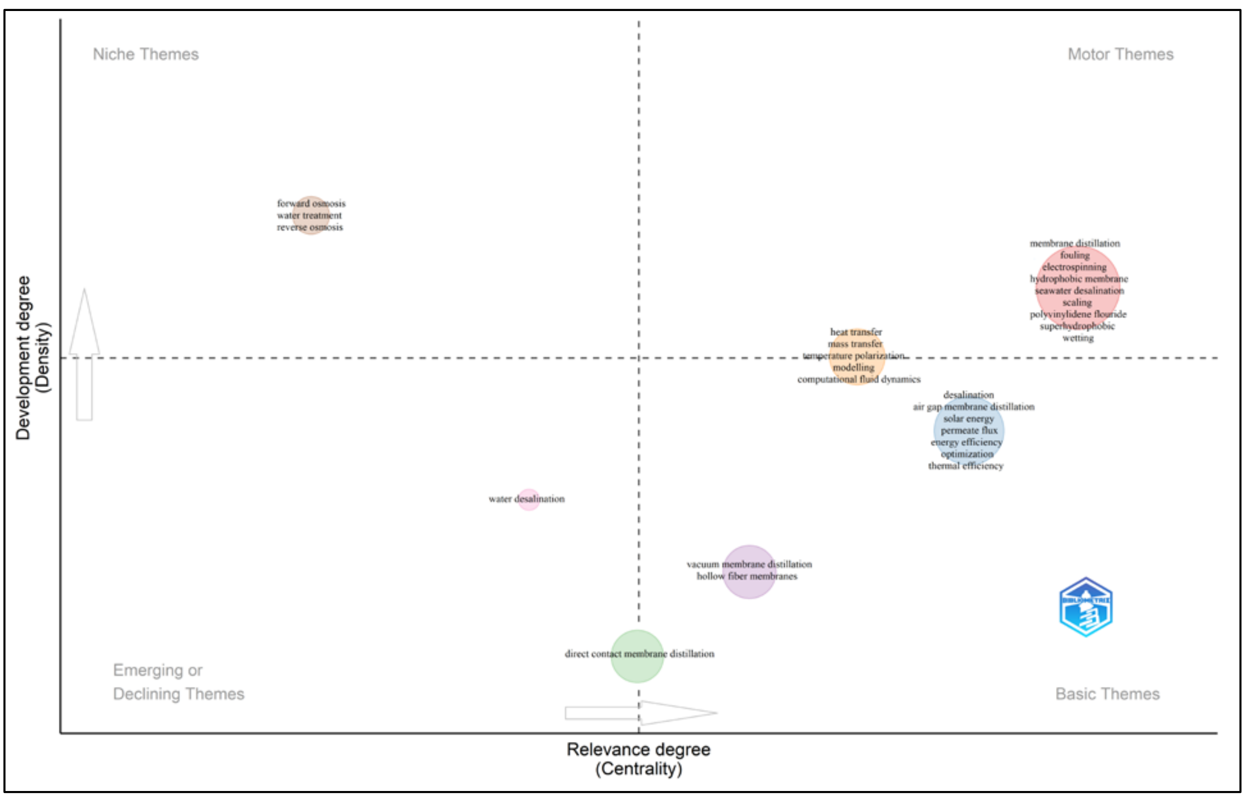
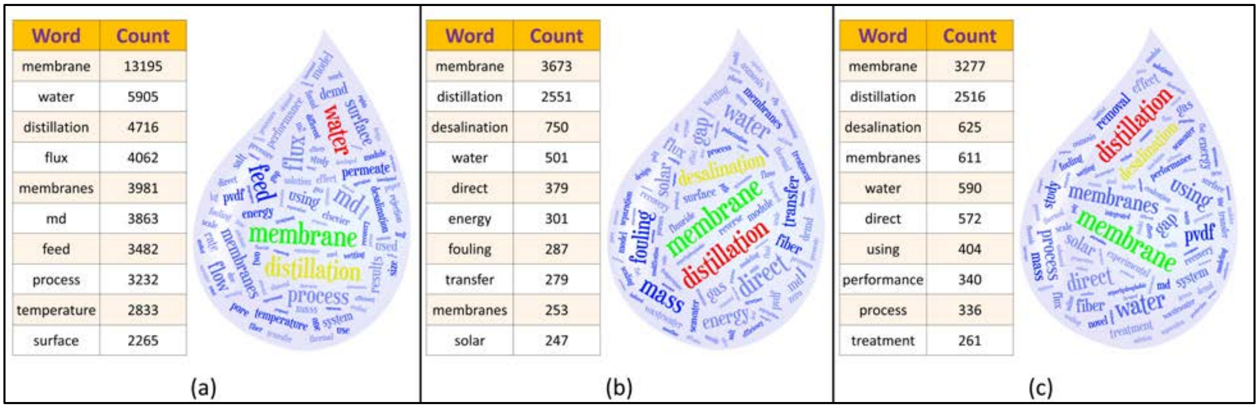


Figure 25. MD thematic map based on authors´ keywords (Clustering algorithm = Louvain, min. word freq. = 10).

As can be seen in Fig. 25, the well-developed and significant themes (motor themes) for MD technology are membrane distillation, fouling, electrospinning, hydrophobic membranes, seawater desalination, scaling, polyvinylidene fluoride, superhydrophobic and wetting. The cluster that has the authors keywords like heat transfer, mass transfer temperature polarization, modelling, computational fluid dynamics is located just in the intersection of basic and motor zones. This cluster has high development degree but moderate relevance degree, which indicates that these research areas may join the motor zone in the coming years boosting up MD even more intensely in the future. Clusters in the basic zone (cluster 1: desalination, air gap membrane distillation, solar energy, permeate flux, energy efficiency, optimization, thermal efficiency; and cluster 2: vacuum membrane distillation, hollow fiber membranes) have high centrality and low density. Both are heavily connected to the rest of the themes indicating that there is a room for these themes to be studied more intensively. In the niche zone there are themes like forward osmosis, water treatment and reverse osmosis, that are detected niche MD applications highly developed and specialized themes with low connections with the rest of the clusters. Emerging and declining zone has low centrality and low density, which indicates that they are marginal and weakly developed areas. Water desalination is localized in this zone showing that this key word is declining in MD because it also appears in the motor themes as indicated previously indicating that this MD application is well developed. In addition, water desalination may also refer to other emerging fields of MD desalination rather than seawater desalination. The author keyword "direct contact membrane distillation" is situated on the center of basic and emerging or declining zones indicating that this MD configuration could be a basic theme or a declining theme in the coming years according to the importance attributed by the researchers to this configuration. Taking into account that DCMD is the most considered MD configuration by practically all research groups, we do believe that it is a basic theme and will be continuously used at least at laboratory scale.

Fig. 26 summarizes the developed word cloud analysis of abstracts, authors` keywords and titles of papers in the collected dataset.



Word	Count
membrane	13195
water	5905
distillation	4716
flux	4062
membranes	3981
md	3863
feed	3482
process	3232
temperature	2833
surface	2265

Word	Count
membrane	3673
distillation	2551
desalination	750
water	501
direct	379
energy	301
fouling	287
transfer	279
membranes	253
solar	247

Word	Count
membrane	3277
distillation	2516
desalination	625
membranes	611
water	590
direct	572
using	404
performance	340
process	336
treatment	261

Figure 26. Word clouds of **(a)** abstracts **(b)** authors` keywords and **(c)** titles of published MD papers.

It is understood that authors use a lot the words "*membrane*", "*membranes*", "*water*", "*distillation*", "*desalination*". This is an indication that the authors employed the same terms more than once in the three parts of MD published papers indicated in Fig. 26. It must be pointed out that when readers see the same words or sentence structures repeated in a paper, it may give them the feeling that they have already read the paper so they may stop looking at it. Use of unique or unusual words in the title, abstract and keyword fields in papers may attract more the attention of the readers. When Fig. 26(a) is examined, it is seen that the words "*flux*", "*temperature*" and "*surface*" are also among the most frequently used words in abstracts. This shows that MD researchers are interested in optimizing the permeate flux, in studying the effects of the feed and/or permeate temperature of the MD permeate flux, and in improving the membrane surface properties. The words "*solar*" and "*energy*" that appear in Fig. 26(b) presenting keywords indicate that MD scientists also care about solar applications of MD technology as well as energy recovery and energy consumption. Again, the term "*performance*" that comes up in Fig 26(c) illustrating word clouds in titles shows that MD researchers are interested to improve the MD performance (i.e. flux and separation factor). The term "*direct*" appearing in Figs. 26(b) and 26(c) proves that DCMD is the most used membrane configuration as mentioned earlier.

3.8. Sentiment analysis of MD technology

In this section the authors´ sentiments on MD technology can be revealed from the abstracts of the published papers. This sentiment analysis is critical for detecting the feelings of authors, either optimistic, neutral or pessimistic about their MD research study for the development and future industrial implementation of MD technology. Fig. 27 shows the sentiments of authors as a histogram plot with density curve in a scale ranging from -1 (negative sentiment/pessimistic) to 1 (positive sentiment/optimistic). It must be pointed out that 15 papers have missing abstracts so sentiment analysis was conducted based on the rest of collected papers 3212 of the dataset.

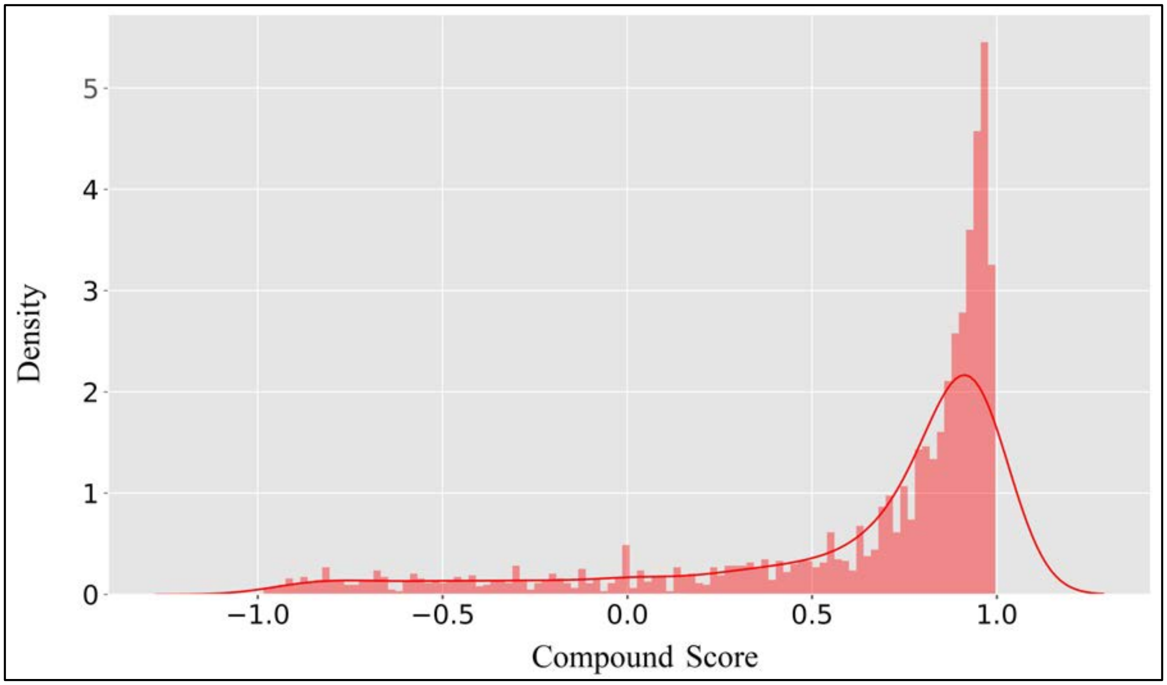


Figure 27. Sentiment analysis from the abstracts of MD published papers.

The highest ratio of sentiment compound score was found in the range 0.96 – 0.98 (i.e. corresponding to 347 papers). The sentiment scores higher than 0.50, indicates that the authors of 75.34% of the abstracts (i.e. corresponding to 2420 papers) were very optimistic about their developed MD research studies. This is a proof of how successful MD technology has been applied for separation and water treatment as shown throughout the present manuscript.

4. Conclusions

MD is one of the attractive technologies of emerging interests for desalination and the treatment of aqueous solutions containing non-volatile solutes. The first MD paper was published 55 years ago and since 2012 the interest on this separation process has grown exponentially because of various advantages such as the treatment of high saline water up to saturation, the production of distilled and ultra-pure water, and the possibility to use renewable and industrial waste heat among others. This study examines MD from a different perspective with data analysis, machine learning approaches and bibliometric methods. As a result of the analysis carried out, remarkable statistics about MD were revealed. By using Scopus database as per 02.02.2022, we collected 3227 published MD papers from 382 sources in the collection. The number of publications about MD is increasing tremendously every year especially after 2012. It was found that worldwide MD researchers have a low degree of collaboration index (1.64). There is also an increase in the number of pages and references counts of the published papers every year. On the contrary, there is a decrease in the number of citations. Among the seven different MD configurations, DCMD is the most preferred one by MD researchers followed by VMD. Considering the increasing number of papers published every year on membrane preparation and modification, recently membranologists show a great interest on MD membrane engineering. Word cloud of membrane engineering papers indicates that the hydrophobic PVDF polymer is used more than other polymers and MD researcher studies are mainly focused on membrane surface modification, improvement of permeate flux and enhancement of the hydrophobic character of the membrane surface. Polymeric phase inversion flat sheet membrane preparation or modification is the most proposed type of membrane for MD.

MD technology has been combined with various other separation processes being the preferred combination multi stage/multi effect distillation/multi effect evaporation (MEE) followed by forward osmosis (FO). While MD is mainly applied in the field of desalination being DCMD the most used one followed by VMD, the second most considered MD application is the removal of volatile compounds from water.

Journal of Membrane Science and Desalination are the locomotives of MD technology (the most publishing, the most cited and the highest metrics journals). Drioli E. (Università della Calabria, Italy) is the researcher contributing with most of published MD papers since 1985, Gryta M. (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie, Poland) reaches the highest articles fractionalized value, while Khayet M. (University Complutense of Madrid, Spain) receives the highest local citations, *h*-index and *g*-index metrics. The paper entitled "*Potential of membrane distillation in seawater desalination: Thermal efficiency, sensitivity study and cost estimation*" [84] is the most cited publication globally with 597 citations, while the most cited paper from the collected dataset is entitled "*Heat and mass transfer in membrane distillation*" [86]. China has the highest number of published papers while Thailand receives the highest average citation per paper metric. The top three universities showing more interest on MD are King Abdullah University of Science and Technology (KAUST), Victoria University (VU) and University Complutense of Madrid (UCM).

Trend topic analysis reveals that membrane scientists give more importance to membrane fouling, superhydrophobic membranes, anti-fouling and membrane distillation crystallization topics. Thematic map analysis reveals motor, basic, emerging and declining, and niche themes for MD technology. Word cloud analysis indicate that MD authors mostly use the same words in the title/authors' keywords/abstract sections of their articles. The sentiment analysis shows that MD researchers express very positive feelings in their paper abstracts and are optimistic about MD technology.

Acknowledgements

Authors acknowledge that the basis of the graphical abstract was created with OpenAI's text-to-image-generation architecture, DALL-E. Upon generating draft image, the authors edited the image, and they took the ultimate responsibility for the content. Authors are thankful to OpenAI. Dr. Ersin Aytaç would like to express his acknowledgement for the postdoctoral grant received from The Scientific and Technological Research Council of Turkey (TUBITAK) at the University Complutense of Madrid (UCM) with the grant number of 1059B191900618.

References

[1] M. Pagliero, M. Khayet, C. García-Payo, L. García-Fernández, Hollow fibre polymeric membranes for desalination by membrane distillation technology: A review of different morphological structures and key strategic improvements, Desalination, 516 (2021) 115235.

[2] M. Essalhi, M. Khayet, S. Tesfalidet, M. Alsultan, N. Tavajohi, Desalination by direct contact membrane distillation using mixed matrix electrospun nanofibrous membranes with carbon-based nanofillers: A strategic improvement, Chemical Engineering Journal, 426 (2021) 131316.

[3] M. Khayet, M.P. Godino, J.I. Mengual, Study of Asymmetric Polarization in Direct Contact Membrane Distillation, Separation Science and Technology, 39 (2005) 125-147.

[4] X. Li, M.C. García-Payo, M. Khayet, M. Wang, X. Wang, Superhydrophobic polysulfone/polydimethylsiloxane electrospun nanofibrous membranes for water desalination by direct contact membrane distillation, Journal of Membrane Science, 542 (2017) 308-319.

[5] M. Al-Salmi, M. Laqbaqbi, S. Al-Obaidani, R.S. Al-Maamari, M. Khayet, M. Al-Abri, Application of membrane distillation for the treatment of oil field produced water, Desalination, 494 (2020) 114678.

[6] M.E. Findley, Vaporization through Porous Membranes, Industrial & Engineering Chemistry Process Design and Development, 6 (1967) 226-230.
[7] H.R. Mortaheb, M. Baghban Salehi, M. Rajabzadeh, Optimized hybrid PVDF/graphene membranes for enhancing performance of AGMD process in water desalination, Journal of Industrial and Engineering Chemistry, 99 (2021) 407-421.
[8] J. Xu, Y.B. Singh, G.L. Amy, N. Ghaffour, Effect of operating parameters and membrane characteristics on air gap membrane distillation performance for the treatment of highly saline water, Journal of Membrane Science, 512 (2016) 73-82.
[9] Q. Ma, L. Tong, C. Wang, G. Cao, H. Lu, J. Li, X. Liu, X. Feng, Z. Wu, Simulation and Experimental Investigation of the Vacuum-Enhanced Direct Membrane Distillation Driven by a Low-Grade Heat Source, Membranes, 12 (2022) 842.
[10] M. Wen, M. Chen, G.-K. Ren, P.-L. Li, C. Lv, Y. Yao, Y.-K. Liu, S.-J. Deng, Z. Zheng, C.-G. Xu, D.-L. Luo, Enhancing the selectivity of hydrogen isotopic water in membrane distillation by using graphene oxide, Journal of Membrane Science, 610 (2020) 118237.
[11] A. Boubakri, A. Hafiane, S.A.T. Bouguecha, Application of response surface methodology for modeling and optimization of membrane distillation desalination process, Journal of Industrial and Engineering Chemistry, 20 (2014) 3163-3169.
[12] M. Khayet, C. Cojocaru, Artificial neural network modeling and optimization of desalination by air gap membrane distillation, Separation and Purification Technology, 86 (2012) 171-182.
[13] M. Khayet, A.O. Imdakm, T. Matsuura, Monte Carlo simulation and experimental heat and mass transfer in direct contact membrane distillation, International Journal of Heat and Mass Transfer, 53 (2010) 1249-1259.
[14] M. Essalhi, M. Khayet, 10 - Fundamentals of membrane distillation, in: A. Basile, A. Figoli, M. Khayet (Eds.) Pervaporation, Vapour Permeation and Membrane Distillation, Woodhead Publishing, Oxford, 2015, pp. 277-316.
[15] K. Gul, A. Hasanoglu, Concentration of skim milk using a hybrid system of osmotic distillation and membrane distillation, Desalin. Water Treat., 149 (2019) 11-22.
[16] J.L. Cartinella, T.Y. Cath, M.T. Flynn, G.C. Miller, K.W. Hunter, A.E. Childress, Removal of natural steroid hormones from wastewater using membrane contactor processes, Environ. Sci. Technol., 40 (2006) 7381-7386.
[17] Z.W. Ding, L.Y. Liu, H.F. Yu, R.Y. Ma, Z.R. Yang, Concentrating the extract of traditional Chinese medicine by direct contact membrane distillation, Journal of Membrane Science, 310 (2008) 539-549.
[18] S. Simone, A. Figoli, A. Criscuoli, M.C. Carnevale, S.M. Alfadul, H.S. Al-Romaih, F.S. Al Shabouna, O.A. Al-Harbi, E. Drioli, Effect of selected spinning parameters on PVDF hollow fiber morphology for potential application in desalination by VMD, Desalination, 344 (2014) 28-35.
[19] S. Al-Gharabli, W. Kujawski, Z.A. El-Rub, E.M. Hamad, J. Kujawa, Enhancing membrane performance in removal of hazardous VOCs from water by modified fluorinated PVDF porous material, Journal of Membrane Science, 556 (2018) 214-226.
[20] A. Alkhudhiri, N. Darwish, N. Hilal, Membrane distillation: A comprehensive review, Desalination, 287 (2012) 2-18.
[21] B.S. Lalia, V. Kochkodan, R. Hashaikeh, N. Hilal, A review on membrane fabrication: Structure, properties and performance relationship, Desalination, 326 (2013) 77-95.
[22] E. Drioli, A. Ali, F. Macedonio, Membrane distillation: Recent developments and perspectives, Desalination, 356 (2015) 56-84.

[23] M. Khayet, Membranes and theoretical modeling of membrane distillation: A review, Advances in Colloid and Interface Science, 164 (2011) 56-88.
[24] M.S. El-Bourawi, Z. Ding, R. Ma, M. Khayet, A framework for better understanding membrane distillation separation process, Journal of Membrane Science, 285 (2006) 4-29.
[25] P. Otlet, Traité de Documentation, Eds Mundaneum, Bruxelles-Belgium, 1934.
[26] A. Pritchard, Statistical Bibliography or Bibliometrics, Journal of Documentation, 25 (1969) 348-349.
[27] H.M. Ashraf, S.A. Al-Sobhi, M.H. El-Naas, Mapping the desalination journal: A systematic bibliometric study over 54 years, Desalination, 526 (2022) 115535.
[28] Y. Chen, M. Lin, D. Zhuang, Wastewater treatment and emerging contaminants: Bibliometric analysis, Chemosphere, (2022) 133932.
[29] M. Khayet, E. Aytaç, T. Matsuura, Bibliometric and sentiment analysis with machine learning on the scientific contribution of Professor Srinivasa Sourirajan, Desalination, 543 (2022) 116095.
[30] P.M. Iftikhar, F. Ali, M. Faisaluddin, A. Khayyat, M. De Gouvia De Sa, T. Rao, A Bibliometric Analysis of the Top 30 Most-cited Articles in Gestational Diabetes Mellitus Literature (1946-2019), Cureus, 11 (2019) e4131-e4131.
[31] E. Aytac, Exploring Electrocoagulation Through Data Analysis and Text Mining Perspectives, Environmental Engineering and Management Journal, 21 (2022) 671-685.
[32] G.A. Borges, C. Dini, M.M.D.d. Medeiros, R.C.M. Rodrigues Garcia, V.A.R. Barão, M.F. Mesquita, Funding assistance and global productivity in the field of implant overdentures: A bibliometric analysis of 35 years, The Journal of Prosthetic Dentistry, (2021).
[33] Y. Dai, Y. Song, H. Gao, S. Wang, Y. Yuan, Bibliometric analysis of research progress in membrane water treatment technology from 1985 to 2013, Scientometrics, 105 (2015) 577-591.
[34] T. Pang, J. Shen, Visualizing the landscape and evolution of capacitive deionization by scientometric analysis, Desalination, 527 (2022) 115562.
[35] W. Suwaileh, N. Pathak, H. Shon, N. Hilal, Forward osmosis membranes and processes: A comprehensive review of research trends and future outlook, Desalination, 485 (2020) 114455.
[36] Y. Tang, X. Long, M. Wu, S. Yang, N. Gao, B. Xu, S. Dutta, Bibliometric review of research trends on disinfection by-products in drinking water during 1975–2018, Separation and Purification Technology, 241 (2020) 116741.
[37] P. Martin-Rodilla, J.I. Panach, C. Gonzalez-Perez, O. Pastor, Assessing data analysis performance in research contexts: An experiment on accuracy, efficiency, productivity and researchers' satisfaction, Data & Knowledge Engineering, 116 (2018) 177-204.
[38] C. Ge, C. Yin, Z. Liu, L. Fang, J. Zhu, H. Ling, A privacy preserve big data analysis system for wearable wireless sensor network, Computers & Security, 96 (2020) 101887.
[39] L. Ogiela, Transformative computing in advanced data analysis processes in the cloud, Information Processing & Management, 57 (2020) 102260.
[40] F.Z. Fagroud, L. Ajallouda, E.H. Ben Lahmar, H. Toumi, K. Achtaich, S.E. Filali, IOT Search Engines: Exploratory Data Analysis, Procedia Computer Science, 175 (2020) 572-577.
[41] R. Indrakumari, T. Poongodi, S.R. Jena, Heart Disease Prediction using Exploratory Data Analysis, Procedia Computer Science, 173 (2020) 130-139.
[42] E. Aytaç, Unsupervised learning approach in defining the similarity of catchments: Hydrological response unit based k-means clustering, a demonstration on Western Black Sea Region of Turkey, International Soil and Water Conservation Research, 8 (2020) 321-331.

[43] E. Aytaç, Modeling Future Impacts on Land Cover of Rapid Expansion of Hazelnut Orchards: A Case Study on Samsun, Turkey, European Journal of Sustainable Development Research, 6 (2022) em0193.
[44] C.S.H. Yeo, Q. Xie, X. Wang, S. Zhang, Understanding and optimization of thin film nanocomposite membranes for reverse osmosis with machine learning, Journal of Membrane Science, 606 (2020) 118135.
[45] C.A. Graham, H. Shamkhalichenar, V.E. Browning, V.J. Byrd, Y. Liu, M.T. Gutierrez-Wing, N. Novelo, J.-W. Choi, T.R. Tiersch, A practical evaluation of machine learning for classification of ultrasound images of ovarian development in channel catfish (Ictalurus punctatus), Aquaculture, 552 (2022) 738039.
[46] P. Xiong, S. Buffett, S. Iqbal, P. Lamontagne, M. Mamun, H. Molyneaux, Towards a robust and trustworthy machine learning system development: An engineering perspective, Journal of Information Security and Applications, 65 (2022) 103121.
[47] E. Aytaç, Forecasting Turkey's Hazelnut Export Quantities with Facebook's Prophet Algorithm and Box-Cox Transformation, ADCAIJ-Advances in Distributed Computing and Artificial Intelligence Journal, 10 (2021) 33-47.
[48] E.H. Houssein, Z. Abohashima, M. Elhoseny, W.M. Mohamed, Machine learning in the quantum realm: The state-of-the-art, challenges, and future vision, Expert Systems with Applications, 194 (2022) 116512.
[49] E. Aytaç, Havzaların Benzerliklerini Tanımlamada Alternatif Bir Yaklaşım: Hiyerarşik Kümeleme Yöntemi Uygulaması, Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi, 21 (2021) 958-970.
[50] S. Kumar, A.K. Kar, P.V. Ilavarasan, Applications of text mining in services management: A systematic literature review, International Journal of Information Management Data Insights, 1 (2021) 100008.
[51] H. Hashimi, A. Hafez, H. Mathkour, Selection criteria for text mining approaches, Computers in Human Behavior, 51 (2015) 729-733.
[52] V. Ahuja, M. Shakeel, Twitter Presence of Jet Airways-Deriving Customer Insights Using Netnography and Wordclouds, Procedia Computer Science, 122 (2017) 17-24.
[53] H. Deng, D. Ergu, F. Liu, Y. Cai, B. Ma, Text sentiment analysis of fusion model based on attention mechanism, Procedia Computer Science, 199 (2022) 741-748.
[54] R.P. Pratama, A. Tjahyanto, The influence of fake accounts on sentiment analysis related to COVID-19 in Indonesia, Procedia Computer Science, 197 (2022) 143-150.
[55] M. Ghiassi, S. Lee, S.R. Gaikwad, Sentiment analysis and spam filtering using the YAC2 clustering algorithm with transferability, Computers & Industrial Engineering, 165 (2022) 107959.
[56] M.T. Islam, L. Xing, Geometry and statistics-preserving manifold embedding for nonlinear dimensionality reduction, Pattern Recognition Letters, 151 (2021) 155-162.
[57] V.P.G. Bretas, I. Alon, Franchising research on emerging markets: Bibliometric and content analyses, Journal of Business Research, 133 (2021) 51-65.
[58] S. Kumar, N. Pandey, W.M. Lim, A.N. Chatterjee, N. Pandey, What do we know about transfer pricing? Insights from bibliometric analysis, Journal of Business Research, 134 (2021) 275-287.
[59] M. Losse, M. Geissdoerfer, Mapping socially responsible investing: A bibliometric and citation network analysis, Journal of Cleaner Production, 296 (2021) 126376.
[60] C. Leong, Narratives and water: A bibliometric review, Global Environmental Change, 68 (2021) 102267.

[61] M. Khayet, T. Matsuura, Chapter 1 - Introduction to Membrane Distillation, in: M. Khayet, T. Matsuura (Eds.) Membrane Distillation, Elsevier, Amsterdam, 2011, pp. 1-16.
[62] M. Aria, C. Cuccurullo, bibliometrix: An R-tool for comprehensive science mapping analysis, Journal of Informetrics, 11 (2017a) 959-975.
[63] Exploratory-Inc., (2020). https://exploratory.io (accessed February 18, 2022)
[64] H. Alyuruk, Chapter 1 - Introduction to R and Python, in: H. Alyuruk (Ed.) R and Python for Oceanographers, Elsevier, 2019, pp. 1-21.
[65] M. Aria, C. Cuccurullo, Biblioshiny : Bibliometrix for no Coders, (2017b). https://www.bibliometrix.org/biblioshiny/assets/player/KeynoteDHTMLPlayer.html#1 (accessed March 30, 2022)
[66] B. Elango, P. Rajendran, Authorship Trends and Collaboration Pattern in The Marine Sciences Literature : A Scientometric Study, International Journal of Information Dissemination and Technology, 2 (2012) 166-169.
[67] O. Kilicoglu, H. Mehmetcik, Science mapping for radiation shielding research, Radiation Physics and Chemistry, 189 (2021) 109721.
[68] M. Aria, C. Cuccurullo, A brief introduction to bibliometrix, (2017c). https://www.bibliometrix.org/vignettes/Introduction_to_bibliometrix.html (accessed March 30, 2022)
[69] G.F. Ferreira, K.C. Stéfani, A Global Bibliometric Analysis of Hallux Valgus Research (1999-2019), The Journal of Foot and Ankle Surgery, 60 (2021) 501-506.
[70] S. Kumar, P. Sharma, K.C. Garg, Lotka's law and institutional productivity, Information Processing & Management, 34 (1998) 775-783.
[71] C. Forliano, P. De Bernardi, D. Yahiaoui, Entrepreneurial universities: A bibliometric analysis within the business and management domains, Technological Forecasting and Social Change, 165 (2021) 120522.
[72] K.M. Kwayu, V. Kwigizile, K. Lee, J.-S. Oh, T. Nelson, Automatic topics extraction from crowdsourced cyclists near-miss and collision reports using text mining and Artificial Neural Networks, International Journal of Transportation Science and Technology, (2021).
[73] M. D'Orazio, E. Di Giuseppe, G. Bernardini, Automatic detection of maintenance requests: Comparison of Human Manual Annotation and Sentiment Analysis techniques, Automation in Construction, 134 (2022) 104068.
[74] A. Borg, M. Boldt, Using VADER sentiment and SVM for predicting customer response sentiment, Expert Systems with Applications, 162 (2020) 113746.
[75] I. Bueno, R.A. Carrasco, R. Ureña, E. Herrera-Viedma, A business context aware decision-making approach for selecting the most appropriate sentiment analysis technique in e-marketing situations, Information Sciences, 589 (2022) 300-320.
[76] C. Lai, H. Zhang, Z. Hu, Performance analyses on a concentrated photovoltaics driven direct contact membrane distillation coupled system, Desalination, 531 (2022) 115695.
[77] O. Koll, S. von Wallpach, B. Uzelac, Stakeholder-oriented brand management: A Venn-diagram approach to monitor brand associations, European Management Journal, (2022).
[78] M. Khayet, Membrane surface modification and characterization by X-ray photoelectron spectroscopy, atomic force microscopy and contact angle measurements, Applied Surface Science, 238 (2004) 269-272.
[79] C. Feng, K.C. Khulbe, T. Matsuura, R. Gopal, S. Kaur, S. Ramakrishna, M. Khayet, Production of drinking water from saline water by air-gap membrane distillation using polyvinylidene fluoride nanofiber membrane, Journal of Membrane Science, 311 (2008) 1-6.

[80] L.D. Tijing, J.-S. Choi, S. Lee, S.-H. Kim, H.K. Shon, Recent progress of membrane distillation using electrospun nanofibrous membrane, Journal of Membrane Science, 453 (2014) 435-462.
[81] F. Yalcinkaya, A review on advanced nanofiber technology for membrane distillation, Journal of Engineered Fibers and Fabrics, 14 (2019) 1558925018824901.
[82] N.B. Goodman, R.J. Taylor, Z. Xie, Y. Gozukara, A. Clements, A feasibility study of municipal wastewater desalination using electrodialysis reversal to provide recycled water for horticultural irrigation, Desalination, 317 (2013) 77-83.
[83] T.O. Omotehinwa, Examining the developments in scheduling algorithms research: A bibliometric approach, Heliyon, 8 (2022) e09510.
[84] S. Al-Obaidani, E. Curcio, F. Macedonio, G. Di Profio, H. Al-Hinai, E. Drioli, Potential of membrane distillation in seawater desalination: Thermal efficiency, sensitivity study and cost estimation, Journal of Membrane Science, 323 (2008) 85-98.
[85] A.M. Alklaibi, N. Lior, Membrane-distillation desalination: Status and potential, Desalination, 171 (2005) 111-131.
[86] R.W. Schofield, A.G. Fane, C.J.D. Fell, Heat and mass transfer in membrane distillation, Journal of Membrane Science, 33 (1987) 299-313.
[87] A. Razmjou, E. Arifin, G. Dong, J. Mansouri, V. Chen, Superhydrophobic modification of TiO2 nanocomposite PVDF membranes for applications in membrane distillation, Journal of Membrane Science, 415-416 (2012) 850-863.
[88] T. Liskiewicz, G. Liskiewicz, J. Paczesny, Factors affecting the citations of papers in tribology journals, Scientometrics, 126 (2021) 3321-3336.
[89] K. Smolders, A.C.M. Franken, Terminology for Membrane Distillation, Desalination, 72 (1989) 249-262.