5Ws of Green and Sustainable Software

Coral Calero, Javier Mancebo*, Félix García, María Ángeles Moraga, José Alberto García Berná,
José Luis Fernández-Alemán, and Ambrosio Toval

Abstract: Green and Sustainable Software has emerged as a new and highly active area in the software community. After several years of research and work, we believe that it is now necessary to obtain a general snapshot of how the research in this area is evolving. To do so, we have applied the 5Ws (why, when, who, where, and what), a formula for getting the complete story on a subject. We have therefore carried out a study, using 542 publications related to Green and Sustainable Software research; these were recovered using SCOPUS. The results obtained allow us to conclude that it is important to identify key elements of the research to allow researchers be fully aware of the state of the research on Green and Sustainable Software *(why)*; the study uses papers published between 2000 and the beginning of November 2018 *(when)*; the most prolific authors are mainly from Europe, although the USA is the most active country, Green and Sustainable Software being a very interactive area with a good number of multinational publications *(who)*; the top five keywords related to sustainable aspects are Green Software, Green IT, Software Sustainability, Energy Consumption, and Energy Efficiency *(what)*; finally, as regards the places authors prefer to publish in, there is almost a complete balance between conferences and journals, with a trend towards an increase in the number of publications *(where)*.

Key words: Green Software; Sustainable Software; Software Engineering; Software Sustainability; Energy Efficiency; Energy Utilization

1 Introduction

Sustainability is gaining importance in every aspect of life, including that of technology. However, whereas hardware has been improved constantly so as to be energy efficient, concerns about the efficiency of software have appeared only more recently.

Software sustainability needs to be applied in several areas: software systems, software products, web applications, data centres, etc.[1] The way to achieve sustainable software is principally by improving its power consumption.

Software life cycle processes require three kinds of resources: human resources, economic resources, and energy resources, and these allow the three dimensions of software sustainability to be defined[2]: Human Sustainability (how software development and maintenance affect the sociological and psychological aspects of the software development community and its individuals, encompassing topics such as labour rights, psychological health, social support, social equity, and liveability); Economic Sustainability (how the software lifecycle processes protect stakeholders' investments, ensure profits, reduce risks, and maintain assets), and

- Coral Calero, Javier Mancebo, Félix García, and María Ángeles Moraga are with the Institute of Technology and Information Systems, University of Castilla-La Mancha, Ciudad Real 13071, Spain. E-mail: Javier.Mancebo@uclm.es; Coral.Calero@uclm.es; Felix.Garcia@uclm.es; MariaAngeles.Moraga@uclm.es.
- José Alberto García Berná, José Luis Fernández-Alemán, and Ambrosio Toval are with the Department of Informatics and Systems, University of Murcia, Murcia 30003, Spain. E-mail: JoseAlberto.Garcia1@um.es; Aleman@um.es; AToval@um.es.
- *To whom correspondence should be addressed.

Environmental Sustainability (how software product development, maintenance, and use affect energy consumption and the usage of other resources). This is known as *Green Software* or *Software Greenability*.

The definitions of Green Software found in literature are rather diverse as regards concepts, however, and it is possible to find terms such as Green Software, Green Through Software, Green in Software, etc.[2] We must differentiate between Green BY (when IT is the tool used to support sustainability goals) and Green IN (when the term "Green" is related to the IT, software, or hardware themselves). In general, the definitions of Green Software tend to mix these two perspectives. As part of Green in Software, Green in Software Engineering aims to include green considerations in the activities that form part of Software Engineering.

In recent years, as stated previously, several researchers have started to work on all the aspects to do with Green and Sustainable Software that were mentioned in the previous paragraphs. As a sufficient number of years have now passed since the research on Green and Sustainable Software began, we believe that it is time to obtain an overview and perspective of how the topic has evolved, and then try to determine what the next steps could be. A helpful starting point for identifying the key elements of a research story can be the 5Ws[3]:

• **Why:** Why did this research happen? Why was there a need for it?

• **When:** When did this study take place, when did the project start, and when did it finish?

• **Who:** Who did the research?

• **Where:** Where were the results published?

• **What:** What were the results of this research?

The 5Ws have the potential to allow researchers to explore key topics. They may also help translate research to a wide audience without much effort, and allow the researcher to retain control of what they say and how they say it, unlike in some traditional media reporting[3]. To answer the 5Ws, we need to determine the key aspects of the Green and Sustainable Software area from a quantitative point of view. To that end, we have performed a bibliometric analysis of the publications on Green and Sustainable Software. Bibliometrics is a statistical analysis that is used in the case of written publications[4].

Some related work dealing with bibliometric studies on software is presented in Section 2. Section 3 provides the main information as regards the study that was undertaken, along with the answers to the 5Ws of Green and Sustainable Software. Section 4 sets out the limitations of the work, while Section 5 presents the sensibility analysis performed to discover the correctness of the results obtained. Section 6 contains the discussion of the bibliometric study that was carried out, together with the conclusion obtained from it.

2 Related Work

There are several works that conduct a Systematic Literature Review (SLR) on Green and Sustainable Software topics[5–7]. However, the goal of an SLR is to identify, evaluate, and interpret all relevant research to answer a particular research question[8]. That is different from the objective proposed in this work, which is to obtain a perspective on the evolution of Green and Sustainable Software. So, for our work we are going to conduct a bibliometrics study.

As indicated in Güneş et al.[9], bibliometrics can be used to obtain statistical results regarding research performance, contributions of countries to international research, the status of journals, etc. We therefore consider that it is a good analysis technique to use in answering the 5Ws on Green and Sustainable Software.

Several pieces of work related to the investigation of research performance using bibliometrics may be found; the two main kinds of bibliometric indicators used to measure research performance in literature are number of publications and citation count[10]. Güneş et al.[9] presented examples of the use of bibliometric measurements of research performance in several areas, such as Abramo et al.[11] or Anninos[12] in the area of higher education, Davarpanah[13] in that of social sciences publications, or Pendlebury[14] in that of research performance evaluation.

Kumar et al.[15] conducted a bibliometric and scientific publication mining-based study focused on software and software engineering, in an effort to discover how the Asia-Pacific Software Engineering Conference (APSEC) evolved from 2010 to 2016. We mention only this work, although there are other similar studies concerning other conferences. Examples that deal with journals can also be found, such as the work of Vijayanathan and Kaliyamoorthi[16] in which the articles published in the open software engineering journal are examined to know the pattern of authorship or geographical distribution of the works, or the paper by Merigó et al.[17] dealing with *International Journal of Intelligent Systems*.

Garousi and Ruhe[18] presented a bibliometric study of software engineering research from 1969 to 2009. Fenandes in Ref. [19] looked at the perspective of authorship in software engineering. In 2016, Garousi and Mäntlylä[20] carried out a bibliometric study of citations, research topics, and countries that are active in software engineering. This paper also includes a list of existing bibliometric studies regarding Software Engineering (SE). One of these is the contribution of Cai and Card[1], which attempts to identify the main topics on software engineering. Neither Green Software nor Sustainable Software appears as a topic in this study.

The following authors also conducted bibliometric studies: Tavares et al.[21] on risk management in scrum projects, Blanco-Mesa et al.[22] on fuzzy decision making research; Koumaditis and Hussain[23] on human computer interaction research, and Heradio et al.[24] on software product lines. Garousi et al.[25], for their part, reviewed UML-driven software performance engineering.

In the specific case of green aspects, de Souza and Borsato[26] employed a bibliometric approach to review sustainable product development and its interface with economic and customer perception issues. They found that there is a growing number of articles being published, year by year, and provided a list of the journals with the largest amount of publications; they also pointed to new trends to be explored in product development.

To the best of our knowledge, there are no bibliometric studies on Green Software research activity, nor any work that establishes the characteristics of this research area. Furthermore, as Garousi and Mäntlylä[20] noted, this kind of studies is needed regularly in the quest to keep up with the most recent research developments. This being so, we decided to carry out a bibliometric study on Green and Sustainable Software, seeking to obtain the information needed to answer the 5Ws.

3   Answering the 5Ws on Green and Sustainable Software

In this section, we present the answers to each one of 5Ws on Green and Sustainable Software. To achieve the ultimate objective of answering these 5Ws, we developed our own specific research questions, each one of them corresponding to the appropriate W.

- **Why**
– RQ1. Why is the field of research relevant?
- **When**
– RQ2. When did this study take place?
– RQ3. What are the general descriptive statistics related to the data set of the study?
- **Who**
– RQ4. Who are the main contributors in this area?
– RQ5. What are the statistics related to authorship?
- **Where**
– RQ6. Which journals are the most effective (in terms of number of publications) as regards Green and Sustainable Software?
– RQ7. Which conferences are the most effective (in terms of number of publications) as regards Green and Sustainable Software?
- **What**
– RQ8. What are the most common keywords in the field of Green and Sustainable Software?
– RQ9. What are the most relevant domains?
– RQ10. In which Software Engineering Body Of Knowledge (SWEBOK) areas have most research efforts been undertaken?

3.1   Why

RQ1. Why is the field of research relevant?

As enough time has passed since the beginning of research on Green and Sustainable Software, the goal of this study is to conduct a bibliometric assessment of Green and Sustainable Software research, in order to determine the key features of the research literature in this field.

3.2   When

RQ2. When did this study take place?

The search took place at the beginning of November 2018, and attained a total of 542 papers. The dataset used in the study was obtained from the computer science category of SCOPUS between 2000 and 2018, for work written in English. We recovered the following information for each paper: authors, year, title, source, type of source, country, keywords，and number of citations. The search string used is shown in Fig. 1.

The graphical and tabular information obtained from the data recovered was processed using VOSviewer (http://www.vosviewer.com/), Tagul (www.wordart. com), and Microsoft Excel.

```
( TITLE-ABS-KEY ( "green* software" )  OR
TITLE-ABS-KEY ( "software green*" )  OR
TITLE-ABS-KEY ( "green in software" )  OR
TITLE-ABS-KEY ( "green by software" )  OR
TITLE-ABS-KEY ( "green through software" )  OR
TITLE-ABS-KEY ( "green requirement* engineer*" )  OR
TITLE-ABS-KEY ( "engineer* green requirement*" )  OR
TITLE-ABS-KEY ( "green* software develop*" )  OR
TITLE-ABS-KEY ( "software develop* green*" )  OR
TITLE-ABS-KEY ( "software energy*" )  OR
TITLE-ABS-KEY ( "sustain* software" )  OR
TITLE-ABS-KEY ( "software sustain*" )  OR
TITLE-ABS-KEY ( "sustain* of software" )  OR
TITLE-ABS-KEY ( "sustain* in software" )  OR
TITLE-ABS-KEY ( "software for sustain*" )  OR
TITLE-ABS-KEY ( "sustain* for software" )  OR
TITLE-ABS-KEY ( "Software Engineering for Sustain*" )  OR
TITLE-ABS-KEY ( "sustain* in require* engineer*" )  OR
TITLE-ABS-KEY ( "sustain* with software" )  OR
TITLE-ABS-KEY ( "sustain* non-function* require*" )  OR
TITLE-ABS-KEY ( "sustain* requirements" )  OR
TITLE-ABS-KEY ( "sustainability indicator software development"))
```

Fig. 1　Search string used.

RQ3. What are the general descriptive statistics related to the data set of the study?

In order to make the study replicable, we will set out the main information contained in it.

Table 1 shows the forums in which papers are most frequently published. The majority of publications are conference papers (62.2%) followed by journal articles (21.4%). The remaining forums cover only 16.4% of the total amount.

The growth in the number of publications is shown in Table 2. As will be noted, the number of papers published has increased considerably since 2011, from 81 publications until 2010 (an average of 7 per year), to 461 between 2011 and 2018 (an average of 57 per year, if 2018 is considered to be a complete year). This confirms what is indicated in Ref. [27], Calero identifies 2011 as the point at which Green in Software Engineering in particular, and Green in Software in general, began to be dealt with as research topics, with the publication of the GREENSOFT model by

Table 1　Main forums used in the area.

No.	Form of publication	Publication	Percentage (%)
1	Conference paper	337	62.2
2	Article	116	21.4
3	Conference review	41	7.6
4	Book chapter	21	3.9
5	Review	14	2.6
6	Article in press	7	1.3
7	Book	2	0.4
8	Editorial	2	0.4
9	Note	1	0.2
10	Short surey	1	0.2
	Total	542	100

Table 2　Growth in publications.

Year	Publication	Percentage (%)
2018	69	12.7
2017	72	13.3
2016	67	12.4
2015	72	13.3
2014	67	12.4
2013	48	8.9
2012	37	6.8
2011	29	5.4
2010	10	1.8
2009	15	2.8
2008	9	1.7
2007	11	2.0
2006	6	1.1
2005	4	0.7
2004	1	0.2
2003	7	1.3
2002	13	2.4
2001	4	0.7
2000	1	0.2
Total	542	100

Naumann et al.[28]

Although it is true that when we carried out the search, the number of contributions in 2018 was less than in 2017, we consider that by the end of the year (almost two months from the date of writing, when all the information will be available) these figures will change and the increase tendency in the number of published works will be confirmed.

The Annual Growth Rate (AGR) defines the total number of publications achieved in comparison to a previous year. This factor is calculated using the number of publications in one year and the number of publications from the previous year (Eq. (1)).

$$\text{AGR} = 100 \times \left(\frac{\text{N. publ}_{\text{year}} - \text{N. publ}_{\text{year}-1}}{\text{N. publ}_{\text{year}-1}} \right) \quad (1)$$

Table 3 shows the AGR of our data, from 2011 (in which the number of papers increases) to 2017 (we have

Table 3　Annual growth rate of publications.

Year	Publication	AGR
2017	72	7.5
2016	67	−6.9
2015	72	7.5
2014	67	39.6
2013	48	29.7
2012	37	27.6
2011	29	190
2010	10	

removed 2018, because the year had not finished when the search was conducted). As can be seen, the positive values are maintained, except in 2016, with a slight decrease (less than 7%).

The Compound Annual Growth Rate (CAGR) provides a comparison of the annual growth rate between different periods of time. This parameter is obtained by considering the number of publications produced in a year, the cumulative number of publications from a year, the year of reference, and the number of years (see Eq. (2)).

$$ \text{CAGR} = 100 \times \left[\left(\frac{\text{Cum. publ}_{\text{year}}}{\text{N. publ}_{\text{year}-1}} \right)^{\frac{1}{\text{year}-\text{ref. year}}} - 1 \right] \tag{2} $$

Table 4 shows the CAGR for our data. 2010 has been used as the year of reference because, as explained previously, the number of contributions increased significantly after that year. As can be observed, the CAGR is always positive.

The Relative Growth Rate (RGR) shows literature hikes related to the number of publications per unit of time. This factor is determined by the Napierian logarithm (Ln) of the publications in a year (W2), the Napierian logarithm of the publications in a year of reference (W1), and the number of years (Eq. (3)).

$$ \text{RGR} = \frac{\text{Ln(N. publ}_{\text{year}}) - \text{Ln(N. publ}_{\text{ref. year}})}{\text{year} - \text{ref. year}} \tag{3} $$

This parameter also allows us to obtain the Doubling Time (DT) factor, which expresses the time that is required to attain double the number of publications at the moment being studied (Eq. (4)).

$$ \text{DT} = \frac{\text{Ln2}}{\text{RGR}} \tag{4} $$

Table 5 shows how the DT has increased each year. At the beginning of the decade, the parameter had a value of around 0.5, whilst in recent years it has increased to over 4.

Table 4   CAGR.

Year	Publication	Cumulative	CAGR
2018	69	471	27.14
2017	72	402	27.85
2016	67	330	30.44
2015	72	263	29.58
2014	67	191	29.94
2013	48	124	37.21
2012	37	76	43.32
2011	29	39	34.48
2010	10	10	

Table 5   RGR & DT.

Year	W1	W2	RGR	DT
2018	6.00	6.15	0.16	4.38
2017	5.80	6.00	0.20	3.51
2016	5.57	5.80	0.23	3.05
2015	5.25	5.57	0.32	2.17
2014	4.82	5.25	0.43	1.60
2013	4.33	4.82	0.49	1.42
2012	3.66	4.33	0.67	1.04
2011	2.30	3.66	1.36	0.51
2010		2.30		

The least squares method was used as a basis to perform a trend analysis in our effort to estimate the number of publications that may appear in the future. A straight line was calculated using the data for the last decade (from 2008 to 2017), resulting in Eq. (5):

$$ Y = \frac{1366}{330}X + \frac{365}{10} \tag{5} $$

where Y is the estimated number of publications each year, and X is an input of the equation selected as appropriate. In order to make it easier to attain the coefficients, X values of the period are selected in such a way as to obtain zero when all their values are added up. Table 6 shows the publication trend calculated. As can be seen, the estimation indicates that 8 additional publications will be produced each year in comparison to the previous year.

3.3   Who

In this section, we present the different figures related to the authorship of the topic. We will thus answer the

Table 6   Computation of straight line trend using the least squares method.

Year	Y	X	$X \times Y$	X^2	Publication trend
2023		21			123
2022		19			115
2021		17			107
2020		15			99
2019		13			90
2018	69	11			82
2017	72	9	648	81	
2016	67	7	469	49	
2015	72	5	360	25	
2014	67	3	201	9	
2013	48	1	48	1	
2012	37	−1	−37	1	
2011	29	−3	−87	9	
2010	10	−5	−50	25	
2009	15	−7	−105	49	
2008	9	−9	−81	81	

RQ4 and RQ5 that have been proposed.

RQ4. Who are the main contributors in this area?

Table 7 shows the most prolific authors as regards Green and Sustainable Software, and Fig. 2 provides a graphical representation of their interaction. The figure shows the different groups of authors, along with the name of the most prolific author in each group. From this information, we can observe that there are four clusters: Cluster 1: Lago (with Procaccianti), Cluster 2: Penzenstadler (with Betz, Duboc, Richardson, and Venters), Cluster 3: Calero (with Piattini and Moraga), Cluster 4: Kern (with Johann and Naumann), and Cluster 5: Hindle. It can also be observed that there are the following relationships: between Cluster 1 and Clusters 2 and 3; between Cluster 2 and Clusters 1 and 3; between Cluster 3 and Clusters 1, 2, and 3; and between Cluster 4 and Cluster 3. We can thus conclude that Cluster 2 is the one with most researchers,

Cluster 3 is the most interrelated, and Cluster 5 has no relationships with the others.

As regards the countries represented, Fig. 3 shows the distribution of papers, together with the specific number of publications for the top ten countries. As may be seen, the USA is the most prolific country, followed by Germany, with a third of the quantity of contributions of the former. Finally, if we analyze the most prolific institutions, we obtain the data shown in Table 8.

When the information in Table 8 is compared with the data about distribution of papers provided in the paragraph above, it would appear that, although that there are several institutions in the USA publishing papers on Green and Sustainable Software, they are dispersed; this dispersion amongst different institutions is smaller in the remaining countries.

RQ5. What are the statistics related to authorship?

Once we know who the most prolific authors are, we can also calculate some other figures related to Green and Sustainable Software authorship. Table 9 shows the degree of collaboration among authors. As can be seen,

Table 7 Most prolific authors.

Author	Number of papers	Author	Number of papers
P. Lago	25	M. Piattini	8
B. Penzenstadler	25	C. C. Venters	8
A. Hindle	14	S. Betz	7
G. Procaccianti	14	S. Naumann	7
C. Calero	13	M. A. Moraga	7
L. Duboc	8	T. Johann	7
E. Kern	8	D. Richardson	7

Fig. 2 Most interactive authors.

Table 8 Most prolific institutions in sustainable software.

No.	Institution	Number of publications	Percentage (%)
1	Vrije Universiteit Amsterdam	29	5.35
2	Universidad de Castilla-La Mancha	20	3.69
3	University of Alberta	14	2.58
4	University of California, Irvine	13	2.40
5	Politecnico di Torino	10	1.85
6	University of Leicester	10	1.85
7	Universidad de Malaga	9	1.66
8	Technical University of Munich	9	1.66
9	California State University Long Beach	9	1.66
10	University of Huddersfield	8	1.48

Country	Number of publications	Country	Number of publications
United States	131	Netherlands	39
Germany	48	Italy	38
United Kingdom	44	China	29
Spain	42	Brazil	24
Canada	41	India	21

Fig. 3 Distribution of contributions by countries.

Table 9   Number of publications and number of authors.

Number of authors	Number of publications	Number of authors	Number of publications
Anonymous	38	7	10
1	55	8	10
2	128	9	3
3	137	10	2
4	97	15	1
5	40	16	1
6	20		

more than a half of the publications have 2, 3, or 4 authors.

This information can be used to calculate Author Participation Productivity (APP), a calculation which is done by employing the mean of the author participants per paper (Eq. (6)).

$$APP = \frac{N.\ authors}{N.\ papers} \qquad (6)$$

Moreover, the inverse value provides the mean value of Productivity Per Author (PPA) (Eq. (7)):

$$PPA = \frac{1}{APP} \qquad (7)$$

From 2000 to 2006, the APP has extreme values (greater than 3, or less than 2). The APP remains at around 2.5 in the period between 2007 and 2018 (with the exception of 2017). However, the PPA appears to follow a pattern of growth from 2011, when there is a large increase in the number of contributions (once more, except for 2017). A similar pattern as regards the number of authors with an increasing value of PPA could therefore be interpreted as the achievement of a sufficient level of maturity in the area. It is worth emphasizing that results for 2018 must be considered with precaution for this interpretation, because the year had not yet finished at the time of the search (see Table 10).

We can also calculate the Collaboration Index (CI) of authors by applying the formula shown in Eq. (8).

$$CI = \frac{N.\ authors\ in\ multi\text{-}authored\ publication}{N.\ multi\text{-}authored\ publications} \qquad (8)$$

Table 11 displays the collaboration index between the authors with regard to Green and Sustainable Software. In most years, the CI is over 3, and there are even three entries that are greater than, or equal to, 4 (2003, 2006, and 2017). Only one year (2007) has a CI that is less than 3 (2004 also has a CI that is less than 3, but that is because in that year there was only one publication, with one single author). The table reflects a high collaborative pattern in the area.

Table 10   Author productivity.

Year	Number of papers	Number of authors	APP	PPA
2018	69	201	2.91	0.34
2017	72	225	3.13	0.32
2016	67	159	2.37	0.42
2015	72	158	2.19	0.46
2014	67	158	2.36	0.42
2013	48	118	2.46	0.41
2012	37	95	2.57	0.39
2011	29	81	2.79	0.36
2010	10	26	2.60	0.38
2009	15	38	2.53	0.39
2008	9	24	2.67	0.38
2007	11	28	2.55	0.39
2006	6	20	3.33	0.30
2005	4	12	3.00	0.33
2004	1	1	1.00	1.00
2003	7	20	2.86	0.35
2002	13	25	1.92	0.52
2001	4	13	3.25	0.31
2000	1	3	3.00	0.33

Table 11   Collaboration index.

Year	Multi-authored publication	Number of authors	CI
2018	63	239	3.79
2017	61	246	4.03
2016	56	200	3.57
2015	59	198	3.36
2014	55	193	3.51
2013	36	122	3.39
2012	31	104	3.35
2011	27	89	3.30
2010	7	23	3.29
2009	10	39	3.90
2008	7	22	3.14
2007	11	30	2.73
2006	4	18	4.50
2005	3	11	3.67
2004	0	0	0
2003	5	20	4
2002	8	27	3.38
2001	4	13	3.25
2000	1	3	3

Table 12 shows the percentages of single-authored publications and multi-authored publications by years. Anonymous publications (38) have not been taken into account in this table. The vast majority of publications are multi-authored, with values of between 62% and 100%. The single-authored publications account for only 11% of the total. As already mentioned, the

Table 12   Single and multi-authored publications.

Year	Single-authored publication	Percentage (%)	Multi-authored publication	Percentage (%)
2018	4	6	63	91
2017	5	7	61	85
2016	7	10	56	84
2015	8	11	59	82
2014	7	10	55	82
2013	6	13	36	75
2012	3	8	31	84
2011	0	0	27	93
2010	3	30	7	70
2009	4	27	10	67
2008	2	22	7	78
2007	0	0	11	100
2006	2	33	4	67
2005	1	25	3	75
2004	1	100	0	0
2003	0	0	5	71
2002	3	23	8	62
2001	0	0	4	100
2000	0	0	1	100
Total	56	11	448	89

number of publications has increased considerably since 2011, and from that year on, the number of multi-authored publications follows a pattern of ongoing increase. This is another piece of data that supports the idea of the high degree of collaboration in the area.

Table 13 shows the number of multinational papers (those written by authors from more than one country). As can be seen, more than half are contributions written by authors from the same country, while around 15% are multinational contributions.

3.4   Where

In this section, we will present the result of the analysis from the viewpoint of publication forums. Table 14 shows the top ten publication forums. As can be observed, 60% are journals (highlighted in grey) and 40% are conferences, signifying that there is a balance between both kinds of publications. This pattern is

Table 13   Multinational collaboration.

Number of countries	Number of papers
1	333
2	102
3	11
4	7
5	1
7	4
Unknown	84

Table 14   Top ten publication forums.

Forum	Number of publications
CEUR Workshop Proceedings	44
Proceedings International Conference on Software Engineering	23
ACM International Conference Proceeding Series	20
Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics (LNCS, LNAI, LNBI)	20
IEEE Software	11
Crosstalk	8
Advances in Intelligent Systems and Computing	7
Communications in Computer and Information Science	7
Information and Software Technology	7
IT Professional	5

almost the opposite when comparing the number of papers published in the top 10 forums, with 70% of papers being published at conferences and 30% in journals.

The CEUR Workshop Proceedings occupies first place in this top ten, with twice the number of contributions of the second on the list. This could be explained by the existence of specific workshops related to Green and Sustainable Software that appeared when the first efforts related to these topics began to be undertaken. Workshops are normally used to present first ideas and emerging results, and seem to be the logical way to start disseminating a new topic. With the passage of time, the work has attained a sufficient level of maturity for it to be published in other forums, such as in the main conferences and journals; these provide an outlet for more formal and complete results.

RQ6.   Which journals are the most effective (in terms of number of publications) as regards Green and Sustainable Software?

Table 15 shows the top ten journals, meaning those with most publications on Green and Sustainable Software in the period being studied. As is evident in the table, the first place is occupied by *IEEE Software*, a specific journal on software. We consider that this is an interesting result which reflects the fact that Green and Sustainable Software is an issue that is considered as relevant by the software community.

The first specific journal on sustainability aspects is in third place: "*Sustainable Computing-Informatics & Systems* (SUSCOM)". SUSCOM is gaining impact

Table 15   Most effective journals.

Forum	Number of publications
IEEE Software	11
Crosstalk	8
Sustainable Computing Informatics and Systems	7
Communications in Computer and Information Science	7
Advances in Intelligent Systems and Computing	7
Information and Software Technology	7
Journal of Systems and Software	7
Journal of Software Evolution and Process	7
IT Professional	5
Empirical Software Engineering	4
Total	70

each year, and is currently in the second quartile in the JCR series. It is worth emphasizing that SUSCOM first appeared in 2011, and is therefore a relatively new journal. Figure 4 displays the evolution of the number of papers published in journals from 2000 to 2018. From this figure we may observe a trend towards increase, which can in turn be interpreted as an increase in the maturity level in relation to Green and Sustainable Software research.

If we focus on the evolution of the number of journal papers published by the most prolific authors shown in Table 7, we obtain the results shown in Fig. 5, which confirm the trend towards the maturity of the research.

RQ7. Which conferences are the most effective (in terms of number of publications) as regards Green and Sustainable Software?

Figure 6 shows the evolution of the number of papers that have been presented in conferences and workshops since 2008.
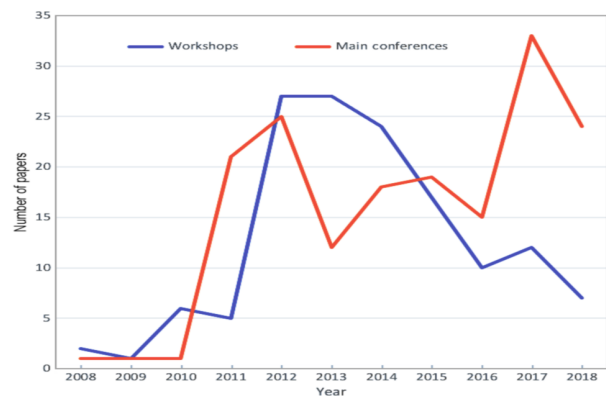
Before 2011, when this was an incipient topic, most of the publications were in workshops. From 2011, the tendency started to change, and currently there are more
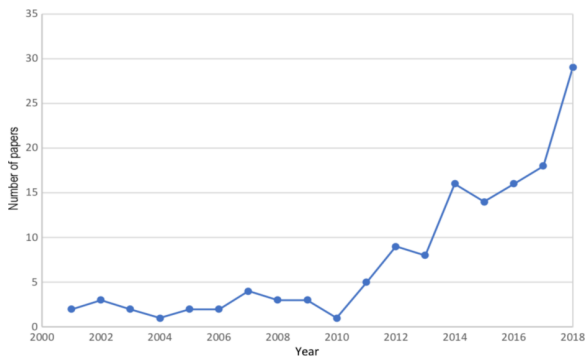
Fig. 4   Number of papers published in journal by years.
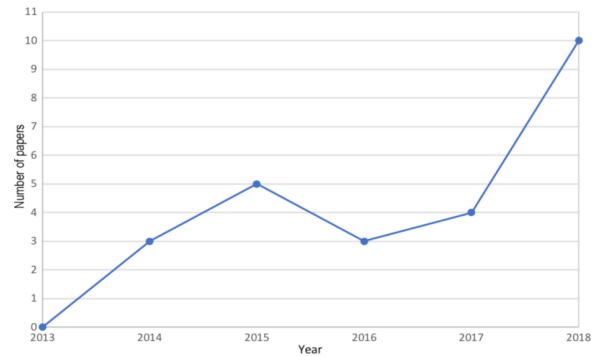
Fig. 5   Number of papers of the most prolific authors published in journal by years.

Fig. 6   Number of papers presented in workshops or main conferences by years.

publications in conferences.

Apart from CEUR Workshop Proceedings (see Table 16), which occupies the first place in the top ten

Table 16   Ten most effective conferences.

Forum	Number of publications
CEUR Workshop Proceedings	44
International Conference on Software Engineering	23
ACM International Conference Proceeding Series	20
Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics	20
Proceedings of the ACM International Conference on Digital Libraries	4
Lecture Notes in Business Information Processing	3
Proceedings of the ACM Symposium on Applied Computing	3
Autotestcon Proceedings	2
IEEE International Conference on Software Maintenance	2
Proceedings of the IEEE International Conference on VLSI Design	2
Total	123

conferences on Green and Sustainable Software, the first specific conference on software is in second place: the "International Conference on Software Engineering (ICSE)". Moreover, it is the first forum on the list that represents an event in itself (as opposed to CEUR workshops or LNCS, LNAI, LNBI, which publish work from several events). ICSE is the best and most important conference on Software Engineering, and we therefore believe that the fact that it publishes a large number of papers on Green and Sustainable Software is proof of the importance that these topics are acquiring in the community.

Another aspect that can be highlighted from the list of top conferences is the great amount of different perspectives that they afford: digital libraries, business processes, applied computing, or Very Large Scale Integration (VLSI) design. This variety can be explained by the search string used, in which we included Green IN, Green BY, and different keywords related to software.

3.5 What

RQ8. What are the most common keywords in the field of Green and Sustainable Software?

This section shows the keywords used most frequently in the papers selected. Table 17 shows those keywords related to Green and Sustainable Software which occurred more than 20 times in the selected papers recovered using VOSViewer. Table 18 shows the ten keywords related to Green and Sustainability, based on the keywords provided by the authors. Figures 7 and 8 present the same information by means of a keyword cloud. The common keywords in both sets, which are those used most frequently, are Green Software, Green IT, Sustainability, Software Sustainability, Sustainable Software, Energy Consumption, and Energy Efficiency.

Table 17 Most common Green and Sustainable Software keywords (provided by VOSViewer).

VOSViewer Green and Sustainability keyword	Ocurrence
Sustainable Development	167
Energy Utilization	137
Energy Efficiency	112
Sustainability	68
Sustainable Software	66
Green Software	43
Software Energy Consumption	33
Software Sustainability	30
Green IT	26
Energy Efficient	22

Table 18 Most common Green and Sustainable Software keywords (provided by authors).

Authors' Green and Sustainability keyword	Ocurrence
Sustainability	70
Green Software	49
Energy Efficiency	43
Software Sustainability	30
Energy Consumption	30
Sustainable Software	23
Green	14
Software Energy Consumption	14
Green IT	14
Energy	13

Fig. 7 VOSViewer Green and Sustainable Software keywords cloud.

Fig. 8 Authors' Green and Sustainable Software keywords cloud.

RQ9. What are the most relevant domains?

We have used the information provided by VOSViewer and by authors (by means of the keywords) to discover the most relevant domains in which Green and Sustainable Software are applied. Table 19 sets out the results regarding the domains obtained from the VOSViewer keywords and Table 20 shows the results obtained from the authors' keywords.

Both tables enable us to observe that the coincidences are Sustainable Development, Computer Software, Requirements Engineering, Embedded Systems / Embedded Software, and Hardware.

Table 19 Most common Green and Sustainable Software domains (obtained from the VOSViewer keywords).

VOSViewer topic keyword	Ocurrence
Sustainable Development	165
Software Engineering	155
Computer Software	101
Software Design	57
Application Programs	46
Software Systems	42
Requirements Engineering	42
Embedded Systems	31
Software Testing	29
Hardware	25

Table 20 Most common Green and Sustainable Software domains (obtained from the authors' keywords).

Authors' topic keyword	Ocurrence
Sustainable Development	105
Requirements Engineering	32
Computer Software	25
Hardware	23
Embedded Software	15
Economic	13
Algorithm	13
Software Product	10
Big Data	9
Mobile Application	7

RQ10. In which SWEBOK areas more research efforts have been under taken?

In order to have another perspective about the papers on Green and Software Sustainability, we have classified the VOSViewer domains shown in Table 19 into SWEBOK (IEEE, 2014) areas. SWEBOK includes 15 areas related to software engineering; we have used neither the foundations areas nor the management ones, because they are too general to be useful for classification purposes.

Table 21 displays the classification. From this table,

Table 21 Classification of VOSViewer domains on SWEBOK.

SWEBOK area	Ocurrence
Requirements	42
Design	57
Construction	46
Testing	29
Maintenance	0
Process	0
Models and Methods	0
Quality	165
Professional Practice	0

we can conclude that around 49% of occurrences are domains related to one of the SWEBOK areas. If we add to this figure the 155 occurrences of the generic domain "Software Engineering", this percentage increases to 71%. Despite this good result, there are four SWEBOK areas without contributions: Maintenance, Process, Models and Methods, and Professional Practice.

4 Limitations of the Study

Although the study was performed in a methodological manner, there are some limitations associated with it.

• The study is limited to papers that can be accessible for reading. However, as we have included all kinds of sources (including open libraries), we believe that this effect, if present, has had a minimum impact on the results.

• We have considered only documents written in English. As this is the language used in software research, we do not believe that this aspect has had a remarkable impact on the study.

• The period chosen was from 2000 to November 2018. Although it might have been possible to find some works before this period, there would be few of these, because it is an area that has been expanding only more recently. In fact, of the topics identified by Ref. [1], Green/Sustainable Software was not then identified as a software engineering topic.

• We have used only SCOPUS. Although work just with it may make us to lose contributions, the set provided by SCOPUS is the most accessible for the general audience because it is the largest bibliometric database. So, we believe that this impact is not of great importance and the results obtained from SCOPUS are representative enough to be used in a bibliometric study as the one we have performed. In fact, SCOPUS is one of the most commonly-used sources for citation data in bibliometric analyses[29].

5 Sensibility Analysis

In order to discover whether the search string had been built correctly, we carried out a sensibility analysis of the publications obtained from it. This was done by randomly selecting some of the papers, and by studying whether they fitted into the topics searched for with our search string. To calculate how many papers it was necessary to select and check, we used Cochran's sample size formula (see Eq. (9))[30].

$$\frac{N \times Z^2 \times p \times (1-p)}{(N-1 \times e^2 + Z^2 \times p + (1-p)} \tag{9}$$

where N is the total number of publications obtained from the Scopus database using the search string (542), Z is the deviation of the mean value that is accepted for the level of confidence (we work with a level of confidence of 95%, which implies that $Z = 1.96$), e is the error margin (0.05 for a level of confidence of 95%), and p is the proportion of results that is expected to be invalid (we have fixed this as 8% of papers that do not fit the topics). By applying the formula, we obtained a value of 93. We have therefore randomly selected 93 papers from our total of 542, and have subsequently checked whether or not they fit the desired topics.

Of the 93 publications randomly selected, 83 (89%) were about Green and Sustainable Software; only 10 publications had no relationship with these topics. We therefore have around 11% of invalid results, as opposed to the 8% expected. We believe that this is a good value, implying that the search has returned the expected results.

6 Conclusion and Future Work

Several years have passed since the first research efforts dealing with Green and Sustainable Software began; we thus believe that it is time to get a snapshot of how this area has evolved. To do that, we decided to answer the 5Ws (why, when, who, where, and what) of Green and Sustainable Software that allow researchers to explore key topics. In our endeavour to answer the five questions, we have used the results of a bibliometric analysis of Green and Sustainable Software, based on a search carried out in SCOPUS, from 2000 to November 2018, which came up with 542 papers.

The main results obtained allow us to conclude that:

• Green and Sustainable Software is a highly active area of research;

• It is a truly interactive area with a good number of multinational publications;

• Some parts of the world are still not working on this topic. The USA, Europe, and Canada are very active;

• Although researching Green Hardware/Green IT was already a trend, Green and Sustainable Software has achieved a good level of maturity, and is now a stable line of research; and

• The most frequently-used keywords related to Green and Sustainable Software aspects are Sustainable Development, Green Software, Green IT, Software Sustainability, Energy Consumption, and Energy Efficiency. Around 71% of the VOSViewer keywords are related to an SWEBOK area, although there are four SWEBOK areas without contributions.

As future work we plan to use other digital libraries to extend this study, in an effort to consolidate the results obtained. We also intend to study other dimensions of Software Sustainability, including topics such as human or economic sustainability.

References

[1] K. Y. Cai and D. Card, An analysis of research topics in software engineering–2006, *J. Syst. Softw.*, vol. 81, no. 6, pp. 1051–1058, 2008.

[2] C. Calero and M. Piattini, Puzzling out software sustainability, *Sust. Comput.: Inform. Syst.*, vol. 16, pp. 117–124, 2017.

[3] A. Tattersall, Who, what, where, when, why: Using the 5 Ws to communicate your research, https://blogs.lse.ac.uk/impactofsocialsciences/2015/04/08/using-the-5-ws-to-communicate-your-research/, 2015.

[4] A. Pritchard, Statistical bibliography or bibliometrics, *J. Document.*, vol. 25, no. 4, pp. 348–349, 1969.

[5] H. Anwar and D. Pfahl, Towards greener software engineering using software analytics: A systematic mapping, in *Proc. 43rd Euromicro Conf. on Software Engineering and Advanced Applications*, Vienna, Austria, 2017, pp. 157–166.

[6] R. Verdecchia, F. Ricchiuti, A. Hankel, P. Lago, and G. Procaccianti, Green ICT research and challenges, in *Advances and New Trends in Environmental Informatics*: *Stability, Continuity, Innovation*, V. Wohlgemuth, F. Fuchs-Kittowski, and J. Wittmann, eds. Springer, 2017, pp. 37–48.

[7] V. Wohlgemuth, F. Fuchs-Kittowski, and J. Wittmann, *Advances and New Trends in Environmental Informatics*: *Stability, Continuity, Innovation*. Springer, 2017.

[8] B. Kitchenham and S. Charters, Guidelines for performing systematic literature reviews in software engineering, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.117.471&rep=rep1&type=pdf, 2007.

[9]   E. Güneş, M. T. Üstündağ, H. Yalçın, and M. Safran, Investigating educational research articles (1980–2014) in terms of bibliometric indicators, *Int. Online J. Educ. Sci.*, vol. 9, no. 1, pp. 101–117, 2017.

[10] A. Diem and S. C. Wolter, The use of bibliometrics to measure research performance in education sciences, *Res. Higher Educ.*, vol. 54, no. 1, pp. 86–114, 2013.

[11] G. Abramo, T. Cicero, and C. A. D'Angelo, The dispersion of research performance within and between universities as a potential indicator of the competitive intensity in higher education systems, *J. Inform.*, vol. 6, no. 2, pp. 155–168, 2012.

[12] L. N. Anninos, Research performance evaluation: Some critical thoughts on standard bibliometric indicators, *Stud. Higher Educ.*, vol. 39, no. 9, pp. 1542–1561, 2014.

[13] M. R. Davarpanah, The international publication productivity of Malaysian in social sciences: Developing a scientific power index, *J. Sch. Publish.*, vol. 41, no. 1, pp. 67–91, 2009.

[14] D. A. Pendlebury, Bibliometrics, research performance evaluation, and beyond: Towards actionable intelligence for science administrators, policymakers, and funders, https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=1002&context=etra, 2012.

[15] L. Kumar, S. Sripada, and A. Sureka, A review of six years of Asia-pacific software engineering conference, in *Proc. 23rd Asia-Pacific Software Engineering Conf.*, Hamilton, New Zealand, 2016, pp. 341–344.

[16] R. Vijayanathan and P. Kaliyamoorthi, The open software engineering journal: Bibliometrics study, *Int. J. Sci. Res.*, vol. 3, no. 7, pp. 1095–1097, 2014.

[17] J. M. Merigó, F. Blanco-Mesa, A. M. Gil-Lafuente, and R. R. Yager, Thirty years of the international journal of intelligent systems: A bibliometric review, *Int. J. Intell. Syst.*, vol. 32, no. 5, pp. 526–554, 2017.

[18] V. Garousi and G. Ruhe, A bibliometric/geographic assessment of 40 years of software engineering research (1969–2009), *Int. J. Softw. Eng. Knowl. Eng.*, vol. 23, no. 9, pp. 1343–1366, 2013.

[19] J. M. Fernandes, Authorship trends in software engineering, *Scientometrics*, vol. 101, no. 1, pp. 257–271, 2014.

[20] V. Garousi and M. V. Mäntylä, Citations, research topics and active countries in software engineering: A bibliometrics study, *Comput. Sci. Rev.*, vol. 19, pp. 56–77, 2016.

[21] B. G. Tavares, C. E. S. da Silva, and A. D. de Souza, Risk management in scrum projects: A bibliometric study, *J. Commun. Softw. Syst.*, vol. 13, no. 1, pp. 1–8, 2017.

[22] F. Blanco-Mesa, J. M. M. Lindahl, and A. M. Gil-Lafuente, A bibliometric analysis of fuzzy decision making research, in *Proc. 2016 Ann. Conf. of the North American Fuzzy Information Processing Society*, El Paso, TX, USA, 2016, pp. 1–4.

[23] K. Koumaditis and T. Hussain, Human computer interaction research through the lens of a bibliometric analysis, in *Proc. Int. Conf. on Human-Computer Interaction. User Interface Design, Development and Multimodality*, Vancouver, Canada, 2017, pp. 23–37.

[24] R. Heradio, H. Perez-Morago, D. Fernandez-Amoros, F. J. Cabrerizo, and E. Herrera-Viedma, A bibliometric analysis of 20 years of research on software product lines, *Inform. Softw. Technol.*, vol. 72, pp. 1–15, 2016.

[25] V. Garousi, S. Shahnewaz, and D. Krishnamurthy, Uml-driven software performance engineering: A systematic mapping and trend analysis, in *Progressions and Innovations in Model-Driven Software Engineering*, V. G. Diaz, J. M. C. Lovelle, B. C. P. Garcia-Bustelo, and O. S. Martinez, eds. Hershey, PA, USA: IGI Global, 2013, pp. 18–64.

[26] V. de Souza and M. Borsato, Sustainable consumption and ecodesign: A review, in *Advances in Transdisciplinary Engineering*, J. Stjepandić and R. Curran, eds. Clifton, VA, USA: IOS Press, 2015, pp. 492–499.

[27] C. Calero, Five years of green in software engineering: The number of the beast, in *Proc. 2nd International Workshop on Green and Sustainable Software*, 2016.

[28] S. Naumann, M. Dick, E. Kern, and T. Johann, The GREENSOFT model: A reference model for green and sustainable software and its engineering, *Sustainable Computing: Informatics and Systems*, vol. 1, no. 4, pp. 294–304, 2011.

[29] P. Mongeon and A. Paul-Hus, The journal coverage of web of science and Scopus: A comparative analysis, *Scientometrics*, vol. 106, no. 1, pp. 213–228, 2016.

[30] W. G. Cochran, *Sampling Techniques*, 3rd ed. New York, NY, USA: John Wiley, 1977.

Coral Calero is a full professor in the Department of Information Technologies and Systems at the University of Castilla-La Mancha (Spain). She hold PMP certification. Her research interests include software quality, software quality models, software measurement, data quality, and software sustainability. She is a member of the Alarcos Research Group.

Javier Mancebo is a PhD student in computer science at the University of Castilla-La Mancha. His research interests are software sustainability and business process management. He is a member of the Alarcos Research Group. He holds the following professional certifications: PMP, CISA, ITIL foundation, and Scrum Manager.

Félix Garcia is currently an associate professor in the Department of Information Technologies and Systems at the Unversity of Castilla-La Mancha. He is a member of the Alarcos Research Group and his research interests include business process management, software processes, software measurement, research methods, and agile methods. He holds the following professional certifications: PMP, CISA, and Scrum Manager.

María Ángeles Moraga is an associate professor at the University of Castilla-La Mancha, Spain. She is a member of the Alarcos Research Group. Her research interests are software quality, measures, process quality, and software sustainability.

José Alberto García Berna is a PhD student at the Department of Computer Science and Systems of University of Murcia. His research interests are requirements engineering and project management, specifically green software engineering and sustainability in information and communication technologies.

José Luis Fernández Alemán is currently an associate professor at University of Murcia, where he is a member of the Software Engineering Research Group. He has published more than 50 JCR papers in the areas of software engineering and requirements engineering and their application to the fields of e-health and e-learning. Currently, his main research interest is continuous requirements engineering, privacy, usability, sustainability processes, and their application to e-health and e-learning.

Ambrosio Toval received the BS degree from University Complutense of Madrid, Madrid, Spain, in 1983, and the PhD degree from Technical University of Valencia, Valencia, Spain, in 1994. He is currently a full professor with University of Murcia, Spain, where he is the head of the Software Engineering Research Group. He has conducted a variety of research and technology transfer projects in the areas of requirements engineering processes and tools, privacy and security requirements, sustainable requirements, and applications in the e-health, e-learning, and mobile development domains. He has published in the same topics in international journals, such as *IEEE Software, Information and Software Technology, Requirements Engineering, Computer Standards & Interfaces, IET Software, International Journal of Information Security*, etc.