IOSR Journal of Business and Management (IOSR-JBM)
e-ISSN:2278-487X, p-ISSN: 2319-7668. Volume 25, Issue 7. Ser. 8 (July. 2023), PP 56-72
www.iosrjournals.org

Systematic Mapping on Clustering Techniques applied to Electricity Consumers

Thiago Eliandro de Oliveira Gomes
Federal University of Santa Maria, Postgraduate Program in Production Engineering, Santa Maria, Rio Grande do Sul, Brazil
Vinícius Jacques Garcia
Federal University of Santa Maria, Technology Center, Department of Production and Systems Engineering, Santa Maria, Rio Grande do Sul, Brazil

Abstract
The objective of this research is to map in a systematic and exploratory way the scientific articles that approach techniques for cluster analysis applied to electricity consumers. The mapping was done based on the literature published in major journals between the period 2015 to 2020. The methodology for this study was considering the preparation of a workflow established for the investigation, covering the mapping of candidate articles, the screening of studies, characterized by the selection of documents through indicators and inclusion and exclusion criteria, the quantitative and qualitative analysis of data, with a hybrid approach based on bibliometric laws and metrics to assess the impact of publications and, finally, the selection of publications related to the research topic. Most studies addressed methodologies used to identify patterns and assess the behavior of individual consumption of electricity, with a tendency to use the Fuzzy and K-Means methods, as they are traditional and efficient methods for the cluster analysis. The results made it possible to compose a detailed and unique portfolio of concepts usually accepted in the literature, as well as previous trends in customer segmentation were explored and the main algorithms for grouping consumers were presented, giving rise to further discussions on the subject.
Keywords: *Systematic Mapping; Cluster Analysis; Electricity Consumers. Bibliometric Analysis*

Date of Submission: 20-07-2023 Date of Acceptance: 30-07-2023

I. Introduction

Currently, the unprecedented volume of data generated from the most diverse sectors has shown exponential growth, which makes the data analysis process a challenging task without the use of specific and adequate techniques for processing during selection, analysis and transforming data into information aimed at better decision making and identifying market opportunities (GARZA-FABRE; HANDL; KNOWLES, 2018; SHUKLA; MUHURI; ABRAHAM, 2020).

Cluster analysis is one of the data mining methods, fundamental for transforming this data into multidimensional information. The grouping of data is a tool for exploratory data analysis, which aims to identify patterns or groups of objects homogeneous with each other, supported by the concept of similarity between the data (Garza-Fabre et al., 2018; Kassambara, 2017; Rajabi et al., 2020; Ramos et al., 2015)

In this way, several groups can be determined in order to capture the common characteristics that best define a specific context, through the attribution of objects, physical or abstract, similar to a set of data of interest, using cluster analysis tools as a procedure for analysis (Faria et al., 2016; Lin et al., 2020; Y. Wang et al., 2017).

Cluster analysis is widely used in several areas of knowledge (H. Zhou et al., 2019), especially for the electric power segment, clustering techniques are used to find similarity between the behaviors of electricity consuming users, classifying them into different groups, offering advantages to electricity concessionaires during their control and decision-making (Rajabi et al., 2020).

With population growth, there is an increasing demand for electricity on a daily basis (Sharma & Singh, 2015), causing the energy industry to have accumulated an enormous amount of data and, therefore, to be able to characterize and understand the behavior of its consumer market, using important support tools during its intelligent energy management process (Cembranel et al., 2019; Ramos et al., 2015; K. Zhou et al., 2016).

Particularly important for the global context, the electricity sector faces new challenges for the development of a model to respond to demand with confidence when facing uncertainties and the quest to maintain an adequate level of service. Essential for measuring and improving the performance of internal operational management systems and thus establishing a differential when compared to other competitors (Cembranel et al., 2019; Jiang & Seidmann, 2014; Ramos et al., 2015; Sharma & Singh, 2015; D. Wang et al., 2013).

From this perspective, (Cembranel et al., 2019), (Lv et al., 2019) and (Rajabi et al., 2020) believe that the implementation of appropriate grouping techniques, allows electric energy distributors, an intelligent management that can be applied to the analysis of the user's behavior in the electric grid, by gathering characteristics and factors that determine the profiles of users and their electricity consumption and, therefore, define a better categorization of the customer.

Given the relevance of the topic, this article presents a systematic mapping of the literature (SML) published in the main journals, between the period 2015 to 2020, through the electronic databases of IEEE Xplore, Scopus and Web of Science within the domain of the research. The choice of the time cut, aimed at, not exhausting all scientific production, but allowing to know and discuss the different aspects related to the theme (Snyder, 2019).

Systematic Review and Systematic Mapping (also called scope review) share a number of similar processes. However, both differ in their purpose. While the purpose of the Systematic Review is to summarize the available research to answer deep questions about certain particular topics, based on defined procedures, the Systematic Literature Mapping is used when it is not necessary to search for answers in depth and thus provide a broad view about the subject. of a given question, without the rigidity defined in a review (FELIZARDO et al., 2017; MOHER; STEWART; SHEKELLE, 2015; OKOLI, 2015; PHAM et al., 2014).

Although they share similar processes, systematic reviews are not always the best strategic alternatives and, therefore, in this study the choice falls to Systematic Mapping, as it identifies literature on a given topic in a broad way, without the need for a rigorous quality assessment, although it is useful, during data extraction, as performed in the reviews, which allows the presentation of the literature under an updated overview (Pham et al., 2014; Snyder, 2019)

Unlike the study carried out by Rajabi et al. (2020), who presented a comparison between different techniques for grouping electricity consumers based on their daily charge patterns, the main objective of this research is to map systematically and exploratory scientific articles that address techniques for cluster analysis applied to consumers of electricity, but, in a comprehensive way to establish a global panorama of the researches carried out on the topic and, thus, to identify trends and research gaps for future investigations.

From the point of view of this discussion, the structural organization of the article is arranged as follows: this introduction, dedicated to the presentation of contextualization on the theme, describing the concepts and discussions on the analysis of groupings found in previous works. Section 2 presents some works considered to be the state of the art for research. The characterization of the methodological procedures takes place in section 3. Subsequently, section 4 discusses the analysis of the results, as well as their discussions. Section 5 concludes the article by presenting the main conclusions in relation to the results found.

II. Related Studies

Cluster analysis is a mathematical tool whose main function is to classify variables according to the similarities in the characteristics of the objects under study (Cembranel et al., 2019; Ueda et al., 2020). In the literature, studies that use cluster analysis are common for problem-solving (Biscarri et al., 2017), as well as a wide variety of clustering algorithms available and applied by different researchers (Sharma & Singh, 2015), there is also the occurrence of comparison between different types of algorithms (Panapakidis & Moschakis, 2019).

Some of these algorithms fall into the following categories:a) algorithms of partitions such as K-means (Cembranel et al., 2019; Pan & Tan, 2019; Panapakidis & Moschakis, 2019), K-medoids (Cembranel et al., 2019; Panapakidis & Moschakis, 2019), G-means, X-means (Cembranel et al., 2019) and others, b) hierarchical algorithms, such as single link, complete link, Ward (Cembranel et al., 2019; Granell et al., 2015) and others, c) Fuzzy algorithms, such as Fuzzy C-Means, (Cembranel et al., 2019; Panapakidis & Moschakis, 2019; Rajabi et al., 2020) d) algorithms based on neural network, such as the Self-Organizing Map (SOM) (Panapakidis & Moschakis, 2019), or e) density-based nonparametric algorithms, such as DBSCAN (Pan & Tan, 2019; Ramos et al., 2015; Y. Wang et al., 2017) among others, in addition to the combinations between traditional algorithms forming contemporary hybrid algorithms and suitable for certain researches.

The different application varieties can be observed in different examples as in Lorentz et al. (2016), where the technique for understanding market strategies, examining the strategic configurations of small Finnish companies and their performance in terms of macroeconomic changes. Ueda, Souza and Menezes (2020), with the combination of cluster analysis and autoregressive vector model, explored the association in the use of macroeconomic variables linked to the Brazilian electronics sector, specifically making use of the number of people admitted and dismissed.

Asvatourian et al. (2018) applied the clustering method to explore dietary patterns between food intakes and their behavior. Jin et al. (2018) sought to save energy resources, and consequently alleviate environmental pollution with the rational use of explosion gas in the steel industry, while the precision related to the forecast of solar generation was studied by Pan and Tan (2019), when selecting similar days based on cluster analysis.

Particulary, on data clustering in the electrical area, in this article, the mapping carried out identified a context of groupings applied to the behavior of different categories of electricity consumers, of which parameters such as economic size, economic activity and energy consumption are used (Biscarri et al., 2017; Piao & Ryu, 2017), and that the grouping serves to assign consumers of similar electricity in the same class (Piao & Ryu, 2017).

However, the grouping of consumers can be approached in different ways. As for energy quality, as highlighted in Jasiński et al. (2019), who directed their research to the improvement of energy quality assessments, or else, to detect abnormal uses of energy which may compromise the reliability of the electrical network (Sharma & Singh, 2015).

Regarding the exploitation of the load profile, Panapakidis and Moschakis (2019), while mentioning the popularity of the K-means algorithm, also make reference to the success of the application in other researches. The authors proposed a modified form of the algorithm for analyzing the daily load profile based on cluster.

Biscarri et al. (2017) sought in their research to develop a set of rules based on the identification of customers' load profiles, to then automatically classify new consumers, in a way that allows for a tariff diversification due to the possible tariff diversification by the electricity concessionaire. For this, the authors resorted to the K-means, Diana, PAM, Clara, Fanny and SOM algorithms comparing them and selecting the most appropriate.

Sharma and Singh (2015), seeking to find the best cluster algorithm, included in their study the hierarchical K-means and Fuzzy C-means algorithms for load profile generation. Another use of the k-means clustering algorithm, but now with the addition of probit regression, was in the study by Rhodes et al. (2014), to explore variations in electricity use across time scales and find consumption-based residential electricity use profiles and thus form seasonal groups.

From the same perspective, Benítez et al. (2014) applied the static K-means algorithm to identify energy consumption patterns in Spanish homes and also to assess their general consumption trends more quickly according to the applied technique.

It is noted that when mining customer data, it is recognized by researchers as an important field in the investigation of several problems inherent to the segment of electricity, given that, when adopting techniques to extract and form data sets from which they have similarities, by the most varied parameters, it results in faster and more viable responses, which directly impacts the elaboration of valuable strategies for the management of electricity retailers as well as beneficial to end users (Biscarri et al., 2017; Cembranel et al., 2019; Panapakidis & Moschakis, 2019; Yang et al., 2019).

III. Research Methodology

Whichever approach is used in an MSL, a workflow must be run to make decisions about its development. Specifically, this research considered the one proposed by Okoli (2015), Heradio et al.(2016), Felizardo et al. (2017), Sivarajah et al. (2017) and Snyder (2019), going through the stages that refer to (1) planning mapping, (2) driving, (3) analysis, and finally, (4) writing the report to disseminate the results, as shown in Fig. 1.

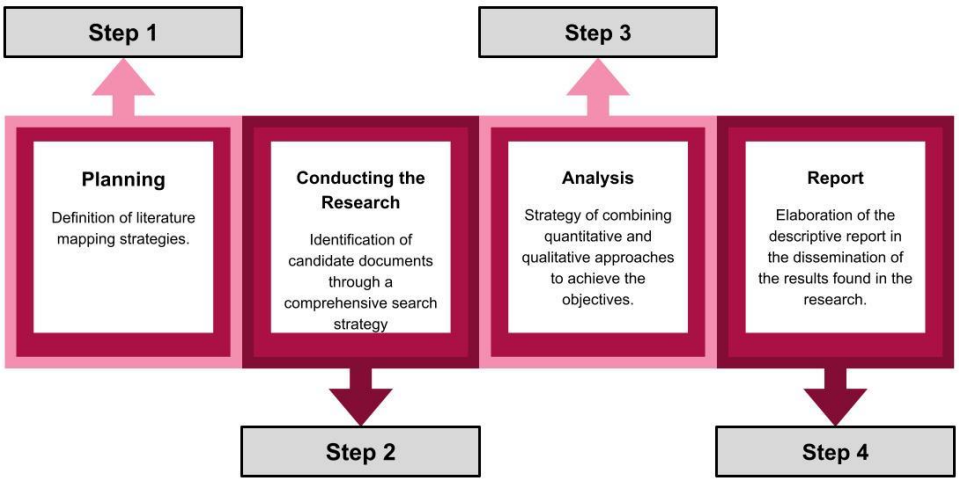


Fig. 1 Stages of development of the Systematic Mapping of Literature.

It is important to highlight that this research does not have the main characteristic of providing details about the SML process, but rather as a methodological strategy adopted for its development. The realization of the planning for the development of this mapping was part of Stage 1.

Stage 2: Conducting the research

Once the stages of development were defined, the candidate documents were previously identified, used a comprehensive search strategy reflected in the *string* and in its criteria for inclusion and exclusion of documents, aiming at results that express the state of the art of the research topic in question. Otherwise, studies relevant to the objectives intended in this mapping could not be contemplated.

The construction of the *string* search was based on the inclusion of terms related to the field of Cluster Analysis, applied to the electronic databases of *IEEE Xplore, Scopus* and *Web of Science*, which are providers of information that bring together the main research institutions of the world with about 20 million researchers and more than 20 million publications of the highest quality in the world, from which they create reliable research data (Heradio et al., 2016) as shown in Table 1, as well as the results regarding the number of publications in each database.

Table 1 Terms *(String)* search for access to documents

Database	Search Terms	results
IEEE Xplore	("Cluster Analysis" OR "Clustering heuristic" OR "Cluster algorithms" OR "Clustering Methods" OR "Consumer Clustering") AND (“Cluster * Analysi *” OR “Cluster * heuristic” OR “Cluster algorithm *” OR “Cluster * Method *”)	2,000
Scopus	(TITLE-ABS-KEY (“Cluster Analysis” OR “Clustering heuristic” OR “Cluster algorithms” OR “Clustering Methods” OR “Consumer Clustering”) AND (“Cluster * Analyzes *” OR “Cluster * heuristic” OR “Cluster algorithm *” OR “Cluster * Method *”))	1,532
Web of Science	TOPIC: (“Cluster Analysis” OR “Clustering heuristic” OR “Cluster algorithms” OR “Clustering Methods” OR “Consumer Clustering ”) AND TOPIC: (“ Cluster * Analysi *” OR “Cluster * heuristic” OR “Cluster algorithm *” OR “Cluster * Method * ”)	543
Total		**4,075**

For the study, inclusion and exclusion criteria were considered. The studies were candidates for inclusion in the mapping in (1) primary studies and (2) peer-reviewed studies, (3) published between January 2015 and October 2020 in the form of (4) articles, classified in the “Engineering” study area and (5) published in open access journals, of which (6) made use of clustering methods to support research.

Studies were excluded if they were (7) duplicated (only one copy of each study was included) and that had a (8) number of citations less than one and (9) that do not use the search terms in the title, abstract and / or keywords.

Once the candidate documents were identified, retrieved from the electronic databases, these were organized and executed in an automated way. For the automation of the workflow during the analyses, the was used package *bibliometrix*, which is a tool anchored in the R software with the nature of open source and that allows several forms of bibliometric analysis. Among the benefits that the R language provides, there is the integration with other packages, as well as quick update to meet the requirements of bibliometric analysis (Aria & Cuccurullo, 2017).

As a complement to assist in the development of the workflow, spreadsheets Excel belonging to the Microsoft Office package also were used as a computational resource, allowing other relevant information to be stored and organized, for later automation in the *bibliometrix.*

Study Identification

A total of 4,075 articles were returned by all search sources considered, 2,000 in the IEEE Xplore database, 1,532 in the Scopus database and 543 on the Web of Science database, according to the strings shown in the Chart 1. It is worth noting that the export of IEEE Xplore data was limited to the 2,000 best results, according to the policy applied by the platform itself.

For the screening of the studies, first there was the union of the bases and automatic removal of duplicate titles with the use of the tool *bibliometrix* of R. Then with the *software* Excel, a better analysis of the studies was carried out identifying the titles, abstracts and keywords that did not meet the inclusion criteria and removed them according to the exclusion criteria (7), (8) and (9). Only those who met the inclusion criteria remained in the research, making up a record of 836 studies that passed to the analysis phase, combining a quantitative and qualitative approach.

Identification of studies in line with the theme

The main objective of this study was to select articles that addressed the application of methods for grouping electricity consumers. Therefore, in order to identify studies that contemplated this approach, terms were applied for the selection of studies. The constructs used were: "*energy*" OR "*energy consumption*" OR "*electricity industry*" OR "*electricity demands*" OR "*power system*" OR "*electric power systems*" OR "*electricity customers*" OR "*distributed generation*" OR "*customer's power consumption*" OR "*electric load*".

Fig. 2 shows the flowchart of the methodological process carried out during the selection of the studies, of which its taxonomy regarding the classification, including all stages, was composed as follows: identification of the studies, screening, selection related to the theme and sample to analyze.

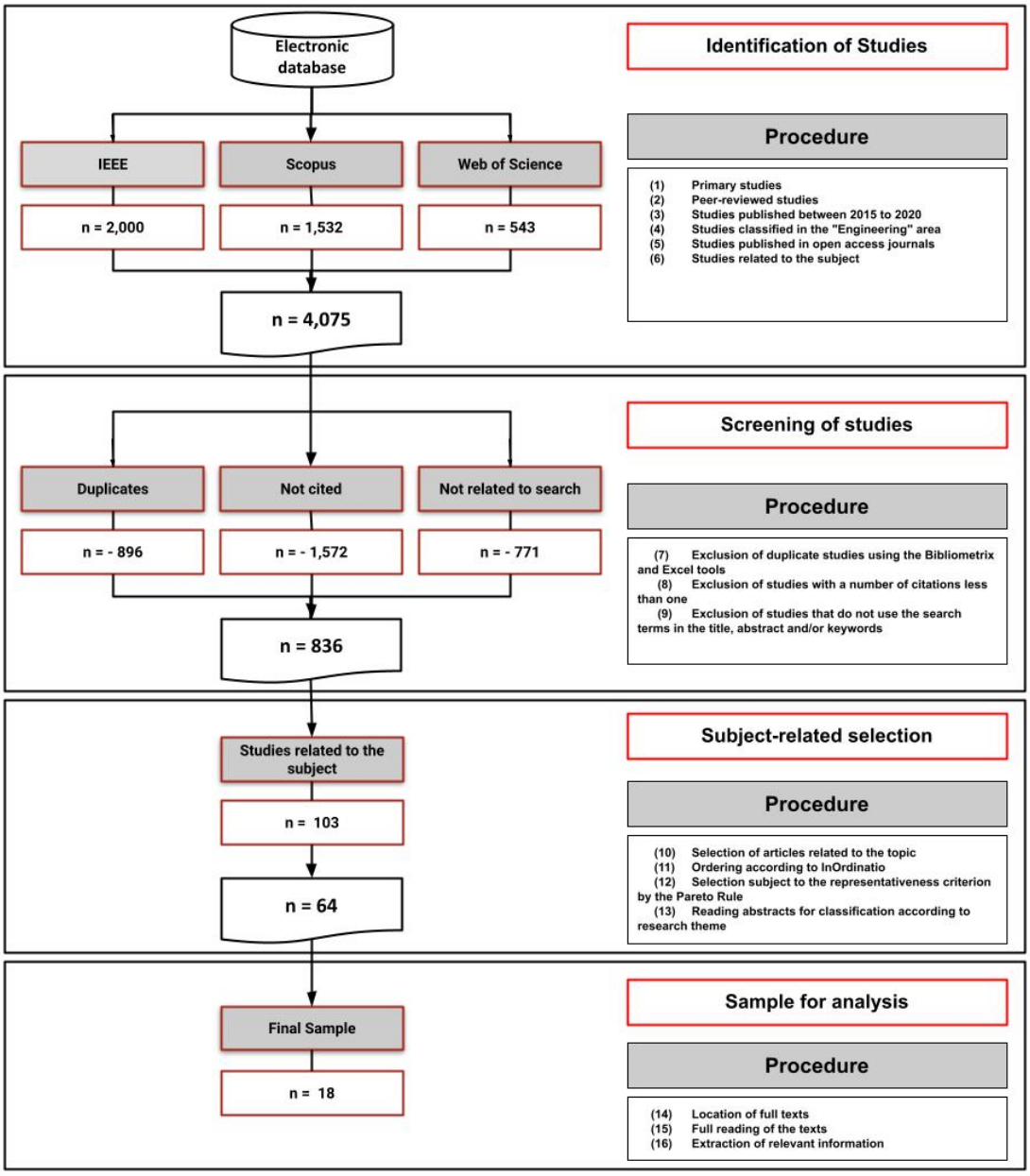


Fig 2 Flowchart of the methodological process for the selection of studies

After the application of the filters for the screening of studies (step 10), 103 documents were obtained in line with the theme, of which the Methodi Ordinatio (step 11) was ordered. Called the Ordinatio Index (InOrdinatio), this index refers to a methodology developed for methodological ordering of articles regarding their scientific relevance, combining impact factor, year of publication and the number of citations attributed to the document (Pagani et al., 2018).

Then, in step 12, the selection was subject to the criterion of representativeness based on the Pareto principle (rule 80-20) to select impactful articles. That is, all publications whose relative frequency was above 80% (de Carvalho et al., 2020) were selected, generating a set of 64 articles selected by both methods concurrently.

Once the selection was made, the abstracts of the 64 documents were read (step 13) and of these, 18 were pre-selected, considered aligned with the object of the research, from which they were read in full (steps 14 and 15).

In preparing the material for extracting the information (step 16), two databases were built to store information. The first had a bank of 836 documents generated in the screening stage, and the second, with the 18 documents related to research. The definition of these bases, aimed at expanding the discussion and its potential regarding theoretical and practical contributions, and the identification of the main published studies on the topic.

While the first database underwent a process of fluctuating reading and subsequent quantitative analysis, the second database underwent a qualitative approach by reading the documents in full, in order to identify relevant studies on the topic addressed in the research, resulting in 6 related articles to the proposed theme. The last step involved the extraction and analysis of relevant information regarding the grouping and consumers of electricity.

Stage 3: Analysis of the information After

Retrieving the records of the considered bases, different approaches were used to examine the information, since each approach gathers presuppositions of which must be weighed, and it is up to the researcher to choose those that respond to his research objectives (Pagani et al., 2018; Snyder, 2019).

Descriptive analysis is applied to the content, seeking to identify trends and patterns in the literature, as well as analysis of the temporal evolution of publications, number of authors and citations, sources of publications, journals, geographic distribution of authors and institutions.

In view of the data, the results were based on the combined use of the bibliometric laws of Lotka and Bradford, originating in 1926 and 1934, respectively, which allow the interpretation of performance, respectively, by assessing the productivity of the authors according to the number of articles published and, of the journals, estimating their degree of relevance, as well as in the Zipf Law, originating in 1949, which measures the frequency of the occurrence of the words presented in the texts (Lima & Carlos Filho, 2019; Machado Junior et al., 2016; Rons, 2018).

This article also uses indicators to assess the impact resulting from the publication, in particular, the H index for the analysis of citations received (Aria et al., 2020; Heradio et al., 2016). Subsequently, another metric used was the analysis of co-citation, given by the connections between the authors through the citation and evaluated by the presence of citations of two articles by a third party (Aria & Cuccurullo, 2017; Rodríguez-Soler et al., 2020).

The hybrid approach, the results of combining different dimensions, capture the contributions with better accurately will when referring to performance analysis and quantification of the impact of scientific data in the field of research (Heradio et al., 2016; Petersen et al., 2015; Rons, 2018).

Limitations of the mapping

One factor that caused limitation during the development of the SML, was regarding the treatment of the data retrieved from the databases online. The tool *bibliometrix* automatically executes the metadata for the scientific production of Web of Science and Scopus, but does not execute data retrieved from the IEEE Xplore database (Aria & Cuccurullo, 2017). Therefore, a pre-processing before the analyses was necessary. With Software R, the bases were converted to a file in *software* Excel, for later manual inclusion of IEEE data, and simultaneously the final cleaning of the data and the standardization of different information were carried out, as well as recognizing and eliminating other documents. Duplicated after the inclusion of the new data (Heradio et al., 2016).

IV. Results and Discussions

This section presents the results found after performing the Systematic Mapping from the selected primary works, in the same way that it describes the analyses carried out as a consequence of the application of the referred mapping of the literature. To demonstrate the results, the analysis combined a quantitative approach and content analysis with a qualitative approach.

General information about the works

As shown in Fig. 2, the initial consultation identified 4,075 articles, resulting in 836 articles eligible for quantitative analysis after the screening procedures. The results expressed in this section are organized with a combined use of the bibliometric laws of Lotka, Bradford and Zipf. The distribution of the selected articles is presented graphically, in Fig. 3, according to their publication by year.

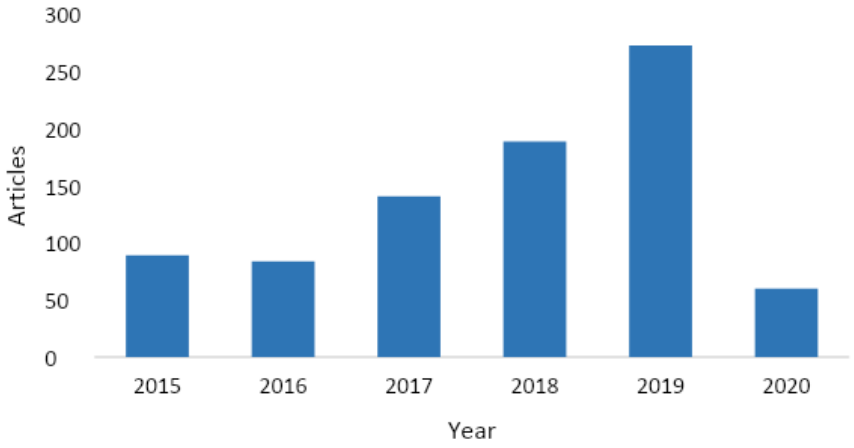


Fig. 3 Annual distribution of publications during the period considered

The Figure 3 shows the total production, during the analyzed period, and shows the distribution of publications presented by all authors and co-authors during the studied time cut. With the number of articles published, evidence of exponential behavior can be seen in studies on the subject in question, characterized by accelerated growth in the years 2017 to 2019. However, as noted, there is a drop in 2020, but this is because the year was still in effect during the survey and therefore the data was not complete. The highest concentration of publication was observed in 2019 with 273 articles published, representing 33% of the total production recovered. Chart 2 presents the main information about bibliometric analysis.

Table 2 Main information on bibliometric analysis

Description	Results
MAIN INFORMATION ABOUT DATA	
Sources (Journals, Books, etc.)	258
Documents / Article	836
Average years from publication	2,22
Average citations per documents	7,657
Average citations per year per doc	2,209
References	22870
DOCUMENT CONTENTS	
Keywords Plus (ID)	4307
Author's Keywords (DE)	3016
AUTHORS	
Authors	2368
Author Appearances	3309
Authors of single-authored documents	26
Authors of multi-authored documents	2342
AUTHORS COLLABORATION	
Single-authored documents	26
Documents per Author	0.353
Authors per Document	2.83
Co -Authors per Documents	3.96

The data shown in Table 2 shows that the 836 documents received together, over the period under study, an average number of citations of approximately 7.65 per document and the average annual value of 2.21 citations per document published among the authors and co-authors, with 0.35 documents per author. The documents were written by 2,368 authors. On average, each document was written by 2.83 authors, of which approximately 3% of the authors wrote articles without the participation of co-authors.

High contribution, articles and keywords

Highlight for the works published by journals Nakhleh et al. (2017), published in ACS Nano, cited 159 times, and followed by D'Oca S (2015) with 137 citations during the analyzed period. From this number of citations, it can be inferred that it is due to the time elapsed from publication, since such publication refers to the year 2017 and, therefore, a longer period of time can cause a high number of citations, however, this fact is not a rule, as seen, given that the second-highest number of citations belongs to a document published in 2015.

The main publication journals on the topic covered were also analyzed, which presented 258 different newspapers, Chart 2. The results indicate that 334 (40.10%) primary works were published in IEEE Access and another 13 (1.56%) in IFAC-PAPERSONLINE. The rest are below 9 articles published as shown in Table 3.

Table 3 Ten main journals and their indicators

Journal	Articles	Cite Score	FI JCR
IEEE Access	334	3.745	3.9
IFAC-Papersonline	13	**	1.6
Indonesian Journal of Electrical Engineering and Computer Science	9	**	1.4
International Journal of Electrical and Computer Engineering	9	**	2.3
Sensors (Basel)	8	**	**
Turkish Journal of Electrical Engineering and Computer Sciences	8	0.682	1.6
Advances in Electrical and Computer Engineering	7	1.102	1.8
European Journal of Enterprise Technologies	7	**	1.9
Journal of Systems and Electronics Engineering	7	0.907	2
Tehnicki Vjesnik	7	**	1.4

In addition to the number of articles published in each source, their classification was raised regarding the Cite Score and Journal Citation Reports Impact Factor (JCR) indicators, metrics accepted worldwide for the evaluation of scientific journals (Shukla et al., 2020), corroborating that IEEE Access is not only the one that published the most articles, but also the one that presents the best indicators.

With Fig. 4, it is possible to check the graphical demonstration of the performance of the journals in highlights, distributed by year, where the peak occurs in 2019.

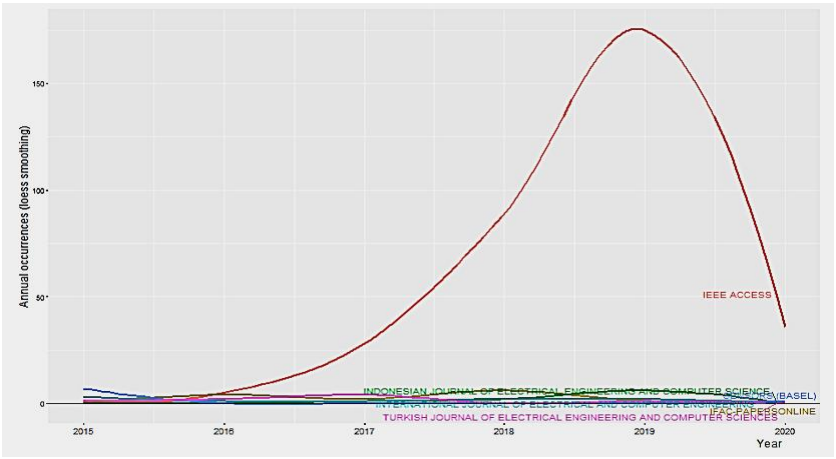


Fig. 4 Graphical demonstration of the Journals highlighted

Interestingly, ACS Nano, despite the work of Nakhleh M. (2017) leading the number of citations, it is the IEEE ACCESS that is on the cutting edge with the highest H Index (16), which is a measurement index based on received citations (Aria et al., 2020; Aria & Cuccurullo, 2017) with 1723 citations. Sensors (Basel) and Energy and Buildings are other sources that appear shortly thereafter, both with 6 and in their H Index, cited, respectively, 227 and 216 times.

The analysis of the 3016 keywords, Table 2, attributed by the authors showed that, after the exclusion of the keywords used as search terms (String) for accessing the documents, Table 1, the ones that stood out the most were "clustering" (n = 119), “data mining” (n = 32), “machine learning” (n = 30), “k-means” (n = 25) and “k-means clustering” (n = 21). Fig. 5 presents the graphic tree of the 25 main keywords.

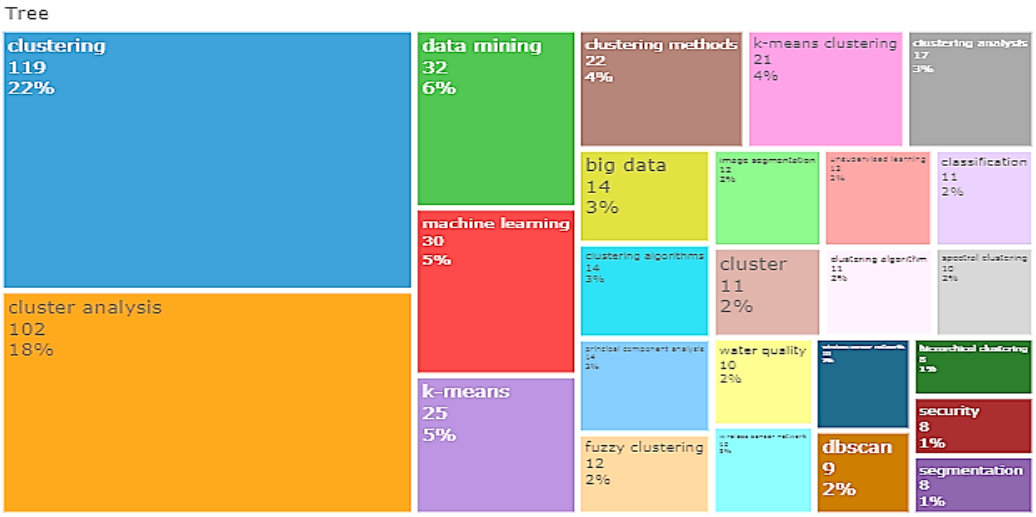


Fig. 5 Graphical tree of the 25 main keywords

Considering that some words do not seem to be important, when analyzed in isolation, the combination of them in a general context, may give evidence of the identification and effective construction of trends and specialties in the literature (Rons, 2018). Applied to Zipf's law, the results provide evidence of its validity, given that the analysis of the keywords assigned by the authors provided the support for identifying topics that are most addressed in the articles.

Geographical distribution of the Productivity of Countries and institutions

This analysis is nothing more than the geographical distribution of articles published by the respective authors. The Figure. 6 shows the origin of the articles published in the countries that make up the collection of research data, showing the degree of the color shade of blue (the stronger the shade of blue, the greater the number of publications). In the survey of the sample data, 76 different countries were identified.

To assess collaboration between countries, the articles were divided into two groups: Single Country Publications (SCP) and Multiple Country Publications (MCP). According to Aria, Misuraca and Spano (2020), the first refers to the number of articles produced by authors from a single country and, the second, the number of articles published in scientific collaboration with other different countries. Fig. 6 shows the SCP, represented in green, while in orange are the articles published in MCP.

The analysis of the country's scientific production showed that the geographic location of the articles, Fig. 6, indicates that China is the most relevant country by number of published articles related to the theme. China stands out, with a total of 42 articles, 11 of these articles published only by Chinese authors and 31 publications with the collaboration of authors from other countries. There are also 32 articles published in the United States of America (USA), of which 9 of them have established international collaboration and 23 without scientific collaboration between different countries.

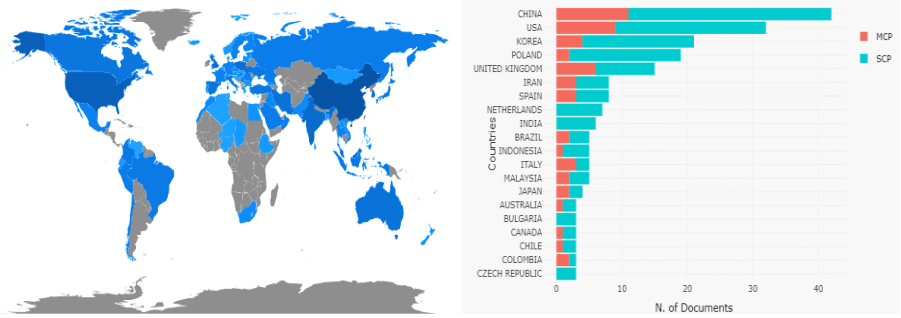


Fig. 6 Geographic location of the origin of articles

However, there is an inversion of positions when referring to the total number of citations (TC). While China is in the first position as the most productive country, the number of citations has been attributed to the USA (n = 481 with an average of citations per article (ACA) of 15.03) is higher than that presented by China (n = 436 with ACA of 10.38) and the United Kingdom (n = 266, with an MCA of 17.73).

As Aria, Misuraca and Spano (2020), this study used as a metric to assess the most collaborative country, the proportion between the number of articles published by authors from different countries, the MCP, and the total number of publications from each country. Thus, it was found that the United Kingdom is the country that contributes the most in the element of international collaboration, revealed by the rate of 40% (6 MCP of 15 articles), followed by the USA and China, respectively with 28% (9 MCP of 32 articles) and 26% (11 MCP of 42 articles). In addition, the three countries together received 1183 citations in 89 published articles.

Still, 1285 different institutions of origin of the respondents were found, which among them, the National Natural Science Foundation of China, ranked first in terms of their productivity of the institutions, with 92 (8.8%) articles published, followed by Fundamental Research Funds for the Central Universities and Jilin University, both also from China, with 22 and 12 articles published, respectively.

Impact of annual production during the period

During the analyzed period, considering the total number of publications, the database presented 3,309 authors and co-authors who, together, wrote 10,045 pages, representing approximately 3 pages for each researcher and an average value of 12 pages for each published document.

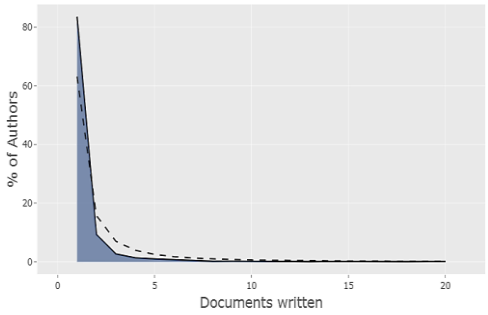
With regard to the distribution of annual scientific production and its average number of citations in 2015, despite having the fourth position in relation to the number of articles, still, it was in first place with regard to the average of citations per article (16.55). Such data can be seen in Table 5.

Table 5 Distribution per year in scientific production and its average number of citations

Year	Articles	Authors	Average Cit. / Art.
2015	89	331	16,55
2016	84	292	12,67
2017	141	578	11,93
2018	189	747	5,92
2019	273	1123	3,41
2020	60	238	2,13
Total	836	3309	--

Regarding the authors' productivity based on Lotka's Law (Fig. 7, a), the researchers “Li Y.” and "Wang J." published 20 articles each, representing less than 0.08%, while 1979 authors published 1 (article) equivalent to 84% of the total number of authors, as shown in Fig. 7.

(a)



(b)

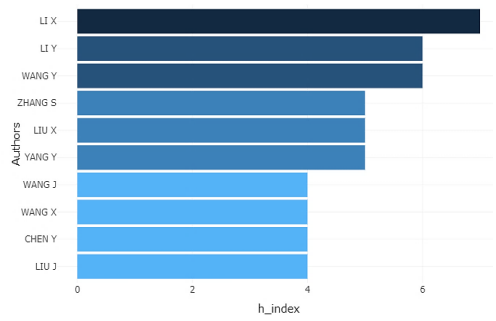


Fig. 7 Frequency distribution of scientific productivity (a) and h index of the authors (b)

The evaluation of the authors' productivity, makes Lotka's bibliometric law valid, since it was noticed that 84% of authors and co-authors wrote a single scientific article, while less than 0.25% of the authors and co-authors produced between 15 and 20 works, a fact which allows to identify that, as few authors have a high production, simultaneously many appear with low production (Machado Junior et al., 2016).

Similar results are found in the study developed by (Fortuna et al., 2020), most of the authors included published between 1 (999 authors) and 2 articles (188 authors), while the most productive researchers published between 10 and 19 articles.

Regarding the indicators related to the author's local impact (Fig. 7, b), the authors with the highest indexes are "Li X." with a h index of 7, with a total citation of 197 and total citation per article (TCA) of 15. This author was followed by “Li Y.” and “Wang Y.”, both with a h index of 6 and a TC of 131 and 144, respectively. These data are valid to measure the merit of an article that, although there are other aspects that may influence it, the amount of citations received directly reflects the prestige of the article before the scientific society (Merigó et al., 2015). Fig. 8 shows the Dendrogram for the ten most frequent keywords.

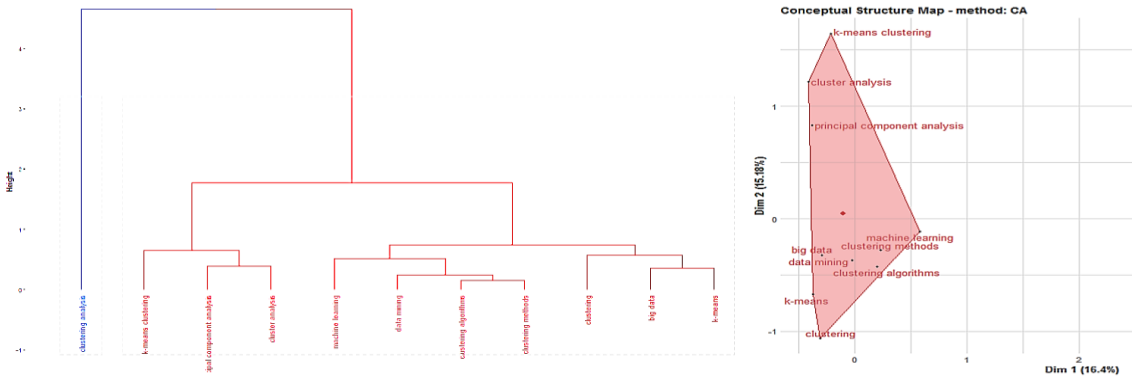


Fig. 8 Dendrogram for the ten most frequent keywords

Analyzing the dendrogram of similarity between more frequent terms, Fig. 8, it can be seen that the production related to most words remained similar, forming more homogeneous groups (in red color), but, on the other hand, that the words “Clustering Analysis ”(in blue) formed a group distinct from the others, that is, during the analyzed period, this production had its own characteristic. However, a variation of the distinct term remained in homogeneity, through the term “Cluster Analysis”.

This analysis gave exclusive attention to the ten most frequent keywords attributed by the authors, and the other words were not represented in the dendrogram, due to their enormous number (3016). Fig. 9 shows the diagram of the Co-citation network.

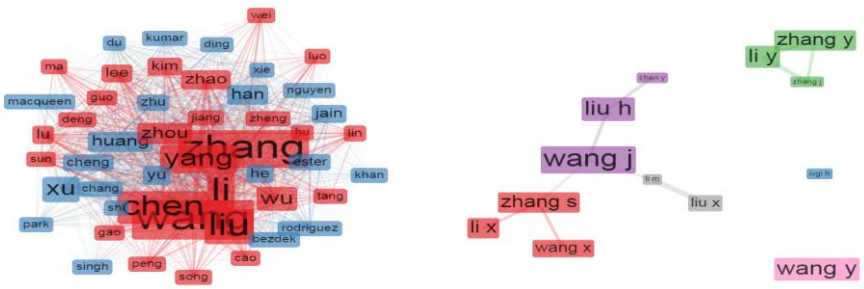


Fig. 9 Diagram of the Co-citation and Collaboration network presented between the main authors

In the field of intellectual and social structure, the analysis is based on two networks of scientific collaboration and their partnerships formed between the authors. The first refers to the specific field of the co-citation, Fig. 9, (a), due to the links between the authors through the citation, the second, shows which study groups are regular, Fig. 9, (b), through the analysis of the relationship between authors. That deal with the research topic. Nevertheless, cooperation networks are also tools that can discover the main scholars of a given subject, allowing the expansion and / or initiating new relationships with the main authors (Aria & Cuccurullo, 2017).

Thus, the formation of four groups of authors was highlighted, highlighted in green, violet, red and gray. The groups formed among the researchers stand out (Fig. 9, b) "Li Y." and "Zhang Y" in green, which are leaders in terms of productivity, both with 20 articles each. This fact reveals the existence of gaps and possible new formations, since the author "Wang Y.", the third-largest author of articles (18) is isolated, and that partnerships with him would be relevant for the enhancement of publications. However, in general, the groups formed explain the reason for how influential the authors are, since in the composition of all the collaboration groups, there are the ten main authors, when analyzed under their productivity potential and regarding their respective indexes h.

Given the main objective of this study, which was to map which methods were applied to group electricity consumers, to investigate trends, the distribution in the form of the most frequent terms relative to their respective methodologies employed over the period in question is shown in Fig. 10. For the construction, the information was extracted, which presents a structure with the cluster formation techniques identified in the abstracts of the 836 studies.

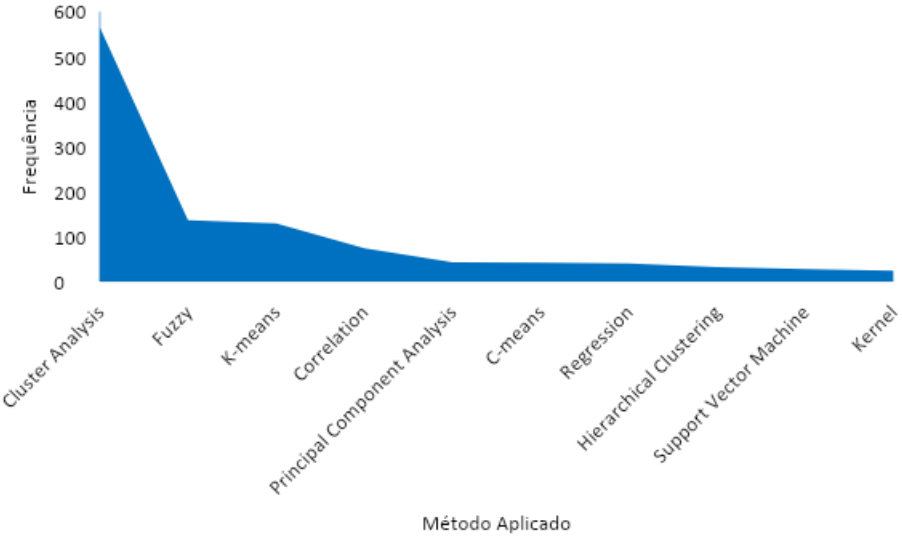


Fig. 10 Most relevant applied methods

In order to investigate the trends regarding the most applied methodologies, methods and / or algorithms, the similar Cluster Analysis classification (Cluster Analysis; Clustering Analysis; Clustering Method) was gathered, which comprises all those methods where the abstract of the article did not define your method explicitly. The methods employed are the most diverse, commonly related to the formation of data clusters.

Based on Fig. 10, and considering the studies, with regard to cluster analysis, the studies focus on 41% of the frequency of occurrence of applications, mentioning the use of cluster-based analyzes. However, individually the authors of the articles resorted to Fuzzy groupings; K-Means, coming close to 10%, followed by Correlation, with 5% of occurrences. It is worth remembering that in some documents there was a combination of two or more methods applied to arrive at the results of the research.

These data shown in Fig. 10, when put to the test, comparing them to the most frequent words found in the keywords assigned by the authors, is shown in Fig. 11.

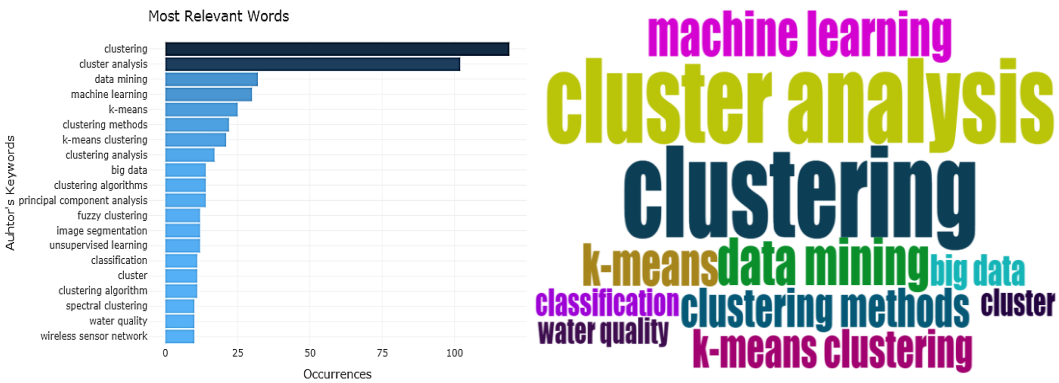


Fig. 11 Most frequent keywords of the studies considered and their word cloud

In Fig. 11, there is confirmation of the data found in Fig. 10. It is verified that the group of Cluster Analysis, formed by its similar ones (Cluster Analysis; Clustering Analysis; Clustering Method) appears among the most recurrent words, as well such as "fuzzy", "clustering k-means" and "main component analysis".

It is perceived that in the studies, the cluster analysis, or specifically the Fuzzy and K-Means methods, are clustering algorithms that are gaining attention in the middle of the research (Shukla et al., 2020), and which are the most used by researchers.

Qualitative analysis based on reading

The complete reading of the articles, sought to identify the particularities of each study and which theoretical directions were being discussed, through the analysis of the results obtained. The mentioned works were selected considering not only the content worked on, but also its alignment and its correlation with the research theme.

After the application of methods concomitantly provided in the literature for the selection of documents, of which they served as criteria, 18 articles were considered pre-selected in line with the object of the research. Ahead, all 18 papers were located and read in full, of which 9 (nine) were directly related to the topic, had their relevant information extracted and subsequently analyzed.

After following the steps described, the articles are presented in chronological order, in order to follow the evolution of research in the area.

Table 6 Pre-selected articles after applying the selection criteria

Title	Authors, Year, Journal	Selection
Factors influencing residents energy use: a study of energy related behaviour in 57 Swedish homes	HILLER, 2015, Energy and Buildings	No
Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data	HSU, 2015, Applied Energy	Yes
Impacts of raw data temporal resolution using selected clustering methods on residential electricity load profiles	GRANELL; AXON; WALLOM, 2015, IEEE Transactions on Power Systems	Yes
Aggregation and remuneration of electricity consumers and producers for the definition of demand-response programs	FARIA; SPÍNOLA; VALE, 2016, IEEE Transactions on Industrial Informatics	No
Analysis and clustering of residential customers energy behavioral demand using smart meter data	HABEN; SINGLETON; GRINDROD, 2016, IEEE Transactions on Smart Grid	Yes
Occupancy data analytics and prediction a case study	LIANG; HONG; SHEN, 2016, Building and Environment	No
A comparison study on node clustering techniques used in target tracking WSNs for efficient data aggregation	MAHDI et al., 2016, Wireless Communications and Mobile Computing	No
Multi-layered clustering for power consumption profiling in smart grids	AL-JARRAH et al., 2017, IEEE Access	Yes
K-means based load estimation of domestic smart meter measurements	AL-WAKEEL; WU; JENKINS, 2017, Applied Energy	Yes
Energy aware multi-hop routing protocol for WSNs	CENGIZ, 2017, IEEE Access	No
A generalised model of electrical energy demand from small household appliances	SANCHO-TOMAS; SUMNER; ROBINSON, 2017, Energy and Buildings	No
C-Vine copula mixture model for clustering of residential electrical load pattern data	SUN; KONSTANTELOS; STRBAC, 2017, IEEE Transactions on Power Systems	Yes
An unequal clustering algorithm concerned with time-delay for internet of things	FENG et al., 2018, IEEE ACCESS	No
Developing urban residential reference buildings using clustering analysis of satellite images	LI et al., 2018, Energy and Buildings	Yes
Predicting the energy consumption of residential buildings for regional electricity supply-side and demand-side management	CAI et al., 2019, IEEE Access	Yes
Energy-efficient routing in WSN: a centralized cluster-based approach via grey wolf optimizer	DANESHVAR et al., 2019, IEEE Access	No
An energy-efficient clustering algorithm combined game theory and dual-cluster-head mechanism for WSNs	LIN; WANG, 2019, IEEE Access	No
An optimizing and differentially private clustering algorithm for mixed data in SDN-based smart grid	LV et al., 2019, IEEE Access	Yes

The complete reading of the articles, sought to identify the particularities of each study and which theoretical directions were being discussed, through the analysis of the results obtained. The works mentioned here were selected considering not only the content worked on, but also its alignment and its correlation with the topic covered in the research. With the individual analysis of primary works, it was possible to identify aspects regarding the use of grouping methods to support the search for results.

Cluster analysis aims to gather objects based on similar characteristics to each other and has its use appropriate to the most diverse areas of study and for different purposes, providing accurate and reliable predictions (Garza-Fabre et al., 2018; Hsu, 2015; Lin et al., 2020; Y. Wang et al., 2017).

Particularly for this research, the cluster analysis, when dedicated to the electric power segment, was present in a relatively small number. In the research developed by (Granell et al., 2015), cluster analysis was applied to group electricity load profiles in order to help distinguish the categories of users for project and tariff switching, fault detection and fraud, demand side management and energy efficiency measures.

In the same way, (Al-Wakeel et al., 2017) proposed an algorithm derived from the K-means method to group the load profiles of different types of Irish customers in distribution networks, based on historical measurements of energy consumption extracted from meters and subsequent analysis of peak demand behavior and general electricity consumption.

In other research, Cai et al. (2019), Li et al. (2018) and Hsu (2015) used cluster analysis to model energy consumption in residential buildings. Cai et al., (2019) used a model based on a combined algorithm to predict the electricity use of residential buildings, which later served as a reference for a possible decision to rationally allocate energy supply in a given region, while in the study by Li et al. (2018) residential building clusters were formed based on data of geometric characteristics constructed from the images extracted from satellite in order to investigate energy consumption based on the shape of the building.

For the study by Li et al. (2018) two partial grouping techniques are employed, K-means and K-medoids, and their results indicated that the K-medoids grouping technique performed best, as did Hsu (2015) who employed two K-means models to 4 and 9 clusters, combined with the Clusterwise regression model, which when comparing them with each other, provided results related to the accurate and stable forecast of data referring to the energy consumption of buildings.

Haben et al. (2016) and Al-Jarrah et al. (2017) carried out the analysis of the behavioral demand of residential customers in Ireland and the United States, respectively. Although, researches point to the use of different algorithms to form clusters, both concluded that cluster analysis provides stable and reliable clusters.

In the research by Lv et al. (2019), the preservation of users' privacy was addressed during the collection and analysis of personal information to form clusters. A private optimization and differential clustering algorithm was proposed for mixed data and, finally, its results were compared to the Normalized Intra-Cluster Variance (NICV), with lower values.

From the same perspective as Li et al. (2018), the study by Sun et al. (2017) dealt with charge pattern clustering, by grouping residential electricity customers according to the consumption behavior of each customer. For that, a new grouping algorithm was proposed to assign new consumers to the existing classes.

The references of the articles related to the study and application of the cluster analysis were analyzed as a secondary source, without considering the period, but rather the theme, adding the collection of articles linked to the theme, the article by (Chicco et al., 2006), totaling 10 articles.

In this article, customer groups are formed according to their consumption patterns in order to build a tariff structure. The authors compare the results obtained with various clustering algorithms, including K-means and fuzzy K-means and self-organizing maps, while applying the main component analysis and curvilinear component analysis in order to reduce the set of data and thus streamline the calculations for the formation of the cluster (Chicco et al., 2006).

V. Final Considerations

In this article, the objective was to map the panorama about the techniques for cluster analysis when applied to electricity consumers. The proposed mapping went through two moments. First, bibliometric techniques were applied to studies related to the topic, where previous trends in the grouping of electricity consumers were explored and, thus, ensure prior knowledge of the works that addressed the use of algorithmic applications on grouping.

In this study, analyzes were performed under a quantitative approach and a content analysis to describe the effects established during the planning phase of the research by adopting a comprehensive strategy making the search criteria linked to more flexible choices, so that more results on the topic were contemplated.

Of the studies analyzed, China was the country with the largest number of articles published, with researchers “Li Y.” and "Wang J." making them the most productive on the subject. This is corroborated when it was observed that the main institutions that work with the application of cluster analysis in their research are Chinese, as they appear at the top of the institutions, such as the National Natural Science Foundation of China, Fundamental Research Funds for the Central Universities and Jilin University.

An evident counterpoint was related to the study by Nakhleh et al. (2017), which used cluster analysis to classify diseases in the same categories, using variables associated with the field of medicine, published in the ACS Nano, which was between the articles cited in the analyzed period, and that this one, belongs to an Israeli institution, but are under the same perspective of the preliminary studies carried out, where it is found the different varieties of the application of the grouping techniques to other fields of research.

Despite the Chinese prominence, an important issue is noted. The United States has the most influential production by having the position of leader in the works cited, indicating that its research has been widely accepted

and that it has a strong influence in the academic environment, and that the focus of American researchers is on publishing studies that address the topic addressed In this job.

In a second step, a qualitative analysis of the studied content sought to identify the theoretical directions of the selected studies and their correlations to the theme. From a general aspect, the results obtained agree with the previous knowledge presented, since, for the most part, the qualitatively evaluated studies presented methodologies used to the consumption behavior of each client to identify patterns.

Although there is a high degree of dispersion in this field of research, with numerous techniques for forming clusters of clients being used, when analyzing their respective methodologies used during the development of research, it was clear that there is a trend regarding the use of cluster analysis, or specifically, attention focused on the Fuzzy and K-Means methods. It is believed that this frequent use is due to the fact that they are already traditional methods and thanks to the efficient results presented over time.

Although this article has been limited by the scarcity of relevant studies that address the clustering techniques applied to electricity consumers, in the analysis of the articles related to the theme, it allowed the identification of the main contexts of the application of the models for grouping electricity customers, whose its parameters served to compose a detailed and unique portfolio of concepts usually accepted by the literature, as well as exploring previous trends in customer segmentation and the main algorithms for grouping consumers were presented, giving rise to further discussions on the topic.

Acknowledgements

To the R&D Program regulated by ANEEL and executed by the State Electricity Distribution Company via Public Call No. CEEE-D 001/2018.

References

[1]. Al-Jarrah, O. Y., Al-Hammadi, Y., Yoo, P. D., & Muhaidat, S. (2017). Multi-Layered Clustering For Power Consumption Profiling In Smart Grids. IEEE Access, 5, 18459–18468. Https://Doi.Org/10.1109/ACCESS.2017.2712258

[2]. Al-Wakeel, A., Wu, J., & Jenkins, N. (2017). K-Means Based Load Estimation Of Domestic Smart Meter Measurements. Applied Energy, 194, 333–342. Https://Doi.Org/10.1016/J.Apenergy.2016.06.046

[3]. Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-Tool For Comprehensive Science Mapping Analysis. Journal Of Informetrics, 11(4), 959–975. Https://Doi.Org/10.1016/J.Joi.2017.08.007

[4]. Aria, M., Misuraca, M., & Spano, M. (2020). Mapping The Evolution Of Social Research And Data Science On 30 Years Of Social Indicators Research. Social Indicators Research, 149(3), 803–831. Https://Doi.Org/10.1007/S11205-020-02281-3

[5]. Asvatourian, V., Craig, T., Horgan, G. W., Kyle, J., & Macdiarmid, J. I. (2018). Relationship Between Pro-Environmental Attitudes And Behaviour And Dietary Intake Patterns. Sustainable Production And Consumption, 16, 216–226. Https://Doi.Org/10.1016/J.Spc.2018.08.009

[6]. Benítez, I., Quijano, A., Díez, J. L., & Delgado, I. (2014). Dynamic Clustering Segmentation Applied To Load Profiles Of Energy Consumption From Spanish Customers. International Journal Of Electrical Power And Energy Systems, 55, 437–448. Https://Doi.Org/10.1016/J.Ijepes.2013.09.022

[7]. Biscarri, F., Monedero, I., García, A., Guerrero, J. I., & León, C. (2017). Electricity Clustering Framework For Automatic Classification Of Customer Loads. Expert Systems With Applications, 86, 54–63. Https://Doi.Org/10.1016/J.Eswa.2017.05.049

[8]. Cai, H., Shen, S., Lin, Q., Li, X., & Xiao, H. (2019). Predicting The Energy Consumption Of Residential Buildings For Regional Electricity Supply-Side And Demand-Side Management. IEEE Access, 7, 30386–30397. Https://Doi.Org/10.1109/ACCESS.2019.2901257

[9]. Cembranel, S. S., Lezama, F., Soares, J., Ramos, S., Gomes, A., & Vale, Z. (2019). A Short Review On Data Mining Techniques For Electricity Customers Characterization. 2019 IEEE PES GTD Grand International Conference And Exposition Asia, GTD Asia 2019, 194–199. Https://Doi.Org/10.1109/Gtdasia.2019.8715891

[10]. Chicco, G., Napoli, R., & Piglione, F. (2006). Comparisons Among Clustering Techniques For Electricity Customer Classification. IEEE Transactions On Power Systems, 21(2), 933–940. Https://Doi.Org/10.1109/TPWRS.2006.873122

[11]. De Carvalho, G. D. G., Sokulski, C. C., Da Silva, W. V., De Carvalho, H. G., De Moura, R. V., De Francisco, A. C., & Da Veiga, C. P. (2020). Bibliometrics And Systematic Reviews: A Comparison Between The Proknow-C And The Methodi Ordinatio. Journal Of Informetrics, 14(3), 101043. Https://Doi.Org/Https://Doi.Org/10.1016/J.Joi.2020.101043

[12]. Faria, P., Spínola, J., & Vale, Z. (2016). Aggregation And Remuneration Of Electricity Consumers And Producers For The Definition Of Demand-Response Programs. IEEE Transactions On Industrial Informatics, 12(3), 952–961. Https://Doi.Org/10.1109/TII.2016.2541542

[13]. Felizardo, K. R., Nakagawa, E. Y., Fabbri, S. C. P. F., & Ferrari, F. C. (2017). Revisão Sistemática Da Literatura Em Engenharia De Software: Teoria E Prática (1st Ed.). Elsevier.

[14]. Fortuna, G., Aria, M., Piscitelli, A., Mignogna, M. D., & Klasser, G. D. (2020). Global Research Trends In Complex Oral Sensitivity Disorder: A Systematic Bibliometric Analysis Of The Structures Of Knowledge. Journal Of Oral Pathology & Medicine, 49(6), 565–579. Https://Doi.Org/10.1111/Jop.13077

[15]. Garza-Fabre, M., Handl, J., & Knowles, J. (2018). An Improved And More Scalable Evolutionary Approach To Multiobjective Clustering. IEEE Transactions On Evolutionary Computation, 22(4), 515–535. Https://Doi.Org/10.1109/TEVC.2017.2726341

[16]. Granell, R., Axon, C. J., & Wallom, D. C. H. (2015). Impacts Of Raw Data Temporal Resolution Using Selected Clustering Methods On Residential Electricity Load Profiles. IEEE Transactions On Power Systems, 30(6), 3217–3224. Https://Doi.Org/10.1109/TPWRS.2014.2377213

[17]. Haben, S., Singleton, C., & Grindrod, P. (2016). Analysis And Clustering Of Residential Customers Energy Behavioral Demand Using Smart Meter Data. IEEE Transactions On Smart Grid, 7(1), 136–144. Https://Doi.Org/10.1109/TSG.2015.2409786

[18]. Heradio, R., Perez-Morago, H., Fernandez-Amoros, D., Javier Cabrerizo, F., & Herrera-Viedma, E. (2016). A Bibliometric Analysis Of 20 Years Of Research On Software Product Lines. Information And Software Technology, 72, 1–15. Https://Doi.Org/10.1016/J.Infsof.2015.11.004

[19]. Hsu, D. (2015). Comparison Of Integrated Clustering Methods For Accurate And Stable Prediction Of Building Energy Consumption Data. Applied Energy, 160, 153–163. Https://Doi.Org/10.1016/J.Apenergy.2015.08.126

[20]. Jasiński, M., Sikorski, T., & Borkowski, K. (2019). Clustering As A Tool To Support The Assessment Of Power Quality In Electrical Power Networks With Distributed Generation In The Mining Industry. Electric Power Systems Research, 166, 52–60. Https://Doi.Org/10.1016/J.Epsr.2018.09.020

[21]. Jiang, Y., & Seidmann, A. (2014). Capacity Planning And Performance Contracting For Service Facilities. Decision Support Systems, 58(1), 31–42. Https://Doi.Org/10.1016/J.Dss.2013.01.010

[22]. Jin, F., Zhao, J., Sheng, C., & Wang, W. (2018). Causality Diagram-Based Scheduling Approach For Blast Furnace Gas System. IEEE/CAA Journal Of Automatica Sinica, 5(2), 587–594. Https://Doi.Org/10.1109/JAS.2017.7510715

[23]. Kassambara, A. (2017). Practical Guide To Cluster Analysis In R: Unsupervised Machine Learning (1st Ed.). STHDA. Https://Books.Google.Com.Br/Books?Hl=Pt-BR&Lr=&Id=Pleydwaaqbaj&Oi=Fnd&Pg=PP2&Dq=Practical+Guide+To+Cluster+Analysis+In+R&Ots=Xdezjcllur&Sig=O2vru6wqmjzsarasb2zysnj7rmc#V=Onepage&Q=Practical Guide To Cluster Analysis In R&F=False

[24]. Li, X., Yao, R., Liu, M., Costanzo, V., Yu, W., Wang, W., Short, A., & Li, B. (2018). Developing Urban Residential Reference Buildings Using Clustering Analysis Of Satellite Images. Energy And Buildings, 169, 417–429. Https://Doi.Org/10.1016/J.Enbuild.2018.03.064

[25]. Lima, S., & Carlos Filho, F. De A. (2019). Bibliometric Analysis Of Scientific Production On Sharing Economy. Revista De Gestão, 26(3), 237–255. Https://Doi.Org/10.1108/Rege-01-2019-0018

[26]. Lin, S. C., Chen, C. J., & Lee, T. J. (2020). A Multi-Label Classification With Hybrid Label-Based Meta-Learning Method In Internet Of Things. IEEE Access, 8, 42261–42269. Https://Doi.Org/10.1109/ACCESS.2020.2976851

[27]. Lorentz, H., Hilmola, O. P., Malmsten, J., & Srai, J. S. (2016). Cluster Analysis Application For Understanding SME Manufacturing Strategies. Expert Systems With Applications, 66, 176–188. Https://Doi.Org/10.1016/J.Eswa.2016.09.016

[28]. Lv, Z., Wang, L., Guan, Z., Wu, J., Du, X., Zhao, H., & Guizani, M. (2019). An Optimizing And Differentially Private Clustering Algorithm For Mixed Data In SDN-Based Smart Grid. IEEE Access, 7, 45773–45782. Https://Doi.Org/10.1109/ACCESS.2019.2909048

[29]. Machado Junior, C., Souza, M. T. S. De, Parisotto, I. R. D. S., & Palmisano, A. (2016). As Leis Da Bibliometria Em Diferentes Bases De Dados Científicos. Revista De Ciências Da Administração, 1(1), 111. Https://Doi.Org/10.5007/2175-8077.2016v18n44p111

[30]. Merigó, J. M., Gil-Lafuente, A. M., & Yager, R. R. (2015). An Overview Of Fuzzy Research With Bibliometric Indicators. Applied Soft Computing Journal, 27, 420–433. Https://Doi.Org/10.1016/J.Asoc.2014.10.035

[31]. Moher, D., Stewart, L., & Shekelle, P. (2015). All In The Family: Systematic Reviews, Rapid Reviews, Scoping Reviews, Realist Reviews, And More. Systematic Reviews, 4. Https://Doi.Org/10.1186/S13643-015-0163-7

[32]. Okoli, C. (2015). A Guide To Conducting A Standalone Systematic Literature Review. Communications Of The Association For Information Systems, 37(1), 879–910. Https://Doi.Org/10.17705/1cais.03743

[33]. Pagani, R. N., Kovaleski, J. L., & Resende, L. M. M. De. (2018). Avanços Na Composição Da Methodi Ordinatio Para Revisão Sistemática De Literatura. Ciência Da Informação, 46(2), 161–187. Http://Revista.Ibict.Br/Ciinf/Article/View/1886

[34]. Pan, C., & Tan, J. (2019). Day-Ahead Hourly Forecasting Of Solar Generation Based On Cluster Analysis And Ensemble Model. IEEE Access, 7, 112921–112930. Https://Doi.Org/10.1109/ACCESS.2019.2935273

[35]. Panapakidis, I. P., & Moschakis, M. N. (2019). Consumer Load Profile Determination With Entropy-Based K-Means Algorithm. International Journal Of Electrical, Electronic And Communication Sciences, 13(3), 144–149. Https://Doi.Org/10.5281/ZENODO.2643537

[36]. Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines For Conducting Systematic Mapping Studies In Software Engineering: An Update. Information And Software Technology, 64, 1–18. Https://Doi.Org/10.1016/J.Infsof.2015.03.007

[37]. Pham, M., Rajić, A., Greig, J., Sargeant, J., Papadopoulos, A., & Mcewen, S. (2014). A Scoping Review Of Scoping Reviews: Advancing The Approach And Enhancing The Consistency. Research Synthesis Methods, 5(4), 371–385. Https://Doi.Org/10.1002/Jrsm.1123

[38]. Piao, M., & Ryu, K. H. (2017). Local Characterization-Based Load Shape Factor Definition For Electricity Customer Classification. IEEJ Transactions On Electrical And Electronic Engineering, 12, S110–S116. Https://Doi.Org/10.1002/Tee.22424

[39]. Rajabi, A., Eskandari, M., Ghadi, M. J., Li, L., Zhang, J., & Siano, P. (2020). A Comparative Study Of Clustering Techniques For Electrical Load Pattern Segmentation. Renewable And Sustainable Energy Reviews, 120, 109628. Https://Doi.Org/10.1016/J.Rser.2019.109628

[40]. Ramos, S., Duarte, J. M., Duarte, F. J., & Vale, Z. (2015). A Data-Mining-Based Methodology To Support MV Electricity Customers' Characterization. Energy And Buildings, 91, 16–25. Https://Doi.Org/10.1016/J.Enbuild.2015.01.035

[41]. Rhodes, J. D., Cole, W. J., Upshaw, C. R., Edgar, T. F., & Webber, M. E. (2014). Clustering Analysis Of Residential Electricity Demand Profiles. Applied Energy, 135, 461–471. Https://Doi.Org/10.1016/J.Apenergy.2014.08.111

[42]. Rodríguez-Soler, R., Uribe-Toril, J., & Valenciano, J. D. P. (2020). Worldwide Trends In The Scientific Production On Rural Depopulation, A Bibliometric Analysis Using Bibliometrix R-Tool. Land Use Policy, 97, 104787. Https://Doi.Org/10.1016/J.Landusepol.2020.104787

[43]. Rons, N. (2018). Bibliometric Approximation Of A Scientific Specialty By Combining Key Sources, Title Words, Authors And References. Journal Of Informetrics, 12(1), 113–132. Https://Doi.Org/10.1016/J.Joi.2017.12.003

[44]. Sharma, D. D., & Singh, S. N. (2015). Aberration Detection In Electricity Consumption Using Clustering Technique. International Journal Of Energy Sector Management, 9(4), 451–470. Https://Doi.Org/10.1108/IJESM-11-2014-0001

[45]. Shukla, A. K., Muhuri, P. K., & Abraham, A. (2020). A Bibliometric Analysis And Cutting-Edge Overview On Fuzzy Techniques In Big Data. Engineering Applications Of Artificial Intelligence, 92, 103625. Https://Doi.Org/10.1016/J.Engappai.2020.103625

[46]. Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical Analysis Of Big Data Challenges And Analytical Methods. Journal Of Business Research, 70, 263–286. Https://Doi.Org/10.1016/J.Jbusres.2016.08.001

[47]. Snyder, H. (2019). Literature Review As A Research Methodology: An Overview And Guidelines. Journal Of Business Research, 104, 333–339. Https://Doi.Org/Https://Doi.Org/10.1016/J.Jbusres.2019.07.039

[48]. Sun, M., Konstantelos, I., & Strbac, G. (2017). C-Vine Copula Mixture Model For Clustering Of Residential Electrical Load Pattern Data. IEEE Transactions On Power Systems, 32(3), 2382–2393. Https://Doi.Org/10.1109/TPWRS.2016.2614366

[49]. Ueda, R. M., Souza, A. M., & Menezes, R. M. C. P. (2020). How Macroeconomic Variables Affect Admission And Dismissal In The Brazilian Electro-Electronic Sector: A VAR-Based Model And Cluster Analysis. Physica A: Statistical Mechanics And Its Applications, 557, 124872. Https://Doi.Org/10.1016/J.Physa.2020.124872

[50]. Wang, D., Tang, O., & Huo, J. (2013). A Heuristic For Rationing Inventory In Two Demand Classes With Backlog Costs And A Service Constraint. Computers And Operations Research, 40(12), 2826–2835. Https://Doi.Org/10.1016/J.Cor.2013.06.001

[51]. Wang, Y., Wu, Z., Li, Q., & Zhu, Y. (2017). A Model Of Telecommunication Network Performance Anomaly Detection Based On Service Features Clustering. IEEE Access, 5, 17589–17596. Https://Doi.Org/10.1109/ACCESS.2017.2685182
[52]. Yang, J., Zhao, J., Wen, F., & Dong, Z. (2019). A Model Of Customizing Electricity Retail Prices Based On Load Profile Clustering Analysis. IEEE Transactions On Smart Grid, 10(3), 3374–3386. Https://Doi.Org/10.1109/TSG.2018.2825335
[53]. Zhou, H., Xi, B., Zhang, Y., Li, J., & Zhang, F. (2019). A Graph Clustering Algorithm Using Attraction-Force Similarity For Community Detection. IEEE Access, 7, 13683–13692. Https://Doi.Org/10.1109/ACCESS.2018.2889312
[54]. Zhou, K., Fu, C., & Yang, S. (2016). Big Data Driven Smart Energy Management: From Big Data To Big Insights. In Renewable And Sustainable Energy Reviews (Vol. 56, Pp. 215–225). Elsevier Ltd. Https://Doi.Org/10.1016/J.Rser.2015.11.050