# The Evolution of Decision Support Systems for Agriculture: A Bibliometric Network Approach

**Dimitris Kremmydas, Alvertos Konstantinis, and Stelios Rozakis**

**Abstract** We use the Scopus database and naïve Bayes text classification to identify almost a thousand and a half DSS papers targeting problems in agriculture during the last three decades. We then use bibliometric network analysis to establish the chronological trends regarding the methodologies, the technologies, the topics, and their interrelation. We also provide insights into the evolution of international research and academic community cooperation and specialization.

**Keywords** DSS · Agriculture · Naïve Bayes · Network analysis

## 1 Introduction

Decision Support Systems (DSS) are human–computer systems that assist stakeholders to make effective decisions. This usually involves the presentation of data from heterogeneous sources in a more intuitive way, and quite often, scientific models that use them in order to provide further insights [1]. In the agricultural domain, the term appears in the late 1980s and the number of publications follows an increasing trend. The applications cover diverse topics, for example, water management, environmental modeling, climate change, crop protection, farm management, agricultural policy, and precision agriculture [2, 3].

DSS have facilitated the exchange and transfer of knowledge between the scientific community and the stakeholders or practitioners [4, 5]. However, this exchange is not trouble-free, and the DSS performance depends on several factors.

D. Kremmydas (✉)
Department of Agricultural Economics and Rural Development, Agricultural University of Athens, Athens, Greece

European Commission, Joint Research Centre, Seville, Spain
e-mail: Dimitrios.KREMMYDAS@ec.europa.eu

A. Konstantinis · S. Rozakis
School of Environmental Engineering, Technical University of Crete, Crete, Greece



J. Papathanasiou et al. (eds.), *EURO Working Group on DSS*, Integrated Series in Information Systems, https://doi.org/10.1007/978-3-030-70377-6_6

The literature mentions quite a few. For example, the degree of user-centered design, the quality of the human–computer interface, the skills of the end-users, etc. [6–8]. Thus, due to the unequal performance of the DSSs developed by the research community, the scene is not homogeneous.

The purpose of this chapter is to outline and discuss the trends of the last 30 years regarding the methodologies, the technologies, and the application domains of DSS in agriculture and their interrelation. We also aim to provide insights into the evolution of the international research and academic community cooperation and highlight any specialization that has emerged. The contribution is twofold. Firstly, we identify a vast amount of literature related to DSS in agriculture that can be further used by other researches for more focused reviews. Secondly, we facilitate the discussion opened in this volume, about the assessment of what has occurred during the last three decades and what can be the implications for the future of the field.

To accomplish those objectives, we resort to a bibliometric approach. The published literature is a reliable measure of the research trends and can thus be used to sketch the evolution of any discipline. However, in order to utilize the information found in the bibliographic databases, we need to resolve two issues.

The first relates to the fact that the term "DSS" refers at the same time to the research domain of "decision support systems" and the implementation of a "decision support system" to other science domains. Thus, a search with the keyword "DSS," returns publications of both kinds. Yet, we are interested in the latter group of publications, and thus we need a way to filter them efficiently. The second issue relates to the fact that agriculture is a relatively large, heterogeneous, and interdisciplinary scientific domain. So a search with the keyword "agriculture," returns but a portion of the actual documents in the field since many of them will include keywords not containing the root "agricult*." On the other hand, it is not possible to enumerate all keywords relevant to the domain. Thus, we need an efficient way to identify the literature related to agriculture.

We overcome the difficulty of efficiently identifying the relevant literature by starting from a broad query that contains "DSS" in the title. This query returns 6000 documents. Then, we narrow down the candidate publications by utilizing a naïve Bayes text classification algorithm, as this method has been used for similar tasks in the past [9, 10].

Then, in order to identify the evolution of the literature and their interrelations, we utilize a network analysis approach. We use the VOSviewer tool [11] to construct two bibliometric networks, one with the author keywords and one with countries. For the first type, we use the co-occurrence analysis in which the proximity of the keywords in the network is determined by "the number of documents in which they occur together" [12]. For the countries' network, we use the co-authorship analysis according to which the spatial relation of the country affiliations in the network is determined by the degree of collaboration for producing research.

The aforementioned networks allow us to identify how the scientific community combined different terms, and not only which key terms that the authors were most interested in during different periods but also how the transnational scientific collaboration evolved.

## 2 Methodology

### *2.1 Naïve Bayes Text Classification*

Naïve Bayes is a simple machine learning technique widely used for text classification with satisfying accuracy [13–15].

The core idea of the naïve Bayes classifier is that we update our prior belief on the probability that a document belongs to a class using the likelihood of the set of words of this document given the class. We estimate the likelihood of using a set of explicitly classified documents (training set). We can utilize either the presence or absence of words using a binomial Bayes classifier or we can use their frequencies and apply a multinomial naive Bayes.

In mathematical terms, we express the naïve Bayes as

$$P\left(\text{class}_\text{j} | \mathbf{words_i}\right) = \frac{P\left(\mathbf{words}_{\boldsymbol{i}} | \text{class}_j\right) \bullet P\left(\text{class}_j\right)}{P\left(\mathbf{words}_{\boldsymbol{i}}\right)}$$

Where

- $\text{class}_j$ are the classes we want to classify the documents into. In our case, we have two classes: the *valid* class, for agriculture domain-related documents; and the *invalid* class for non-agriculture domain-related documents.
- $\mathbf{words}_{\boldsymbol{i}}$ are the words of document-i.
- $P(\text{class}_j)$ is the prior probability of class-j. In our case, we used the frequency of the documents of the training set that belong to the agricultural domain.
- $P(\mathbf{words}_{\boldsymbol{i}})$ is the probability of a certain bag of words to appear.
- $P(\mathbf{words}_{\boldsymbol{i}} | \text{class}_j)$ is the likelihood of observing a certain bag of words for class-j. In naïve Bayes, we assume the conditional independence of the words of the documents. That is, the probability of each word appearing in a document does not depend from the other words of the document. This is a naïve assumption and the reason that the method is called *naive*. Anyhow, given this assumption, we are allowed to estimate this likelihood easily as $P(\mathbf{words}_{\boldsymbol{i}} | \text{class}_j) = P(\text{word}_{i,1} | \text{class}_j) \bullet P(\text{word}_{i,2} | \text{class}_j) \bullet \cdots \bullet P(\text{word}_{i,n} | \text{class}_j)$
- $P(\text{class}_j | \mathbf{words}_{\boldsymbol{i}})$ is the posterior probability, i.e., the probability of the document-i belonging to class-j given its words. We compute the posterior probability of a document for all classes and classify the document to the class that has the higher one.

In order to prepare the data for the classifier, several preprocessing steps are essential [16–18]. Primarily, the documents must be broken down into the set

**Table 1** Document feature matrix, as shown from the quanteda R package [19]

| Document-feature matrix of: 6 documents, 1637 features (98.3% sparse), and 6 docvars | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | Features | | | | |
| docs | object | knowledg | enhanc | abil | maker | task | provid |
| 2-s2.0-0025505218 | 1 | 1 | 1 | 1 | 1 | 2 | 2 |
| 2-s2.0-0038176956 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2-s2.0-0025700733 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2-s2.0-0025444756 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2-s2.0-0025431677 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 2-s2.0-7044990132 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| [reached max_nfeat ... 1627 more features] | | | | | | | |

of individual words (tokenization). Tokens can also include consecutive word combinations (n-grams). For example, the sentence "I read many books" will include the 1-g tokens {"I," "read," "many," "books"} and the 2-g tokens {"I read," "read many," "many books"}. After tokenization, the words that provide little information are removed (at minimum, the so-called stopwords, e.g., "a," "and," and "if"). Letters can be converted to lower case and punctuation or numbers are removed if they do not convey information. Finally, most often, the words are transformed into their root form, e.g., the words "ability" or "abilities" were converted to "abil" (this process is called *stemming*).

The preprocessing step will result in a *document feature matrix* (DFM) structure. A DFM is a matrix where documents are in the rows and their containing words (their features) in the columns and the values are the frequency of each word for each document. In the example below (Table 1), we give an excerpt of a DFM of the abstracts we downloaded from SCOPUS. Each row is a different paper abstract, each column is a word detected and the numerical values of the table show the frequency of each word in each abstract. In Table 1, in the first row, the abstract of the first publication (id=2-s2.0-0025505218) contains the words "object" one time, "knowledge" one time, "enhanc" one time, etc. In the second row, the second document (id=2-s2-0.-0038176956) contains the word "object" two times, the word "knowledg" zero times, etc.

## 2.2 *Bibliometric Network Analysis*

The basic concept behind network analysis is that the construction of a network allows the exploration of complex and multi-factorial subjects or fields by the visualization of their interconnections with nodes and vertices. Its combination with bibliometric analysis is termed bibliometric network analysis.

Initially, the bibliometric analysis was limited to extracting mostly descriptive statistical indices to evaluate the progress of the academic research, based on the creation of simple productivity indicators of the authors or countries. Gradually, more complex indexes were introduced which allowed the researchers to be able to identify the emergence of new multi- and trans-disciplinary fields.

Nowadays, with more advanced visualization techniques available, the most contemporary statistical indices can be implemented in a mapping procedure of the networks of the scientific literature. In this chapter, we have used the visualization of similarities (VOS) developed by Van Eck and Waltman [11]. There, the mapping is combined with clustering methods to transcend the two-dimensional constraint of the former [11]. There is a sufficient amount of relevant literature in which the reader can get familiar with the VOS technique implemented in different scientific fields [20–22]. Although the VOS technique is, in principle, similar to the multidimensional scaling technique [23]; it is more visual-oriented as it is a distance-based mapping tool, which implies that the distance in which the nodes-terms are placed in the network, represents their relative relatedness which is defined by the method of analysis we have chosen.

Our choice for the VOS technique was based on four (4) central criteria [24]. At first, it is a broadly used technique. Secondly, it is accepted as a reliable mapping technique. Thirdly, the VOS viewer tool is user friendly and; fourthly, it is an open source software. The four pre-mentioned factors increase both the accessibility and the repeatability of the results and thus, increase the validity of the present study.

## 3 Identification of the Relevant Literature

As already mentioned, we are interested in publications that (1) are focused on the implementation side of a Decision Support System and (2) respond to a problem in the agricultural domain.

We selected publications that complied with the first criterion by requiring the terms *decision support system*, *decision support systems,* or *dss* to be explicitly included in their title. Since the title of a publication signals the focal subject of the chapter, this requirement excludes all publications where DSS is incidental. This, although does not separate DSS-domain papers from DSS-implementation papers, must include the vast majority of papers that focus on presenting a DSS.

The following query in the SCOPUS database returns 13,747 documents.[1]

| |
|---|
| (TITLE ("DECISION SUPPORT SYSTEMS") OR TITLE ("DECISION SUPPORT SYSTEM") OR TITLE ("DSS")) AND PUBYEAR > 1989 |
| 13,747 documents |

To distinguish only the documents related to the agricultural/environmental domain, we initially remove items classified by the SCOPUS database to profoundly

[1]The same query on the title or abstracts or the keywords, returns 111,569 documents.

irrelevant fields (e.g., Medicine, Psychology). The refined SCOPUS query returns 9779 documents.[2,3] Furthermore, since we use the abstract and the keywords to facilitate our analysis, we also exclude the items that are missing either the abstract or the keywords and thus conclude to 6725 documents.

| |
|---|
| (TITLE ("DECISION SUPPORT SYSTEMS") OR TITLE ("DECISION SUPPORT SYSTEM") OR TITLE ("DSS")) AND PUBYEAR > 1989 AND (EXCLUDE (SUBJAREA, "MEDI") OR EXCLUDE (SUBJAREA, "BIOC") OR EXCLUDE (SUBJAREA, "HEAL") OR EXCLUDE (SUBJAREA, "ARTS") OR EXCLUDE (SUBJAREA, "PSYC") OR EXCLUDE (SUBJAREA, "PHYS") OR EXCLUDE (SUBJAREA, "MATE") OR EXCLUDE (SUBJAREA, "IMMU") OR EXCLUDE (SUBJAREA, "PHAR") OR EXCLUDE (SUBJAREA, "CENG") OR EXCLUDE (SUBJAREA, "NURS") OR EXCLUDE (SUBJAREA, "CHEM") OR EXCLUDE (SUBJAREA, "NEUR") OR EXCLUDE (SUBJAREA, "DENT")) |
| 9779 documents in the query/6725 documents with metadata on both abstract and author keywords |

Due to the high number of documents, the use of a semi-automated method for identifying the documents related to the agriculture domain is beneficial. Thus, we use the *Naive Bayes* classifier to accelerate the filtering of the publications.[4] The algorithm will be applied to the title, the abstract, and the author keywords of the 6725 documents.

The first step is to estimate the probability of a random document of the 6725 documents being related to agriculture. This will be the prior for the naïve Bayes. In order to do so, we randomly selected 1689 documents and manually inspected and classified them as either "related to agriculture" or "not related to agriculture".[5] We found that 18.85% of the sample (352 documents) was related to agriculture while the rest did not. The prior probability of the naïve Bayes is thus set to 0.1885.

Then, based on the above-classified sample (i.e., each paper classified as "related to agriculture" or "not related to agriculture"), we computed the likelihood for each word of the abstracts to appear on each of the classes ("related to agriculture"; "not

[2]The excluded subjects were: Medicine; Biochemistry, Genetics and Molecular Biology; Health Professions; Arts and Humanities; Psychology; Physics and Astronomy; Materials Science; Immunology and Microbiology; Pharmacology; Toxicology and Pharmaceutics; Chemical Engineering; Nursing; Chemistry; Neuroscience; Dentistry.

[3]The included subjects were: Computer Science; Engineering; Environmental Science; Business, Management, and Accounting; Decision Sciences; Social Sciences; Agricultural and Biological Sciences; Earth and Planetary Sciences; Energy; Economics, Econometrics, and Finance; Multidisciplinary; Veterinary.

[4]For processing the text and applying the naïve Bayes, we used the *quanteda* R package [19].

[5]We opted for randomly inspecting 1689 documents (25% of the 6725) for two reasons: (a) the higher the number, the most accurate the estimator of the prior; (b) on the other hand, we want to minimize the effort of manually classifying documents; inspecting abstracts and other metadata for 1689 documents is a reasonable effort for an extended literature review.

related to agriculture"). We also compute the probability of observing each word (the frequency a word appears in the corpus of documents).

Next, using the above information, we ran three naïve Bayes classifiers for the remaining non-classified documents (5036 out of 6725); one for titles, one for abstracts, and one for author keywords.[6] Thus, it was possible that a document is classified as "related to agriculture" based on the abstract, but not based on the title. In order to consolidate our findings, we used the following rules:

1. We classify a document as "related to agriculture" if the result of the naïve Bayes is "related to agriculture" in the following cases: in all three fields, i.e., title, abstract, keyword (rule 1.1); or in both title and abstract (rule 1.2); or solely in title (rule 1.3); or only in abstract (rule 1.4). The number of documents classified under these rules is given below:

| Classified as | Rule 1.1: Title AND Abstract, AND Keyword | Rule 1.2: Title AND Abstract | Rule 1.3: Title | Rule 1.4: Abstract |
|---|---|---|---|---|
| "related to agriculture" | 509 | 121 | 166 | 384 |

2. If a document is classified in all three fields (title, abstract, keyword) as "not related to agriculture," then we classify it as "not related to agriculture".

| Classified as | Rule 2.1: Title AND Abstract, AND Keyword |
|---|---|
| "not related to agriculture" | 3552 |

3. The remaining documents are classified as "unknown class".

| Classified as | Rule 2.1: Title AND Abstract, AND Keyword |
|---|---|
| "unknown class" | 115 |

Based on the above results, we manually inspected all documents classified as "related to agriculture" due to rule 1 and all documents classified as "unknown class." We also did a random sampling manual inspection on the "not related to agriculture" documents derived from rule 2. The results are presented in Table 2.

The results of the classification are presented in more detail in the next section.

[6]We did so because the information load may differ for each of the above document properties and those three document properties are of different nature and cannot be concatenated.

**Table 2** Naive Bayes results

| | | Rule 1 | | | | Rule 2 | Rule 3 | |
|---|---|---|---|---|---|---|---|---|
| Initial naïve Bayes classification: | | 1.1: Title AND Abstract, AND Keyword | 1.2: Title AND Abstract | 1.3: Title | 1.4: Abstract | Title AND Abstract, AND Keyword | Remaining documents | Already classified manually for the training set |
| Manual classify-cation | “related to agriculture” | 498 | 97 | 75 | 288 | 1 | 3 | 352 |
| | “not related to agriculture” | 11 | 24 | 91 | 96 | 3551 | 112 | 1337 |
| Accuracy of rule | | 97.8% | 80.1% | 45.1% | 75% | 99.2%[a] | – | – |

[a]Notes: For this estimation, since we reviewed manually a sample of 120 documents of this category, the accuracy was calculated as the number of correct guesses (119) to the sample size

# 4 Results

## *4.1 Naïve Bayes Classification*

The distribution of the identified documents in time is shown in Fig. 1 and in Table 3. We observe that there is a steady increase in the absolute number of publications, especially for journal papers. It is also interesting that 95% of the journals has published five or fewer papers in total in the 1990–2019 period and 75% of the journals have published 2 or fewer papers in the same period. Two journals seem to publish papers in the field regularly (Table 4).

In Fig. 2, we also provide the *word cloud* per 5-year period of the most frequent words in the identified publications' abstracts. The size of the word in each group denotes its frequency.

The metadata (titles, abstracts, and author keywords) of the initial data set, the results of the training manual classification, the results of the naïve Bayes, and the final classification are provided in the "01.naive_bayes_results.xlsx" file of the supplementary material.

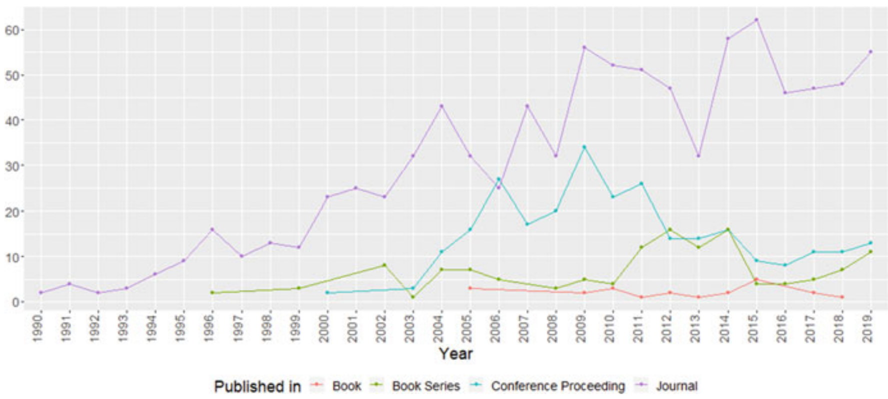


**Fig. 1** Evolution of DSS publications for agriculture, 1990–2019 (source: identified papers according to Sect. 2)

**Table 3** Number of DSS for agriculture publications for 5-year periods

| Period | Books | Book Series | Conference Proceeding | Journal |
|---|---|---|---|---|
| 1990/1994 | 0 | 0 | 0 | 17 |
| 1995/1999 | 0 | 5 | 0 | 60 |
| 2000/2004 | 0 | 16 | 16 | 146 |
| 2005/2009 | 5 | 20 | 114 | 188 |
| 2010/2014 | 9 | 60 | 93 | 240 |
| 2015/2019 | 8 | 31 | 52 | 258 |

**Table 4** Number of identified publications in most frequent journals

| Journal | 1990–1999 | 2000–2009 | 2010–2019 |
|---|---|---|---|
| Computers and Electronics in Agriculture | 25 | 27 | 30 |
| Environmental Modelling and Software | 4 | 38 | 25 |
| Agricultural Systems | 2 | 11 | 6 |
| Transactions of the Chinese Society of Agricultural Engineering | 0 | 18 | 3 |
| Water Resources Management | 0 | 8 | 9 |

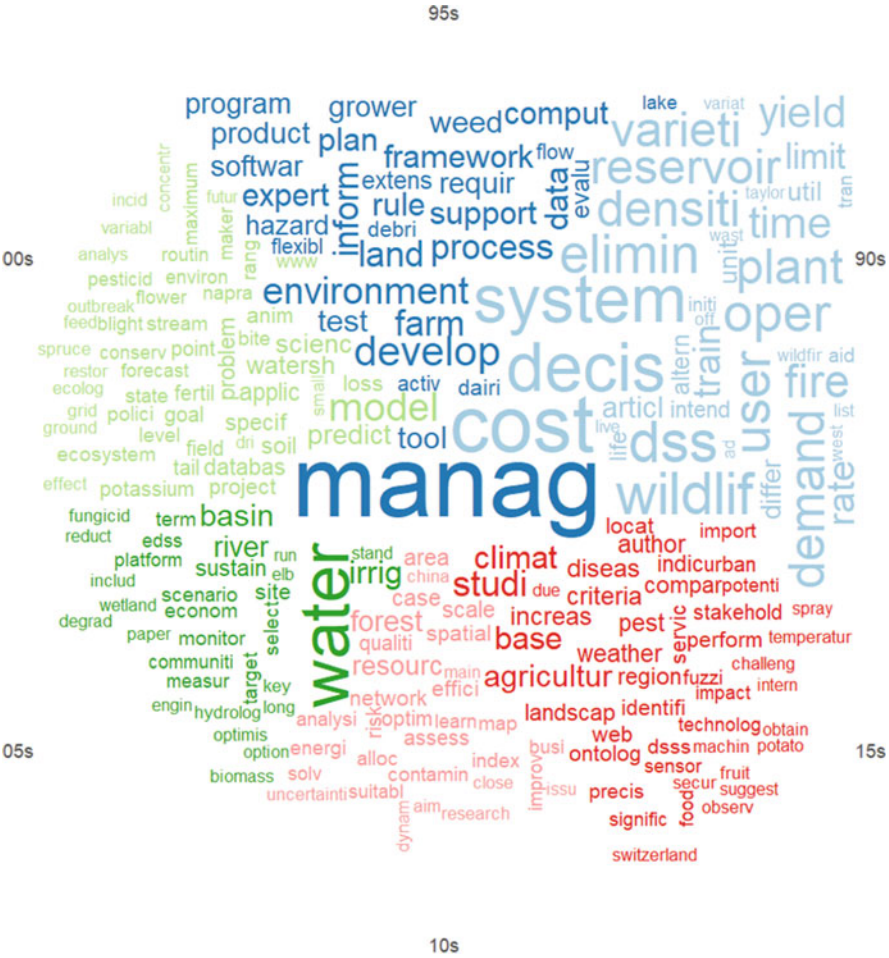


**Fig. 2** Word cloud of author-keywords quinquennially

## 4.2 Bibliographic Network Analysis

Using the metadata of the identified "DSS in agriculture" publications, we constructed two sets of networks.[7]

The first network contains the relationships of the selected papers according to the author keywords. We have constructed two such networks, one for 2000–2009 and one for 2010–2019 (Figs. 2 and 3). The nodes of the network represent author keywords. The size of the node indicates the number of occurrences of this keyword in our data set (i.e., in how many papers it appears). If there is an arc connecting two nodes, it means that those two keywords appeared in the same paper at least once. The width of the arc's line is a measure of the frequency those two keywords appear together in papers (more thick line, more times the keywords appear together). The distance between two nodes is a sign of their *relatedness*. The relatedness of two nodes is determined by the number of times the two keywords occur together, considering the relatedness of all other nodes that are connected with them. Thus, while the weight of the arc is a direct sign of the number of co-appearances in papers, the relatedness is a more holistic measure that displays the relation of the

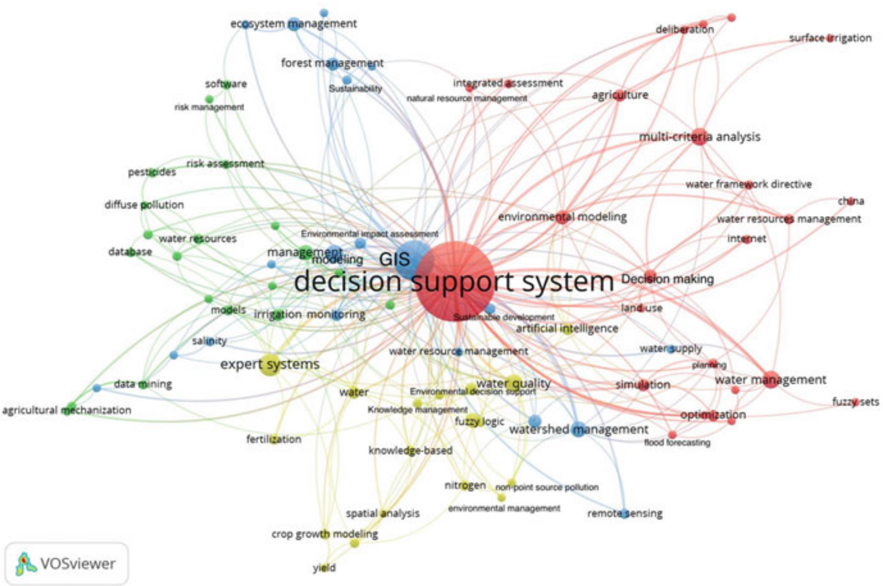


**Fig. 3** Network for the decade 1999–2008

[7]Certain quantitative properties of those networks (degree, betweenness, and closeness centrality) are provided in the supplementary material.

**Table 5** Relative frequency of selected keywords

| Period | 1990/94 (%) | 1995/99 (%) | 2000/04 (%) | 2005/09 (%) | 2010/14 (%) | 2015/19 (%) |
|---|---|---|---|---|---|---|
| **GIS** “*gis*”,“*geog*inf*” | 11.8 | 29.2 | 26.3 | 29.1 | 28.3 | 18.9 |
| **Machine Learning** “*mach*learn”, “*neural*”, “*genet*algo*”, “big*data*” | 0.0 | 3.1 | 2.8 | 2.1 | 2.5 | 7.4 |
| **Mobile technologies** “*mobil*”, “pda*”, “android”,“*wsn*” | 0.0 | 0.0 | 0.0 | 2.1 | 2.7 | 7.1 |

two nodes without ignoring the big picture. Finally, the colors of the nodes are a cluster of keywords that form “sub-networks” within the principal network.

The second network contains the countries, based on the papers’ affiliations, for the 2000–2009 and 2010–2019 periods (Figs. 5 and 6). The nodes are the individual countries and the arcs and the relatedness refer to the volume of co-authoring between countries. We have chosen this type of analysis as we believe, in principle, that the higher the number of co-authored documents between two countries, the higher the collaboration and the scientific proximity between them is.

## 5 Discussion

In both 1999/2008 and 2009/2019 author-keyword networks, the position of GIS is very central. That reveals that the initiation and the evolution of DSS are very much connected to GIS technologies. This could be attributed to the fact that spatial dimension is integral to agriculture, whether on the farm or on the policy level. Thus, the perspective of the GIS technologies naturally fitted into the agricultural decision-making framework. In turn, the rise in the availability of spatial data and the user-friendliness of the GIS interface resulted in the central position of this technology in the DSS in agriculture. If we look at the relative frequency of the GIS-related terms, it seems that the relative frequency in journal papers decline in the last 5 years (Table 5). In contrast, the share of emerging technologies, like machine learning and mobile networks is increasing. However, as shown from the network and is confirmed quantitatively by network metrics (see supplementary material), is that its central role is maintained.

The apparent changes of the last decade are the decrease in the frequency and the centrality of the ‘expert systems” and the appearance of the terms “climate change” and “precision agriculture” in a relatively central place. Also water-related keywords, like “watershed management” or “water,” decrease in centrality in the new decade, possibly because they are not anymore examined isolated but rather

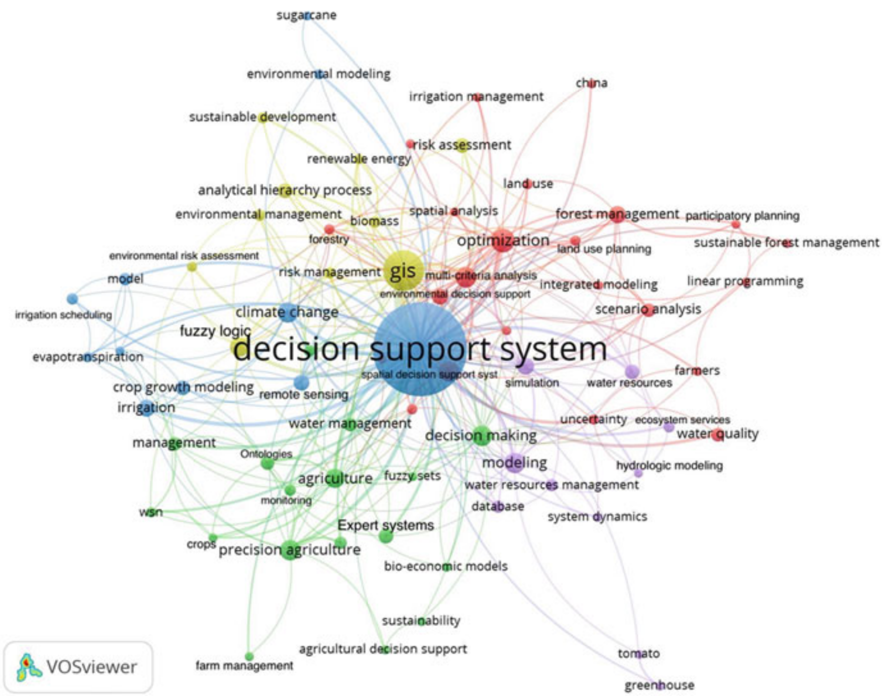


**Fig. 4** Network for the decade 2009–2019

under the “climate change” perspective. Regarding the methodologies, “simulation” and “multicriteria analysis” are becoming more central.

Regarding the clusters of the author-keywords network, we observe the following: In the 1999/2008 period (Fig. 3), there are two apparent clusters. The red cluster contains topics related to water management and environment, having the multicriteria methods at its center, while simulation, optimization, and integrated assessment are part of it. The yellow cluster is related to the topics of “water,” “fertilization,” “nitrogen,” and “yield,” and the “expert systems” methodology/technology. The existence of two clusters, containing both a water-related topic can be attributed to the fact that the second cluster targets the farm level while the second a more generic level (resources in general). This indicates that a different focus level affects the methodologies used. This becomes more apparent in the author-keyword of the second decade (Fig. 4). The green cluster is a farm level-related cluster, with many “management” keywords and with technologies like sensor networks (wsn) and precision agriculture.

In general, for both decades, there does not seem to be a very clear separation of keyword clusters. If this was the case, one would observe distinct groups of keywords, without nodes of one cluster positioned inside another cluster. This can be attributed to either the need for multidisciplinary solutions to actual decision-making problems.

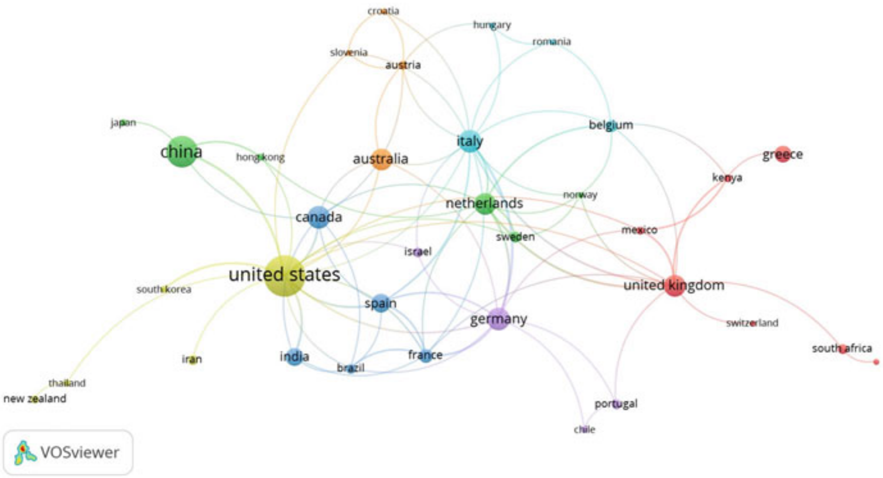


**Fig. 5** Country Network for the decade 1999–2008

Regarding the country co-authorship networks (Figs. 5 and 6), there are significant changes between the two decades. In the 1999/2008 period, the Netherlands take a central position, which means that authors of that country seem to have the most connections with authors of other countries. The United States, China, and the United Kingdom, although relatively significant in terms of publications, lie in more isolated locations of the network. This structure changes completely in the current decade (2010–2019). The group of European countries with dense connections between them emerges. European countries that in the previous decade were far from the other European countries (e.g., Greece, Austria), in the current one become part of this European network.

## 6 Conclusions

We have used a naïve Bayes text classification algorithm to identify around 1600 agricultural DSS papers. We then have constructed the author keywords co-occurrence network and the co-authorship network for countries, one for each of the 1999/2008 and 2009/2019 periods. The methodology applied accelerated the review process and accurately identified the relevant literature. It can be easily extended to other bibliographic databases (Google Scholar, Web of Science) and can be used to efficiently identify the literature of other subjects too.

We have found that the Geographical Information System technology has a central position in the discipline for both decades. New terms appear and take a central position in the current decade's network, like "climate change" and "precision agriculture." However, in all author-keyword networks, there are

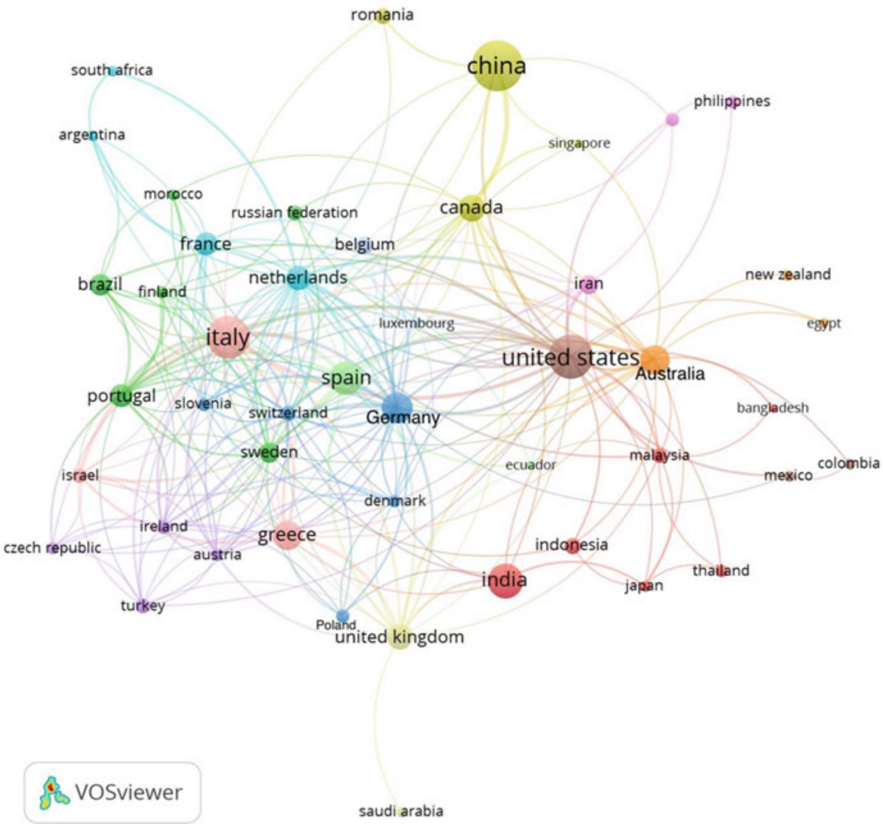


**Fig. 6** Country Network for the decade 2009–2019

no clear clusters, probably denoting the need for multidisciplinary solutions to actual decision-making problems. Regarding the evolution of the discipline in the country's dimension, the European countries emerge as a distinct group with dense connections between its members.

Regarding the future trends, DSS literature is increasingly connected to the new technological advances of mobile applications, machine learning, and the internet of things. DSS have a great potential for bringing these technologies in the farm. However, the concern of low intake from end-users of the DSS applications in agriculture remains and more research is required regarding the user-centered design of DSS [7].

## References


1. Airinei, D., & Homocianu, D. (2009). DSS vs. business intelligence. *Revista Economica*, 2009, Available at SSRN: https://ssrn.com/abstract=2381821.
2. Manos, B. D., Ciani, A., Bournaris, T., Vassiliadou, I., & Papathanasiou, J. (2004). A taxonomy survey of decision support systems in agriculture. *Agricultural Economics Review, 5*, 80–94.
3. Sun, Y. Y., & Shen, S. H. (2019). Research progress in application of crop growth models. *Chinese Journal of Agrometeorology, 40*(7), 444–459.
4. Zasada, I., Piorr, A., Novo, P., Villanueva, A. J., & Valánszki, I. (2017). What do we know about decision support systems for landscape and environmental management? A review and expert survey within EU research projects. *Environmental Modelling and Software, 98*. 63–74. https://doi.org/10.1016/j.envsoft.2017.09.012.
5. Zhai, Z., Martínez, J. F., Beltran, V., & Martínez, N. L. (2020). Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture, 170*, 105256.
6. Gutiérrez, F., Htun, N. N., Schlenz, F., Kasimati, A., & Verbert, K. (2019). A review of visualisations in agricultural decision support systems: An HCI perspective. *Computers and Electronics in Agriculture, 163*, 104844.
7. Rose, D. C., Parker, C., Fodey, J., Park, C., Sutherland, W. J., & Dicks, L. V. (2018). Involving stakeholders in agricultural decision support systems: Improving user-centred design. *International Journal of Agricultural Management, 6*(3–4), 80–89.
8. Rose, D. C., Sutherland, W. J., Parker, C., Lobley, M., Winter, M., Morris, C., et al. (2016). Decision support tools for agriculture: Towards effective design and delivery. *Agricultural Systems, 149*, 165–174.
9. Adeva, J. G., Atxa, J. P., Carrillo, M. U., & Zengotitabengoa, E. A. (2014). Automatic text classification to support systematic reviews in medicine. *Expert Systems with Applications, 41*(4), 1498–1508.
10. Gulo, C. A., Rúbio, T. R., Tabassum, S., & Prado, S. G. (2015). Mining scientific articles powered by machine learning techniques. In *2015 Imperial College Computing Student Workshop (ICCSW 2015)*. Schloss: Dagstuhl-Leibniz-Zentrum fuer Informatik.
11. Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538.
12. Van Eck, N. J., & Waltman, L. (2013). VOSviewer manual. *Leiden: Univeristeit Leiden, 1*(1), 1–53.
13. Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing, 36*(1), 20–38.
14. Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval. Online Edition* (p. 258). Cambridge: Cambridge University Press. ISBN: 0521865719.
15. Pranckevičius, T., & Marcinkevičius, V. (2017). Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Baltic Journal of Modern Computing, 5*(2), 221.
16. Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications, 36*(3), 5432–5435.
17. Raschka, S. (2014). Naive bayes and text classification i-introduction and theory. arXiv preprint arXiv:1410.5329.
18. Schneider, K. M. (2005, February). Techniques for improving the performance of naive bayes for text classification. In *International conference on intelligent text processing and computational linguistics* (pp. 682–693). Berlin, Heidelberg: Springer.
19. Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software, 3*(30), 774. https://doi.org/10.21105/joss.00774, https://quanteda.io.

20. Cancino, C., Merigó, J. M., Coronado, F., Dessouky, Y., & Dessouky, M. (2017). Forty years of Computers & Industrial Engineering: A bibliometric analysis. *Computers & Industrial Engineering, 113*, 614–629.
21. Sweileh, W. M., Al-Jabi, S. W., AbuTaha, A. S., Sa'ed, H. Z., Anayah, F. M., & Sawalha, A. F. (2017). Bibliometric analysis of worldwide scientific literature in mobile-health: 2006–2016. *BMC Medical Informatics and Decision Making, 17*(1), 72.
22. Van Eck, N. J., & Waltman, L. (2011). Text mining and visualization using VOSviewer. arXiv preprint arXiv:1109.2058.
23. Borg, I., & Groenen, P. J. (2005). *Modern multidimensional scaling: Theory and applications*. Berlin: Springer Science & Business Media.
24. Konstantinis, A., Rozakis, S., Maria, E. A., & Shu, K. (2018). A definition of bioeconomy through bibliometric networks of the scientific literature. *AgBioforum, 21*(2), 64–85.