

## A Bibliometric Analysis of the 35th anniversary of the paper “The Statistical Analysis of Compositional Data” by John Aitchison (1982)

**Carolina  
Navarro-Lopez**  
University of the  
Balearic Islands

**Silvia  
Gonzalez-Morcillo**  
University of the  
Balearic Islands

**Carles  
Mulet-Forteza**  
University of the  
Balearic Islands

**Salvador  
Linares-Mustarós**  
University of  
Girona

---

### Abstract

This study presents a comprehensive bibliometric analysis of the paper published by John Aitchison in the Journal of the Royal Statistical Society. Series B (Methodological) in 1982. Having recently reached the milestone of 35 years since its publication, this pioneering paper was the first to illustrate the use of the methodology “Compositional Data Analysis” or “CoDA”. By October 2019, this paper had received over 780 citations, making it the most widely cited and influential article among those using said methodology. The bibliometric approach used in this study encompasses a wide range of techniques, including a specific analysis of the main authors and institutions to have cited Aitchison’s paper. The VOSviewer software was also used for the purpose of developing network maps for said publication. Specifically, the techniques used were co-citations and bibliographic coupling. The results clearly show the significant impact the paper has had on scientific research, having been cited by authors and institutions that publish all around the world.

*Keywords:* John Aitchison, compositional data analysis, bibliometrics, Web of Science, author ranking, institution ranking, country ranking, VOSviewer.

---

## 1. Introduction

Nowadays compositional data are defined as arrays of strictly positive numbers for which ratios between them are considered to be relevant (Egozcue and Pawłowsky-Glahn (2019)). Despite warnings about the problems involved in not using specific methods for such data (Pearson (1897), Chayes (1948) and Vistelius and Sarmanov (1961)), it was not until the 1980s that the first general methods were proposed as appropriate methods for their analysis (Aitchison (1982) and Aitchison (1986)). This methodology received the name of compositional data analysis, CoDa analysis or simply CoDA. It is usually written CoDa when it refers to “compositional data” and CoDA when it refers to “Compositional Data Analysis”. That same terminology also encompassed methods that allow the analysis of data with positive values, whereby although the data do not have to fulfill the characteristic of constant sum, they do need to meet the requirement that the study of certain ratios of this study is considered

as relevant in the study of the problem. Sample applications of data that do not represent parts of any whole can be found in [Rodrigues, Daunís-I-Estadella, Mateu-Figueras, and Thió-Henestrosa \(2011\)](#), [Ortells, Egozcue, Ortego, and Garola \(2015\)](#) and [Linares-Mustarós, Coenders, and Vives-Mestres \(2018\)](#).

The scientific production related to compositional data analysis has increased constantly over the years, and in the last ten years especially started to flourish in very different fields to the ones where it was initially employed ([Kogovšek, Coenders, and Hlebec \(2013\)](#), [Ferrer-Rosell, Coenders, and Martínez-García \(2015\)](#), [Batista-Foguet, Ferrer-Rosell, Serlavós, Coenders, and Boyatzis \(2015\)](#), [Belles-Sampera, Guillen, and Santolino \(2016\)](#), [Morais, Thomas-Agnan, and Simioni \(2018\)](#), [Blasco-Duatis, Coenders, Saez, García, and Cunha \(2019\)](#), [Creixans-Tenas, Coenders, and Arimany-Serrat \(2019\)](#), [Carreras Simó and Coenders \(2020\)](#) and [Coenders and Ferrer-Rosell \(2020\)](#)). This growth and expansion to new fields can be related to four easily identifiable events. Normally, after beginning its wanderings in an intuitive manner, every new scientific theory enters a period of mathematic axiomatic formalization. Set theory is a clear example of this. In this sense, the theory has not been an exception and the purely mathematic works in which the theory is axiomatized may have assigned a greater degree of confidence given by the scientific community towards the new methods ([Pawlowsky-Glahn and Egozcue \(2001\)](#), [Martín-Fernández, Olea-Meneses, and Pawlowsky-Glahn \(2001\)](#) and [Egozcue, Pawlowsky-Glahn, and Gloor \(2018\)](#)). The second event is the proliferation of tutorials and textbooks on the theory ([Aitchison \(1986\)](#), [Aitchison \(1994\)](#), [Valls \(2018\)](#), [Pawlowsky-Glahn and Buccianti \(2011\)](#), [Van den Boogaart and Tolosana-Delgado \(2013\)](#), [Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado \(2015\)](#), [Greenacre \(2017\)](#) and [Filzmoser, Hron, and Templ \(2018\)](#)), which ensure its growth by facilitating the inclusion of new researchers. The third event that seems to have facilitated the expansion of CoDA consists in the development of multiple libraries with R ([Van den Boogaart and Tolosana-Delgado \(2013\)](#), [Palarea-Albaladejo and Martín-Fernández \(2015\)](#) and [Filzmoser \*et al.\* \(2018\)](#)) and the creation of various softwares, such as the CoDaPack ([Thió-Henestrosa and Martín-Fernández \(2005\)](#) and [Comas-Cufí, Thió-Henestrosa, Egozcue, Tolosana-Delgado, and Ortego \(2011\)](#)), which allows operations to be performed without any previous knowledge of programming. The fourth event likely to have triggered the expansion of the CoDa theory is the creation of different working groups, the multiple actions they have undertaken helping to disseminate this new theory. As examples of this, we can mention the establishing of a biannual congress, different introductory courses on CoDA and the creation of websites offering users, among other features, the option to download working material. Finally, it is also worth mentioning that the work done by the groups has led to the creation of a CoDa Association, spreading CoDa theory even wider.

To celebrate 35 years since publication of the seminal article on CoDa analysis “The Statistical Analysis of Compositional Data” ([Aitchison \(1982\)](#)), the main purpose of this paper is to carry out an exhaustive bibliometric analysis of all publications to have cited the paper based on data taken from the Web of Sciences (WoS). This analysis serves the purpose of gathering information on trends in research using CoDa analysis.

The bibliometric analysis is divided into two parts. The first presents an analysis of the academic structure used in the documents that have cited Aitchison’s 1982 paper, while the second focuses on an analysis of the main authors, institutions and countries to have cited it. The information we expect to gather from the analysis should answer the following research questions (RQ):

- RQ1: What is the evolution of the number of citation of Aitchison’s 1982 paper?
- RQ2: Who are the authors that most cite Aitchison’s paper?
- RQ3: What are the institutions that most cite Aitchison’s paper?
- RQ4: What are the countries that most cite Aitchison’s paper?

To meet our aim, the WoS database and the VOSviewer Software (Van Eck and Waltman (2010)) were used. The VOSviewer Software was employed with the aim of graphically mapping the bibliographic material used. Specifically, the following techniques were considered in this paper: bibliographic coupling and co-citation. The reason for using the WoS database is that it is considered the most influential in the world (Merigó, Gil-Lafuente, and Yager (2015)).

The rest of the document is divided into the following sections: the second section presents the bibliometric methods used in this paper; the third section provides a complete bibliometric study of Aitchison's work "The Statistical Analysis of Compositional Data" (Aitchison (1982)); and the fourth section summarizes the main conclusions, limitations and future lines of research.

## 2. Methodology

The term bibliometrics was introduced by Pritchard (1969) as "the application of mathematical and statistical methods to books and other means of communication". Currently, although many other definitions exist (see Yuan, Gretzel, and Tseng (2015) and Köseoglu, Sehitoglu, Ross, and Parnell (2016)), they all describe it as an instrument for analyzing the evolution of scientific disciplines based on intellectual, social and conceptual structures (Zupic and Čater (2015)) in order to identify trends and patterns in scientific research (Merigó, Blanco-Mesa, Gil-Lafuente, and Yager (2017)). Therefore, bibliometrics is one of the most widely used approaches for analyzing how a scientific field develops (Bar-Ilan (2008)).

For this bibliometric study, data were gathered from the WoS database in October 2019 using "The Statistical Analysis of Compositional Data" as a keyword in the field "title" and "Aitchison, J." as a keyword in the field "author". These searches returned Aitchison's 1982 paper as the only result. Subsequently, the information was refined based on the total number of citations obtained by the paper, which resulted in 784 publications for analysis.

Given that no consensus exists in the literature on which methods are best or most appropriate, we used several bibliometric indicators to present the data. Firstly, we considered the number of publications and citations, these methods being considered the most popular according to Ding, Rousseau, and Wolfram (2016). The former indicates productivity, while the latter quantifies the influence of these publications (Svensson (2010)). Other common indicators include the most productive authors, institutions and countries, and number of publications and citations per person (Mulet-Forteza, Salvá, Monserrat, and Amores (2020)). For the analysis of institutions, we also included general university rankings. The results in the tables are sorted by total number of publications (TP).

In addition, we used the VOSviewer software (Van Eck and Waltman (2010)) to graphically map the bibliographic data (Sinkovics (2016)) for co-citations (Small (1973)) and bibliographic coupling (Kessler (1963)). Co-citation assumes that there is some kind of relationship between two documents cited jointly by a third document (McCain (1990), Ramos-Rodríguez and Ruíz-Navarro (2004) and Small (1973)). According to McCain (1986) and McCain (1991), these documents allow the academic structure of a scientific discipline to be determined. Bibliographic coupling measures the similarity of the subject analyzed among the documents considered from the frequency in which certain references are shared. A bibliographic coupling occurs when two documents include the reference to a third document (Young (1983)), so there is a possibility that these documents are linked (Martyn (1964)). Bibliographic coupling is usually applied to perform the graphic mapping of institutions and countries (Small (1999) and Boyack and Klavans (2014)), while co-citation is usually used to perform the graphic mapping of authors (Glänzel and Thijs (2012) and Zupic and Čater (2015)).

The combination of methods used to collect data from the WoS database, along with use of the VOSviewer software, allowed us to incorporate both the "full counting" and "fractional counting" techniques. The difference between these methods is that "full counting" assigns one point to each participant of a paper, whereas "fractional counting" takes into account co-

authorship of the paper (Mulet-Forteza, Genovart-Balaguer, Merigó, and Mauleon-Mendez (2019b)).

### 3. Bibliometric study of Aitchison's paper (1982)

In this section, we will address the different research questions posed.

#### 3.1. Evolution of number of citations received by Aitchison's paper (1982)

Regarding the first question (RQ1), Figure 1 presents the evolution of the citations received by Aitchison's 1982 paper.

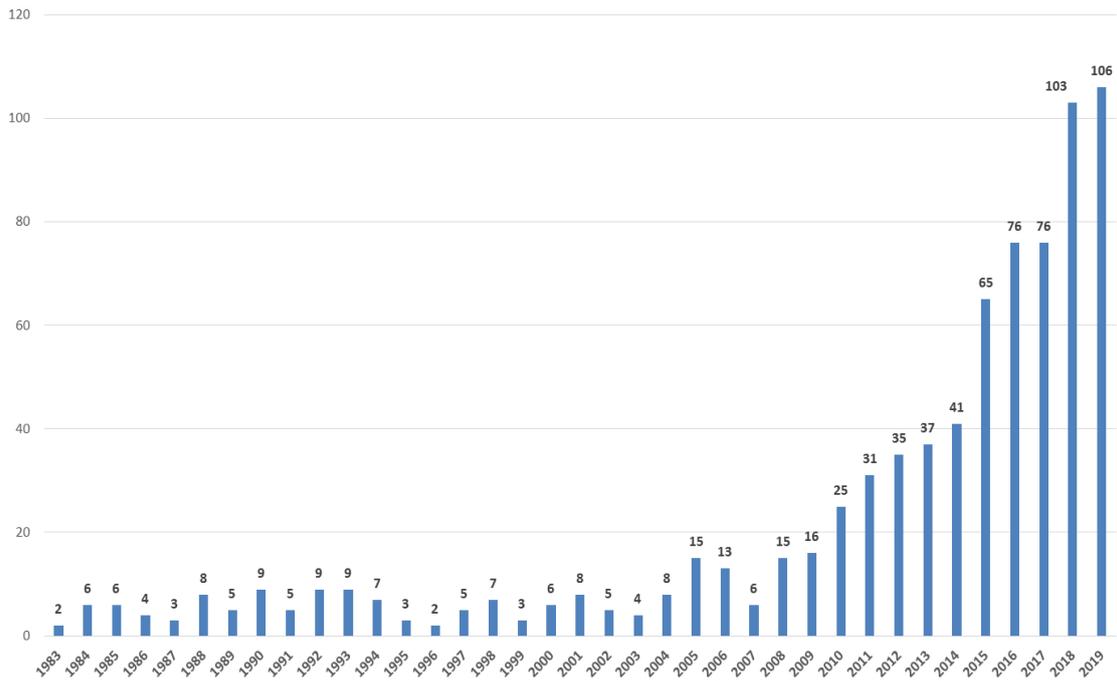


Figure 1: Annual number of citations received by Aitchison's 1982 paper. Source: own elaboration, compiled from WoS database.

Figure 1 shows that the paper has received uninterrupted citations since its publication in 1982. It also indicates how the number of citations received has evolved over different periods. In this sense, with few exceptions, the number of annual citations received by the paper between 1983 and 2007 did not exceed 10 per year. On the other hand, since 2008 annual citations have exceeded the previous value every year, following an expected exponential growth (Price (1986)). Likewise, a very significant increase in the number of citations received can be observed since 2015, and this increased still further in the years 2018 and 2019.

We have analyzed some of the reasons why Aitchison's 1982 paper has received a significant number of citations, especially since 2011. To this effect, in Table 1 we examine the evolution over time of the main research areas where the citations for the said work have been provided.

Table 1 clearly shows that the majority of citations received by Aitchison's 1982 paper come from three research areas, i.e. Mathematics, Geology and Environmental Sciences Ecology. Nevertheless, the interest of these research areas in Aitchison's 1982 paper has been aroused only in the last decade. Almost 60% of the citations from the area of Mathematics belong to this period, while the percentage goes up to over two-thirds in the areas of Geology and Environmental Sciences Ecology. Other areas that have also provided a great number of citations of Aitchison's 1982 paper are Geochemistry Geophysics, Mathematical Computational Biology, Engineering, Biochemistry Molecular Biology and Agriculture.

Table 1: Main research areas that have cited Aitchison's paper (1982). Source: own elaboration, WoS database, 1982 through December 2019. Abbreviations: R = Ranking that occupies the research area during the period 1983-2000; R1 = Ranking that occupies the research area during the period 2001-2009; R2 = Ranking that occupies the research area during the period 2010-2019; R3 = Ranking that occupies the research area during the whole period. Note: The same journal may be indexed in two or more research areas at the same time. Some examples are Mathematical Geology and Mathematical Geosciences, both indexed in the Geosciences and Mathematics research areas, and Bioinformatics, indexed in the Biotechnology and Mathematics research areas.

R1	R2	R3	R	Research Areas	1983-2000	2001-2009	2010-2019
1	1	1	1	Mathematics	55	36	130
2	2	2	2	Geology	27	25	103
3	3	3	3	Environmental Sciences Ecology	18	17	86
6	9	4	4	Computer Science	11	4	56
13	4	5	5	Geochemistry Geophysics	5	7	49
5	13	8	6	Mathematical Computational Biology	13	3	34
14	6	7	7	Engineering	5	6	36
8	33	6	8	Biochemistry Molecular Biology	9	0	37
15	7	10	9	Agriculture	5	6	27
4	5	16	10	Zoology	13	6	18
24	39	9	11	Science Technology Other Topics	4	0	33
23	18	11	12	Physical Geography	4	2	25
9	14	18	13	Chemistry	8	3	17
37	12	13	14	Water Resources	2	4	21
29	21	14	15	Public Environmental Occupational Health	3	1	19
78	78	12	16	Microbiology	0	0	23
38	19	17	17	Business Economics	2	2	18
17	15	20	18	Life Sciences Biomedicine Other Topics	5	3	12
51	57	15	19	Biotechnology Applied Microbiology	1	0	19
7	32	31	20	Nutrition Dietetics	11	0	7
22	11	33	22	Marine Freshwater Biology	4	4	7
16	10	40	24	Plant Sciences	5	4	5
48	16	26	25	Operations Research Management Science	1	3	10
19	20	28	26	Behavioral Sciences	5	1	8
47	8	36	28	Biodiversity Conservation	1	5	6
18	17	41	29	Paleontology	5	2	5
10	34	49	30	Endocrinology Metabolism	8	0	4
20	37	32	31	Physiology	5	0	7
11	35	89	42	Physical Sciences Other Topics	8	0	0
12	36	90	43	Reproductive Biology	7	0	0

On the other hand, Table 1 also illustrates a wide range of research areas that provided the largest amount of citations of Aitchison's 1982 paper when it was first published but are not so relevant today, including Zoology, Chemistry, Nutrition Dietetics, Plant Sciences, Behavioral Sciences, Paleontology, Endocrinology Metabolism, Physiology, Physical Sciences Other Topics, Reproductive Biology, among others.

It can therefore be seen that there has been a shift in interest in the research carried out by Aitchison in 1982, and that areas related to Statistics, Geosciences, Mathematics, Computer Science, Biochemistry and Economics, among others, have replaced those initially used by CoDA. Consequently, the journals that have cited Aitchison's 1982 paper the most are those indexed in these research areas. Just as an example, it is noteworthy that the Journal of Geochemical Exploration, indexed in the Geochemistry Geophysics research area, is the one that has cited Aitchison's 1982 paper most often, with a total of 23 papers citing the said document during the last decade. Other journals indexed in the research areas that cited very often Aitchison's 1982 paper during the last 10 years include:

- In the Mathematics research area: Bioinformatics, Mathematical Geosciences, Environmental and Ecological Statistics and Environmetrics, among others.
- In the Environmental Sciences Ecology: Environmental Earth Sciences and International Journal of Environmental Research and Public Health, among others.
- In the Statistics research area: Journal of the American Statistical Association, Annals of Applied Statistics, Stochastic Environmental Research and Risk Assessment, Biometrics and Austrian Journal of Statistics, among others.
- In the Geosciences research area: Quaternary International and Geoderma, among others.

Therefore, it can be stated that the interest aroused in these research areas by Aitchison's 1982 paper has caused a genuinely growing interest in this publication, especially during the last decade.

It has also been possible to confirm that the authors who have most often cited Aitchison's 1982 paper during the last decade match those at the top of Table 2. In fact, only some positions have been exchanged. Thus, for example, Antonella Buccianti and Vera Pawlowsky-Glahn would exchange their positions, while Andrea Bloise, who occupies position 11 in Table 2, if we consider only the citations made to Aitchison's 1982 paper during the last decade, would occupy the ninth position in this new ranking, relegating John Aitchison from the TOP 10, who would be left out of the list of authors who have cited Aitchison's 1982 paper the most. This is not surprising, considering that John Aitchison died in 2016 at the age of 90. Finally, we also analyzed the original source of the 784 citations received by Aitchison's paper. In this regard, 90.7% of citations were from documents published as papers, 4.6% from proceedings papers, 2.2% from books, 2% from reviews, and the remaining 0.5% from notes and letters. Thus, 93.2% of citations came from papers that had passed a strict process of arbitration; in other words, articles, reviews, letters and academic notes.

### 3.2. Most productive authors citing Aitchison's paper (1982)

In this section, we address the second question (RQ2) posed in our paper. Firstly, Table 2 lists those authors who have cited Aitchison's paper (1982) the most.

Vera Pawlowsky-Glahn (University of Girona, Spain) is the author who has cited the Aitchison paper (1982) the most, followed by Antonelle Buccianti (Università degli Studi di Firenze, Italy) and Juan José Egozcue (Polytechnic University of Catalonia, Spain). As Table 2 shows, the three main authors in this ranking have cited Aitchison's paper (1982) a total of 86 times, although it should be noted that this value, when obtained by means of a full counting

Table 2: Main authors who have cited Aitchison's paper (1982). Source: own elaboration, WoS database, 1982 through October 2019. Abbreviations: R = Ranking; TP = Total citing papers. Note: There are 25 authors with 4 papers.

R	Name	Institution (Country)	TP
1	Pawlowsky-Glahn, V	University of Girona (Spain)	32
2	Buccianti, A	Università degli Studi di Firenze (Italy)	29
3	Egozcue, JJ	Polytechnic University of Catalonia (Spain)	25
4	Tolosana-Delgado, R	Helmholtz-Zentrum Dresden-Rossendorf (Germany)	18
5	Mateu-Figueras, G	University of Girona (Spain)	14
6	Aitchison, J	University of Glasgow (UK)	13
7	Martin-Fernandez, JA	University of Girona (Spain)	13
8	Li, HZ	University of Pennsylvania (USA)	11
9	Dumuid, D	University of South Australia (Australia)	9
10	Olds, T	University of South Australia (Australia)	9
11	Bloise, A	University of Calabria (Italy)	8
12	Gallo, M	Università degli Studi di Napoli L'Orientale (Italy)	8
13	Miriello, D	University of Calabria (Italy)	8
14	van den Boogaart, KG	Helmholtz-Zentrum Dresden-Rossendorf (Germany)	8
15	Woronow, A	Exxon Mobil Corporation (USA)	8
16	Palarea-Albaladejo, J	Biomathematics & Statistics Scotland (UK)	7
17	Weltje, GJ	KU Leuven (Belgium)	7
18	Crisci, GM	University of Calabria (Italy)	6
19	De Luca, R	University of Calabria (Italy)	6
20	Ortego, MI	Polytechnic University of Catalonia (Spain)	6
21	Wang, HW	Beijing Advanced Innovation Center for Big Data and Brain Computing (China)	6
22	Zuo, RG	China University of Geosciences, Wuhan (China)	6
23	Bagneres, AG	Université de Tours (France)	5
24	Blei, DM	Columbia University in the City of New York (USA)	5
25	Chaput, JP	University of Ottawa (Canada)	5
26	Gloor, GB	Western University (Canada)	5
27	Heslop, D	Australian National University (Australia)	5
28	Hron, K	Palacký University in Olomouc (Czech Republic)	5
29	Klotz, S	Helmholtz Zentrum für Umweltforschung (Germany)	5
30	Kuhn, I	Helmholtz Zentrum für Umweltforschung (Germany)	5
31	Lorenzi, MC	Université Paris 13 (France)	5
32	Love, KM	BEB Erdgas und Erdörl GmbH (Germany)	5
33	Maher, C	University of South Australia (Australia)	5
34	Mueller, U	Edith Cowan University, Joondalup (Australia)	5
35	Parent, LE	Université Laval (Canada)	5
36	Scealy, JL	Australian National University (Australia)	5
37	Szava-Kovats, RC	Estonian Institute of Ecology (Estonia)	5
38	Tjallingii, R	Deutsches GeoForschungsZentrum (GFZ) (Germany)	5
39	Tremblay, MS	Children's Hospital of Eastern Ontario (Canada)	5
40	Tsagris, M	Panepistimio Kritis (Greece)	5
41	Wang, J	Chengdu University of Technology (China)	5
42	Welsh, AH	Australian National University (Australia)	5

method, does not take into account co-authors among these authors. This bias will be eliminated later when performing the graphic analysis of the main authors through a fractional counting method.

It is also interesting to observe the decreasing number of authors producing an increasing number of citing papers, as predicted by the bibliometric law of authors' productivity (Lotka (1926)).

The University of Calabria (Italy) counts four authors and is the most repeated institution among the authors who lead the ranking in Table 2, followed by the Australian National University, the University of Girona and the University of South Australia, with three authors each. With two authors, we find the Helmholtz Zentrum für Umweltforschung (Germany), the HZDR - Helmholtz-Zentrum Dresden-Rossendorf (Germany) and the Polytechnic University of Catalonia (Spain). The rest of the institutions have only one author represented in Table 2 (23).

Finally, the authors in Table 2 work in 13 different countries. Australia (with seven authors) leads this ranking, followed by Germany and Italy, with six authors each. Next, we find Spain (five authors), Canada (four authors), China and the US (with three authors each), France and the UK (with two authors each), while Belgium, Czech Republic, Estonia and Greece only have one author in Table 2.

Figure 2 shows a graphic map of the co-citations among the most influential authors to have cited Aitchison's paper.

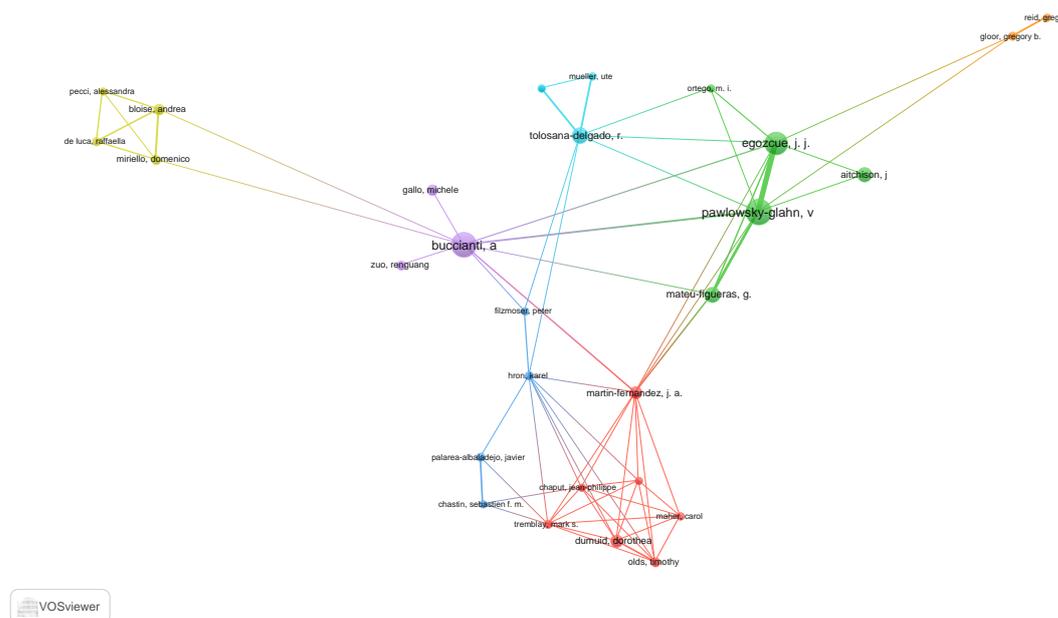


Figure 2: Co-citation of authors who have cited Aitchison's paper (1982). Node size = the number of citations received by an author; line thickness indicates multiple connections; line length is not significant. Citation threshold of 5 and showing the 100 most representative co-citation connections. Source: own elaboration, based on WoS database; figure created using VOSviewer Software.

Figure 2 reveals four main node clusters, indicating networks of connections between authors working on similar topics, and three secondary node clusters. The largest group, with seven authors, is focused around the figures of Dorothea Dumuid and Tim Olds, both from the University of South Australia. The second group contains five authors focused on the topics of Vera Pawlowsky-Glahn (University of Girona), Juan José Egozcue (Polytechnic University of Catalonia), Glòria Mateu-Figueras (University of Girona) and John Aitchison (University



Table 3: Main institutions that have cited Aitchison's paper (1982). Source: own elaboration, WoS database, 1982 through October 2019. Abbreviations: R = Ranking; TP = Total citing papers; QS = Quacquarelli Symonds (QS) World University Rankings (Symonds (2019)); ARWU = Academic Ranking of World Universities (Consultancy (2019)). Note: There are 21 institutions with seven papers.

R	Institution (Country)	TP	QS	ARWU
1	University of Girona (Spain)	46	-	801-900
2	Polytechnic University of Catalonia (Spain)	33	275	601-700
3	University of Florence (Italy)	31	501-510	201-300
4	Helmholtz Association (Germany)	29	-	-
5	Centre National de La Recherche Scientifique (France)	26	-	-
6	University of Sao Paulo (Brazil)	18	-	101-150
7	University of London, City (UK)	17	351	901-1000
8	Consiglio Nazionale Delle Ricerche CNR (France)	16	-	-
9	University of Hong Kong (China)	15	25	101-150
10	Australian National University (Australia)	14	24	76
11	China University of Geosciences (China)	14	-	401-500
12	Universite Confederale Leonard de Vinci (France)	14	-	-
13	Centre Val de Loire Comue (France)	13	-	-
14	Commonwealth Scientific Industrial Research Organisation (Australia)	13	-	-
15	Harvard University (USA)	13	3	1
16	James Hutton Institute (UK)	12	-	-
17	University of Pennsylvania (USA)	12	19	17
18	University of Sydney (Australia)	12	42	80
19	Duke University (UK)	11	26	28
20	Laval University (Canada)	11	402	201-300
21	University of Melbourne (Australia)	11	39	41
22	Communauté d'Universités et Établissements D'Aquitaine Comue (France)	10	-	-
23	United States Department of Energy Doe (USA)	10	-	-
24	University of Bremen (Germany)	10	501-520	501-600
25	University of Calabria (Italy)	10	-	801-900
26	University of California San Diego (USA)	10	41	18
27	University of South Australia (Australia)	10	264	501-600
28	CEA (France)	9	-	-
29	Institute for Humanities Social Sciences (France)	9	-	-
30	Institute of Ecology Environment (France)	9	-	-
31	Helmholtz Zentrum Dresden Rossendorf (Germany)	9	-	-
32	Institut National de La Recherche Agronomique (France)	9	-	-
33	University of Edinburgh (UK)	9	18	31
34	University of Southampton (UK)	9	96	101-150
35	University of Turin (Italy)	9	571-580	201-300
36	University of Washington (USA)	9	66	14
37	University of Washington Seattle (USA)	9	-	-
38	Istituto di Geoscienze E Georisorse (Italy)	8	-	-
39	Nerc Natural Environment Research Council (UK)	8	-	-
40	United States Department of the Interior (USA)	8	-	-
41	Universite Bourgogne Franche Comte Comue (France)	8	-	-
42	Universite Paris Saclay (France)	8	-	-
43	University College London (UK)	8	-	15
44	University of Barcelona (Spain)	8	166	151-200
45	University of Houston (USA)	8	651-700	210-300
46	University of Naples L'Orientale (Italy)	8	-	-
47	University of Oxford (UK)	8	5	7
48	University of Zurich (Switzerland)	8	78	61

Figure 3 shows four main node groups and five secondary node groups. The largest group of 14 institutions is focused around English-speaking institutions, including Harvard University, Duke University and the University of California San Diego. The second group contains 13 institutions, including the University of South Australia, Victoria University and the University of Zurich. The third is composed of 11 institutions, among which the University of Sao Paulo, the University of Turin and the University of Bremen stand out, and the fourth main node, with nine institutions, is led by the University of Girona, the Polytechnic University of Catalonia, and the University of Florence. The latter three lead the first three positions in Table 3. In turn, this group of institutions is the one with the largest network of connections, both with one another and with institutions in the other nodes. Therefore, with some exceptions, it can be observed that most of the institutions present in Figure 3 also occupy relevant positions in Table 3, which indicates that there are no significant differences between the analysis performed by the WoS database using the “full counting” method and the VOSviewer Software using “fractional counting”.

### **3.4. Most productive countries citing Aitchison's paper (1982)**

Regarding the fourth question (RQ4), Table 4 shows the countries that have most frequently cited Aitchison's 1982 paper.

Table 4 shows that the countries with the highest populations are not those that have cited Aitchison's 1982 paper most, with the exception of the US. As a matter of fact, only three countries in the top 10 (US, China and Brazil) have over 100 million inhabitants. In contrast, Table 4 shows how countries where English is widely spoken, especially among academics, are those where Aitchison's paper (1982) tends to be cited the most. This trend is especially present in countries such as Australia, Norway and New Zealand. These countries would lead the rankings in Table 4 if we ordered it based on number of papers cited by population.

Like the previous tables, Table 4 uses the full counting method, which skews countries where several authors write articles together rather than working independently. We therefore implemented a fractional counting method in Figure 4, which shows a bibliographic coupling of the countries that have cited Aitchison's 1982 paper most.

Figure 4 shows eight clusters. The first cluster by number of countries (10) is led by Germany and is composed of European countries, with the exception of Taiwan. The second cluster (seven countries) is led by Australia. This cluster is basically made up of non-European (6) countries. The third cluster by number of countries (6) is led by the United Kingdom. The United States leads the fourth cluster and is the most productive country, with the broadest network of connections on the map. The fifth cluster is led by Spain, while the last is led by Italy. In general, it can be observed that the results obtained under both the full counting system (WoS database) and the fractional counting system (VOSviewer Software) are very similar. Figure 4 shows a very diverse network of connections, where we find cultural connections between different countries such as Colombia and Ireland, or Australia and Iran.

## **4. Conclusions**

Adopting a bibliometric approach and based on data obtained from the WoS database, in this paper we have carried out an analysis of all the publications that have cited the paper entitled “The Statistical Analysis of Compositional Data” published by John Aitchison in the *Journal of the Royal Statistical Society. Series B (Methodological)* in 1982. Having recently reached the milestone of 35 years since its publication, the paper is considered to be the seminal article on the CoDa analysis.

In this paper, we have met all of our established aims. Specifically, we have answered the four research questions we asked at the beginning. As for the first (RQ1), we have analyzed how the number of citations of this paper has evolved, showing how the paper has received uninterrupted citations since its publication and that over the past four years the number of

Table 4: Main countries that have cited Aitchison's paper (1982). Source: own elaboration, WoS database, 1982 through October 2019. Abbreviations: R = Ranking; TP = Total citing papers; POP = population in millions; TP/POP = total publications per millions of inhabitants Note: There are six countries with six papers.

R	Country	TP	POP	TP/POP
1	USA	207	327.2	0.63
2	UK	115	67.5	1.70
3	Spain	90	46.7	1.93
4	Italy	87	60.4	1.44
5	Australia	78	24.6	3.17
6	Germany	77	82.8	0.93
7	Canada	63	37.1	1.70
8	China	59	1403.4	0.04
9	France	53	67	0.79
10	Brazil	33	210.1	0.16
11	Netherlands	30	17.3	1.73
12	Belgium	21	11.4	1.84
13	Sweden	21	10	2.10
14	Switzerland	18	8.42	2.14
15	Norway	15	5.3	2.83
16	Czech Republic	13	10.6	1.23
17	New Zealand	13	4.9	2.65
18	Denmark	12	5.8	2.07
19	South Africa	12	56.7	0.21
20	Austria	10	8.8	1.14
21	Japan	9	126.8	0.07
22	Finland	8	5.5	1.45
23	Iran	8	82.6	0.10
24	Ireland	8	6.6	1.21
25	India	7	1372.1	0.01
26	Mexico	7	129.2	0.05
27	Russia	7	146.8	0.05
28	South Korea	7	51.5	0.14
29	Taiwan	7	23.8	0.29

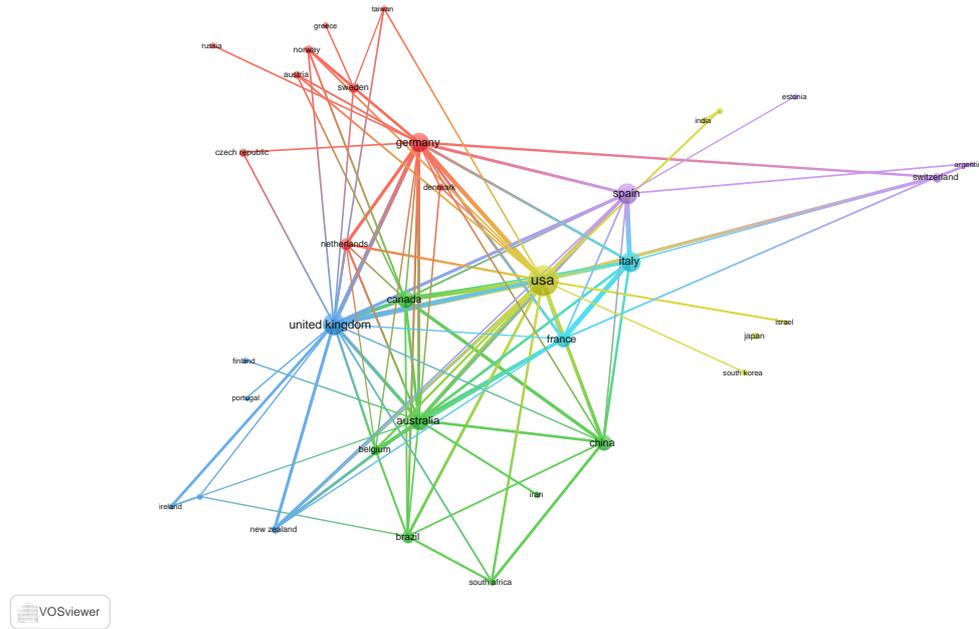


Figure 4: Bibliographic coupling of countries citing Aitchison's paper (1982). Node size = number of citations received by a country; line thickness indicates multiple connections; line length is not significant. Citation threshold of five and showing the 100 most representative co-citation connections. Source: own elaboration, based on WoS database; figure created using VOSviewer Software.

citations has increased more significantly in line with typical expected exponential growth. We have also corroborated that almost 95% of the 784 citations received by Aitchison's 1982 paper have come from documents that have undergone a strict arbitration process.

Regarding the second research question (RQ2), Vera Pawlowsky-Glahn and Glòria Mateu-Figueras (University of Girona), Antonella Bucciante (Università degli Studi di Firenze), Juan José Egozcue (Polytechnic University of Catalonia) and Raimon Tolosana-Delgado (HZDR - Helmholtz-Zentrum Dresden-Rossendorf) are the authors who have cited Aitchison's 1982 paper most. Furthermore, authors of the University of Girona, Polytechnic University of Catalonia, University of Florence, Helmholtz Association and Center National of La Recherche Scientifique CNRS are the ones who have cited the paper (RQ3) the most, while by country, the authors from the United States, the United Kingdom, Spain, Italy, Australia and Germany are the ones who have cited the paper (RQ4) the most. Our analysis indicates that there are no significant differences between the analysis of the WoS database using the "full counting" method and the "fractional counting" method used with the VOSviewer Software. Although this document provides a description of the structure of citations, leading authors, institutions and countries that have cited Aitchison's 1982 paper, it does have some limitations. For example, since data were collected from the WoS database, the limitations of this database also apply to this analysis. As we have indicated previously, the WoS database collects information under a "full counting" method, meaning that documents with many co-authors generally have more weight than documents produced by a single author (Mulet-Forteza, Genovart-Balaguer, Mauleon-Mendez, and Merigó (2019a)). To resolve this limitation, we also employed the "fractional counting" method, using the VOSviewer software to identify co-citations and bibliographic coupling. A further limitation is that the results are dynamic and will inevitably change over time.

Despite the above limitations, this paper represents a starting point for future bibliometric studies in this field. In this respect, future lines of research should aim to carry out a bibliometric analysis focusing on all publications that have included the methodology of

“Compositional Data Analysis”, firstly in the field of the social sciences, and then increasing the number of publications by also covering papers indexed in the WoS under “Science Citation Index Expanded”.

Although we recognize the limitations of our analysis, the main aim of this paper was to analyze the academic structure of the papers, authors, institutions and countries that have cited Aitchison’s 1982 paper. We believe it does this in a sufficiently rigorous and complete manner, while also presenting an overview of the most important data related to Aitchison’s 1982 paper, which has recently celebrated the 35th anniversary since its publication.

## Acknowledgments

The article was supported by the Spanish Ministry of Science, Innovation and Universities (grant RTI2018-095518-B-C21). Acknowledgements are due to Germà Coenders for his comments on previous drafts. We also would like to thank the anonymous reviewers for valuable comments that have improved the quality of the paper.

## References

- Aitchison J (1982). “The Statistical Analysis of Compositional Data.” *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**(2), 139–160.
- Aitchison J (1986). *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability*. Chapman & Hall (Reprinted in 2003 with additional material by The Blackburn Press).
- Aitchison J (1994). “Principles of Compositional Data Analysis.” *Lecture Notes-Monograph Series*, **24**, 73–81.
- Bar-Ilan J (2008). “Informetrics at the Beginning of the 21st Century A Review.” *Journal of Informetrics*, **2**(1), 1–52.
- Batista-Foguet JM, Ferrer-Rosell B, Serlavós R, Coenders G, Boyatzis RE (2015). “An Alternative Approach to Analyze Ipsative Data. Revisiting Experiential Learning Theory.” *Frontiers in Psychology*, **6**, 1742.
- Belles-Sampera J, Guillen M, Santolino M (2016). “Compositional Methods Applied to Capital Allocation Problems.” *Journal of Risk*, **19**, 15–30.
- Blasco-Duatis M, Coenders G, Saez M, García NF, Cunha IF (2019). “Mapping the Agenda-setting Theory, Priming and the Spiral of Silence in Twitter Accounts of Political Parties.” *International Journal of Web Based Communities*, **15**(1), 4–24.
- Boyack KW, Klavans R (2014). “Creation of a Highly Detailed, Dynamic, Global Model and Map of Science.” *Journal of the Association for Information Science and Technology*, **65**(4), 670–685.
- Carreras Simó M, Coenders G (2020). “Principal Component Analysis of Financial Statements: A Compositional Approach.” *Revista de Métodos Cuantitativos para la Economía y la Empresa*, **29**, 18–37.
- Chayes FA (1948). “A Petrographic Criterion for the Possible Replacement Origin of Rocks.” *American Journal of Science*, **246**(7), 413–425.
- Coenders G, Ferrer-Rosell B (2020). “Compositional Data Analysis in Tourism: Review and Future Directions.” *Tourism Analysis*, **25**(1), 153–168.

- Comas-Cufi M, Thió-Henestrosa S, Egozcue J, Tolosana-Delgado R, Ortego M (2011). "Co-DaWork11: 4th International Workshop on Compositional Data Analysis." *Sant Feliu de Guíxols*.
- Consultancy SR (2019). "QS World University Ranking 2019."
- Creixans-Tenas J, Coenders G, Arimany-Serrat N (2019). "Corporate Social Responsibility and Financial Profile of Spanish Private Hospitals." *Heliyon*, **5**(10), e02623.
- Ding Y, Rousseau R, Wolfram D (2016). *Measuring Scholarly Impact*. Springer.
- Egozcue JJ, Pawlowsky-Glahn V (2019). "Compositional Data: The Sample Space and Its Structure." *TEST*, **28**(3), 599–638.
- Egozcue JJ, Pawlowsky-Glahn V, Gloor GB (2018). "Linear Association in Compositional Data Analysis." *Austrian Journal of Statistics*, **47**(1), 3–31.
- Ferrer-Rosell B, Coenders G, Martínez-García E (2015). "Determinants in Tourist Expenditure Composition The Role of Airline Types." *Tourism Economics*, **21**(1), 9–32.
- Filzmoser P, Hron K, Templ M (2018). *Applied Compositional Data Analysis*. Springer.
- Glänzel W, Thijs B (2012). "Using Core Documents for Detecting and Labelling New Emerging Topics." *Scientometrics*, **91**(2), 399–416.
- Greenacre M (2017). *Correspondence Analysis in Practice*. CRC press.
- Kessler MM (1963). "Bibliographic Coupling between Scientific Papers." *American Documentation*, **14**(1), 10–25.
- Kogovšek T, Coenders G, Hlebec V (2013). "Predictors and Outcomes of Social Network Compositions: A Compositional Structural Equation Modeling Approach." *Social Networks*, **35**(1), 1–10.
- Köseoglu MA, Sehitoglu Y, Ross G, Parnell JA (2016). "The Evolution of Business Ethics Research in the Realm of Tourism and Hospitality." *International Journal of Contemporary Hospitality Management*, **28**, 1598–1621.
- Linares-Mustarós S, Coenders G, Vives-Mestres M (2018). "Financial Performance and Distress Profiles. From Classification according to Financial Ratios to Compositional Classification." *Advances in Accounting*, **40**, 1–10.
- Lotka AJ (1926). "The Frequency Distribution of Scientific Productivity." *Journal of the Washington Academy of Sciences*, **16**(12), 317–323.
- Martín-Fernández JA, Olea-Meneses RA, Pawlowsky-Glahn V (2001). "Criteria to Compare Estimation Methods of Regionalized Compositions." *Mathematical Geology*, **33**(8), 889–909.
- Martyn J (1964). "Bibliographic Coupling." *Journal of Documentation*, **20**(4), 236–236.
- McCain KW (1986). "Cocited Author Mapping as a Valid Representation of Intellectual Structure." *Journal of the American Society for Information Science*, **37**(3), 111–122.
- McCain KW (1990). "Mapping Authors in Intellectual Space: A Technical Overview." *Journal of the American Society for Information Science*, **41**(6), 433.
- McCain KW (1991). "Mapping Economics through the Journal Literature: An Experiment in Journal Cocitation Analysis." *Journal of the American Society for Information Science*, **42**(4), 290.

- Merigó JM, Blanco-Mesa F, Gil-Lafuente AM, Yager RR (2017). “Thirty Years of the International Journal of Intelligent Systems: A Bibliometric Review.” *International Journal of Intelligent Systems*, **32**(5), 526–554.
- Merigó JM, Gil-Lafuente AM, Yager RR (2015). “An Overview of Fuzzy Research with Bibliometric Indicators.” *Applied Soft Computing*, **27**, 420–433.
- Morais J, Thomas-Agnan C, Simioni M (2018). “Interpretation of Explanatory Variables Impacts in Compositional Regression Models.” *Austrian Journal of Statistics*, **47**(5), 1–25.
- Mulet-Forteza C, Genovart-Balaguer J, Mauleon-Mendez E, Merigó JM (2019a). “A Bibliometric Research in the Tourism, Leisure and Hospitality Fields.” *Journal of Business Research*, **101**, 819–827.
- Mulet-Forteza C, Genovart-Balaguer J, Merigó JM, Mauleon-Mendez E (2019b). “Bibliometric Structure of IJCHM in Its 30 Years.” *International Journal of Contemporary Hospitality Management*, **31**(12), 4574–4604.
- Mulet-Forteza C, Salvá AS, Monserrat S, Amores A (2020). “80th Anniversary of Pure and Applied Geophysics: A Bibliometric Overview.” *Pure and Applied Geophysics*, **177**(2), 531–570.
- Ortells R, Egozcue JJ, Ortego MI, Garola A (2015). “Relationship between the Popularity of Key Words in the Google Browser and the Evolution of Worldwide Financial Indices.” In *International Workshop on Compositional Data Analysis*, pp. 145–165. Springer.
- Palarea-Albaladejo J, Martín-Fernández JA (2015). “zCompositionsR Package for Multivariate Imputation of Left-censored Data under a Compositional Approach.” *Chemometrics and Intelligent Laboratory Systems*, **143**, 85–96.
- Pawlowsky-Glahn V, Buccianti A (2011). *Compositional Data Analysis: Theory and Applications*. John Wiley & Sons.
- Pawlowsky-Glahn V, Egozcue JJ (2001). “Geometric Approach to Statistical Analysis on the Simplex.” *Stochastic Environmental Research and Risk Assessment*, **15**(5), 384–398.
- Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015). *Modeling and Analysis of Compositional Data*. John Wiley & Sons.
- Pearson K (1897). “Mathematical Contributions to the Theory of Evolution on a Form of Spurious Correlation which May Arise when Indices Are Used in the Measurement of Organs.” *Proceedings of the Royal Society of London*, **60**, 489–498.
- Price DJ (1986). *Little Science, Big Science... and Beyond*. Columbia University Press.
- Pritchard A (1969). “Statistical bibliography or bibliometrics.” *Journal of Documentation*, **25**(4), 348–349.
- Ramos-Rodríguez AR, Ruíz-Navarro J (2004). “Changes in the Intellectual Structure of Strategic Management Research: A Bibliometric Study of the Strategic Management Journal, 1980–2000.” *Strategic Management Journal*, **25**(10), 981–1004.
- Rodrigues LA, Daunís-I-Estadella J, Mateu-Figueras G, Thió-Henestrosa S (2011). *Flying in Compositional Morphospaces: Evolution of Limb Proportions in Flying Vertebrates*. Wiley Online Library.
- Sinkovics N (2016). “Enhancing the Foundations for Theorising through Bibliometric Mapping.” *International Marketing Review*, **33**(3), 327–350.

- Small H (1973). "Co-citation in the Scientific Literature: A New Measure of the Relationship between Two Documents." *Journal of the American Society for Information Science*, **24**(4), 265–269.
- Small H (1999). "Visualizing Science by Citation Mapping." *Journal of the American Society for Information Science*, **50**(9), 799–813.
- Svensson G (2010). "SSCI and Its Impact Factors: A Prisoner's Dilemma?" *European Journal of Marketing*, **44**, 23–33.
- Symonds Q (2019). "QS World University Ranking 2019."
- Thió-Henestrosa S, Martín-Fernández JA (2005). "Dealing with Compositional Data: The Freeware CoDaPack." *Mathematical Geology*, **37**(7), 773–793.
- Valls R (2018). *Why and How Should Geologists Use Compositional Data Analysis*. Valls Geoconsultant.
- Van den Boogaart KG, Tolosana-Delgado R (2013). *Analyzing Compositional Data with R*. Springer.
- Van Eck NJ, Waltman L (2010). "Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping." *Scientometrics*, **84**(2), 523–538.
- Vistelius AB, Sarmanov OV (1961). "On the Correlation between Percentage Values: Major Component Correlation in Ferromagnesium Micas." *The Journal of Geology*, **69**(2), 145–153.
- Young H (1983). *The ALA Glossary of Library and Information Science*. American Library Association.
- Yuan Y, Gretzel U, Tseng YH (2015). "Revealing the Nature of Contemporary Tourism Research: Extracting Common Subject Areas through Bibliographic Coupling." *International Journal of Tourism Research*, **17**(5), 417–431.
- Zupic I, Čater T (2015). "Bibliometric Methods in Management and Organization." *Organizational Research Methods*, **18**(3), 429–472.

**Affiliation:**

Carolina Navarro-Lopez  
Department of Economics  
Faculty of Tourism  
University of the Balearic Islands  
E-mail: [carolina.navarro@uib.es](mailto:carolina.navarro@uib.es)  
URL: <https://www.uib.eu/personal/ABTE2NjI2/>

Silvia Gonzalez-Morcillo  
Department of Economics  
Faculty of Tourism  
University of the Balearic Islands  
E-mail: [s.gonzalez-morcillo1@estudiant.uib.eu](mailto:s.gonzalez-morcillo1@estudiant.uib.eu)

Carles Mulet-Forteza  
Department of Economics  
Faculty of Tourism  
University of the Balearic Islands  
E-mail: [carles.mulet@uib.es](mailto:carles.mulet@uib.es)  
URL: <https://www.uib.eu/personal/ABTE1NDI1/>

Salvador Linares-Mustaros  
Department of Business Administration  
University of Girona  
E-mail: [salvador.linares@udg.edu](mailto:salvador.linares@udg.edu)  
URL: <https://www.udg.edu/es/directori/pagina-personal?ID=2002049>