

Energy efficiency in buildings: analysis of scientific literature and identification of data analysis techniques from a bibliometric study

Talita Mariane Cristina¹ · Antonio Faria Neto¹  ·
Antonio Fernando Branco Costa¹

Received: 3 August 2017 / Published online: 14 December 2017
© Akadémiai Kiadó, Budapest, Hungary 2017

Abstract This study uses bibliometrics methods to analyze the specialized literature of energy efficiency in buildings, including the Scopus database during the period of time ranging from 1980 to 2016, to identify the most relevant publications, authors, researcher groups, the evolution of the theme over the years, journals, geographical areas and eventually data analysis techniques employed. The countries with the most contributions have been the USA, China and the UK, where the Lawrence Berkeley National Labor, Hong Kong Polytechnic University and City University of Hong Kong were the three institutions with the most publications in this area. The publications have been concentrated primarily in thirty-three journals. The three most important journals are *Energy and Buildings*, *Applied Energy*, and *Energy* and are categorized primarily in engineering, energy and environmental sciences. The key terms may be divided into seven clusters: Buildings and Energy Uses; Building Energy Conservation; Energy Consumption; Energy Consumption Forecasting and Computational Intelligence; Energy Efficiency and Climate Effects; Building Energy Efficiency and Multivariate Statistics; and Building Energy Analysis and Stochastic Processes. The Data Analysis Techniques contained seven groups: Regression Analysis, Descriptive Statistics, Multivariate Analysis, Computational Intelligence, Stochastic Processes, Inferential Statistics and Design of Experiments. The data analysis techniques identified in this article raise the possibility of reformulation and adequacy of the curricula of the undergraduate and graduate courses in the area of energy and smart buildings. The results of this research have shown a general perspective regarding the energy efficiency in buildings, which can be useful in showing relevant themes for further research.

✉ Antonio Faria Neto
antfarianeto@gmail.com

Talita Mariane Cristina
talitamaryane@hotmail.com

Antonio Fernando Branco Costa
fbranco@feg.unesp.br

¹ São Paulo State University (UNESP), School of Engineering, Guaratinguetá, Brazil

Keywords Bibliometrics methods · Energy efficiency · Buildings · Data analysis techniques

Introduction

Energy is considered a key factor for development (Sepúlveda 2016) and an important element contributing to the achievement of a world economy (Zhenxing and Jing 2007). According to the International Energy Agency, the development of effective actions becomes important to solve problems related to climate change in the energy sector. The CO₂ emissions in this sector correspond to two-thirds of the total emissions; these levels have increased over the last years (IEA 2015). Therefore, increase of energy efficiency can be considered one of the primary strategies to reduce energy consumption and CO₂ emissions (Nagy et al. 2015). The UNEP (United Nation Environment Programme) (2016) shows that commercial and residential buildings consume approximately 60% of the global energy. According to the International Energy Agency, these buildings represent 10% of total CO₂ emissions (Soares et al. 2017). To reduce this index, it is important to modernize buildings by applying energy efficiency measures. This approach is a logical way to increase the life cycle of buildings, resulting in improved living conditions, reduced energy bill of the occupants (Dascalaki et al. 2016) and reduction of environmental impacts caused by building construction, which are the primary objectives for world energy policy (Soares et al. 2017).

Since the fuel crisis, which occurred in the 1970s, energy efficiency has been a significant factor to reduce the cost of energy and ensure sustainability in the world (Kazanasmas et al. 2014). Attention has been given to the development of technologies capable of improving the energy performance of buildings; however, only a few of the technologies have been applied (Mardoockhy et al. 2014).

According to Zhang et al. (2015) the improvement of energy efficiency in buildings is still one of the easiest, immediate and economic ways to reduce a country's energy consumption. Those advantages are why energy efficiency has received considerable attention (Zorita et al. 2016).

Several countries have created legislation that imposes the construction of buildings with a positive energy balance. For example, in Europe the buildings consume more than 40% of the total electricity production. Because of that the EPDB (The Directive on Energy Performance in Buildings) was created. It is a legislative instrument that affects energy use and efficiency in this sector of the EU and it is expected that in 2020 new buildings should consume “nearly zero” from the power grid; a large part of the demand will be generated locally by renewable sources. The new buildings are known as “very low energy building” or “nearly zero energy building” by its characteristics of very high energy performance (ECEEE 2010).

Many articles have been written relating to several factors that affect the use of energy and energy efficiency in buildings. Therefore, the most diverse tools for data analysis have been employed (Engvall et al. 2014).

This research aims to identify the most relevant publications and authors and the importance of the subject over the years, as well as the most frequent data analysis procedures employed in this area.

This research is justified to the extent that it contributes to the formulation and adequacy of curricula of undergraduate and postgraduate courses in the area of energy and smart buildings regarding the data analysis disciplines to be offered.

A bibliometrics analysis of articles extracted from the SCOPUS database was performed, which involved evaluating several types of information, such as evolution of publications, research areas, number of publications divided by region and country, and most cited key terms. Based on this, it was possible to identify the most relevant publications in the area as well as the most important authors and research centers. Thus, it was possible to evaluate the relevance of the subject over the last years. It was also possible to see the timing and the sequence of the data analysis procedures introduced in these research areas. The main journal that published and the most relevant papers were identified as well.

The article is organized in three sections: the methods, the results and findings, and the conclusions. In the method is presented, in detail, the methodological approach adopted. In the results and findings is presented the data extracted from the database, as well as the discussion of the main findings. In addition, the conclusions will present the most relevant comments on the most important points of the research.

Method for bibliometrics analysis

The method employed for this research is illustrated in Fig. 1.

Research subject

This research started off by defining the key terms to be used. It is worth recalling that the main interest of this research relied on identifying as many papers as possible dealing with data analysis techniques among papers concerned with energy efficiency in buildings which, in turn, have been extracted from a pull of all papers related to energy efficiency. Figure 2 represents the research strategy.

The SCOPUS was the chosen database. Only papers from journals from 1980 to 2016 were considered. To get the desired effect, suggested by Fig. 2, the research initially retrieved papers containing research terms related to “Energy Efficiency”. The query of these terms in the title resulted in 96,400 documents, is available in “Appendix 1”.

The next research also included the term “Building” in the title, and resulted in 3678 publications, according to the query in the “Appendix 2”.

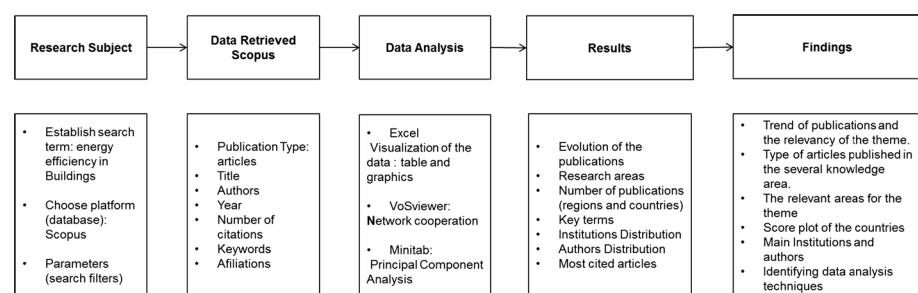


Fig. 1 Research method. (*Source* Li et al. 2015a, b)

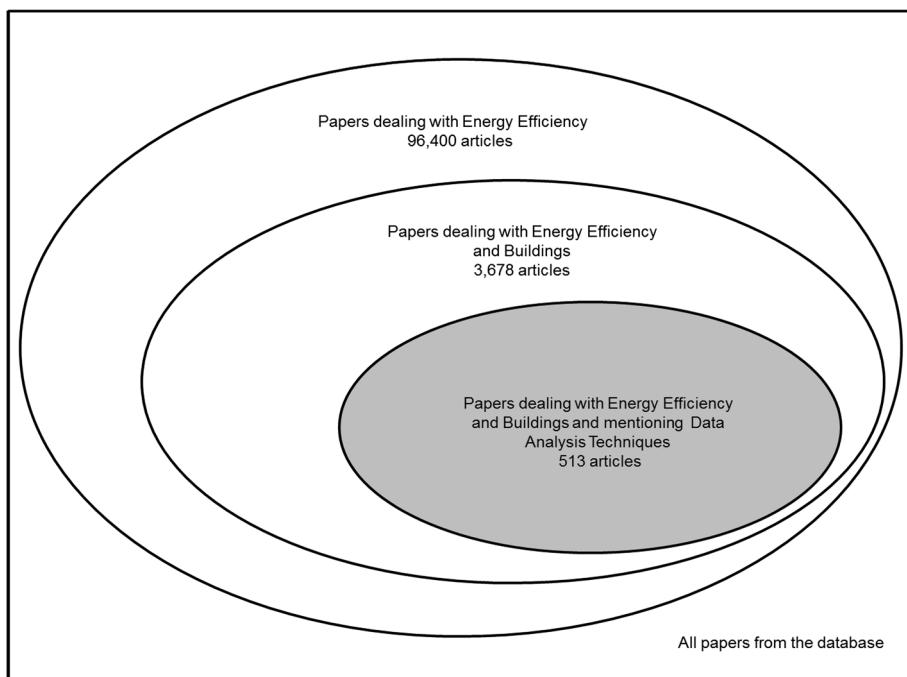


Fig. 2 Defining the research key terms

The final query applied, containing all terms of the previous researches and the terms related to “Data Analysis Techniques” in the title, abstract, and keywords, resulting in 513 articles, is available in “Appendix 3”. All the bibliometric analysis was based on these last results.

Data retrieved

This step retrieved information from the 513 articles as a title, keywords, year of publication, authors, affiliations and number of citations.

Data analysis

The previous step retrieved a large amount of information that was summarized, for further analysis, into tables and several types of graphs by means of such software packages as Microsoft Excel, VoSviewer and Minitab.

Results

Several of the results obtained were the evolution of the publications; the most common research areas; number of publications per area; most cited key terms; institution distribution; author distribution and most cited articles.

Findings

Trend of publications and the relevancy of the theme; type of articles published in the several knowledge areas; the relevant areas for the theme; the most published journal for the knowledge area; the evolution of publications per region; number of publication by country; number of citations per publication; percentage of non-cited papers; number of citations of the most cited paper; score plot of the countries, analysis of the most used key terms; main institutions and authors.

Many of the 513 articles, firstly retrieved, only shallowly mentioned some data analysis technique. The total number of techniques really applied in this amount of articles accounts for 296. In some cases more than one technique have been applied simultaneously.

Results and findings

Analysis of the evolution of publications

The number of publications related with the subject over the period under analysis has been considered a measure of scientific productivity and interest in the subject. Figure 3 shows a clear upward trend from 1980 (2 articles) to 2016 (80 articles). About fifty (53%) percent of the articles over this period were published in the last four years (2013–2016).

According to the Price's law (1963) the growth of scientific production of a field follows an exponential function and Fig. 4 shows that the growth pattern of the subject under research in this paper fits an exponential function. This statement is supported by the statistic R^2 , which is approximately 0.96, what means that such a model can explain 96% of the total variability of the data.

According to Dabi et al. (2016) while “the main hypothesis of Price's law is that the development of science follows an exponential growth. The growth of a scientific domain goes through four phases”. Taking a closer look at Fig. 4 it is possible to identify the first three of these phases as can be seen in Fig. 5.

The first phase, which extends itself from 1980 to roughly 1992, is the precursors' phase, according to Dabi et al. (2016) “during this phase only a small number of researches

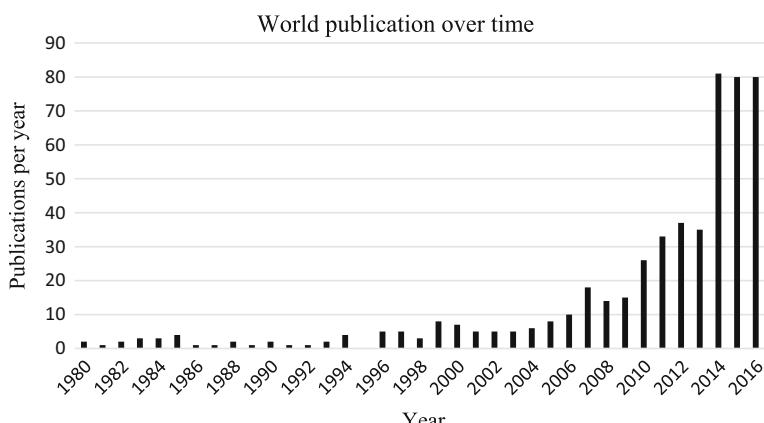


Fig. 3 Evolution of the publications over the period 1980–2016 according to the Scopus database

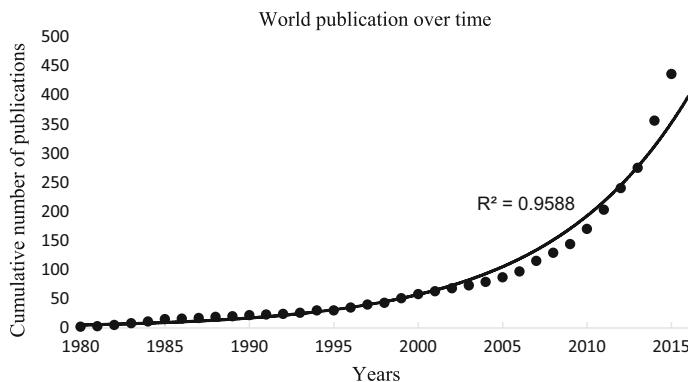


Fig. 4 Growth pattern of the publications over the period 1980–2016 according to the Scopus database

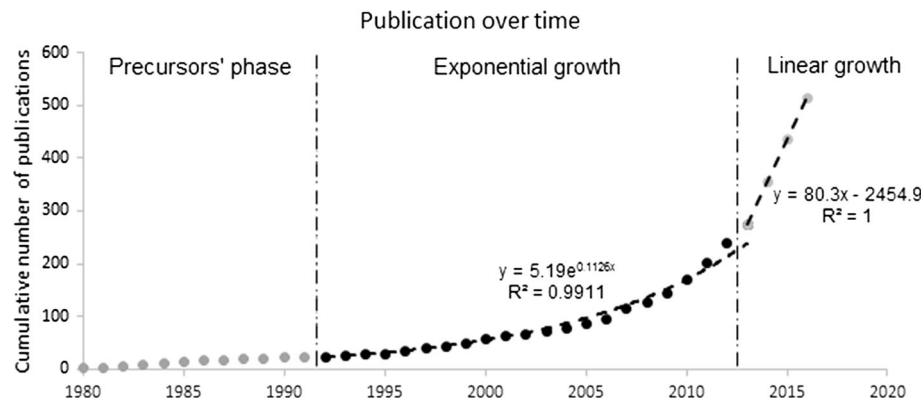


Fig. 5 Phases of growth of a scientific domain according to Price's law (Dabi et al. 2016)

begins publishing". In this study, this phase accounts for about 4.4% of the whole publication body. The second phase (1992–2013) is the proper exponential growth. "During this phase the expansion of the field attracts many researchers as many aspects of the subject still have to be explored" (Dabi et al. 2016). It can be seen that the data fit well an exponential function, since the statistic R^2 is very close to 1.00. It is worth mentioning that the initial and final years of this phase were set in order to maximize the statistic R^2 for this and the subsequent phase. During this phase, the number of publications doubled approximately every each 7 years. The third phase (2013–2016), the body of knowledge is consolidated and the growth of scientific production becomes linear (Dabi et al. 2016). Observing Fig. 5 it can be seen that the data retrieved fit perfectly a linear function. It is important to note that the next phase, according to Dabi et al. (2016), "corresponds to the collapse of the domain and is marked by a decrease in the number of the publications. The aspect of the curve transforms from exponential to logistic", reaching a ceiling value after passing through an inflection point.

Based on the analysis exposed above, it can be concluded that number of publications in this field, from 1980 to 2016, follows the Price's law and an inflection point has not been reached yet.

The evolution of the number of articles related to energy efficiency in Buildings over the analyzed period of time can be explained by the fact that residential, commercial and public buildings account for approximately 60% of global energy (UNEP—United Nations Environment Programme 2016) and for 10% of global greenhouse gas emissions, according to the International Energy Agency (Soares et al. 2017). Therefore, it seems that the development of energy efficient buildings has been a great deal of research on sustainable development what has been attracted researchers' attention worldwide. In Europe, for example, where the energy consumption in this sector is greater than 40%, several incentive programs have been created to promote the rational use of energy in buildings after 2010. For instance, the Energy Performance of Buildings Directive, which aims at increasing the energetic performance of European enterprises, estimating that in 2020 most of the new buildings will supply their electric energy demand without being connected to the power grid. The electric power will be locally produced by renewable sources. This new buildings will be known as very low energy building or nearly zero energy buildings (ECEEE 2010). This set of actions has motivated the increasing number of publications in this area following an exponential function as can be seen in Fig. 6.

Compared to Europe, publication in this field in North America had begun 10 years before due to the great attention paid by the United States to the theme. The expectations of an energetic crisis, which actually arose in California (2000–2001), had incentivized many programs aiming to promote the energetic efficiency in buildings that were followed by several publications whose growth followed the same pattern of world growth, as can be seen in Fig. 7.

The number of publications in Asia have been followed the same pattern in Europe as can be seen in Fig. 8.

Research areas

This section discusses the relevance of the theme for several research areas. Figure 9 presents the number of publication per area for the eighteen most relevant research areas.

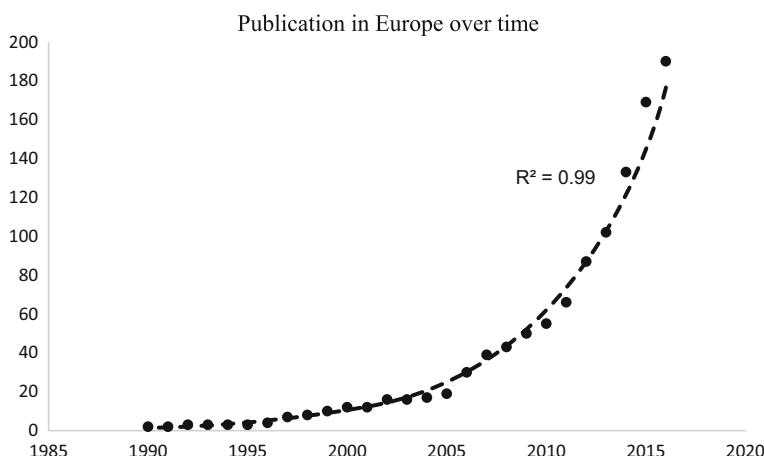


Fig. 6 Growth pattern of the publications over the period 1990–2016 in Europe, according to the Scopus database

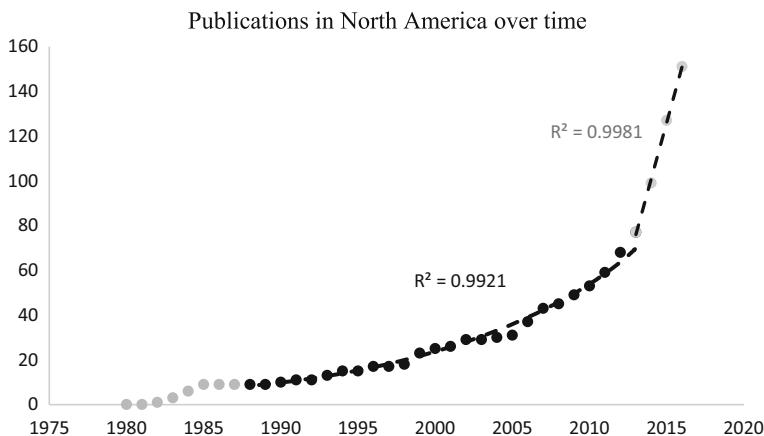


Fig. 7 Growth pattern of the publications over the period 1990–2016 in North America, according to the Scopus database

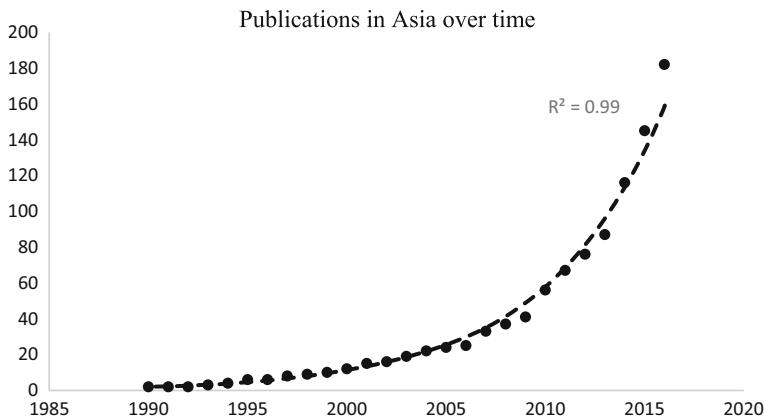


Fig. 8 Growth pattern of the publications over the period 1990–2016 in Asia, according to the Scopus database

The four most relevant areas for the theme are Engineering, Energy, Environmental Science and Computer Science. It can also be seen that the subject has gotten the attention of some peripheral areas, such as Social Science; Business, Management and Accounting; Economics, Econometrics and Finance; Arts and Humanities; and Medicine. One of the reasons for a such intersection is the fact that many publications were classified into more than one area. But scientific interests can also explain the overlap between the areas.

Social Sciences area contains publications dealing with thermal comfort and occupants' behavior. According to Nghana and Tariky (2016) 62% of the building energy use are towards maintaining thermal comfort and a frequent worry is reducing the energy consumption keeping the occupants' comfort (Cetin et al. 2016; Koumoutsos et al. 2015). For achieving this goal, it is important to understand the occupants' behavior. This explain the interest of some social sciences' studies in this area.

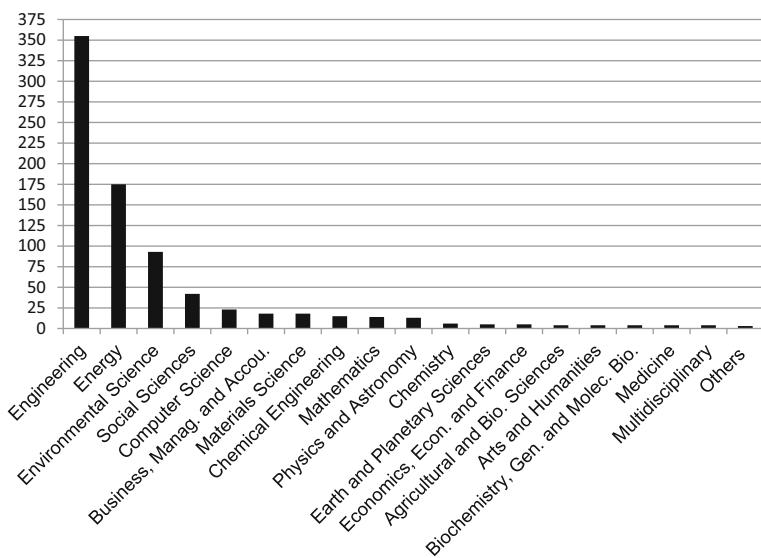


Fig. 9 Research areas where articles were published according to the Scopus database

Business, Management and Accounting area groups articles that address the subject from the stand point of relevant planning and management tools for energy-efficient buildings. These tools simulate energy consumption in buildings, facilitating comparative analyses of the energy consumed in relation to the efficient operation of the buildings and the behavioral changes of the occupants (Kim and Yu 2016; Azizi et al. 2014; Kontokosta 2015).

Economics, Econometrics and Finance area presents papers that are correlated with Market Values of buildings. The increase of energy efficiency requirements for new construction, makes the less efficient buildings subject to greater economic depreciation, with declining Market Values (Surmann et al. 2015; Cajias and Piazolo 2013). This area are interested in studying the relationship between the energetic aspects and demand for energy efficiency buildings.

Arts and Humanities area contains papers that approach the theme related from the historical heritage stand point, the assessment of historical buildings may help to recognize the actions to reduce energy needs, that can reflect in a set of general interventions in the management process of historical architecture. This approach is new in the field of conservation, restoration, and reuse of Cultural Heritage (Magrini and Franco 2016).

Medicine area presents publications dealing with health problems recurrent of CO₂ emissions. There is a growing concern regarding the energy consumption in buildings in many countries, as a part of “Low Carbon Cities” programmes striving for reducing greenhouse-gas emissions, these programmes can reduce energy consumption and also improve health problems of occupants (Norbäck et al. 2014).

Scientific journals

The total amount of papers (513) are produced by 155 journals as shown in Appendix 4. It is worth analyzing the distribution of these articles between the different journals. Bradford's law of scattering (1948 *apud* Palomo et al. 2017) has stated that: “if scientific

journals are arranged in order to decrease the productivity of articles, they may be divided into a nucleus of periodicals, more particularly devoted to the subject, and several groups, or zones, (normally 3 as defined by Bradford), containing the same number of articles as the nucleus. Where the number of periodicals in the nucleus and succeeding zones will be $[1:n:n^2]$. This means that a small set of journals (the nucleus zone) accounts for the majority of the articles, whereas other sets of journals have to be larger in order to account for the same amount of published articles. This suggests that there is an inverse relationship between the number of articles and the amount of journals where they were published (Palomo et al. 2017).

The journals within each zone are identified in the “Appendix 4”. The theoretical number of articles per zone, according to Bradford’s law, should be $513/3 = 171$, or 33% of the total. Table 1 shows that the three zones contain roughly the same number of articles. The number of journals in the three zones are in this proportion $[1:32:122]$, which is roughly equal to $[1:32:128]$ that is close to $[2^0:2^5:2^7]$. Therefore, this distribution does not fit into the original Bradford’s distribution $[1:n:n^2]$, since it is not possible to find out the Bradford multiplier n .

Table 1 presents the distribution of the journals in three zones, as defined originally by the Bradford’s law.

Even though the distribution presented in Table 1 can be changed by moving one or more articles from one zone to another, keeping roughly the same proportion of articles, among zones, it still will not fit into the original Bradford’s distribution.

Although the observed data does not follow the Bradford’s law, it is kept that, “few journals publish a relatively high percent of the articles in the field, and there are many journals that publish only few articles each” (Diodato 2012) as can be seen in Fig. 10, that presents one of the several versions of the Bradford curve, that stands out that 10% of the journals published 60% of the papers and about 60% of the journals published only one paper.

The most productive journal in this area is Energy and Buildings, which published 153 articles, and is the nucleus. The second zone grouped the 32 journals, which published at least three articles each. Within this zone, the most productive is Applied Energy, with 36 papers and the less productive, with 4 papers each, are Architectural Science Review, Building; Services Engineering Research and Technology; Building Simulation; HVAC and R Research; Journal of Building Physics; Renewable and Sustainable Energy Reviews; Tumu Jianzhu Yu Huanjing Gongcheng Journal of Civil Architectural and Environmental Engineering; and Applied Thermal Engineering. The third zone clusters 122 journals, 12 of them with 3 articles each, 24 with 2 articles and 86 with only one paper.

Table 2 shows the journals within the first and second zones.

The Journal with the greatest number of publications is Energy and Buildings (153) followed by Applied Energy (36), Energy (22), Building and Environment (20), and Energy Conservation and Management (16).

Table 1 Distribution of the journals

Zone	# Journals	% Journals	# Articles	% Articles
1	1	0.7	153	29.8
2	32	20.6	177	34.6
3	122	78.7	182	35.6
Total	155	100	513	100

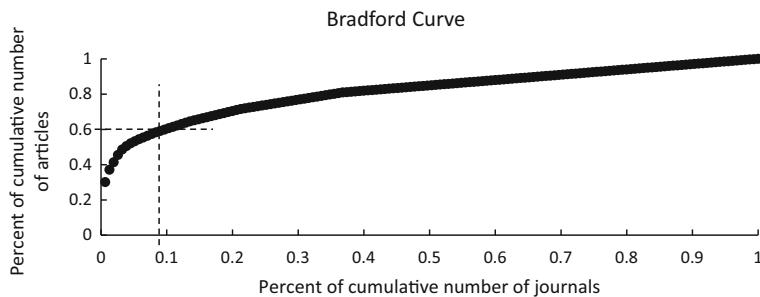


Fig. 10 Bradford curve

According to the number of publications, the Energy and Buildings is the most important journal. Most researchers have chosen it to publish their articles related to data analysis applied to energy efficiency in buildings because this it is concerned with energy use in buildings and it covers a broad range of topics, twenty-six (Appendix 5) ranging from energy conservation to smart buildings passing by thermal comfort, energy management, modeling, energy sustainability, etc. Most of the papers published are linked to laboratory or field measurements, comparisons of results, and replication studies (Elsevier 2017a, b, c, d, e).

The second most important journal is the Applied Energy, which deals with problems of modeling and forecasting, energy conservation strategies, and the environmental, social and economic impacts of energy policies and usage. It covers about 8 topics (Appendix 5) (Elsevier 2017a, b, c, d, e).

The third most important journal is the Energy. It covers twelve topics (Appendix 4) on research in mechanical engineering and thermal sciences, passing by energy analysis, energy modelling and prediction, integrated energy systems, energy planning and energy management (Elsevier 2017a, b, c, d, e).

The fourth most important journal is the Building and Environment, which covers 4 topics (Appendix 5). It publishes papers related to building science and human interaction with built environment (Elsevier 2017a, b, c, d, e).

The fifth most important journal is the Energy Conservation and Management, which publishes papers dealing with modeling, experimental, analysis and optimization issues covering 8 topics (Appendix 5) concerned with interdisciplinary energy subjects related to advanced technologies (Elsevier 2017a, b, c, d, e).

The main metrics for these journals are presented in Table 3.

Energy and Buildings and Applied Energy journals are related to the research area Engineering; Energy and Energy Conservation and Management are linked to the research area Energy; and Building and Environment is concerned with subjects related to the research area Social Science.

Table 4 presents the journals of the first and second zones distributed among five research areas.

It can be seen that Engineering (16) and Energy (10) areas stands out with greater number of journals, followed by Environmental Science (4), Social Sciences (2) and Material Sience (1).

Table 2 Journal within the first and second zones

Journals	Number of publications
Energy and Buildings	153
Applied Energy	36
Energy	22
Building and Environment	21
Energy Conversion and Management	16
Energy Policy	10
Journal of Solar Energy Engineering Transactions of the ASME	8
Journal of the Association of Energy Engineering	6
Jurnal Teknologi	6
Energy Efficiency	5
Journal of Building Performance Simulation	5
Renewable Energy	5
Solar Energy	5
Architectural Science Review	4
Building Services Engineering Research and Technology	4
Building Simulation	4
HVAC and R Research	4
Journal of Building Physics	4
Renewable and Sustainable Energy Reviews	4
Journal of Civil Architectural and Environmental Engineering	4
Applied Thermal Engineering	3
Automation in Construction	3
Bauphysik	3
Frontiers in Energy	3
International Journal of Energy Research	3
International Journal of Ventilation	3
Journal of Environmental Engineering	3
Journal of Harbin Institute of Technology New Series	3
Science and Technology for the Built Environment	3
Strategic Planning for Energy and the Environment	3
Sustainability Switzerland	3
Taiyangneng Xuebao Acta Energiae Solaris Sinica	3
Wit Transactions on Ecology and the Environment	3

Number of publications by country

The descriptive bibliometrics results of the countries that have published more than five articles are seen in Table 5.

Figure 11 shows the first five countries in number of publications.

It can be seen that United States (134), China (82) and the United Kingdom (41) stand out. Germany (20) and Italy (20) share the fifth place of the most productive countries.

Table 3 Journal metrics

Journals	Cite score	Impact factor	5-Year impact factor	SJR	Average acceptance rate ^a
Energy and Buildings	4.64	4.067	4.599	2.093	37.2%
Applied Energy	7.78	7.182	7.500	3.058	–
Energy	5.17	4.520	5.182	1.999	–
Building and Environment	4.51	4.053	4.464	2.015	19.9%
Energy Conserv. and Manag.	6.04	5.589	5.472	2.287	–

^aOver the last 5 years

Table 4 Distribution of journals within the first and second zones according to the research area

Research areas	Journals
Engineering	Energy and Buildings (153); Applied Energy (36); Journal Teknologi (6); Journal of Building Performance Simulation (5); Journal of Civil Architectural and Environmental Engineering (4); Journal of Building Physics (4); HVAC and R Research (4); Building Simulation (4); Building Services Engineering Research And Technology (4); Architectural Science Review (4); Science and Technology for the Built Environment (3); Journal of Harbin Institute of Technology New Series (3); Journal of Environmental Engineering (3); International Journal of Energy Research (3); Automation in Construction (3); Applied Thermal Engineering (3)
Energy	Energy (22); Energy Conservation and Management (16); Journal of Solar Energy Engine (7); Journal of the Association of Energy Engineers (6); Renewable Energy (5); Energy Efficiency (5); Renewable And Sustainable Energy Reviews (4); Taiyangneng Xuebao Acta Energiae Solaris Sinica (3); Frontiers in Energy (3); Bauphysik (3)
Environmental Science	Energy Policy (10), Wit Transactions on Ecology and the Environment (3); Strategic Planning for Energy and the Environment (3); International Journal of Ventilation (3)
Social Sciences	Building and Environment (20); Sustainability Switzerland (3)
Material Science	Solar Energy (5)

The number of publications is in brackets

The profile of citations is shown in Fig. 12.

In this figure, as in the previous chart, the first three positions are occupied by the United States (3103), China (1102) and the United Kingdom (974). The only difference between this chart and that shown in Fig. 11 is the fifth position. Indeed this suggests a correlation between the number of publication and the number of citations, which can be explained by the fact that the authors' collaboration net is mainly composed of authors from the same country.

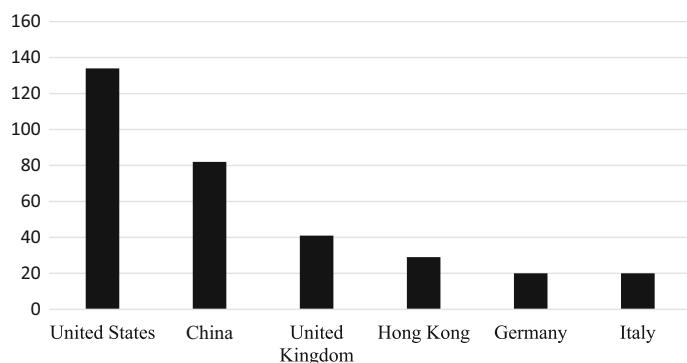
The number of citations per publication (CPP) is presented in Fig. 13.

The chart presents several countries that were not presented in the former charts. The first two positions are filled by Greece (49.84) and Switzerland (49.34), which are the

Table 5 Countries with number of publications ≥ 5 (1980–2016)

Country	Publications	Citations	CPP	PNC	MAX CITES
United States	134	3103	23.16	9.71	642
China	82	1102	13.44	26.61	107
United Kingdom	41	974	23.76	2.44	205
Hong Kong	29	882	30.42	3.45	133
Germany	20	448	22.40	20.00	135
Italy	20	352	17.60	5.00	72
Japan	19	417	21.95	10.53	160
Taiwan	18	342	19.00	11.12	66
Canada	17	580	34.12	5.89	160
South Korea	16	170	10.63	12.50	45
Sweden	16	294	18.38	0	40
France	15	275	18.34	6.67	58
Australia	12	145	12.09	16.67	58
Malaysia	12	122	10.17	33.34	86
Singapore	11	368	33.46	9.10	161
Brazil	10	280	28.00	20.00	185
Finland	9	82	9.12	11.12	29
Saudi Arabia	9	127	14.20	11.12	50
Spain	9	139	15.45	0	34
Switzerland	9	444	49.34	0	135
India	7	48	6.86	28.58	22
Greece	6	299	49.84	16.67	135
Belgium	5	54	10.80	20.00	31
Egypt	5	14	2.80	20.00	7
Norway	5	50	10.00	0	42
Portugal	5	101	20.20	0	40

CPP citations per publication, PNC percentage of non-cited papers, MAX CITES the number of citations of the most cited article in the respective country

**Fig. 11** Number of publications per country

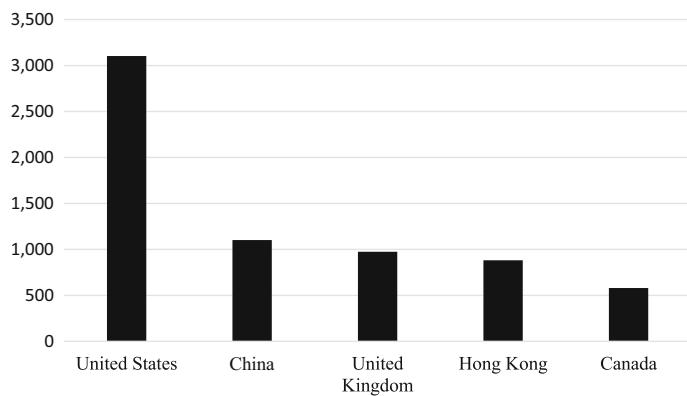


Fig. 12 Number of citations per country

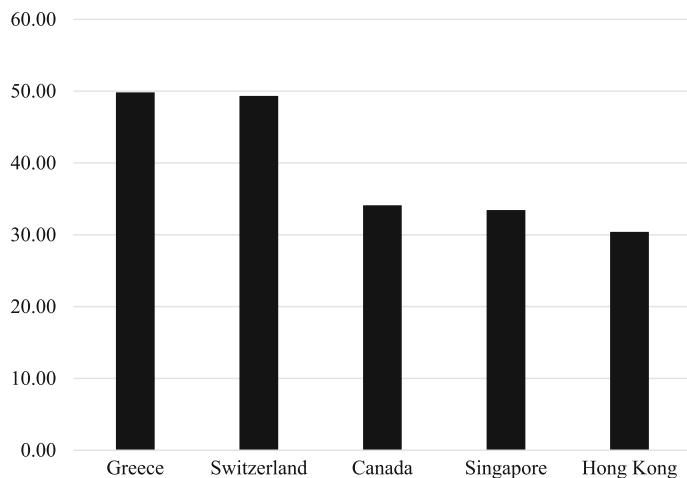


Fig. 13 Number of citations per publication

fourteenth and sixteenth, respectively, in number of publications. Because the number of citations per paper is greater than the number of publications, it is possible to conclude that these publications have been cited by authors abroad. In this case we can conclude that these countries have produced good publications in the area.

The percentage of publications non-cited (PNC) are illustrated in Fig. 14.

The fourth position is occupied by Germany (22.40%), Brazil (20.00%), Belgium (20.00%) and Egypt (20.00%). Approximately 20% of the total number of publications of these countries has not been cited. Another point worth mentioning is the fact that although China has a great number of publications, more than 25% of them have not been cited. Malaysia (33.34%) and India (28.58%) also have a great percentage of publications not cited.

Figure 15 shows the number of citations of the most-cited paper for several countries.

The graph highlights the United States (642) position. The United Kingdom (205) is ranked in the second position. Brazil (185) appears in the third place. The fourth position is

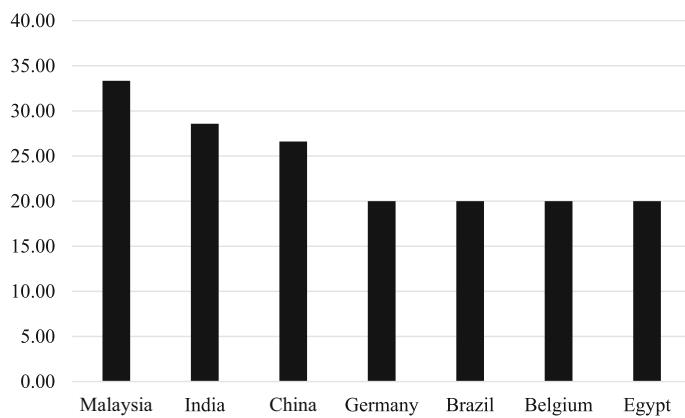


Fig. 14 Percentage of publications non-cited

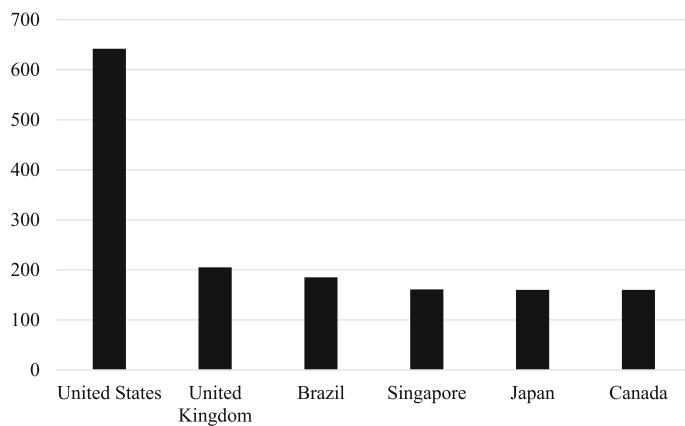


Fig. 15 Number of citations of the most-cited article

occupied by Singapore (161), Japan (160) and Canada (160). The most cited article in The United States is distant from the most cited article in any other country.

Table 6 shows the most-cited articles used to produce Fig. 15.

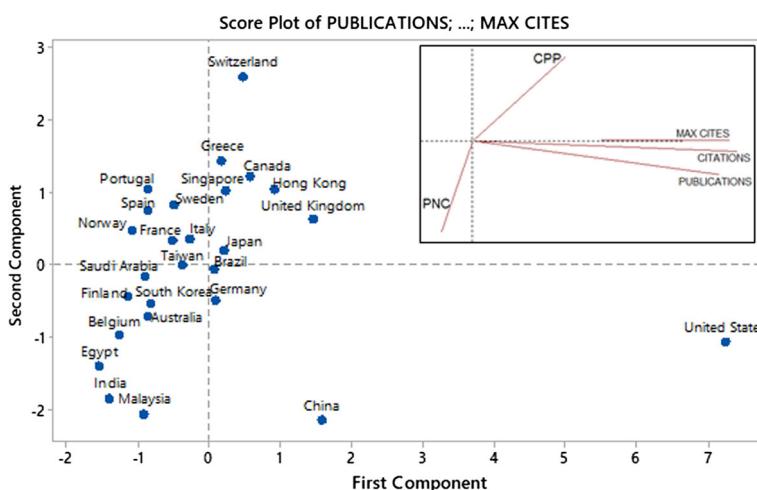
It is remarkable that the most cited articles were published in the same journal, Energy and Building. All of them were published from 2001 to 2016. It is worth mentioning that five of them are concerned with simulation. The one produced by Japan made use of a clustering algorithm to carry out the research.

The data from Table 5 have been submitted to a Principal Components Analysis (PCA), and Fig. 16 presents the first and the second components plotted against each other. Each data point represents a single country.

The positioning of each country on the graph depicted in Fig. 16 depends on a linear combination of all variables presented in Table 5. On the right upper of the graph can be seen the load plot for this analysis, which is useful in explaining the positioning of each country on the graph. The greater publications, citations and max cites, further to the right side of the graph will be the country. The greater the CPP is, the higher the country on the graph will be. The greater the PNC is, the lower the country on the graph will be.

Table 6 The most-cited articles

Country	Paper	Authors	Journal	Year
United States	EnergyPlus: Creating a new-generation building energy simulation program	Crawley, D.B.; Lawrie, L.K.; Winkelmann, F.C.; Buhl, W.F.; Huang, Y.J.; Pedersen, C.O.; Strand, R.K.; Liesen, R.J.; Fisher, D.E.; Witte, M.J.; Glazer, J.	Energy and Buildings	2001
United Kingdom	Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings	Rijal, H.B.; Tuohy, P.; Humphreys, M.A.; Nicol, J.F.; Samuel, A.; Clarke, J.	Energy and Buildings	2007
Brazil	Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption	Neto, A.H; Fiorelli, F.A.S.	Energy and Buildings	2008
Singapore	The effects of rooftop garden on energy consumption of a commercial building in Singapore	Wong, N.H.; Cheong, D.K.W.; Yan, H.; Soh, J.; Ong, C.L.; Sia, A.	Energy and Buildings	2003
Japan	A systematic procedure to study the influence of occupant behavior on building energy consumption	Yu, Z.; Fung, B.C.M.; Haghishat, F.; Yoshino, H.; Morofsky, E.	Energy and Buildings	2011
Canada	Effect of cool roofs on commercial buildings energy use in cold climates	Hosseini, M.; Akbari, H.	Energy and Buildings	2016

**Fig. 16** Principal component analysis

The United States stands out because of the number of publications, citations and the number of citations of the most cited article. Switzerland stands out because of the number of citations per publication. Although China has a great number of publications and citations it also has a high percentage of non-cited publications. India and Malaysia are on the left bottom of the graph because the percentage of non-cited papers is significant.

Analysis of the key terms

The pattern of key terms that appear in publications of a given field of knowledge defines the underlying themes concerned with such a field. Therefore, when a set of publications is analyzed, “the ideas, concepts, and methods that constitute this field of knowledge are defined by clusters of key terms that reflect commonality within a field of scientific research” (Phillips et al. 2015)

According to Phillips et al. (2015) there are several strategies for visualizing textual data. This paper uses bibliometric maps for visualizing keywords associations, generated by the VOS viewer software. The interpretation of such maps are based on:

- The size of the visual representation of terms is related to the frequency they appear in the retrieved articles.
- The relative position of terms in the map reflects their relative association, it means that all possible pairs of terms that are commonly associated are positioned close to one another, and terms with a low degree of association with each other are positioned far from one another.
- The clustering of terms establishes related conceptual domains.

In order to visualize changes in the thematic focus of this research field, three bibliometric maps covering the period of analysis (1980-1990, 1980-2000, 1980-2016), will be

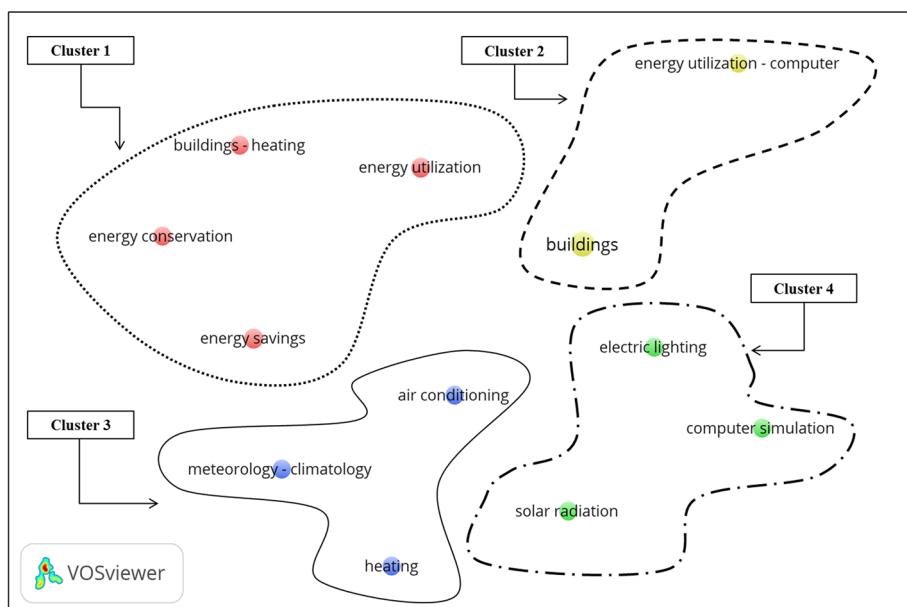


Fig. 17 Bibliometric map ranging from 1980 to 1990

employed. Thus, it will be possible to study the formation and evolution of the clusters, as a way to understand the development of this field.

Figure 17 shows the bibliometric map ranging from 1980 to 1990.

During this period, 22 articles were produced, totaling 81 keywords. In the interest of clarity, Fig. 17 presents only the 12 keywords with frequency greater than or equal to 2, forming 4 clusters.

The clustering process intend to form heterogeneous groups of homogeneous individuals. In this case, the clustering algorithms uses the frequency and strength of connection of the keywords to create the clusters. Since many keywords were cited only once, many clusters containing only one element would arise obscuring the analysis. Because of that, only the keywords that appeared 2 or more times were used.

The oil crisis in the middle of the 1970s turned the world's attention to the rational use of energy what reflected in terms like "energy conservation", "energy savings", and "energy utilization" in the publications at the beginning of 1980s, which were grouped in Cluster 1.

Thermal comfort systems and lighting were, and still are, important energy loads then it was expected that such a theme would be explored. Cluster 3 groups terms related to thermal comfort like "air conditioning", "heating", and "meteorology-climatology". Cluster 4 puts together terms concerned with lighting, like "electric lighting" and "solar radiation" this last term is in accordance with a trend at that time of improving the use of natural lighting in order to save energy.

With the advent of the personal computers in the 1980s, it was also expected that softwares of data analysis were introduced in this research area. In cluster 4, one can see the keyword "computer simulation", which is perfectly correlated to the other elements of this cluster as it could mean the use of a dedicated software to lighting design.

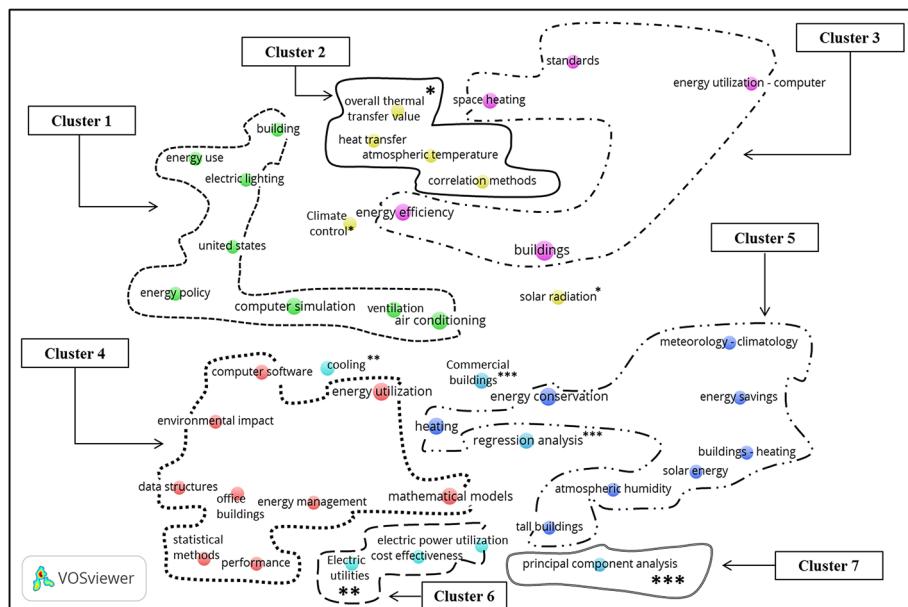


Fig. 18 Bibliometric map ranging from 1980 to 2000

Since this was the start-up of this research area, the number of clusters and the number of elements per clusters increased as the time passed, as can be seen in Fig. 18 that presents the bibliometric map for 1980 to 2000.

The bibliometric map presented in Fig. 18 shows 43 keywords that appeared two or more times from 58 articles, grouped into 7 clusters with, at least, 2 keywords each. This map shows the introduction of new terms like “data structures”, “mathematical models”, “statistical methods”, “regression analysis”, “correlation methods”(univariate data analysis techniques), and “principal component analysis”(multivariate data analysis technique), then it can be concluded that this decade was marked by the introduction of the univariate data analysis techniques in the publications in this area.

It also can be noted the presence of the first country cited in the articles, the “United States” of America (EUA), which is in accordance with the fact that they were the precursors in this research field as can be seen in Fig. 3. It is also remarkable the very first appearance, in this kind of study, of the term “environmental impact”.

This map shows that the cluster 4, in Fig. 17, aggregated more 6 words and the correlation of the term “solar radiation” to the others became weak, forcing it to group with 6 new terms to form a new cluster, the cluster 7 in Fig. 18. Clusters 1 and 3 from Fig. 17 joined to form the cluster 6 in Fig. 18. The cluster 2 in Fig. 17 increased with the inclusion of three new terms. The clusters 1, 3, and 5 are formed by keywords that did not appear in the preceding map.

From Fig. 18 it is possible to infer the emergence of a pattern in the cluster formation. It seems that some of them are specialized in a topic of the research area. For example, cluster 1 groups mainly terms related to data analysis techniques like “data structures”, “mathematical models”, “statistical methods”, and “computer software”. Cluster 4, the electric loads in a building like “energy use”, “electric lighting”, “ventilation”, and “air

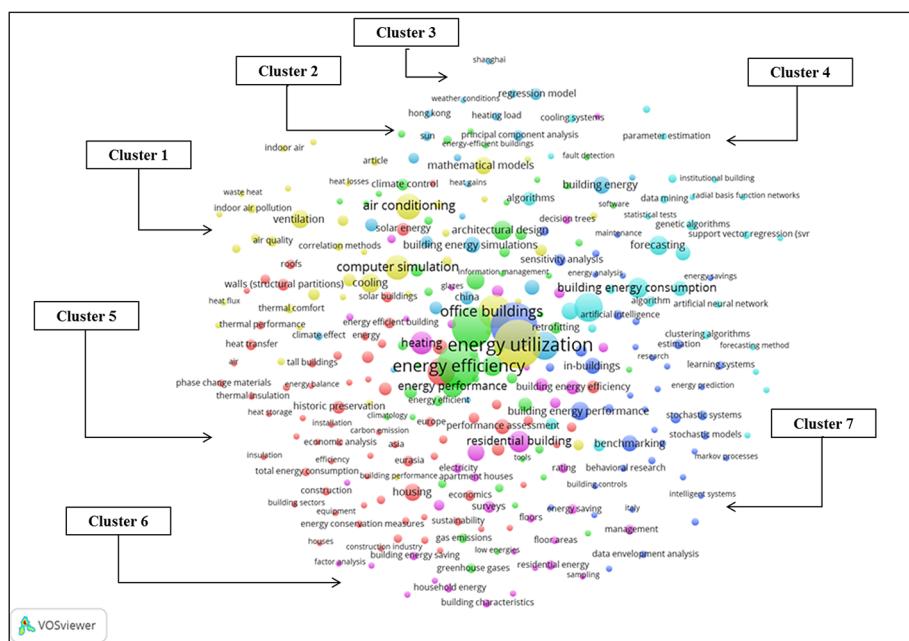


Fig. 19 Bibliometric map ranging from 1980 to 2016

conditioning”. In some of the other clusters, the word association pattern is not clear yet, but the more words are included in the clusters, the clearer the pattern becomes.

The final map covers the period ranging from 1980 to 2016 (Fig. 19). This map shows 300 keywords that appeared four or more times from 513 articles, grouped into 7 clusters.

This map shows the inclusion of 257 new key terms, roughly seven times more words than the presented in the preceding map. Thus, the cluster structure is different, many of the clusters presented in the previous map merged with each other, some disappeared, and others arisen.

The environmental questions has gained more importance after the year 2000, then many authors tried to call the readers’ attention by including among the keywords, terms like “gas emissions”, “global warming”, “climatic index”, “climate models”, and “climate changes”. For the same reason, terms related to renewable energy appeared many times “renewable energy resources”, “wind power”, “wind effect”, “solar power”, and “alternative energy”.

It also can be seen terms related to well-being like “comfort level”, “thermal comfort”, “indoor air quality”, “indoor air pollution”, and “indoor environment”, what means that many articles were produced focusing on the building occupants’ wellness.

Over this period, more attention was paid to the building material and building design to improve the building energy efficiency, thus terms like “building materials”, “recycling”, “phase change materials”, “concrete aggregates”, “structural design”, “architectural design”, and “computer aided design” arose among the keywords.

Although the univariate statistics has been widely applied in the articles, this period was marked by the introduction of data analysis techniques based on computational intelligence, thus many terms like “neural networks”, “artificial neural networks”, “genetic algorithms”, “adaptive algorithms”, “artificial intelligence”, and “back propagation” appeared in the articles.

Many articles were produced mentioning geographical region were the research took place “Far East”, “Asia”, “Middle East”, “Eurasia”, “Europe”, “Southern Europe”, “Northern Europe”, and “Scandinavia”; “Singapore”, “China”, “Hong Kong”, and “Italy”, showing that over this period this subject spread out all over the world.

Terms like “intelligent buildings”, “building management systems”, and “intelligent structures” suggest the rise of the smart building concept.

It is remarkable that terms concerned with cost and technological innovation appeared so few times in the literature “energy cost”, “cost effectiveness”, and “innovation”.

Terms concerned with multivariate analysis like “cluster analysis”, “multiple regression analysis”, “factor analysis”, and “multivariant analysis” are also present among the keywords.

In order to analyze each cluster, it is worth classifying their elements into categories in order to make the comparison between clusters possible. Before proceeding with this classification it will be necessary to create such categories, what is not an easy task.

As the keywords were extracted from technical papers, which have a quite similar structure, i.e., they are built around the same elements, which are the main objective, additional objectives, research’s focus, method, data analysis techniques, and delimitation; these elements were chosen to identify the categories.

The classification process is not straightforward. While some words can be clearly classified into a category, others, only roughly, can be grouped into one. The connection between a term and categories sometimes is made by an underlying sense. For example, the term “meteorological data” is connected to simulation, since they are one of the inputs for simulation software’s.

In the name of the clarity, the clusters identified in Fig. 19 will be presented and analyzed individually. Figure 20 shows cluster 1.

The elements of Cluster 1 can be roughly classified according to Table 7.

It is possible to infer that this cluster groups terms from articles whose main objective is to “reduce energy consumption”, to improve the energetic “building performance”, to increase the rational “energy utilization”, the “building energy use”, and the “building operations”. The main objective is to be reached by taking into account the occupants’

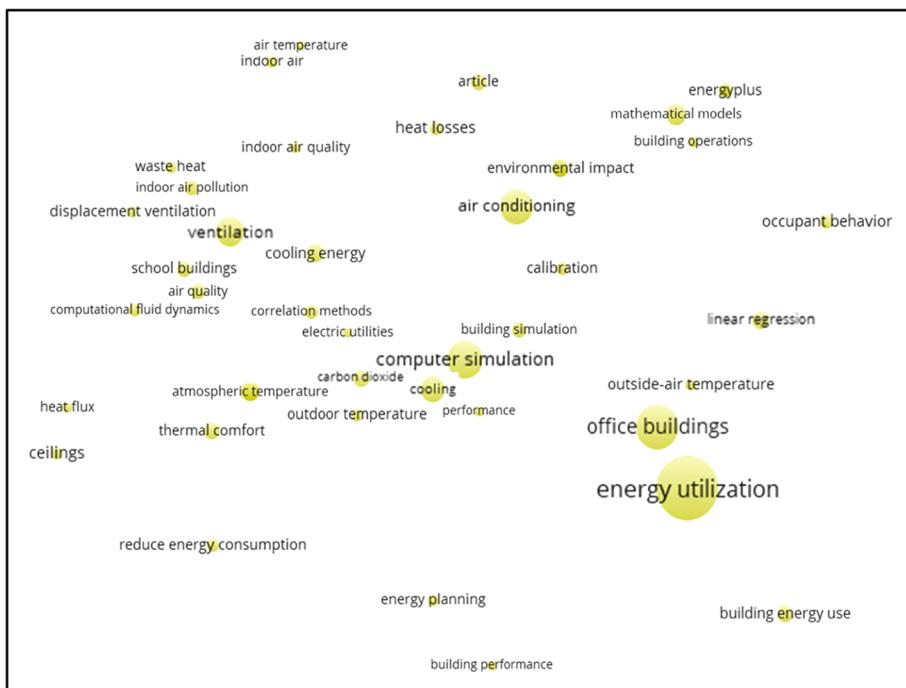


Fig. 20 Cluster 1 from Fig. 19

Table 7 Classification of individuals of cluster 1

Main objective	“energy utilization”, “building energy use”, “building performance”, “performance”, “reduce energy consumption”, “building operations”
Additional objectives	“thermal comfort” (“indoor air”, “indoor air quality”, “indoor pollution”, “air quality”, “carbon dioxide”)
Research’s focus	“heat flux”, “waste heat”, “ventilation”, “air conditioning”, “cooling energy”, “cooling”, “atmosphere temperature”, “outdoor temperature”
Method	“computer simulation”, “building simulation”, “mathematical models”, “computational fluid dynamics”, “EnergyPlus”
Data analysis techniques	“linear regression”, “correlation methods”
Delimitation	“office buildings”, “school buildings”

wellness by dealing with the “indoor air”, “indoor air quality”, “indoor pollution”, “air quality”, and “carbon dioxide”.

Terms like “heat flux”, “waste heat”, “ventilation”, “air conditioning”, “cooling energy”, and “cooling” allow one to conclude that the articles within this cluster focused on Heating, Ventilation, and Air Conditioning (HVAC) systems. The words “atmosphere temperature” and “outdoor temperature” are indirectly related to HVAC systems.

The terms “office buildings”, and “school buildings” suggest that the researches described were delimited to commercial buildings.

The methods employed by the authors were “computer simulation” or “building simulation”. “Mathematics models” are necessary to carry out simulations, and “computational fluid dynamics” is also related to simulation. “EnergyPlus” is a building energy simulation software used to evaluate the energy consumption for heating, cooling, ventilation, lighting and other plug loads. The most cited data techniques were “linear regression” and “correlation methods”.

Figure 21 presents the second key terms cluster.

As done before, the keywords from Fig. 21 are classified according Table 8.

The main objective of the articles within this cluster is the “energy efficiency”, “energy conservation”, “energy performance”, or “energy saving potential” applied to buildings, as can be concluded from the terms “energy efficiency in buildings”, “energy-efficient buildings”, and “energy performance of buildings”. These papers consider environmental questions like “global warming”, “gas emission”, “greenhouse gases”, “sustainable development”, and signal the idea of “sustainable building”. It seems that the researchers

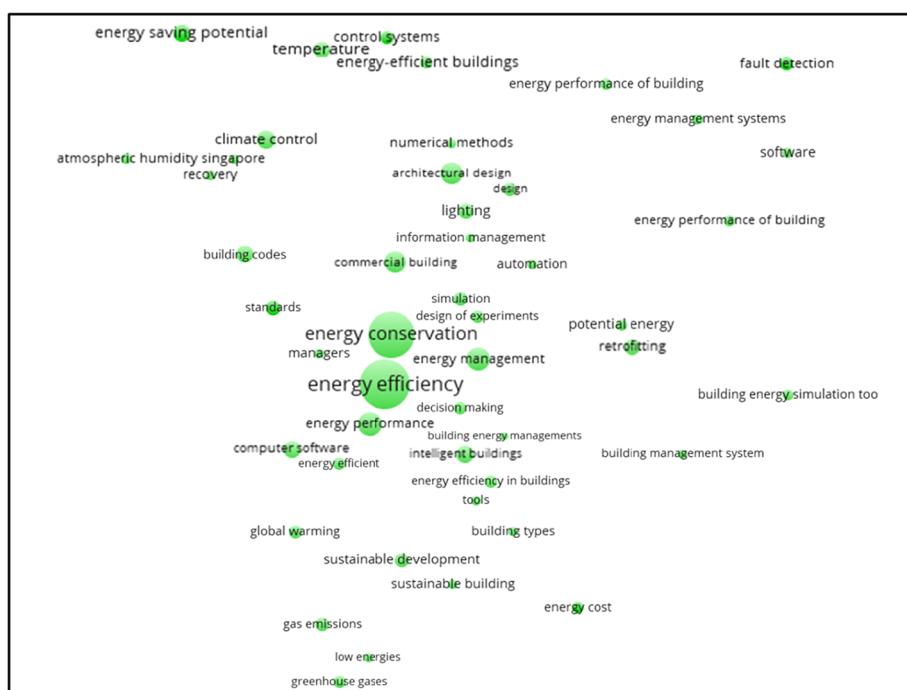


Fig. 21 Cluster 2 from Fig. 19

Table 8 Classification of individuals of cluster 2

Main objective	“energy efficiency”, “energy conservation”, “energy performance”, “energy efficiency in buildings”, “energy saving potential”, “energy-efficient buildings”, “energy performance of buildings”
Additional objectives	“global warming”, “gas emissions”, “greenhouse gases”, “sustainable development”, “sustainable building”
Research’s focus	“energy management”, “building energy managements”, “building management system”, “information management” “automation”, “climate control”, “control systems”, “intelligent buildings” “design”, “architectural design”, “retrofit”
Method	“simulation”, “building energy simulation tools”, “computer software”, “software”
Data analysis techniques	“numerical methods”, “design of experiments”
Delimitation	“commercial buildings” “Singapore”

were delimited to “commercial buildings” with a certain highlight to researches done in Singapore.

The research focus relies on three areas building energy management as suggested by “energy management”, “building management system”, and “information management”. Automation as can be deduced from terms like “climate control”, “control systems”, and “intelligent buildings”. And architectural design.

The preferred methodological approach used by these researches is simulation (“building energy simulation tools”, “computer simulation software”, and “software”) and the design of experiments”.

The third cluster of keywords is presented in Fig. 22.

The classification of the keywords presented in Fig. 22 is shown in Table 9.

The main objective of the papers in this cluster is to evaluate the buildings energy use (“energy use” and “building energy simulation”), taking into account the “climate change”, and “climate effect”. Great part of the researches grouped in this cluster were carried out mainly in China and were not delimited to only a specific type of building.

These researches focused their attention on the thermal comfort systems (“cooling load”, “heating load”, and “heat gains”). The approach used by the researches was mainly the simulation as inferred from the terms “climate models”, “meteorological data”, “weather conditions”, “typical meteorological year”, and “solar radiation”.

Some of the articles, within this cluster, innovate using the principal component analysis, as data analysis techniques.

Figure 23 presents the fourth cluster.

Table 10 presents a classification of the keywords from Fig. 23.

The main objective of these articles in the analysis of “building energy consumption” (“building energy analysis”, “energy use intensities”). Therefore, such papers deal with “forecasting”, “prediction”, and “electric load forecasting”. This focus relied on “forecasting algorithms”, “model validation”, “parameter estimation”, and “optimization”. These researches seem to be conducted mainly in the USA.

In these papers, the authors employed a wide variety of data analysis techniques ranging from the basic univariate statistics (“statistical tests”, “mean square error”, “regression

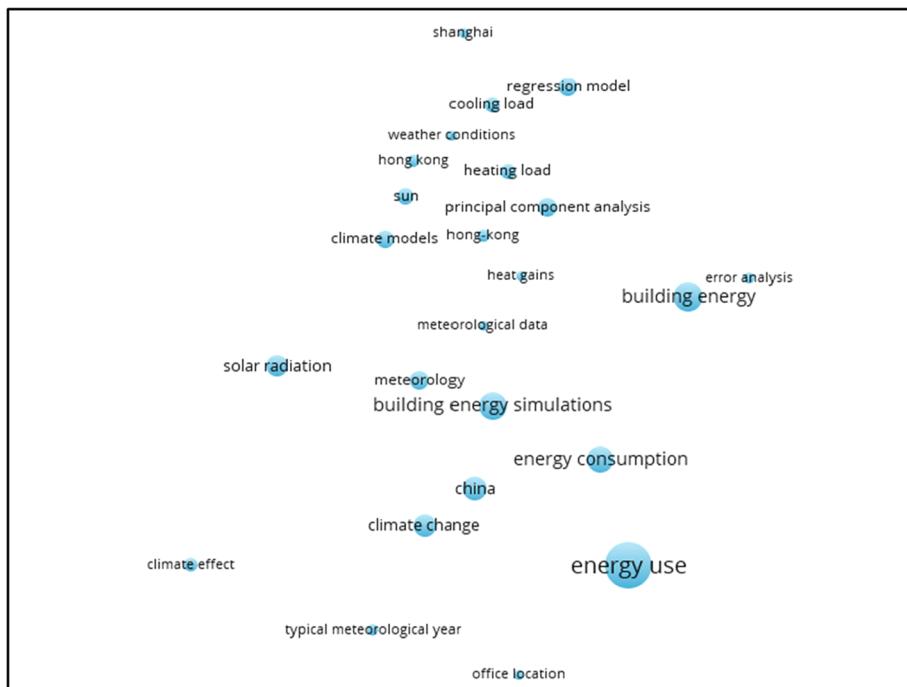


Fig. 22 Cluster 3 from Fig. 19

Table 9 Classification of individuals of cluster 3

Main objective	“energy use”, “building energy”
Additional objectives	“climate change”, “climate effect”
Research's focus	“cooling load”, “heating load”, “heat gains”
Method	“simulation” (“climate models”, “meteorological data”, “weather conditions”, “building energy simulation”, “typical meteorological year”, “solar radiation”)
Data analysis techniques	“regression model”, “principal components analysis”
Delimitation	“China”, “Hong-Kong”, “Shanghai”

analysis”) to computational intelligence (“neural networks”, “fuzzy systems”, “genetic algorithms”, “data mining”), possibly multivariate analysis (“clustering algorithms”).

The fifth cluster is presented in Fig. 24.

The classification of the terms of Fig. 24 is shown in Table 11.

The keywords of this cluster suggest articles interested in energy efficiency in buildings, considering environmental and sustainability question, as well as, historical aspects and urban planning. The focus of these articles relied on four fronts: the energetic performance; thermal performance; constructive techniques; and economic analysis. The method preferred by these paper’s authors was the benchmarking. Such publications did not employ a

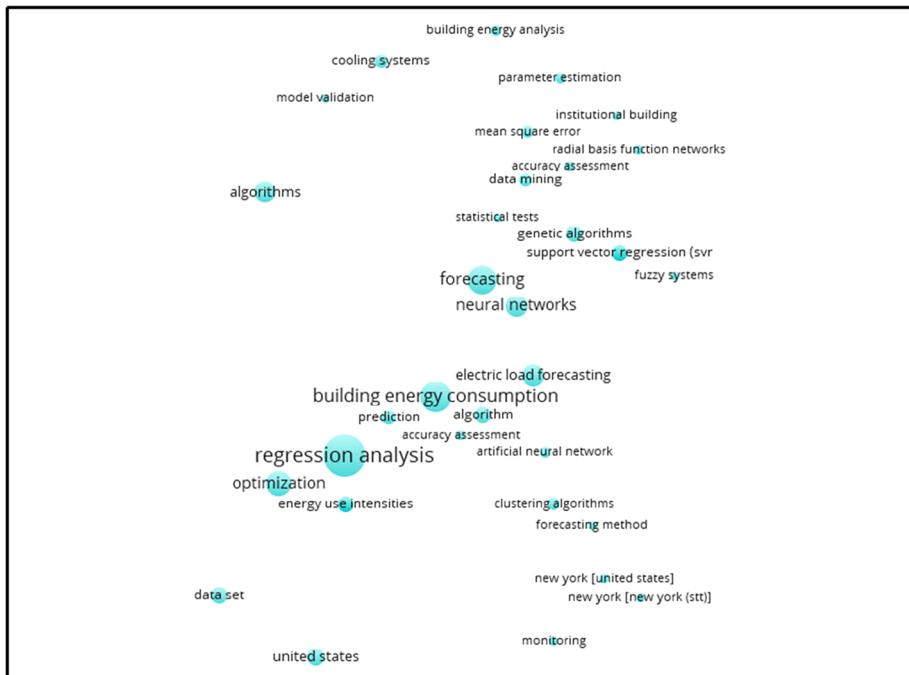


Fig. 23 Cluster 4 from Fig. 19

Table 10 Classification of individuals of cluster 4

Main objective	“building energy consumption”, “building energy analysis”, “energy use intensities”
Additional objectives	“forecasting”, “prediction”, “electric load forecasting”
Research's focus	“forecasting method”, “forecasting algorithms”, “model validation”, “parameter estimation”, “optimization”
Method	
Data analysis techniques	“regression analysis”, “data mining”, “neural networks/artificial neural networks”, “clustering algorithms”, “fuzzy systems”, “genetic algorithms”, “statistical tests”, “mean square error”, “radial basic function network”, “optimization”
Delimitation	“institutional buildings” “United States”, “New York”

diversity of data analysis techniques. These researches were delimited to buildings in general and great part of them took place in Europe, Asia, Eurasia, Asia and Australia.

Figure 25 shows the sixth cluster.

The classification of the keywords from Fig. 25 is shown in Table 12.

The articles within this cluster aim to improve “building energy efficiency”, “building energy saving”, to decrease the “electric consumption” and the “electric power utilization”. They reach their goals by focusing on HVAC system (“heating”, “heating energy”).

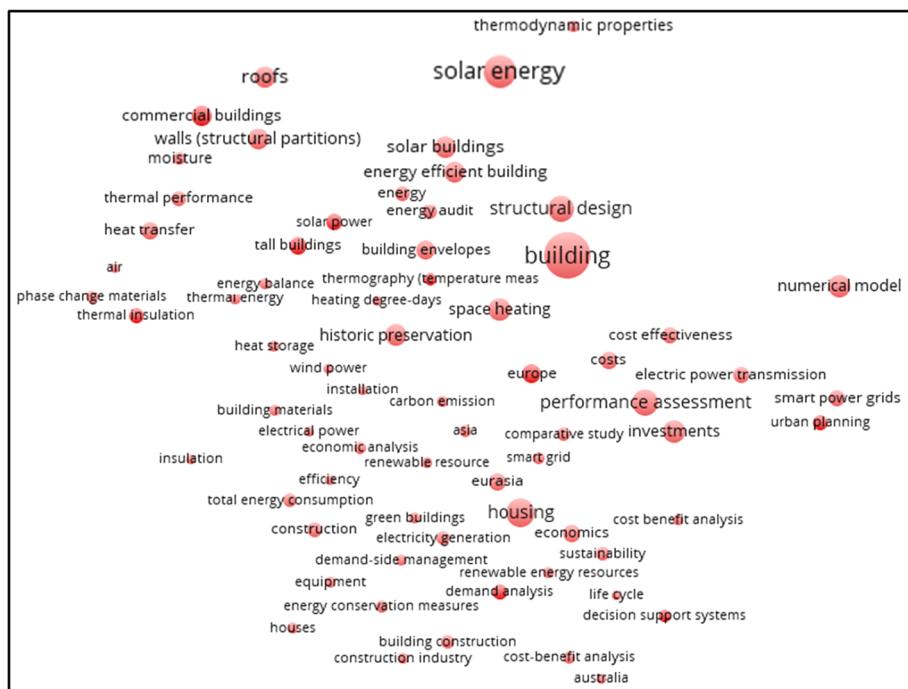


Fig. 24 Cluster 5 from Fig. 19

Table 11 Classification of individuals of cluster 5

Main objective	“energy efficient building”, “energy”, “efficiency”
Additional objectives	“renewable resource”, “sustainability”, “renewable energy resources”, “solar energy”, “solar power”, “carbon emission”, “green buildings”, “wind power”, “solar buildings”, “energy conservation measures”, “demand analysis”, “energy balance”, “historic preservation”, “urban planning”
Research’s focus	“performance assessment”, “total energy consumption”, “energy audit”, “decision support systems”, “buildings envelopes”, “roofs”, “walls (structural partitions)”, “thermal performance”, “thermal insulation”, “thermal energy”, “heat transfer”, “thermography (temperature measures)”, “heating degree-days”, “space heating”, “heat storage”, “thermodynamic properties”, “building construction”, “construction industry”, “structural design”, “building materials”, “insulation”, “phase change materials”, “economics”, “economic analysis”, “cost”, “cost effectiveness”, “cost–benefit analysis”
Method	“comparative study”
Data analysis techniques	
Delimitation	“building”, “commercial buildings”, “tall buildings”, “houses”, “housing”, “Australia”, “Europe”, “Asia”, “Eurasia”

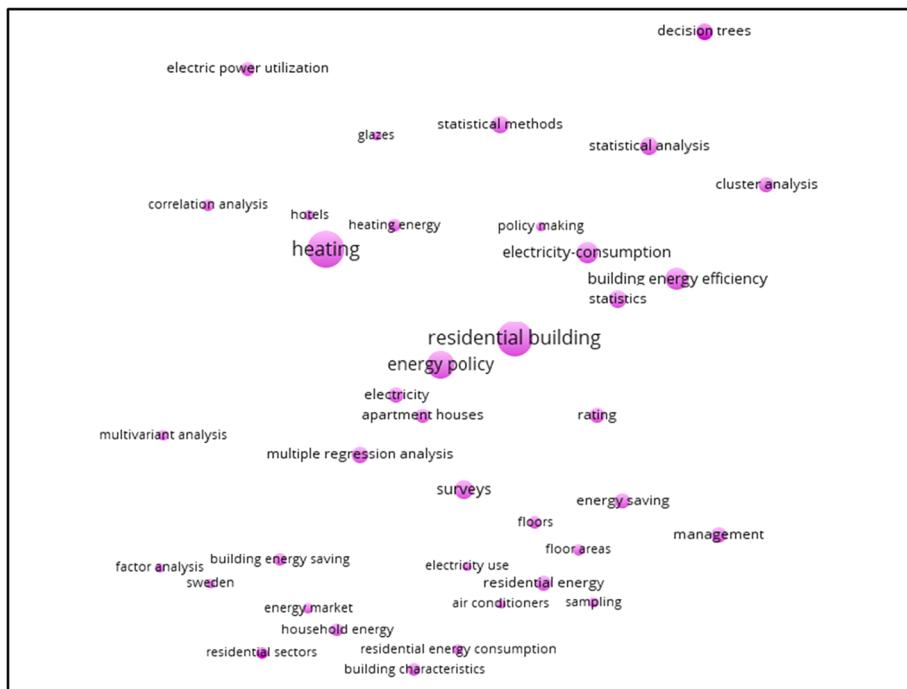


Fig. 25 Cluster 6 from Fig. 19

Table 12 Classification of individuals of cluster 6

Main objective	“building energy efficiency”, “building energy saving”, “energy saving”, “electric consumption”, “electric power utilization”, “electricity use”,
Additional objectives	“energy policy”, “energy market”, “policy making”,
Research's focus	“heating”, “heating energy”, “air conditioners”,
Method	“surveys”
Data analysis techniques	“statistics”, “statistical analysis”, “cluster analysis”, “correlation analysis”, “multiple regression analysis”, “factor analysis”, “multivariate analysis”, “decision trees”
Delimitation	“residential energy”, “household energy”, “residential building”, “residential sectors”, “hotels”, “apartment houses”, “residential energy consumption” “Sweden”

The data used by such papers were gathered by means of surveys and were analyzed by univariate statistics (“statistics”, “statistical analysis”, “correlation analysis”, and “decision trees”) and by multivariate statistics (“cluster analysis”, “multiple regression analysis”, “factor analysis”, and multivariate analysis”).

Figure 26 presents the last cluster.

The keywords from Fig. 26 can be classified according to Table 13.

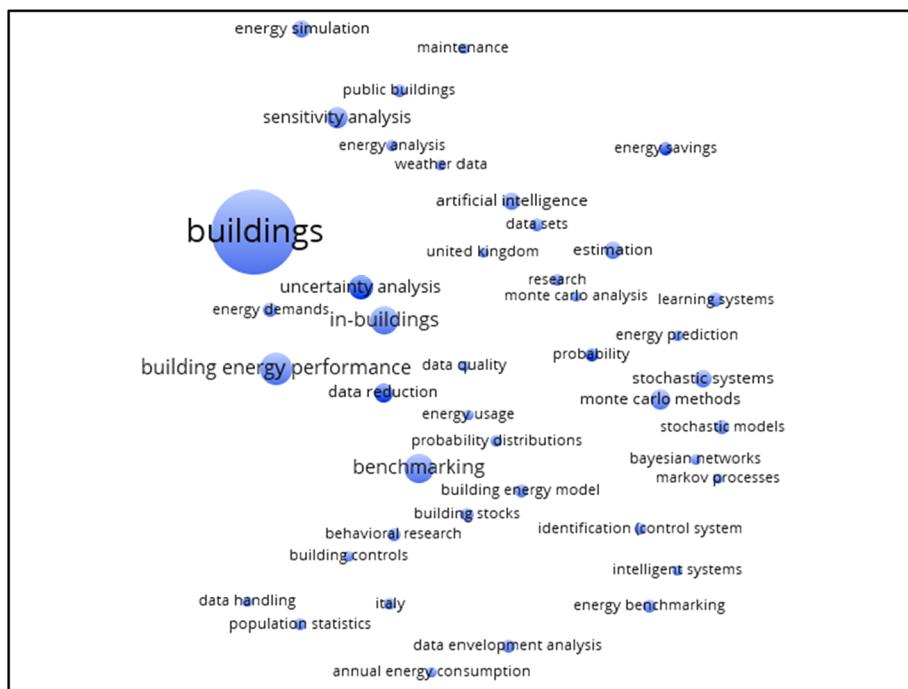


Fig. 26 Cluster 7 from Fig. 19

Table 13 Classification of individuals of cluster 6

Main objective	“energy demand”, “energy usage”, “annual energy consumption”, “building energy performance”, “energy predictions”
Additional objectives	
Research's focus	“building controls”, “identification (control system)” “behavior research”
Method	“benchmarking”, “energy benchmarking”, “energy simulation”, “weather data”
Data analysis techniques	“sensitivity analysis”, “uncertainty analysis” “probability distributions”, “population statistics”, “data envelopment analysis”, “probability” “artificial intelligence”, “learning systems” “stochastic systems”, “stochastic models”, “markov processes”, “bayesian networks” “Monte Carlo analysis”, “Monte Carlo methods”,
Delimitation	“buildings”, “public buildings” “Italy”, “United Kingdom”

Reducing the “energy demand”, the “energy usage”, and the “annual energy consumption” are the main objectives of the articles grouped into this cluster. The researches’ focus is a building automation (“building controls”, “identification control system”) and

“behavior research”. The energy benchmarking” and “energy simulation” are the methods used such researches.

Several data analysis techniques were employed by researches: descriptive statistics (“population statistics”, “probability distributions”); stochastic processes (“stochastic models”, “stochastic systems”, “Markov processes”); computational intelligence (“artificial intelligence”, “learning systems”); and operations research techniques (“data envelopment analysis”). The researches seem to be carried out in buildings in general.

The seven clusters can be characterized, respectively, by the following terms: Buildings and Energy Uses; Building Energy Conservation; Energy Consumption; Energy Consumption Forecasting and Computational Intelligence; Energy Efficiency and Climate Effects; Building Energy Efficiency and Multivariate Statistics; and Building Energy Analysis and Stochastics Processes.

Observing the seven clusters it can be seen that the most frequent method applied in the researches is the simulation and the most used data analysis technique is regression analysis. The cluster 2, 3, and 5 grouped articles worried by the environmental questions.

The clusters 1, 3, and 6 focused on the occupants’ thermal comfort. Only the cluster 5 deals with building materials.

The articles within clusters 4, 6, and 7 employ a large variety of data analysis techniques ranging from basic statistics to computational intelligence.

In cluster 4 most of the data analysis techniques mentioned is related to computational intelligence and most of the articles deals with forecasting and prediction.

In cluster 6, the multivariate statistics were largely cited along with univariate descriptive statistics, which is in accordance with the method used by the articles.

The cluster 7 was dominated by the stochastic processes as data analysis techniques, which seems to be in tune with the researches’ focus.

Institutions analysis

Figure 27 presents the percent of institutions that published the articles, subdivided into regions.

Table 14 presents the percent of institutions that published the articles, subdivided into countries.

Fig. 27 Percentage of institutions that published articles dealing with the theme, distributed by region

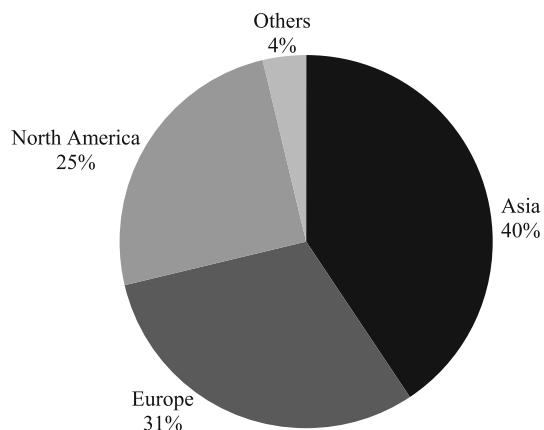


Table 14 Number of institutions that published articles dealing with the theme, distributed by country

Regions	Countries
Asia	China (21); Japan (19); Taiwan (5); South Korea (4); Singapore (2); Malaysia (2); Finland (2); Saudi Arabia (2); Hong Kong (2); India (2); Nigeria (1); Turkey (1); Kuwait (1); Qatar (1)
Europe	United Kingdom (14); Sweden (6); Germany (6); Italy (4); Switzerland (4); France (4); Greece (2); Portugal (2); Norway (1); Belgium (1); Netherlands (1); Estonia (1); Denmark (1); Spain (1); Cyprus (1)
North America	United States (38); Canada (2)

The number of publications is in brackets

It can be seen that Asia stands out with a number of sixty-five institutions, followed by Europe (49) and North America (40), which highlights that these regions contributed the most to publications related to the theme. These same institutions were distributed according to the country, as shown in Table 14.

It can be seen that the most countries that have the largest number of institutions are among the ten countries that presented the largest number of publications, as has been seen, previously, in the analysis of a number of publications by country.

Table 15 presents the number of publications for the twelve most relevant institutions.

The first rank is not taken by a university, but by a Lawrence Berkeley National Labor, that is a Department of Energy (DOE) Office of managed by the University of California, located in the United States. The second, the third, the fourth, the fifth and the seventh go to the Hong Kong Polytechnic University, City University of Hong Kong, Tongji University, Xi'an University of Architecture and Technology and Tianjin University which are all located in the China. The sixth most productive institution is the National University of Singapore which represents the Singapore. The eighth position is occupied by the Texas

Table 15 Analysis of institutions with the highest number of publications

Institution	Country	Documents	Citation count
Lawrence Berkeley National Labor	United States	25	1164
Hong Kong Polytechnic University	China	15	308
City University of Hong Kong	China	15	811
Tongji University	China	14	195
Xi'an University of Architecture and Technology	China	11	287
National University of Singapore	Singapore	10	362
Tianjin University	China	10	76
Texas A and M University	United States	10	128
Universiti Teknologi Malaysia	Malaysia	9	132
Tohoku University	Japan	8	358
University College London (UCL)	United Kingdom	8	160
University of California, Berkeley	United States	8	173

A and M University, United State; followed by Universiti Teknologi Malaysia located in Malaysia, Tohoku University which represents the Japan, UCL in the United Kingdom and UC Berkeley which is located in the United States.

The expectations of an energetic crisis, which actually arose in California (2000–2001), incentivized many programs aiming to promote the energetic efficiency in buildings that have been followed by several publications whose growth followed the same pattern of world energetic crisis, this suggests the prominence of the University of California, so can be seen from Table 15.

It can be seen that the five most prolific institutions are in the United States and China that occupies the two first positions among the most productive countries (Table 5). Curiously among the twelve most productive institutions presented in Table 15, there is solely one European Institution, the University College London (UCL).

By reviewing the departments involved with the published articles by the top twelve institutions, it was noticed that the large part of them are purely related with energy, engineering and buildings like “Department of Building Services Engineering”, “Department of Mechanical Engineering”, “Department of Building” or “Department of Energy”. This is aligned with the authors’ expectation.

Authors' distribution

According Palomo et al. (2017) to analyze the pattern of productivity of the researchers in a subject area the Lotka's law can be applied.

In order to apply the Lotka's law it is necessary to know the frequency distribution of authorship. Table 16 shows the authorship distribution considering multiple authors per article, i.e., including all co-authors.

From Table 16 it can be seen that 1334 authors and co-authors are responsible for the total of publications. The Lotka's law, “describes the productivity distribution among researchers and shows that a small number of authors are responsible for most of literature whereas the contribution of the large majority of researchers is very low in terms of number of publications” (Barrios et al. 2008).

The Lotka's inverse power law (Pao 1985) states that the number of authors (y_x) that produce x articles is inversely proportional to x that is the output of each individual author. Such a relation can be modeled as Eq. (1).

$$x^n y_x = C \quad (1)$$

where n and C are constants to be estimated from the observed data set.

Table 16 Frequency distribution, including all co-authors

X	y_x
1	1153
2	118
3	39
4	13
5	4
6	5
10	2
	1334

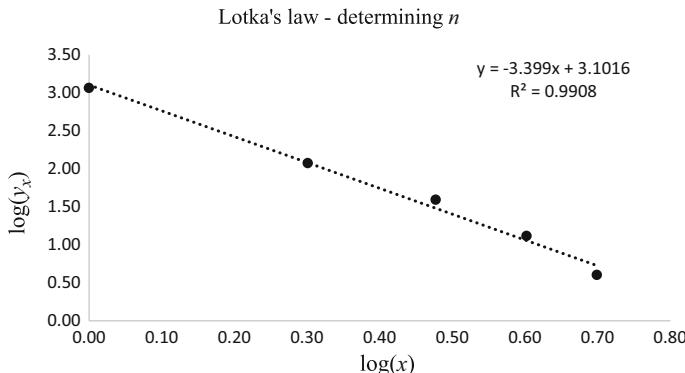


Fig. 28 Determining the n value (the most prolific author was discarded, $x = 6$, and $x = 10$)

Figure 28 presents the procedure described by Pao (1985) to determine the n value, which is the angular coefficient of the straight line depicted in Fig. 28 the n value was calculated using the least-squares method ($n \approx -3.4$).

According to Pao (1985) the C value can be calculated by Eq. (2)

$$C = \frac{1}{\sum x^n} \quad (n < 0) = \frac{1}{\sum \frac{1}{x^n}} \quad (n > 0) \quad (2)$$

Unfortunately, has is no easy formula for computing the sum of the infinite series described by Eq. 2. Although Pao (1985) derived a function approximating such a summation, in this paper the C value was calculated by means of using the Wolfram Alpha by means of the following command string:

$$(\text{sum } x^{-3.4} \text{ from 1 to infinity})^{-1}$$

leading to a C value of approximately 0.88.

Finally, in order to verify that the observed data fitted the estimated distribution, the non-parametric Kolmogorov–Smirnov goodness-of-fit test has been applied according to the procedure described by Pao (1985), Table 17.

From Table 17, it can be seen that the observed data fitted the Lotka's law, since the maximum deviation between the cumulative proportions of observed data and estimated values (0.0157) is lower than the critical value at a significance level of 0.05 (0.0446), which means that about 88% of the authors made a single contribution.

According to Pao (1986) among the sources of errors that lead to questionable parameters estimates for Lotka's equation, is the number of authors per publication considered. Although Coile (1977, *apud* Pao 1986) showed that data counts with senior authors (or first-authors) and those with all authors could produce significantly conclusions, since they would produce different values for the exponent in Eq. (1), this is not the case for this study, since the n value keeps around 3.4 when only one author is considered as well, as illustrated in Fig. 29.

In his original article, Lotka described experiments with two different data sets, from which he derived two different values for n that he considered approximately 2. “That enabled him to draw his often quoted conclusion that the number of persons making n contribution is about $1/n^2$ of those making one and the proportion of contributors that make a single contribution is about 60%” (Lotka 1926, *apud* Pao 1985). This statement is

Table 17 Kolmogorov–Smirnov goodness-of-fit test (Pao 1985)

		<i>n</i> = 3.4		Critical value K_S = 0.0446	
$C = 0.88$	Number of publications X	Number of authors with x publications Y	Proportion of authors $y\%$	Accumulated proportion $\sum y\%$	Expected proportion of authors with x publications $C(1/x')$
1	1153	0.8643	0.8643	0.8800	0.8800
2	118	0.0885	0.9528	0.0834	0.9634
3	39	0.0292	0.9820	0.0210	0.9844
4	13	0.0097	0.9918	0.0079	0.9923
5	4	0.0030	0.9948	0.0037	0.9960
6	5	0.0037	0.9985	0.0020	0.9980
10	2	0.0015	1.0000	0.0004	0.9983
Total	1334				$D_{\max} = 0.0157$

H0: Data come from the same population ($D_{\max} < \text{Critical } K_S$)H1: Data don't come from the same population ($D_{\max} > \text{Critical } K_S$)

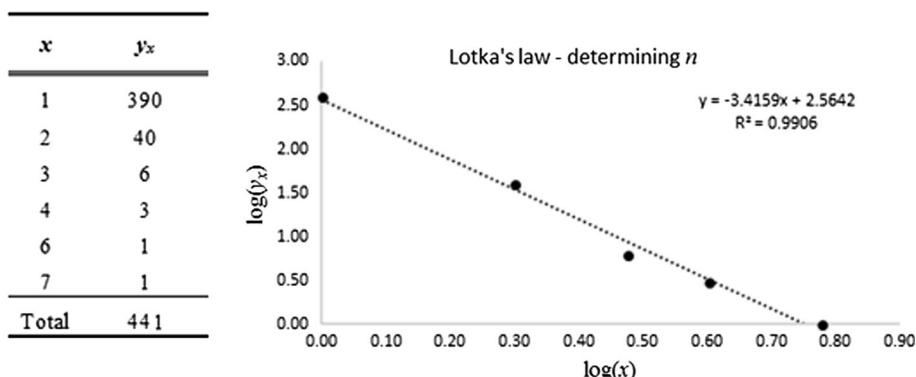


Fig. 29 Determining the n value (the most prolific author was discarded, $x = 7$)

not true for this case, since, according to Kolmogorov–Smirnov goodness-of-fit test, the theoretical distribution considering n value equals 2 does not have adherence to the observed data (considering single or multiple authors).

Another interesting consideration is the Price's square root law which states that “half of the literature on a subject will be contributed by the square root of the total number of authors publishing in that area”. In other words “Price's contention was that half the published output in a subject field will be contributed by a highly productive subset of authors” (Nicholls 1988).

Considering a single author per paper, according to the Price's contention, the 21 most prolific authors are responsible for 256 articles. This result is not supported by the empirical evidences since the 21 most prolific authors are responsible for 63 papers. If multiple authors per article are considered, the 36 most prolific authors should be responsible for 256 articles. Empirical evidences do not support this result either, since the 36 most prolific authors and respective co-authors are responsible for 158 articles. Although it is true, for this area, that a small number of authors are responsible for most of articles and a large number of authors produces only one paper each, the results do not support the Price's square root law, because the Price's original claim was based on Lotka's inverse square law, rather than on the generalized Lotka's inverse power law (Nicholls 1988).

The top eleven most productive authors in the data analysis techniques applied to energy efficiency in buildings, and theirs respective citations are shown in the Fig. 30.

The most productive authors are Lam, J.C. (City University of Hong Kong, Department of Civil and Architectural Engineering, Hong Kong, China) which published ten articles on the subject, was also the most cited (586). Among his publications, three exceeded 100 citations, namely: “Sensitivity analysis of energy performance of office buildings” which was published in 1996 and cited 126 times, “Future trends of building heating and cooling loads and energy consumption in different climates” published in 2011 and cited 107 times; “Energy analysis of commercial buildings in subtropical climates” published in 2000 and cited 101, these articles were published in Building and Environment journal; and Yoshino, H. (Tohoku University, Department of Architecture and Building Science, Sendai, Japan) which published ten articles and was cited 340 times. Two of his publications exceeded 100 citations, namely: “A systematic procedure to study the influence of occupant behavior on building energy consumption” (168) published in 2011 and “A

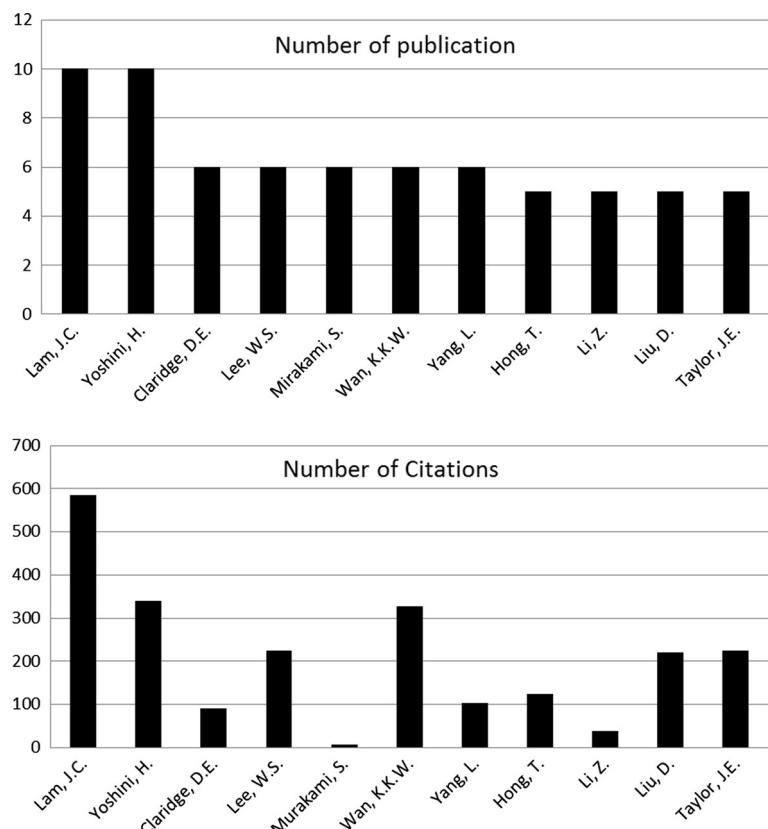


Fig. 30 Authors' production and citations (considering multiple authors per article)

decision tree method for building energy demand modeling” (114) published in 2010. These articles were published in Energy and Buildings journal.

Identifying the data analysis techniques employed

The 513 papers retrieved at the beginning of this research have been read, and only 296 of them really applied any data analysis technique. Some of the data analysis techniques pointed out by the key terms analysis was only mentioned, but not really employed, thus

Table 18 Data analysis techniques

Data analysis techniques	Occurrences
Regression Analysis	139
Descriptive Statistics	74
Multivariate Analysis	40
Computational Intelligence	34
Stochastic Processes	18
Inferential Statistics	30
Design of Experiments	23

they were not taken in consideration. The data analysis techniques used in these 296 articles were roughly grouped into seven categories. Table 18 presents these categories and the number of times they appear. It is worth mentioning that sometimes more than one technique was used in an article.

The most cited category was the Regression Analysis. This category grouped several techniques such as linear regression (77 occurrences of the technique), multivariate linear regression (39), non-linear regression (4), ordinary least squares and other techniques (19). These techniques were applied in articles focusing on almost all types of investigation within this field, as can be seen in the analysis of the key terms topic. For example, to estimate the energy performance of buildings (Melo et al. 2016); to understand the influential characteristics of energy consumption in buildings (Ma and Cheng 2016); and for prediction of the energy consumption of buildings (Zhou et al. 2016).

The second most cited category was the Descriptive Statistics, which brings together techniques such as central tendency and dispersion measures (48), and correlation measures (26). This category is employed in papers describing exploratory researches, and it was also employed concomitantly with other techniques., for example, analysis of the correlation between energy consumption in buildings and greenhouse gas emissions (Wang et al. 2016), and between building design and energy consumption (Liu et al. 2015).

The third most cited group was the Multivariate Data Analysis Techniques presenting methods such as Principal Component Analysis (16), Clustering (16), Structural Equation Modeling (4), Factor Analysis (2), Discriminant Analysis (1) and Multivariate Analysis of Variance (1). This category is present in papers which deal with surveys, for example, to describe the perceptions of individuals in relation to the challenges of building energy efficiency (Addy et al. 2014); and analyze the impacts of household and building characteristics on energy consumption (Estiri 2014).

The fourth most cited group was the Computational Intelligence, which brings together techniques such as Artificial Neural Networks (25), Genetic Algorithms (8) and Machine Learning (1). These techniques were applied in papers focusing on forecasting, for example, for energy savings predictions for a retrofit project in buildings (Yalcintas 2008); and in the prediction of building energy consumption (Li et al. 2015a, b; Buratti et al. 2014).

The fifth most cited category was the Stochastic Processes that consists of a variety of techniques such Monte Carlo (8) and Sensitivity Analysis techniques (10). These analysis techniques are used in articles interested in simulation and modeling, for example, to find a model for lower cost and energy consumption (Orosa 2012; Wang et al. 2012); and to examine the sensitivity of energy performance of buildings (Lam and Hui 1996; Hemsath and Bandhossani 2015).

The sixth most cited category was the Inferential Statistics that brings together parametric tests (25) and nonparametric tests (5), which brought together techniques that are jointly employed with most of the other techniques. For example, comparative analysis of building energy simulation of the different variables (Pernigotto and Gasparella 2013); identification of variables to estimate the energy performance of buildings (Tsanas and Xifara 2012); and to evaluate the frequency distribution of annual energy consumption in buildings (Capozzoli et al. 2016).

The last category was the Design of Experiment, which groups several types of experimental designs (17) and Analysis of Variance (6). The Design of Experiments is used to optimize processes and to assess the influence of input factors in the response variable of a process. For example, to optimize the energy consumption based on environmental factors (Assadi et al. 2016); and to evaluate the effectiveness of climate factors on energy saving in green buildings (Zahraee et al. 2014).

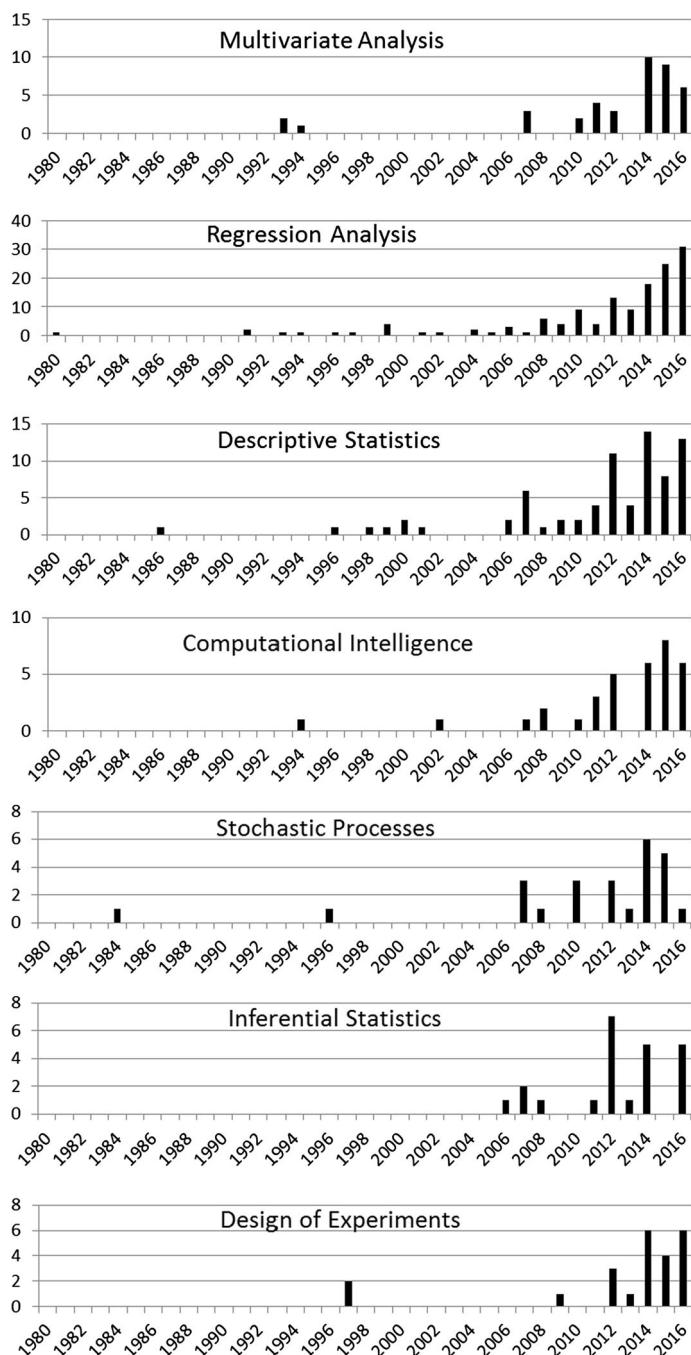


Fig. 31 Evolution of the application of the data analysis techniques

It is worth mentioning that the use of some of these techniques presented an overlap with others. Figure 31 shows the dispersion of the data analysis techniques over the years.

It can be seen from Fig. 31 that all of those data analysis techniques presented in Table 18 have been remarkably applied after 2010. It is also possible to see that Regression Analysis and Descriptive Statistics have been used since the beginning and the middle of the 1990s, respectively.

Although the utilization of all of the techniques has been increasing over the observed period, none of them showed a growth more consistent than the one showed by the Regression Analysis. Its utilization was exceeded by the others techniques only in 6 years: 1984, 1993, 1997, 1998, 2000 and 2007.

It is worth mentioning that from 2012 to 2016 were the period when most of the techniques were applied.

Conclusions

This paper describes a bibliometrics analysis concerned with the data analysis procedures applied to studies of energy performance in buildings. This article investigated the evolution of publications on the theme from 1980 to 2016.

The interest in this subject has been increasing over the period under analysis, this can be interpreted by the growth in the publications related to the theme, it was identified three of the four phases in the growth of publications according to Price's law. The first phase (1980–1998) was the precursors' phase, because a small number of researchers began to publish, this phase accounts for about 4.4% of the whole publication body. The second phase (1992–2013) presented a proper by exponential growth, demonstrating that the number of publications doubled approximately every each 7 years. The third phase (2013–2016), the body of knowledge was consolidated and the growth of scientific production becomes linear. Based on this analysis, it could be concluded that the number of publications in this field, from 1980 to 2016, followed the Price's Law and an inflection point has not been reached yet.

The growth of publications was primarily driven by the Europe, North America and Asia. In Europe, the governmental incentive programs for promoting the rational use of energy and legal determination for the creation of zero energy buildings were responsible for the growing in the number of publications. In comparison, North America had begun 20 years before due to the great attention paid by the United States to the theme, the expectations of an energetic crisis, which actually arose in California, had incentivized many programs aiming to promote the energy efficiency in buildings that were followed by several publications. The number of publications in Asia followed the same pattern in Europe, because the reasons were the same.

The United States is the country with the greatest number of publication and citations. The most cited paper is also published in the United States. Greece and Switzerland have the greatest number of citations per publication (CPP), although they are out of the top ten publishers. Since the CPP is greater than the number of publications, one may conclude that these publications have been cited by authors abroad, so the publications are relevant to the area.

Malaysia, India and China are the countries with the greatest percentage of non-cited publications. Although China is one of the greatest publishers and their papers are much cited, more than 28% of its production has not been cited yet.

A remarkable fact is that the number of citations of the Brazilian most-cited paper occupies the third position at the rank of the most cited publications. Brazil occupies the thirteenth position out of twenty-six in number of publications. In addition, 20% of its whole production is not cited.

According to the Principal Components Analysis (PCA), the United States stood out because of the number of publications, citations and the number of citations of the most cited article. Switzerland stood out because of the number of citations per publication. Although China has a great number of publications and citations it also has a high percentage of non-cited. India and Malaysia appear highlighted because the percentage of non-cited papers is significant.

The most relevant areas related to the theme were Engineering, Energy, Environmental Science and Computer Science. It could also be seen that the subject has gotten the attention of some peripheral areas such as Social Sciences; Business, Management and Accounting; Economics, Econometrics and Finance; Arts and Humanities; and Medicine. One of the reasons for a such intersection is the fact that many publications are classified into more than one area. But scientific interests can also explain the overlap between the areas.

Social Sciences area contained publications dealing with thermal comfort and occupant behavior. Business, Management and Accounting area grouped articles that address the subject from the stand point of relevant planning and management tools for energy-efficient buildings. Economics, Econometrics and Finance area presented papers that was correlated with Market Value of buildings. Arts and Humanities area contained papers that approach the theme related to the buildings of historical heritage. Medicine are presented publications dealing with health problems recurrent of CO₂ emissions.

The 513 articles were distributed among 155 journals. It was observed that this distribution does not follow the Bradford's law, but it is worth mentioning that 10% of the journals published 60% of the papers, and about 60% of the journals published only one paper.

Energy and Buildings was by far the scientific journal that had published most of the paper related to the theme. It was responsible for about 32% of the publications in the period covered by this research, being the nucleus of journals' distribution. The next zone, containing 32 journals, is responsible for about 35% of the publications in the same period. The third and last zone of the journals' distribution (122 journals) accounts for about 36% of the publications.

The first and second zones account for 22% of the journals with 64% of the articles. Considering only these two zones. Engineering is the area with the greatest number of journals (15), followed by the area of Energy (10), Environmental Science (4), Social Sciences (12), and Material Science (1).

The most productive research centers in this area are located in Asia, with a number of sixty-five institutions, followed by Europe and North America.

China, United Kingdom and the United States stood out with a greater number of institutions, according to their respective region. Taking into account all the figures from bibliometrics analysis, it was possible to conclude that these countries are on the vanguard of the application of data analysis techniques for building energy efficiency studies.

The three most important research centers found in these countries were Lawrence Berkeley National Labor, University of California, United States; Hong Kong Polytechnic University, China; and University College London, United Kingdom. And by reviewing the departments involved with the published articles, it was noticed that the large part of

them were purely related with energy, engineering and buildings, what is aligned with authors' expectations.

The total of authors (first author only) involved in the production of these 513 articles is 441. Considering both author and co-authors this number increases to 1334. In both cases a small number of them is responsible for most of the literature, and in both cases, as well, the frequency distribution of articles per authors follow the Lotka's inverse power law with $n = 3.4$ and $C = 0.88$. Since $n = 3.4$ and not 2 as predictive by the Lotka's inverse square law, the observed data does not support the Price's square root law since 88% of the authors are responsible for only one article and the 21 most prolific authors are responsible for 63 articles instead of 256.

The most productive and cited authors was Lam, J.C. and Yoshini, H. They published ten articles each and some of their papers reached more than 100 citations.

In order to visualize changes in the thematic focus of this research field, according to the key terms, three bibliometric maps covering the period of analysis (1980–1990, 1980–2000, 1980–2016), have been employed. Thus, it was possible to study the formation and evolution of the clusters, to understand the development of this field of knowledge.

During the period ranging from 1980 to 1990, 22 articles were produced, totaling 81 keywords, but only 12 words occurred two or more times, forming 4 clusters. Because of the oil crisis in the middle of the 1970s, terms like "energy conservation", "energy savings", and "energy utilization" appeared in the publications at the beginning of 1980s. It was noted that the advent of the personal computers in the 1980s brought terms related to computer simulation.

From 1980 to 2000, 58 articles were produced, totaling 43 keywords, grouped into 7 clusters. Over this period new terms concerned with descriptive statistics were introduced in the publications, suggesting that this period was marked by the inclusion of the univariate data analysis techniques in the publications. It was also remarkable the very first appearance, in this kind of study, of the term "environmental impact".

From 1980 to 2016, 257 new terms have been included, modifying the former 7 clusters. These clusters are characterized, by the following terms: Buildings and Energy Uses; Building Energy Conservation; Energy Consumption; Energy Consumption Forecasting and Computational Intelligence; Energy Efficiency and Climate Effects; Building Energy Efficiency and Multivariate Statistics; and Building Energy Analysis and Stochastic Processes. It was also remarkable that terms concerned with cost and technological innovation appeared so few times in the literature.

After the year 2000, the environmental questions and terms related to renewable energy have gained more importance. Many articles, in this period, were produced focusing on the building occupants' wellness, building material and building design to improve energy efficiency. Although the univariate descriptive analysis has been applied in the articles, terms concerned with multivariate statistics and computational intelligence arose.

The data analysis techniques were roughly grouped into seven categories: Regression Analysis, Descriptive Statistics, Multivariate Analysis, Computational Intelligence, Stochastic Processes, Inferential Statistics and Design of Experiments. Some of the data analysis techniques pointed out by the key terms analysis were only mentioned, but not really employed, thus they were not took in consideration in the formation of the categories.

The most cited category was the Regression Analysis, which grouped techniques that were applied in articles focusing on almost all types of investigation within this field, in accordance to the key terms analysis. The second category was the Descriptive Statistics, which was employed in papers describing exploratory researches, and it was also employed

concomitantly with other techniques. The Multivariate Data Analysis Techniques was the third most cited group, which was presented in articles that dealt with surveys. The fourth most cited group was the Computational Intelligence, which brought together techniques that were applied in papers focusing on forecasting. The fifth category was the Stochastic Processes that consisted of a variety of techniques, which were used in articles interested in simulation and modeling. The sixth most cited category was the Inferential Statistics, which was jointly employed with most of the other techniques. The last category was the Design of Experiment Cluster, which grouped techniques that was used to optimize processes and to assess the influence of input factors in the response variable of a process or phenomenon.

These data analysis techniques have been remarkably applied in the building energy efficiency researches after 2010, maybe because of the popularization of statistical packages, reaching the top between 2012 and 2016. Although the utilization of all of the techniques has been increasing over the observed period, none of them showed a growth more consistent than the one showed by the Regression Analysis. Its utilization was exceeded by the other techniques only in 1984, 1993, 1997, 1998, 2000 and 2007.

The data analysis techniques identified in this article may influence the possibility of reformulation and the adequacy of the curricula of the undergraduate and graduate courses in the area of energy and smart buildings. The results of this research showed a general perspective about the data analysis tools employed in energy efficiency research in buildings, which can be useful in showing relevant themes for further research.

Acknowledgements The authors would like to thank the National Council for Scientific and Technological Development (CNPq) for supporting this research.

Appendix 1: Scopus database query for energy efficiency

TITLE (

(“power-efficient”) OR (“energy efficiency”) OR (“low-power”) OR (“low-energy”) OR (“energy-saving”) OR (“fuel-efficient”) OR (“energy-conscious”) OR (“high-efficiency”) OR (“Low Power-Consumption”) OR (“energy efficient”) OR (“power allocation”) OR (“energy consumption”) OR (“smart grid”) OR (“smart metering”) OR (“enhancement of efficiency”) OR (“Energy Management”) OR (“Energy Performance”) OR (“Rational Use”) OR (“Electricity Consumption”) OR (“System Efficiency”) OR (“Energy Efficient Prosperity”) OR (“Eco-Power”) OR (“Energy-Efficient Window”) OR (“energy analysis”) OR (“efficiency factor”) OR (“Energy-saving awareness”) OR (“energy efficiency measures”) OR (“energy utilization”) OR (“energy conservation”) OR (“intelligent control”) OR (“energy demand”) OR (“sustainable energy”) OR (“energy policy”) OR (“energy efficient systems”) OR (“energy usage”) OR (“maximum efficiency”) OR (“energy efficiency ratio”) OR (“energy efficiency strategies”) OR (“energy simulation”) OR (“energy use”) OR (“energy storage”) OR (“energy forecasting”) OR (“consumption of energy”) OR (“Conscious use”)

)

AND (LIMIT-TO (DOCTYPE, “ar ”))

AND PUBYEAR > 1979 AND PUBYEAR < 2017

Appendix 2: Scopus database query for energy efficiency and buildings

TITLE ((“power-efficient”) OR (“energy efficiency”) OR (“low-power”) OR (“low-energy”) OR (“energy-saving”) OR (“fuel-efficient”) OR (“energy-conscious”) OR (“high-efficiency”) OR (“Low Power-Consumption”) OR (“energy efficient”) OR (“power allocation”) OR (“energy consumption”) OR (“smart grid”) OR (“smart metering”) OR (“enhancement of efficiency”) OR (“Energy Management”) OR (“Energy Performance”) OR (“Rational Use”) OR (“Electricity Consumption”) OR (“System Efficiency”) OR (“Energy Efficient Prosperity”) OR (“Eco-Power”) OR (“Energy-Efficient Window”) OR (“energy analysis”) OR (“efficiency factor”) OR (“Energy-saving awareness”) OR (“energy efficiency measures”) OR (“energy utilization”) OR (“energy conservation”) OR (“intelligent control”) OR (“energy demand”) OR (“sustainable energy”) OR (“energy policy”) OR (“energy efficient systems”) OR (“energy usage”) OR (“maximum efficiency”) OR (“energy efficiency ratio”) OR (“energy efficiency strategies”) OR (“energy simulation”) OR (“energy use”) OR (“energy storage”) OR (“energy forecasting”) OR (“consumption of energy”) OR (“Conscious use”))
)
AND TITLE
 (“building”)
AND (LIMIT-TO (DOCTYPE, “ar ”))
AND PUBYEAR > 1979 AND PUBYEAR < 2017

Appendix 3: Scopus database query for energy efficiency, buildings and data analysis techniques

TITLE ((“power-efficient”) OR (“energy efficiency”) OR (“low-power”) OR (“low-energy”) OR (“energy-saving”) OR (“fuel-efficient”) OR (“energy-conscious”) OR (“high-efficiency”) OR (“Low Power-Consumption”) OR (“energy efficient”) OR (“power allocation”) OR (“energy consumption”) OR (“smart grid”) OR (“smart metering”) OR (“enhancement of efficiency”) OR (“Energy Management”) OR (“Energy Performance”) OR (“Rational Use”) OR (“Electricity Consumption”) OR (“System Efficiency”) OR (“Energy Efficient Prosperity”) OR (“Eco-Power”) OR (“Energy-Efficient Window”) OR (“energy analysis”) OR (“efficiency factor”) OR (“Energy-saving awareness”) OR (“energy efficiency measures”) OR (“energy utilization”) OR (“energy conservation”) OR (“intelligent control”) OR (“energy demand”) OR (“sustainable energy”) OR (“energy policy”) OR (“energy efficient systems”) OR (“energy usage”) OR (“maximum efficiency”) OR (“energy efficiency ratio”) OR (“energy efficiency strategies”) OR (“energy simulation”) OR (“energy use”) OR (“energy storage”) OR (“energy forecasting”) OR (“consumption of energy”) OR (“Conscious use”))
)
AND TITLE
 (“building”)
AND ABS
(

(“Statistic Statistical Methods”) OR (“Statistical Analysis”) OR (“Statistics”) OR (“Statistical Data Analysis”) OR (“Statistical properties”) OR (“Statistical classifier”) OR (“Statistical comparisons”) OR (“statistical forecasting”) OR (“statistical model”) OR (“Statistical information”) OR (“statistical approach”) OR (“Residual analysis”) OR (“statistical evidence”) OR (“ANOVA”) OR (“Analysis Of Similarities”) OR (“Analysis of Variance”) OR (“Anderson–Darling Test”) OR (“Binomial Classification Analysis”) OR (“Cluster Analysis”) OR (“Clustering”) OR (“Cochran’s Q”) OR (“Cohen’s Kappa”) OR (“Correspondence Analysis”) OR (“Descriptive Statistics”) OR (“Design Experiments”) OR (“Design of Experiments”) OR (“Discriminant Analysis”) OR (“DOE”) OR (“Experimental Design”) OR (“Factor Analysis”) OR (“Factorial ANOVA”) OR (“Factorial Design”) OR (“Factorial Experiment”) OR (“Factorial Experiments”) OR (“Friedman Two-Way Analysis Of Variance By Ranks”) OR (“Histogram”) OR (“Hypotheses Test”) OR (“Hypotheses Testing”) OR (“Hypothesis Test”) OR (“Hypothesis Testing”) OR (“Inference Test”) OR (“inferential Statistics”) OR (“Kaplan–Meier”) OR (“Kendall’s Tau”) OR (“Keⁿdall’s W”) OR (“Kolmogorov–Smirnov Test”) OR (“Kruskal–Wallis”) OR (“Kuiper’s Test”) OR (“Latent Variables”) OR (“Least Squares”) OR (“Logrank Test”) OR (“Mann–Whitney U”) OR (“MANOVA”) OR (“MCA”) OR (“McNemar’s Test”) OR (“MDS”) OR (“Multidimensional Scaling”) OR (“Multiple Correspondence Analysis”) OR (“Multiple Regression Analysis”) OR (“Multivariate Analysis”) OR (“Multivariate Analysis of Variance”) OR (“Multivariate Regression”) OR (“Multivariate Statistics”) OR (“Non-Parametric Models”) OR (“Nonparametric Statistics”) OR (“Normal”) OR (“Parametric Models”) OR (“Parametric Statistics”) OR (“Partial Least Squares”) OR (“PCA”) OR (“Pearson”) OR (“Pitman’s Permutation Test”) OR (“Principal Component Analysis”) OR (“Principal Components Analysis”) OR (“Probability Distributions”) OR (“Rank Products”) OR (“Regression”) OR (“Scores”) OR (“Siegel–Tukey Test”) OR (“Sign Test”) OR (“Spearman’s Rank Correlation Coefficient”) OR (“Squared Ranks Test”) OR (“Statistical Bootstrap Methods”) OR (“Statistical Models”) OR (“Statistical Test”) OR (“Statistics Multivariate Data Analysis”) OR (“Taguchi”) OR (“The Tukey–Kramer Method”) OR (“Training Data”) OR (“T Test”) OR (“Tukey”) OR (“Tukey Method”) OR (“Tukey–Duckworth Test”) OR (“Tukey’s Test”) OR (“Tukey’s Honest Significance Test”) OR (“Tukey’s HSD Test”) OR (“Tukey’s Range Test”) OR (“Univariate Regression”) OR (“Wald–Wolfowitz Runs Test”) OR (“Wilcoxon Rank Sum Test”) OR (“Wilcoxon Signed-Rank Test”) OR (“Z-Score”) OR (“Design Response Surface”) OR (“Analysis of Covariance”) OR (“Chi Square Analysis”) OR (“Control Charts”) OR (“Correlation”) OR (“Linear Regression Analysis”) OR (“Multiple Linear Regression”) OR (“Multiple Response”) OR (“Nonparametric Analysis”) OR (“Probability”) OR (“p-values”) OR (“Residual Analysis”) OR (“Simple Linear Regression”) OR (“Statistics Bivariate”) OR (“Statistics Inference”) OR (“Structural Equation”) OR (“Structural Equation Analysis”) OR (“The Paired T-Test”))
AND (LIMIT-TO (DOCTYPE, “ar ”))
AND PUBYEAR > 1979 AND PUBYEAR < 2017

Appendix 4

Journals	Number of pub/zone
Energy and Buildings	153/1
Applied Energy	36/2
Energy	22/2
Building and Environment	21/2
Energy Conversion and Management	16/2
Energy Policy	10/2
Journal of Solar Energy Engineering Transactions of the ASME	8/2
Journal of the Association of Energy Engineering	6/2
Jurnal Teknologi	6/2
Energy Efficiency	5/2
Journal of Building Performance Simulation	5/2
Renewable Energy	5/2
Solar Energy	5/2
Architectural Science Review	4/2
Building Services Engineering Research and Technology	4/2
Building Simulation	4/2
HVAC and R Research	4/2
Journal of Building Physics	4/2
Renewable and Sustainable Energy Reviews	4/2
Journal of Civil Architectural and Environmental Engineering	4/2
Applied Thermal Engineering	3/2
Automation in Construction	3/2
Bauphysik	3/2
Frontiers in Energy	3/2
International Journal of Energy Research	3/2
International Journal of Ventilation	3/2
Journal of Environmental Engineering	3/2
Journal of Harbin Institute of Technology New Series	3/2
Science and Technology for the Built Environment	3/2
Strategic Planning for Energy and the Environment	3/2
Sustainability Switzerland	3/2
Taiyangneng Xuebao Acta Energiae Solaris Sinica	3/2
Wit Transactions on Ecology and the Environment	3/2
ASHRAE Journal	2/3
ENR Engineering News Record	2/3
Energy Educ Science and Technology Part A	2/3
Environment and Planning B Planning and Design	2/3
International Journal for Housing Science and Its Applications	2/3
International Journal of Ambient Energy	2/3
International Journal of Energy Sector Management	2/3
International Journal of Heat and Technology	2/3
International Journal of Sustainable Built Environment	2/3

Journals	Number of pub/zone
Journal of Asian Architecture and Building Engineering	2/3
Journal of Building Engineering	2/3
Journal of Central South University of Technology English Edition	2/3
Journal of Chemical and Pharmaceutical Research	2/3
Journal of Civil Engineering and Management	2/3
Journal of Cleaner Production	2/3
Journal of Computing in Civil Engineering	2/3
Journal of Environmental Engineering Japan	2/3
Journal of Green Building	2/3
Journal of Performance of Constructed Facilities	2/3
Journal of Process Control	2/3
Plos One	2/3
Tongji Daxue Xuebao Journal of Tongji University	2/3
VTT Tiedotteita Valtion Teknillinen Tutkimuskeskus	2/3
Xuebao Journal of Xi'an University of Architecture and Technology	2/3
AEJ Alexandria Engineering Journal	1/3
ASHRAE Transactions	1/3
Acta Montanistica Slovaca	1/3
Advanced Science Letters	1/3
Advances in Mechanical Engineering	1/3
Aerosol Science and Technology	1/3
Aij Journal of Technology and Design	1/3
Applied Mathematics and Computation	1/3
Applied Soft Computing Journal	1/3
Arabian Journal for Science and Engineering	1/3
Architectural Engineering and Design Management	1/3
Archnet Ijar	1/3
Arpn Journal of Engineering and Applied Sciences	1/3
Building Research and Information	1/3
Building Services Engineering Research Technology	1/3
Buildings	1/3
Built Environment Project and Asset Management	1/3
Cailiao Gongcheng Journal of Materials Engineering	1/3
Chung Kuo Chi Hsueh Kung Ch Eng Hsueh Pao Journal of the Chinese Society of Mechanical Engineers	1/3
Communications in Nonlinear Science and Numerical Simulation	1/3
Construction and Building Materials	1/3
Construction Management and Economics	1/3
Consulting Specifying Engineer	1/3
Data in Brief	1/3
Electronic Journal of Structural Engineering	1/3
Elektrotehniski Vestnik Electrotechnical Review	1/3
Energy Economics	1/3
Energy Sources	1/3

Journals	Number of pub/zone
Energy Sustainability and Society	1/3
Energy Systems and Policy	1/3
Engineered Systems	1/3
Engineering Applications of Artificial Intelligence	1/3
European Transactions on Electrical Power	1/3
Facilities	1/3
Federal Register	1/3
Gaswaerme International	1/3
Gcb Bioenergy	1/3
Habitat International	1/3
Hkje Transactions Hong Kong Institution of Engineers	1/3
IEEE Design and Test of Computers	1/3
IEEE Transactions on Antennas and Propagation	1/3
IEEE Transactions on Automation Science and Engineering	1/3
IEEE Transactions on Consumer Electronics	1/3
IEEE Transactions on Industrial Electronics	1/3
IEEE Transactions on Industrial Informatics	1/3
IEEE Transactions on Smart Grid	1/3
Ieej Transactions on Electrical and Electronic Engineering	1/3
Indoor Air	1/3
Intelligent Buildings International	1/3
International Journal of Agricultural and Biological Engineering	1/3
International Journal of Construction Management	1/3
International Journal of Distributed Sensor Networks	1/3
International Journal of Environmental Research and Public Health	1/3
International Journal of Environmental Technology and Management	1/3
International Journal of Hospitality Management	1/3
International Journal of Metrology and Quality Engineering	1/3
International Journal of Sustainable Building Technology and Urban Development	1/3
International Journal of Technology	1/3
International Journal of Thermal Sciences	1/3
Journal of Algorithms and Computational Technology	1/3
Journal of Architectural and Planning Research	1/3
Journal of Architectural Engineering	1/3
Journal of Climate Applied Meteorology	1/3
Journal of Computational Information Systems	1/3
Journal of Corporate Real Estate	1/3
Journal of Cultural Heritage	1/3
Journal of Engineering Science and Technology Review	1/3
Journal of European Real Estate Research	1/3
Journal of Information Technology in Construction	1/3
Journal of Materials for Energy Systems	1/3
Journal of Real Estate Finance and Economics	1/3
Journal of Solid Waste Technology and Management	1/3

Journals	Number of pub/zone
Journal of Statistical Computation and Simulation	1/3
Journal of the Illuminating Engineering Society	1/3
Journal of Thermal Envelope and Building Science	1/3
Jurnal Teknologi Sciences and Engineering	1/3
Key Engineering Materials	1/3
Light and Engineering	1/3
Lighting Design and Application LD and A	1/3
Magazine of Civil Engineering	1/3
Optical Engineering	1/3
Pacific Rim Property Research Journal	1/3
Philosophical Transactions of the Royal Society A	1/3
Prace Naukowe Instytutu Gornictwa Politechniki Wroclawskiej	1/3
Proceedings of the Inst of Mechanical Engineers Part A	1/3
Revista Tecnica De La Facultad De Ingenieria Universidad Del Zulia	1/3
Rocznik Ochrona Srodowiska	1/3
Sensors	1/3
Sensors Switzerland	1/3
Journal of Shenzhen University Science and Engineering	1/3
Simulation Modelling Practice and Theory	1/3
Structural Survey	1/3
Studies in Computational Intelligence	1/3
Sustainable Cities and Society	1/3
Urban Climate	1/3
Wseas Transactions on Environment and Development	1/3
Xiandai Huagong Modern Chemical Industry	1/3
Journal of Central South University Science and Technology	1/3

Appendix 5

Energy and Buildings

Energy demands and consumption in existing and future buildings—prediction and validation;
 Indoor environment quality, including health and thermal comfort vis-à-vis energy;
 Natural, mechanical and mixed ventilation;
 Air distribution in buildings;
 Application of solar and other renewable energy sources in buildings;
 Energy balances in building complexes (residential, commercial, industrial, public and other buildings);
 Energy efficiency improvement measures of HVAC&R and other technical systems in residential, commercial, public and industrial buildings, and semi open built spaces;

Heat recovery systems in buildings;
Buildings and district heating and cooling;
Energy conservation in built environment;
Energy efficient buildings;
Building physics;
Energy sustainability, resilience and climate adaptability of buildings;
Evaluation and control of indoor thermal and lighting systems;
Building's total performance and intelligent buildings;
Links between architectural design, mechanical and lighting systems;
New materials in buildings and their impact on energy demands;
External and internal design conditions for energy efficient buildings;
Building envelope materials and structure energy performance;
Thermal energy storage and thermally active building systems—TABS;
Energy performance of buildings and modeling predictive control;
Zero CO₂ emission—zero energy and energy plus buildings and their smart grid harmonized operation;
Residential/municipal energy refurbishment and renovation;
Life cycle energy efficiency of buildings and embodied energy;
Architectural structure—construction energy efficiency;
Energy related aspects of buildings after catastrophic events.

Applied Energy

Energy conversion and conservation;
Optimal use of energy resources;
Analysis and optimization of energy processes;
Mitigation of environmental pollutants;
Sustainable energy systems;
Problems of modeling and forecasting;
Conservation strategies;
The environmental, social and economic impacts of energy policies and usage.

Energy

Energy analysis
Energy modelling and prediction
Integrated energy systems
Energy planning
Energy management
Energy conservation
Energy efficiency
Biomass and bioenergy
Renewable energy
Electricity supply and demand
Energy storage
Energy in buildings, and on economic and policy issues

Building and Environment

Technologies and integrated systems for high performance buildings and cities;
Thermal, acoustic, visual, air quality building science and human impacts;
Tools for the design and decision-making community, including tested computational, economic, educational and policy tools;

Solutions for mitigating environmental impacts and achieving low carbon, sustainable built environments.

Energy Conservation and Management

Energy generation;

Energy utilization;

Energy conversion;

Energy storage;

Energy transmission;

Energy conservation;

Energy Management;

Sustainability

References

- Addy, M. N., Adinyira, E., & Koranteng, C. (2014). Architect's perception on the challenges of building energy efficiency in Ghana. *Structural Survey*, 32(5), 365–376.
- Assadi, M. K., et al. (2016). Evaluating different scenarios for optimizing energy consumption to achieve sustainable green building in Malaysia. *ARPJ Journal of Engineering and Applied Sciences*, 11(20), 12112–12116.
- Azizi, N. S. M., Wilkinson, S., & Fassman, E. (2014). Management practice to achieve energy-efficient performance of green buildings in New Zealand. *Architectural Engineering and Design Management*, 10(1), 25–38.
- Barrios, M., et al. (2008). A bibliometric study of psychological research on tourism. *Scientometrics*, 77(3), 453–467.
- Buratti, C., Arbanera, M., & Palladino, D. (2014). An original tool for checking energy performance and certification of buildings by means of Artificial Neural Networks. *Applied Energy*, 120(1), 125–132.
- Cajias, M., & Piazolo, D. (2013). Green performs better? Energy efficiency and financial return on buildings. *Journal of Corporate Real Estate*, 15(1), 53–72.
- Capozzoli, A., et al. (2016). A novel methodology for energy performance benchmarking of buildings by means of Linear Mixed Effect Model: The case of space and DHW heating of out-patient Healthcare Centres. *Applied Energy*, 171(1), 592–607.
- Cetin, K. S., Manuel, L., & Novoselac, A. (2016). Effect of technology-enabled time-of-use energy pricing on thermal comfort and energy use in mechanically-conditioned residential buildings in cooling dominated climates. *Building and Environment*, 96(1), 118–130.
- Coile, R. C. (1977). Lotka's frequency distribution of scientific productivity. *Journal of American Society for Information Science*, 28(6), 366–370.
- Dabi, Y., et al. (2016). Publication trends in bariatric surgery: A bibliometric study. *Obesity Surgery*, 26(11), 2691–2699.
- Dascalaki, E. G., et al. (2016). Modeling energy refurbishment scenarios for the Hellenic residential building stock towards the 2020 & 2030 Targets. *Energy and Buildings*, 132, 74–90.
- Diodato, V. P. (2012). *Dictionary of bibliometrics*. Boca Raton: Taylor & Francis Group.
- Elsevier. (2017a). Energy and buildings—An International Journal devoted to investigations of energy use and efficiency buildings. <https://www.journals.elsevier.com/energy-and-buildings/>. Accessed 28 Sept. 2017.
- Elsevier. (2017b). Applied Energy. <https://www.journals.elsevier.com/applied-energy/>. Accessed 28 Sept. 2017.
- Elsevier. (2017c). Energy—The International Journal. <https://www.journals.elsevier.com/energy/>. Accessed 28 Sept. 2017.
- Elsevier (2017d). Building and Environmental—The International Journal of Building Science and its Application. <https://www.journals.elsevier.com/building-and-environment>. Accessed 28 Sept. 2017.
- Elsevier. (2017e). Energy Conservation and Management—An International Journal. <https://www.journals.elsevier.com/energy-conversion-and-management>. Accessed 28 Sept. 2017.
- Engvall, K., et al. (2014). Interaction between building design, management, household and individual factors in relation to energy use for space heating in apartment buildings. *Energy and Buildings*, 81, 457–465.

- Estiri, H. (2014). Building and household x-factors and energy consumption at the residential sector: A structural equation analysis of the effects of household and building characteristics on the annual energy consumption of US residential buildings. *Energy Economics*, 43, 178–184.
- European Council for an Energy Efficient Economy (ECEEE). (2010). The energy performance of buildings directive. https://umanitoba.ca/faculties/engineering/departments/ce2p2e/alternative_village/media/1_eceee_buildings_policybrief2010_rev.pdf. Accessed 5 April 2017.
- Hemsath, T. L., & Bandhossani, K. A. (2015). Sensitivity analysis evaluating basic building geometry's effect on energy use. *Renewable Energy*, 76(1), 526–538.
- International Energy Agency (IEA). (2015). Energy and climate change. <https://www.iea.org/publications/freepublications/publication/WEO2015SpecialReportonEnergyandClimateChange.pdf>. Accessed 25 June 2017.
- Kazanasmas, T., et al. (2014). On the relation between architectural considerations and heating energy performance of Turkish residential buildings in Izmir. *Energy and Buildings*, 72, 38–50.
- Kim, K., & Yu, J. (2016). A process to divide curved walls in IFC-BIM into segmented straight walls for building energy analysis. *Journal of Civil Engineering and Management*, 22(3), 333–345.
- Kontokosta, C. E. (2015). A Market-Specific Methodology for a commercial building energy performance Index. *Journal of Real Estate Finance and Economics*, 51(2), 288–316.
- Koumoutsos, K., et al. (2015). Gathering and processing energy consumption data from public educational buildings over IPv6. *Energy, Sustainability and Society*, 5(1), 1–12.
- Lam, J. C., & Hui, S. C. M. (1996). Sensitivity analysis of the energy performance of office building. *Building and Environmental*, 31(1), 27–39.
- Li, J., et al. (2015a). Bibliometric analysis of fracking scientific literature. *Scientometrics*, 105(2), 1273–1284.
- Li, M., et al. (2015b). Climate impacts on extreme energy consumption of different types of buildings. *PLoS ONE*, 10(4), 01–12.
- Liu, L., Lin, B., & Peng, B. (2015). Correlation analysis of building plane and energy consumption of high-rise office building in cold zone on China. *Building Simulation*, 8(5), 487–498.
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science*, 16(12), 317–323.
- Ma, J., & Cheng, J. C. P. (2016). Identifying the influential features on the regional energy use intensity on residential buildings based on Random Forests. *Applied Energy*, 183(1), 193–201.
- Magrini, A., & Franco, G. (2016). The energy performance improvement of historic buildings and their environmental sustainability assessment. *Journal of Cultural Heritage*, 21, 834–841.
- Mardookhy, M., et al. (2014). A study of energy efficiency in residential buildings in Knoxville, Tennessee. *Journal of Cleaner Production*, 85, 01–09.
- Melo, A. P., et al. (2016). A novel surrogate model to support building energy labeling system: A new approach to assess cooling energy demand in commercial buildings. *Energy and Buildings*, 131(1), 233–247.
- Nagy, Z., et al. (2015). Occupant centered lighting control for comfort and energy efficient building operation. *Energy and Buildings*, 94, 100–108.
- Nghana, B., & Tariky, F. (2016). Phase change material's (PCM) impacts on the energy performance and thermal comfort of buildings in a mild climate. *Building and Environment*, 99(1), 221–238.
- Nicholls, P. T. (1988). Price's Square Root Law: Empirical validity and relation to Lotka's Law. *Information Processing and Management*, 24(4), 469–477.
- Norbäck, D., Lampa, E., & Engvall, K. (2014). Asthma, allergy and eczema among adults in multifamily houses in Stockholm (3-HE Study)—Associations with buildings characteristics, home environment and energy use for heating. *PLoS ONE*, 9(12), 01–17.
- Orosa, J. A. (2012). New methodology to define probability of buildings energy consumption. *Energy Education Science and Technology Part A Energy Science and Research*, 28(2), 891–902.
- Palomo, J., Figueroa-Domecq, C., & Laguna, P. (2017). Women, peace and security state-of-art: A bibliometric analysis in social sciences based on SCOPUS database. *Scientometrics*, 113(1), 123–148.
- Pao, M. L. (1985). Lotka's Law: A testing procedures. *Information Processing and Management*, 21(4), 305–320.
- Pao, M. L. (1986). An empirical examination of Lotka's law. *Journal of the American Society for Information Science*, 37(1), 26–33.
- Pernigotto, G., & Gasparella, A. (2013). Extensive comparative analysis of building energy simulation codes: Heating and cooling energy needs and peak loads calculation in TRNSYS and EnergyPlus for southern Europe climates. *HVAC and R Research*, 19(5), 481–492.
- Phillips, J. F., Sheff, M., & Boyer, C. B. (2015). The astronomy of Africa's Health Systems Literature during the MDG era? Where are the systems clusters? *Global Health: Science and practice*, 3(3), 482–502.

- Price, D. J. S. (1963). *Little Science, big science* (301 p.). New York: Columbia University Press.
- Sepúlveda, J. (2016). Evaluation of research in the field of energy efficiency and MCA methods using publications databases. *International Journal of Environmental, 10*(2), 01–04.
- Soares, N., et al. (2017). Energy efficiency and thermal performance of lightweight steel-framed (LSF) construction: A review. *Renewable and Sustainable Energy Reviews, 78*, 194–209.
- Surmann, M., Brunauer, W., & Bienert, S. (2015). How does energy efficiency influence the Market Value of office buildings in Germany and does this effect increase over time? *Journal of European Real Estate Research, 8*(3), 243–266.
- Tsanas, A., & Xifara, A. (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings, 49*(1), 560–567.
- United Nations Environment Programme. (UNEP). Why buildings? <http://staging.unep.org/sbci/AboutSBCI/Background.asp>. Accessed 09 Dec 2016.
- Wang, X., Huang, C., & Zou, Z. (2016). The analysis of energy consumption and greenhouse gas emissions a large-scale commercial building in Shanghai, China. *Advances in Mechanical Engineering, 8*(2), 01–08.
- Wang, L., Mathew, P., & Pang, X. (2012). Uncertainties in energy consumption introduced by building operations and weather for a medium-size office building. *Energy and Buildings, 53*(1), 152–168.
- Yalcintas, M. (2008). Energy-saving predicting fro building equipment retrofits. *Energy and Building, 40*(12), 2111–2120.
- Zahraee, S. M., et al. (2014). Application of design experiments to evaluate the effectiveness of climate factors on energy saving in green residential buildings. *Jurnal Teknologi (Sciences and Engineering), 6*(5), 107–111.
- Zhang, Y., et al. (2015). Comparisons of inverse modeling approaches for predicting building energy performance. *Building and Environment, 86*, 177–190.
- Zhenxing, J., & Jing, W. Y. L. (2007). Design of energy efficiency supervision system for large-scale public buildings. *Journal HV&AC, 8*, 19–22.
- Zhou, Z., et al. (2016). Establishing energy consumption quota for residential buildings using regression analysis and energy simulation. *Journal of Engineering Science and Tehnology Review, 9*(6), 103–107.
- Zorita, A. L., et al. (2016). A statistical modeling approach to detect anomalies in energetic efficiency of buildings. *Energy and Buildings, 110*, 377–386.