CHAPTER V

DESCRIPTION OF SOURCE DATABASE AND STATISTICAL TECHNIQUES EMPLOYED

5.1 Introduction

This study was aimed to examine India and G7 countries quantitative growth of literature in the field of 'Textile Research' with the help of a source database SciVerse SCOPUS. The study is mainly exploratory and analytical in nature because it uses various statistical tools and techniques in this Chapter, a description about the source databases, the bibliometric/scientometrics indicators, statistical tools and scientometrics indices too have been outlined.

5.2 Source Database

SciVerse SCOPUS database is a comprehensive bibliographic database covering the worldwide literature in engineering and technology. SCOPUS is produced by Elsevier Properties. SciVerse® is a registered trademark of Elsevier Properties, S.A. SCOPUS only indexes serial publications (journals, trade journals, book series and conference material) that have an ISSN (International Standard Serial Numbers) assigned to them. The only exception concerns conference papers, which can be captured via different routes than by being published in a serial publication with an ISSN. Trade journals are included in SCOPUS because users and librarians consider selected articles to be scientifically relevant. A special document type policy for trade journals was introduced in 2008, to ensure that only articles or reviews of scientific relevance are included in SCOPUS. The minimum requirements for items in trade journals to be
captured SCOPUS database are 1) minimum of 1 page, 2) minimum of one mentioned author. The book series in SCOPUS include Science Direct Handbooks whose records are visible as articles or reviews. Conference materials included in SCOPUS in two different ways such as special issues of regular journals and in the form of dedicated conference proceedings. Proceedings can be published as a serial or non-serial and may contain either the full articles of the papers presented or only the abstracts.

The source title usually includes words like ‘proceeding(s)’, ‘meeting(s)’, ‘conference(s)’, ‘symposium/ symposia’, ‘seminar(s)’ or ‘workshop(s)’ (or synonyms in other languages like ‘Tagungsberichte’ etc.), although some Journals also have titles with the word ‘Proceedings’. SCOPUS covers conferences that publish full-text papers, e.g. document type “conference papers” whereas conferences that publish only abstracts (“meeting abstracts”) are not considered for coverage. Conference coverage in SCOPUS is focused primarily on those subject areas where conference papers represent a substantial portion of the published research, e.g. engineering, computer science, and some areas of physics.

A non-serial source is a publication with an ISBN unless it is a report, part of a book series, proceeding (non-serial), or patent. Usually, it is a monograph or composed work. A book can have different physical formats (e.g. print, electronic). There are hundreds of thousands of scientific books that have been published (both in print and out of print), but these come with many challenges when attempting to include them in an A&I database such as SCOPUS. Currently, the policy at SCOPUS does include books.

**Model used**

One of the most obvious features of modern science in recent years has been the spectacular development of scientific discoveries and knowledge producing an
unprecedented growth of literature in different fields. Different scholars have tried to fit different types of growth models to the publication data. The models applied in the past for analyzing the growth of literature in different specialties are linear growth model (Egghe, L., 1992)\(^{28}\) exponential and growth model, (Gompertz model, 1998) power model and logistics growth model (Rao, I.K.R., 1998). However, in the present scenario, growth model has been applied as it gives the best fit, as illustrated by Jain and Garg in their study on textile research.

5.3 **Used Bibliometrics/Scientometrics Indicators and Statistical Techniques Employed**

In this study, the following bibliometric/scientometrics indicators and statistical techniques were employed while analyzing the data on Nanotechnology research output collected from the SCOPUS database.

- Growth Rate (GR)
- Collaboration Index (CI)
- Degree of Collaboration (DC)
- Collaborative coefficient (CC)
- Co-authorship Index (CAI)
- Activity index (AI)
- Impact Factor (IF)
- High Quality Papers (HQP)
- Citation Per Paper (CPP)
- Relative Citation Index (RCI)
- H-Index (HI)
5.3.1 Growth Rate (GR)

One of the most obvious features of science in recent years has been its rate of growth. Scientific growth is also responsible for increase in manpower and finance. The flood of papers represents one aspect of the general growth of scientific communication. Wooster (1970) has estimated the number of journals that existed in the world at any one time, where as some estimates the number of papers published annually at various times was done by Vickery (1968). Martyn (1973) Gottschalk and Desmond (1963) have also estimated the number of scientific and technical journals that exist in the world. Growth studies in other scientific areas included the work of Baker (1976) in Chemistry, Conard (1957) in Biology, May (1966) and Lamb (1971) in Mathematics, Sengupta (1973) in Microbiology, Physiology and Biochemistry.

5.3.2 Collaboration Index (CI)

The simplest of the indices presently employed in the literature is the collaboration index, CI, which is to be interpreted merely as the mean number of authors per paper: and is given by k (4.3).

\[
CI = \frac{\sum_{j=1}^{A} J_{fj}}{N} 
\]

j is the number of co-authored papers appearing in a discipline; N is the total number of papers in the discipline over the same time interval, and k the greatest number of authors per paper in a discipline. As pointed out by Ajiferuke et al. (1988) this is to be
interpreted merely as a mean, for in the absence of an upper limit there is no way of interpreting the numbers generated, and secondly, the method imputed a nonzero weight to single authored papers. To overcome this index referred to as the degree of collaboration is introduced, where single-author papers have zero-weight.

5.3.3 Degree of Collaboration (DC)

Subramanyam’s (1983)\textsuperscript{139} formula (4.4) has been adopted to examine the extent of research collaboration in the study.

$$C = \frac{Nm}{Nm + NS} \quad (4.4)$$

whereas

$$C = \text{degree of collaboration in a subject}$$

$$Nm = \text{number of multiple authored papers}$$

$$Ns = \text{number of single authored papers}$$

5.3.4 Collaborative Coefficient (CC)

The index CC is given to overcome the disadvantages of the collaborative index and to make it possible to draw a comparison between different sub-disciplines. In order to make a relevant comparison, consider the collaboration coefficient. The patterns of co-authorship among different countries have been examined by making use of Collaborative Coefficient (CC) suggested by Ajiferuke (1988). The formula (4.5) used for calculating CC is given below:
\[ \text{CC=I}\sum \]

where as

\( F_j \) = the number of authored papers
\( N \) = total number of research published; and
\( k \) = the greatest number of authors per paper

### 5.3.5 Co-authorship Index (CAI)

To study how the pattern of co-authorship, the use of Co-authorship Index as suggested by Garg and Padhi (2002)\(^{32}\) has been made and is explained below. For calculating Co-authorship Index (CAI) (4.6), the entire data set was divided into five blocks.

\[
\text{CAI} = \left( \frac{\{N_{ij}/N_{io}\}}{\{N_{oj}/N_{oo}\}} \right) \times 100 \quad (4.6)
\]

whereas

\( N_{ij} \): Number of papers having \( j \) authors in block \( i \);
\( N_{io} \): Total output of block \( i \);
\( N_{oj} \): Number of papers having \( j \) authors for all blocks;
\( N_{oo} \): Total number of papers for all authors and all blocks;
\( j = 1, 2, 3, 4 > 5 \)

### 5.3.6 Activity Index (AI)

Activity Index, also known as Priority Index, was first proposed by (Frame, 1997)\(^{29}\) and has been elaborated by (Schubert and Braun, 1986)\(^{121}\). It characterizes the relative research effort a nation or an institution devotes to a given subject field.
Activity Index (AI) is defined as follows:

\[ AI = \left( \frac{\text{The country’s share in world’s publication output in the given field}}{\text{The country’s share in world’s publication output in all the fields}} \right) \times 100 \]

The value of AI=100 indicates that the research effort of a country/institution in a given field corresponds precisely to the world’s average; AI>100 reflects higher than average activity and AI=<100 lower than average effort dedicated to the field.

The major advantage of using activity index over the raw count of publications is that it takes into account both the size of the nation/institution as well as the size of the discipline.

### 5.3.7 Impact Factor

At present, there is no better indicator applicable in practice characterizing the scientific impact of journals than the impact factor suggested by (Garfield, 1960)\(^{31}\). Although in bibliometric, there are other indices as well like influence factor suggested by (Narin, 1976)\(^{88}\) their use has not still become widespread. The Garfield’s impact factor “is basically a ratio of the number of citations a journal receives to the number of papers published over a period of time”.

Since impact factor varies from field to field and the output in textile research and technology is published in journals ranging from optics to medicine, it is better to use normalized impact factor. Use of normalized impact factor for assessment is similar to weighted publication count development by Computer Horizons Inc (CHI), in which each paper in a journal is given an influence based on the citation properties of the journal rather than the citation counts for individual papers.
5.3.8 Properties of High-Quality Papers

(Nagpaul, 1985) has also suggested this measure. Papers having normalized impact factor $> 3$ have been considered as high-quality papers. The value 3 is based on the average of the normalized impact factor of all the journals in which the papers are published.

5.3.9 Citations per Paper (CPP)

Based on the publication output and the number of citations received by these papers, citation per paper for different countries and different institutions have been calculated. Citation per paper has been calculated by using the following formula (4.10)

\[
\text{CPP} = \frac{\text{Total Number of citations for a country or an institution}}{\text{Total number of quality papers}} \times 100
\]

5.3.10 Relative Citation Impact (RCI)

The citation impact for a group of researchers is basically the number of citations per paper that that group has received over a certain time period. A relative citation impact can be calculated for any group of researchers and compared to an appropriate baseline.

The citation impact for a field of research can be calculated for, say, the world, and then the citation impact can be calculated for various institutions around Australia and their impact, relative to the world, can be compared as a measure of quality. By this definition the RCI for the world is always 1.00 for any field and therefore if the RCI for an institution is greater than 1.00, then it is performing above the world average for that
field. Conversely if the RCI is less than 1.00, the institution is performing poorly compared to the average for that field in the world. Often the source of the data may be provided in such a way as to make direct comparisons with groups or individuals quite difficult. This may include inconsistencies in the naming of research fields and subject categories.

5.3.11 H-Index

To calculate your h-index take a list of your publications and sort them from the most cited to the least cited. Scan down the list until you find the minimum number of publications with an equal number of citations.

For example, if you have at least 5 publications with 5 or more citations each then your h-index is 5 (this is regardless of your total number of publications), or if you have at least 30 publications each with 30 or more citations then your h-index is 30. If you have a high h-index you can consider yourself to be a high quality and highly productive researcher.

Note though that the h-index was originally designed for the field of physics research. The h-index works quite well for the physical sciences but may not directly transfer to other fields of research such as the humanities and social sciences. There are also now alternatives to the h-index which take into account the length of your research career.

5.4 Bibliometrics/Scientometrics Indices

Standard bibliometric indicators such as the number of publications (P) during the study period, number of citations (C) during the study period and the average citation per paper (CPP) have a number of limitations. Based on science managers and policy makers request and to support research decisions, it is required to scale-up the
bibliometric studies. Different measures and indices have been developed at this level of analysis. One type of indices, such as the h-index and g-index, describe the most productive core of the output of a researcher and inform about the number of papers in the core. The h-index is supposed to measure the broad impact of an individual scientist and to avoid all the disadvantages. Moreover, the online database such as Web of Science, SCOPUS, Google Scholar provides the h index. Other indices, such as the a-index and m-index, depict the impact of the papers in the core.

Bibliometrics methods are being used more and more often for evaluation purposes (Pritchard 1969). Large electronic database enable a reasonably fast determination of publication lists and corresponding citation records. But for a comparison of different datasets, the dangerous idea to quantify the research output by a single number remains fascinating. Simple indicators as the total number of citations to all papers or the average citation frequency have obvious disadvantages like the difficulty to determine all the citation counts with reasonable accuracy or giving undue weight to highly cited review articles, or taking a possibly large number of irrelevant (not or lowly cited) papers into account. This can be avoided by considering only a small number of relevant or significant papers, but this solution raises the question how to determine this core set of significant papers from a given set of publications. Taking a fixed number or a certain percentage of all publications into consideration would mean a somewhat arbitrary and biased choice.

Hence to solve this problem, Hirsh introduced h-index. The h-index, also known as the Hirsch index, was introduced (Hirsch 2005), as an indicator for lifetime achievement. Considering a scientist’s list of publications, ranked according to the
number of citations received, the h-index is defined as the highest rank such that the first h publications received each at least h citations. The h index is not an average, not a percentile, not a fraction; it is a totally new way of measuring performance impact, visibility, quality, etc., of the career of a scientist. It is a simple measure without any threshold. Based on this h index, various indices are developed for evaluating the career of individual scientists according to their scientific output.

Conclusion

The bibliometrics/scientometrics indicators and statistical tools and techniques described in this chapter are used for data analysis. Besides these indices, there exists a number of bibliometrics/scientometrics software that was not employed in this study. In the next chapter, the data collected from the source database have been analyzed to find the growth of textile research with the application of above-explained bibliometrics / scientometrics indicators and statistical tools and techniques.