Technological Forecasting & Social Change 78 (2011) 408–420

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Bibliometric analysis of distributed generation

Wei Lee Woon [a,*], Hatem Zeineldin [a], Stuart Madnick [b]

[a] Masdar Institute of Science and Technology, MASDAR, PO Box 54224, Abu Dhabi, United Arab Emirates
[b] Sloan School of Management, M.I.T., E53-321, Cambridge MA, 02139, USA

ARTICLE INFO

Article history:
Received 4 April 2010
Received in revised form 30 July 2010
Accepted 30 August 2010

Keywords:
Bibliometrics
Distributed generation
Technology mining

ABSTRACT

This paper presents an application of term frequency (*TF*) as a means of identifying useful trends from text documents. Of particular interest is the relationship between publication patterns, as characterized by *TF*, and the underlying technological developments. To demonstrate the usefulness of our approach, a case study on distributed generation (DG) was conducted. Important sub-domains of DG research were identified and the associated *TF* values were extracted using relevant keywords. The evolution of these values through time helps to highlight key trends in the development of DG-related technologies.

1. Introduction

The planning of research and development is a difficult task which is compounded by the availability of a large amount of data from a variety of channels including publication and patent databases. Access to these databases is widely available with the advent of the Internet while the availability of high performance computers has also increased dramatically. The combination of these two factors strongly motivates the use of quantitative methods to detect technological trends and patterns present in this data.

In this paper, we present a simple process for conducting a bibliometric based study focused on a given domain of research. The case study presented here targets the field of distributed generation, where our objective is to study the evolution of individual sub-fields within DG.

This paper is structured as follows. The overall motivation and context for this research was described in this section. Section 2 briefly reviews the relevant literature, while Section 3 details the data collection and research methodology used. In Section 4 the main results and observations are described. Section 5 analyzes these results in the context of specific developments and trends within the DG domain. Section 6 concludes the paper by summarizing the main findings and presenting suggestions for future research.

2. Background

Using bibliometrics to study the progression of research and technological development is not a new idea and there is already a significant body of research addressing this problem (for a good review, the reader is referred to [1–4]). Interesting examples include visualizing the inter-relationships between research topics [1,5] and technologies [6], identification of important researchers or research groups [3,7], the study of research performance by country [8,9], the study of collaboration patterns [10–12], the study of the innovation process [13,14] and the prediction of future trends and developments [5,15–17]. Nevertheless, given the many difficulties inherent to these undertakings, there is still scope for further development, as well as for testing and fine-tuning these methodologies via appropriately scoped case studies. The research described in this paper was motivated by, and seeks to address this need.

* Corresponding author.
E-mail addresses: wwoon@masdar.ac.ae (W.L. Woon), hzainaldin@masdar.ac.ae (H. Zeineldin), smadnick@mit.edu (S. Madnick).

0040-1625/$ – see front matter
doi:10.1016/j.techfore.2010.08.009

To focus discussions and to provide a concrete domain on which the proposed approach can be demonstrated, a case study was conducted on the area of distributed generation (DG). Broadly speaking, DG is a generation source that is connected to the distribution system close to the load point. Distributed generation could be classified into two main types: renewable such as wind and solar or non-renewable such as diesel generators and micro-turbines. The increasing penetration of DG in distribution systems coupled with the growing interest among the electrical power engineering community to meet environmental and energy efficiency constraints resulted in an increasing interest in DG. Thus, distribution systems world-wide are experiencing significant changes and challenges due to the increasing penetration of DG.

In general, distribution systems are radial in nature, meaning that power flows in one direction from the main substation to the load point. The addition of DG on the distribution system transforms the distribution system into a multi-source system. Topics such as distribution system protection, control and stability, which were considered simple and easy to design for distribution systems, have now become complex and of major importance [18–20]. The planning and reliability of the distribution system will be affected with the addition of DG sources [21–24]. In addition, new concepts and operational issues have emerged from this new distribution system structure which includes Smart-Grids, Micro-Grids and islanding detection [25–27].

As such, while DG research provides a well-defined scope in which to conduct our analysis, it still spans a wide range of technical areas and promises to be a rich and challenging problem domain on which to conduct our analysis. In addition, we note that bibliometric analysis of a variety of energy related issues has been conducted before (for e.g.: energy research in general [28], power sources [29], bio-fuels [30]), but not in the DG domain. As will be demonstrated, the application of bibliometric techniques can be extremely helpful in highlighting the main directions among the power engineering community in the area of DG.

3. Methods

The main aim of this study is to investigate the use of *TF* for monitoring the development of DG-related technologies.

To conduct the case study, we collect a set of records consisting of journal and conference publications relevant to the field of distributed generation. Two databases were considered for use in this study: ISI's Web of Science (WOS) and Elsevier's Scopus database. The following keywords were submitted to both databases through their respective web interfaces (as a title/abstract search): {*distributed generation, dispersed generation, distributed resources, embedded generation, decentralized generation, decentralized energy, distributed energy, on-site generation*}.

WOS returned with a total of 1348 records, while Scopus yielded 5436 records. In addition, the records were compared based on the article titles, and it was found that 1036 titles were common to the two databases ($\approx$77.37% of the records contained in the WOS result set). As such, we used the results from Scopus; these were saved locally and used to represent the body of research in DG.

As mentioned in Section 1, an important task was to identify subsets of this body of research. Focusing on these subsets should allow the general growth and direction of the DG domain to be studied. We discuss this in terms of the following three-stage process:

1. Identification of comparable topics or technologies.
2. Extraction of relevant/related studies.
3. Normalization and pre-processing.

3.1. Distributed generation: topics and themes

To obtain results that are interesting and useful, it is important that the analysis be appropriately framed. For example, comparing the relative growths of two sub-areas of DG would only make sense if they were broadly comparable. As research in DG is extremely diverse, we define three separate dimensions in which to conduct our analysis:

1. DG technology — i.e. part of the attraction of using a DG system is that it allows energy from a variety of localized sources to be integrated into the grid. As such, one way of decomposing the field of DG research is in terms of the technologies for generating this electricity. In [31], a comprehensive survey on the different DG types and technologies was presented. Based on [31], some of the most dominant DG technologies include wind turbines, photovoltaic, fuel cells and micro-turbines. Hence, we choose to study the following four topics: (1) *Wind,* (2) *Micro-turbines,* (3) *Solar PVs* and (4) *Fuel cells.*
2. DG interfaces — another important component of a DG system is the interface to the utility grid or to the customer. As highlighted in [32], almost all DG can be grouped into three main types, according to the type of interface, which include inverter, induction and synchronous. Thus, publication patterns for the following three classes of DG interfaces will be analyzed: (1) *Inverter based,* (2) *Synchronous based* and (3) *Induction based.*
3. General studies — besides the above two categories, research in DG also spans a range of other challenges and issues. The third dimension in our analysis groups a number of these topics: (1) *Control,* (2) *Reliability,* (3) *Islanding detection,* (4) *Planning,* (5) *Stability,* (6) *Power quality,* (7) *Electricity markets/economics,* (8) *Protection,* (9) *Forecasting* and (10) *Micro-Grids/Smart-Grids.*

To provide additional depth and breadth to our study, we supplement our results with two further forms of analysis:

1. To elucidate more detailed trends in the DG domain, the database is further “sliced and diced” by searching for occurrences of terms in the second dimension (inverter, synchronous and induction based) within the context of the topics listed in the third dimension.
2. High level statistics regarding the distribution of publications amongst countries and journal titles are also presented to give a broader perspective of the current state of DG research.

3.2. Data extraction

Once the initial database is constructed, patterns and trends of individual research topics can be easily extracted and graphed using a combination of appropriate SQL queries and regular expressions.

In general, we are interested in the level of research activity or interest in a particular topic. Of course, the true research activity is unobservable, but we postulate that *TF*, the *frequency* at which an associated term appears in the academic literature, can serve as a proxy variable for this interest. Note that this is not a new idea (for relevant examples, see [33–36]) though the exact form by which the term frequency is calculated may differ slightly. Given that the academic literature is the primary channel for disseminating research findings, it is reasonable to expect that the term frequency approximates the rate at which research on a particular topic is generated.

To obtain a measure of term frequency for a term $\mathcal{T}_i$, the number of occurrences of this term for a given year is counted then normalized by the total number of words in all of the abstracts in the year as follows:

$$TF_i(t) = \frac{n_i}{|\mathcal{A}_t|} \tag{1}$$

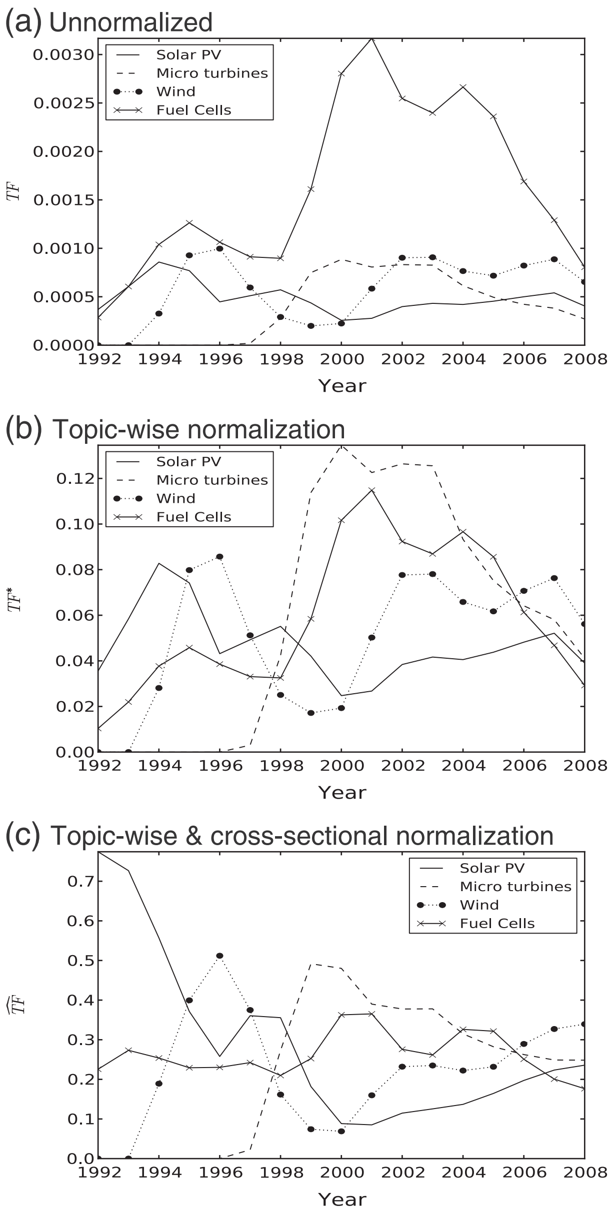

Fig. 1. DG technology.

where n_i is the number of occurrences of term $\mathcal{T}_i$ in all abstracts published in the year t, and $\mathcal{A}_t$ is the string formed by concatenating all the words in the abstracts in that same year.

Regular expressions are used to detect occurrences of terms in the document. Note that, where variations of a phrase are known (for example, "microturbine" and "micro-turbine"), regular expressions that were general enough to detect all variants were used (the list of regular expressions that were used are listed in Appendix A).

3.3. Normalization and post-processing

One problem with using *TF* is that it tends to favor very common terms, such as "Fuel cells", over terms which are more specific like "Micro-turbine". In Fig. 1(a), it can be seen that the lines for "Fuel cells" and "Wind" are a lot higher than the other two lines. However, in many cases we are less concerned with the absolute value of *TF* than with the overall *trend* observed in these figures — a low absolute value for *TF* may be secondary to the fact that these values are doubling every year, for example. To better observe these trends, we define TF^*, a normalized form of *TF*:

$$TF_i^*(t) = \frac{TF_i(t)}{\sum_{j=1992}^{2008} TF_i(j)}. \tag{2}$$

Doing this for the four topics listed above results in the graph in Fig. 1(b). From this sub-figure, it can be seen that, if we disregard the absolute values, it would appear that "Fuel cells" and "Micro-turbines" have broadly similar trends over the last 16 years, as do "Solar PV" and "Wind". Note that, while the two graphs contain essentially the same information, it is evident how appropriate normalization allows different aspects of the data to be more easily noticed.

Finally, an alternative form of normalization was considered:

$$\widehat{TF}_i(t) = \frac{TF_i^*(t)}{\sum_{j=1}^{m} TF_j^*(t)} \tag{3}$$

where m is the number of topics being studied concurrently. This would hopefully allow the growth in the different research areas to be more easily compared. When applied to the four topics from above, the graph in Fig. 1(c) is obtained. Note that, while this graph is quite similar to Fig. 1(b), there is an important difference early on in the graph where, in Fig. 1(c), the plot for solar PV related research is seen to be starting from a high value and slowly decreasing while for wind the curve shows an early rise in $\widehat{TF}$. This highlights the fact that, while both topics can be seen gaining in popularity early on, the $\widehat{TF}$ for "wind" gradually gains on solar PV and exceeds it by around 1995.

In brief, what we seek to demonstrate is not that any one of these graphs are necessarily "correct" or "wrong", but that different forms of normalization highlight different aspects of the data. As such, there is value in considering all the different forms of normalization.

3.4. Implementation

The required computational tools were implemented using the Python programming language and the SQLite database engine, as these facilitated faster development. The former also includes a broad selection of libraries, including those useful for the analysis of text and for data collection from the WWW. Both products are also cross-platform and open-sourced, which allowed applications to be deployed on a range of different operating systems and environments, and at a very low cost.

4. Results

The data collected from the Scopus database is used to generate yearly frequency values for the research areas listed in Section 3.1. In addition, a three-tap Gaussian filter was used to smooth the resulting time series as this was quite noisy and in some cases the number of publications retrieved were quite low. This is a reasonable pre-processing step because the research which results in a publication would have been carried out over a period of time prior to the appearance of the publication; as these publication counts serve as a proxy for the underlying research activities, smoothing the raw data in this way also helps to account for this spread.

Four different sets of graph are presented here, corresponding to the three "dimensions" described in Section 3.1 (dimension three comprised a large number of topics and was split into two sets of topics); these are presented in Figs. 1–4. In the following subsections, the initial observations are noted and discussed. In Section 5 a more detailed analysis will be presented, which will take into account the underlying drivers within the field of DG.

4.1. DG technology (Fig. 1)

1. As noted in Section 3, the results varied significantly when the different normalization schemes are applied. In general, the normalized data tended to show underlying trends with more clarity. However, using cross-sectional normalization appeared to produce noisier curves.

2. One prominent trend which was consistent across the three versions was with micro-turbine research, which exhibited a marked increase in term frequencies between the years 1996 and 1999, before leveling off and finally declining again. Research in fuel cells also showed a similar trend though the term frequencies in this case started from a higher point compared to the plot for micro-turbine.
3. In the case of solar PV, a "V" shaped curve centered around the year 2000 was observed. While this is most prominent in Fig. 1 (c), it can be discerned in all three graphs and is consistent with general trends in the field of solar PV research [37].

4.2. DG interface technologies (Fig. 2)

Two broad trends observed can be seen in the term frequency plots for Inverter and Synchronous interfaces — specifically, over the period of analysis, research in synchronous interfaces seems to have declined somewhat, with the transition point falling somewhere between 1996 and 1998. In contrast, research in inverter based interfaces showed the exact opposite trend — the term frequencies for "inverter" started out low but increased sharply at or around the year 1998.

For induction based interfaces, the trend is less clear but broadly, research in this topic is seen increasing gradually throughout the analysis period, starting from around 1994.

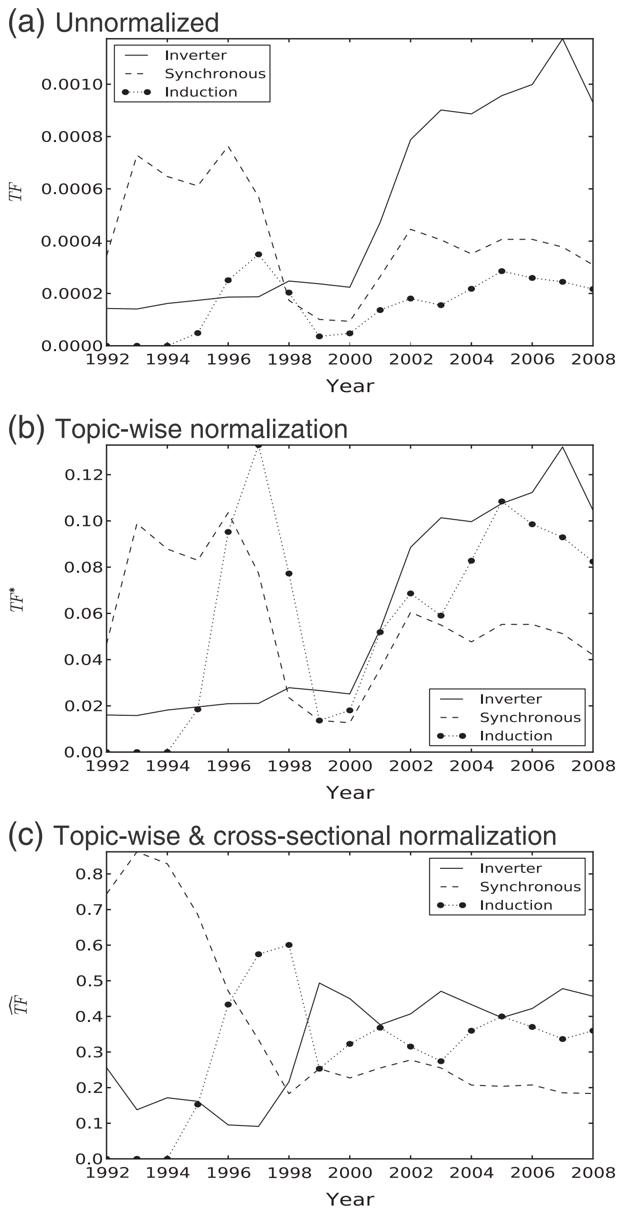

Fig. 2. DG interfaces and technologies.

4.3. General studies (Figs. 3 and 4)

This dimension included a large number of topics, which were further divided into two separate figures (Figs. 3 and 4) to facilitate interpretation and analysis of the results. The division into the two batches was to reduce the clutter within individual graphs.

4.3.1. Batch 1 — {Control, Reliability, Islanding, Planning, Stability} (Fig. 3)

1. There are a number of prominent trends in this first batch of topics. One interesting example is with research in “islanding”; in Fig. 3(a), this plot can hardly be seen as its overall term frequency is quite low. However, when either of the two normalization schemes is applied, it is immediately clear that this is one of the fastest growing research topics in this batch.
2. A similar pattern is observed with “stability”, with the TF^* and $\widehat{TF}$ curves showing sharp increases at or around the year 1997, but decreasing somewhat for the next few years before again starting to increase sharply.

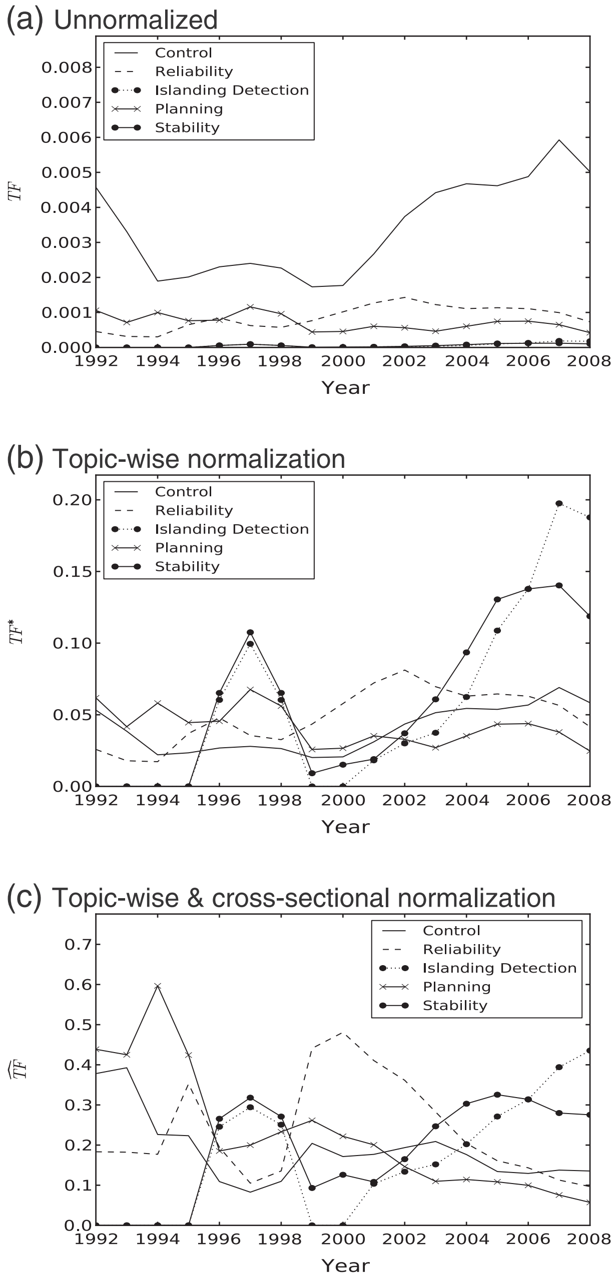

Fig. 3. General studies (batch 1).

3. Research interest in "planning" starts out with relatively high term frequencies, but these gradually decrease over the analysis period.
4. Another interesting example is in the case of "control" — where the values of TF and TF^* are seen increasing quite steadily (Fig. 3(a) and (b)); however, when normalized with respect to the other topics ($\widehat{TF}$ — Fig. 3(c)), it appears that this area is in fact declining gradually.

4.3.2. Batch 2 — {Power Quality, Electricity Markets/Economics, Protection, Forecasting, Micro-Grids/Smart-Grids} (Fig. 4)

1. The most prominent trend in this second batch of topics was for "Micro-Grids/Smart-Grids", which in all three figures can be seen to be increasing rapidly post-2000.
2. Both "Power Quality" and "Protection" gradually increase in popularity. The term frequencies for "Forecasting" can also be observed to be increasing in the initial period; however, after around the year 2000, they start to drop for around 3 years before staging a moderate "rebound" over the remaining years (though never quite regaining the initial highs).
3. "Electricity markets/economics" is another topic with inconsistent results — in Fig. 4(a) and (b), a mixed trend is observed though broadly speaking the popularity of this topic seems to have increased. However, in contrast, the $\widehat{TF}$ plot (Fig. 4(c)) appears to show a decline in this topic.

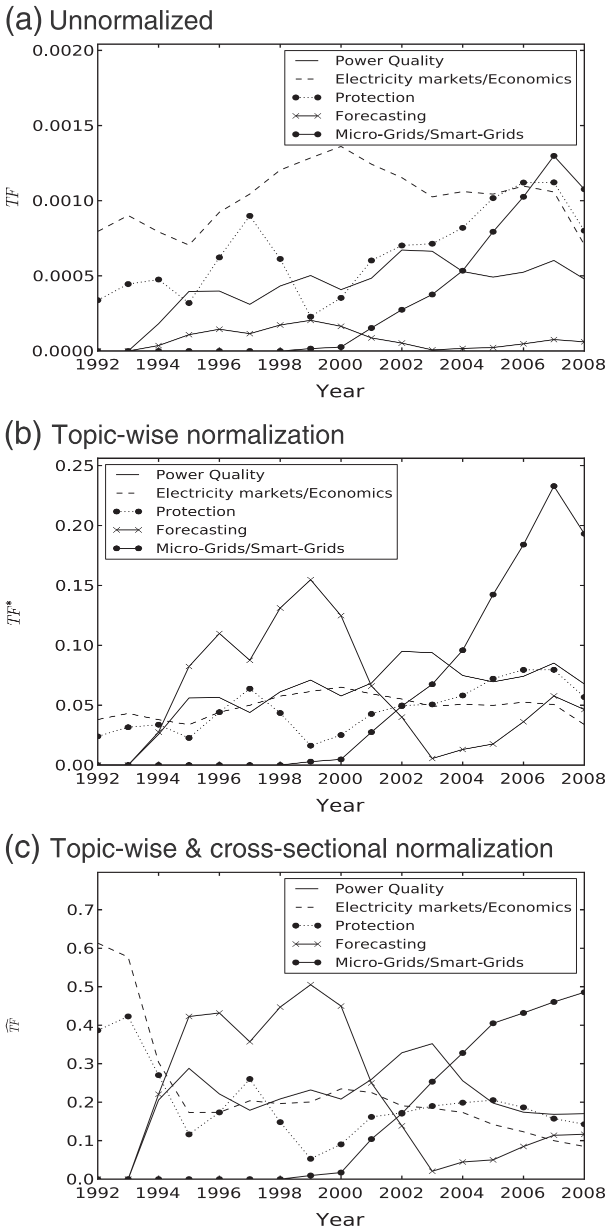

Fig. 4. General studies (batch 2).

4.4. Cross-dimensional analysis

As mentioned in Section 3, we would also like to analyze the data using a “slice and dice” approach involving the topics in dimensions 2 and 3. For each of the “general studies” topics, we are interested in the relative growth potential of each topic for the different interface technologies:{Inverter,Synchronous,Induction}.

However, as occurrences of each of these terms would be very low given the extremely specific conditions, the *TF* terms for each of the topic combinations were averaged over the ranges [1992,2000] and [2000,2008] (inclusive). As an indicator of the growth potential for each combination, we simply use the differences between the mean *TF* for each of these two periods. These results are presented in Table 1.

The most prominent observation from this table was that for 7 out of the 10 topics, the inverter based interface technology had the highest rate of growth. Conversely, induction based interface technologies had the lowest growth rate for those 7 topics. However, it had the *highest* growth rate for the remaining three topics (stability, planning and forecasting), though in the case of planning, it had the same growth rate as for synchronous technologies.

4.5. Publication statistics by country and journals

In addition, publication counts according to the country of origin and journal were generated. These statistics are presented in Table 2 (organized by country), 3 (by journal) and 4 (by journal, for papers published after 2006).

The distribution of publications by country of affiliation is presented in Table 2; this table is divided into two halves — the first provides statistics which take all the papers into account, while the second half only takes “recent” (defined as the last three years) publications into account. The main observations were:

1. As expected the USA had the largest share of the publications. However, its lead over the other countries has been reduced in recent years; while the total number of papers published in the USA is around 2.5 times the number published by it's closest competitor, the ratio for the last three years has reduced to slightly under 1.5 times.
2. Overall, the UK has been ranked second in this list. However, while it appeared to have gained slightly on the USA in the last few years, the largest shift has been in papers published by Chinese researchers — not only has China moved up to second position, the ratio of Chinese to American papers has increased tremendously from around a third to almost three quarters.
3. Besides China, other developing countries also appear to have increased their research output, notably India (from 10th to 7th place) and Iran (15th to 10th place).
4. From amongst the developed countries, Canada made the greatest gains, moving from 6th to 4th place. However, Germany and France suffered significant declines from 7th and 8th to 9th and 11th positions respectively while Australia dropped off the top fifteen list entirely.
5. In terms of publication density (the ratio of number of papers to population), the UK had the highest overall number, followed by The Netherlands and Canada. However, for the last three years, Canada had the largest publication density, while the UK and The Netherlands had dropped to 3rd and 4th position respectively. Unexpectedly, Belgium, a “newcomer” in the top fifteen list, now occupied the 2nd position.

Statistics on the distribution of publications by journal titles have also been divided into general and recent tranches. These are presented in Tables 3 and 4 respectively. Our observations are:

1. The top journal for DG-related research appears to be the IEEE Transactions on Power Systems; overall it had the largest number of relevant papers as well as the largest number of incoming citations. However, when only papers published after 2006 were considered, it was ranked third in terms of number of papers published, but still had the largest number of incoming citations.
2. Besides the above, the list had a large number of other IEEE Transactions. For both lists, five of the top fifteen journals were IEEE Transactions.

Table 1
Growth rates for interface technologies w.r.t. each of general study topics. The highest growth rate per row is printed in bold, and the rows are sorted in order of descending average growth rate.

Rank	Interface/Topic	$TF_{[2000,2008]} - TF_{[1992,2000]}$			
		Inverter	Synchronous	Induction	Mean
1	Micro/Smart-Grids	$\mathbf{2.9\times10^{-3}}$	2.8×10^{-4}	1.1×10^{-5}	1.1×10^{-3}
2	Power quality	$\mathbf{2.2\times10^{-3}}$	3.6×10^{-4}	2.8×10^{-4}	9.4×10^{-4}
3	Islanding detection	$\mathbf{1.8\times10^{-3}}$	4.4×10^{-4}	0.0	7.4×10^{-4}
4	Stability	7.1×10^{-5}	8.3×10^{-4}	$\mathbf{1.1\times10^{-3}}$	6.8×10^{-4}
5	Control	$\mathbf{1.5\times10^{-3}}$	3.4×10^{-4}	1.4×10^{-4}	6.8×10^{-4}
6	Reliability	$\mathbf{6.7\times10^{-4}}$	2.6×10^{-4}	7.9×10^{-5}	3.3×10^{-4}
7	Planning	1.4×10^{-4}	$\mathbf{1.7\times10^{-4}}$	$\mathbf{1.7\times10^{-4}}$	1.6×10^{-4}
8	Forecasting	0.0	5.4×10^{-5}	$\mathbf{4.2\times10^{-4}}$	1.6×10^{-4}
9	Electricity markets/Economics	$\mathbf{1.5\times10^{-4}}$	1.2×10^{-4}	0.0	9.2×10^{-5}
10	Protection	$\mathbf{3.0\times10^{-4}}$	2.2×10^{-5}	-2.4×10^{-4}	2.5×10^{-5}

Table 2
Prolific countries.

Overall statistics				Papers published after 2006			
Country	No. papers	Population (millions)	No. papers/population	Country	No. papers	Population (millions)	No. papers/population
USA	1572	306	5.1	USA	400	306	1.3
United Kingdom	639	62	10.3	China	308	1336	0.2
China	574	1336	0.4	United Kingdom	215	62	3.5
Japan	323	128	2.5	Canada	142	34	4.2
Italy	315	60	5.2	Italy	119	60	2.0
Canada	298	34	8.8	Japan	118	128	0.9
Germany	264	82	3.2	India	99	1161	0.1
France	203	65	3.1	Spain	96	46	2.1
Spain	199	46	4.3	Germany	90	82	1.1
India	185	1161	0.2	Iran	75	70	1.1
Netherlands	150	16	9.4	France	64	65	1.0
Brazil	130	191	0.7	Netherlands	52	16	3.2
Australia	128	22	5.8	South Korea	51	48	1.1
South Korea	125	48	2.6	Brazil	49	191	0.3
Iran	122	70	1.7	Belgium	45	11	4.1

3. One surprising observation was that the journals IET Generation, Transmission and Distribution, IET Renewable Power Generation and IET Power Electronics, which are extremely prominent in the DG field, were initially missing from the overall list of prolific journals. However, upon closer inspection, it turned out that this was because both IET Renewable Power Generation and IET Power Electronics are new journals which have only appeared in the last two or three years, and as such would not fare well in a ranking based on the total number of papers (in fact, it was this observation which originally motivated us to compile a set of tables which only considered papers published after 2006). In the table using recent papers, we see that IET Renewable Power Generation has appeared in the top fifteen list, however, IET Power Electronics is still off the list — subsequent checks to Scopus' source list confirmed that this journal had yet to be added to their database. For IET Generation, Transmission and Distribution, it was noted that the journal title had previously been "IEE Proceedings on Generation, Transmission and Distribution", which had caused papers from the journal to be allocated to different bins, thus diluting their impact. When the two titles are combined, we find that this journal has also moved into the list of top fifteen titles. These are good examples of the limitations of the proposed method — i.e. that results are only as good as the data extracted from the source database, and are a motivating factor for combining query results from more than one database.
4. Note also the emergence of a number of Asian journals — Dianli Xitong Zidonghua and Zhongguo Dianji Gongcheng Xuebao from China, and the Transactions of the Korean Institute of Electrical Engineers. In the overall data, only Dianli Xitong Zidonghua is in the top fifteen list but in the list which reflects recent publication trends we find that the other two have appeared in positions 8 and 9, respectively.

However, upon more careful inspection, we find that the majority of the incoming citations for the first two journals originate from within the same journal, while for the Transactions of the Korean Institute of Electrical Engineers, there are no incoming citations at all. Therefore, while the volume of research published in these journals is increasing, the impact of this research is still relatively low, or is restricted to the respective local contexts.

Table 3
Top fifteen journals (by No. papers). Ranks in brackets are based on the number of incoming citations.

Rank (Weighted)	Journal	No. papers	No. citations	No. citations/No. papers
1 (1)	IEEE Transactions on Power Systems	93	1132	12
2 (3)	IEEE Transactions on Power Delivery	68	478	7
3 (5)	Electric Power Systems Research	47	247	5
4 (9)	Dianli Xitong Zidonghua/Automation of Electric Power Systems	43	168	3
5 (4)	Energy Policy	37	275	7
6 (6)	IEEE Transactions on Energy Conversion	37	235	6
7 (2)	IEEE Transactions on Power Electronics	32	610	19
8 (7)	Journal of Power Sources	28	234	8
9 (15)	BWK — Energie-Fachmagazin	28	3	0
10 (12)	Renewable Energy	23	72	3
11 (8)	IEEE Transactions on Industry Applications	23	176	7
12 (10)	IET Generation, Transmission and Distribution	22	159	7
13 (11)	Energy	17	88	5
14 (14)	IEEJ Transactions on Power and Energy	17	11	0
15 (13)	International Journal of Electrical Power and Energy Systems	16	37	2

Table 4
Top fifteen journals w.r.t. papers published after 2006. Ranks in brackets are based on the number of incoming citations.

Rank (Weighted)	Journal	No. papers	No. citations	No. citations/No. papers
1 (2)	Electric Power Systems Research	37	46	1
2 (9)	IEEE Transactions on Power Delivery	35	26	0
3 (1)	IEEE Transactions on Power Systems	34	55	1
4 (4)	Dianli Xitong Zidonghua/Automation of Electric Power Systems	28	32	1
5 (5)	IEEE Transactions on Energy Conversion	20	32	1
6 (6)	Energy Policy	16	29	1
7 (8)	IEEE Transactions on Power Electronics	14	27	1
8 (15)	Transactions of the Korean Institute of Electrical Engineers	12	0	0
9 (13)	Zhongguo Dianji Gongcheng Xuebao/Proc. Chinese Soc. of Electr. Eng.	10	2	0
10 (11)	IET Renewable Power Generation	10	9	0
11 (10)	Journal of Power Sources	9	14	1
12 (3)	IEEE Transactions on Industrial Electronics	9	37	4
13 (7)	Energy	9	28	3
14 (12)	IET Generation, Transmission and Distribution	9	3	0
15 (14)	Electric Power Components and Systems	9	1	0

5. Analysis

In the previous section, the results of the bibliometric analysis were presented, and broad trends and patterns described. In this section a more detailed discussion will be presented, and attempts will be made to link prominent observations to relevant developments in the DG domain.

By referring to Figs. 1 and 2, it can be seen that fuel cells and inverter tend to follow the same trend. The strong correlation between both topics is due to the fact that inverters are the sole interface option for fuel cells. During earlier years, PV was the most dominant research topic but the other technologies have been starting to attract more attention since the late 90s (refer to Fig. 1). With regards to the type of interface, induction and synchronous were overall the main focus in the early 90s but with new emerging technologies relying mainly on inverters, inverters have become in recent years of major research focus (refer to Fig. 2). In addition, the topics wind and induction share to some extent the same research trend which is obvious since the majority of wind turbines are connected to the power system through induction machines.

The research trend on the various research challenges associated with the different types of technologies and interfaces is given in Figs. 3 and 4. It can be seen that since 2002, the majority of research studies are of equal importance. One unique trend is the Smart-Grid research trend which has been increasing since 1998. The increasing interest in Smart-Grids is as a result of the increasing advancements in communications coupled with the increasing penetration of distributed generation in power systems. Consequently, the IEEE launched in 2009 a new transaction IEEE Transactions on Smart-Grids that will address and focus on research advancements in this area. Recently, most of the research studies, given in Figs. 3 and 4, tend to be of equal importance except for islanding detection and stability which have a higher term frequency factor. As a result, the IEEE launched in 2009 another transaction IEEE Transactions on Sustainable Energy which will include evaluation of power systems that are affected by sustainable energy.

As highlighted earlier, DG could be interfaced to the distribution system through an inverter, synchronous or induction machine. From Table 1, it can be seen that most of the interface topics are of relative importance for inverter based and induction based DG. This could be related to the fact that an inverter interface is the best candidate for DG sources that inherently generate DC such as PV and fuel cells. On the other hand, while existing wind turbine systems can use either synchronous or induction interfaces, the induction machine interface has several advantages which include lower capital cost, lower maintenance cost as well as better transient performance. For this reason, the induction machine interface is widely and extensively used for wind turbines [38]. For brevity, we will focus our discussion on some of the topics presented in Table 1.

The main role of an inverter is to convert the DC power generated by the DG source to AC power necessary for feeding loads. Inverters are composed of a group of switches that are controlled in such a way to synthesize the DC waveform into an AC waveform. In this conversion process, harmonics are generated which results in power quality problems [39]. Synchronous and induction machines generally do not suffer from this problem and for this reason power quality is considered one of the main challenges when designing inverter based DG (refer to Table 1). The second challenge that is of relative importance is islanding detection. Islanding is a condition where the DG is operating in an isolated mode and not directly connected to the utility. The IEEE Standards necessitate the disconnection of a DG once it is islanded to avoid any power quality and safety problems.[1] This topic is a major challenge when implementing inverter and synchronous based DG because both types can sustain stable operation in the absence of the grid connection. On the other hand, induction machines require a grid connection for stable operation and for this reason an islanding condition could be easily detected and a sophisticated islanding detection method is not required [32]. This coincides with the results presented in Table 1 where most of the research work for islanding detection falls under the inverter and synchronous interface area.

[1] IEEE Std. 1547-2003.

Forecasting, stability and planning are three research topics where induction based DG become more dominant. Induction based DG are commonly used for wind generation. Large scale wind installation commonly referred to as wind farms are usually connected to the transmission and sub-transmission levels and their capacity is in the range of a common generation plant. Accurate forecasting techniques and stability analysis becomes an essential requirement to guarantee efficient, reliable and stable power system operation. Taking into account wind generation in planning studies is a result of the increasing reliance on and penetration of wind generation.

Two research topics that have been drawing much attention recently are Micro-Grids and Smart-Grids. DG can form a new type of power system, the so-called Micro-Grid. Micro-Grids can be viewed as a group of DG sources operating either connected or isolated to/from the main grid [40,41]. Smart-Grids are an extension of the Micro-Grid concept. Smart-Grid is a distribution grid with a communication infrastructure with capability to control the different components within the Smart-Grid to achieve efficient, reliable and secure operation [25,26]. For utilities with high penetration of DG sources, the practice of disconnecting all DG during a grid failure is neither practical nor reliable. There is a pressing need for modifications to these policies with regard to DG disconnection after a disturbance. The current IEEE standards do not address the topic of Micro-Grids. However, a new IEEE standard is currently being developed to address Micro-Grids [42]. Inverter based DG provides an attractive option due to its fast transient response and flexibility in interface control design. For this reason, inverter based DG are the most dominant technology for implementing Smart-Grids and Micro-Grids.

Power system protection with embedded DG is one of the important issues associated with the integration of DG. Distribution system protective devices (re-closers, fuses and overcurrent relays) are designed based on the short circuit currents. The integration of DG will affect the magnitude and direction of the fault currents. In literature, several studies highlighted the impacts of distributed generation on protection ([43–48]). Based on that, it was expected that "protection" would be ranked among the top 5 in Table 1 but rather surprisingly it was ranked with the lowest growth rate. Thus, distribution system protection could be a potential area for further improvement, research and development.

6. Conclusions

Below we summarize the main findings and limitations of the analysis, and suggest directions for future research.

6.1. Results

The results presented in this paper demonstrate that the proposed methodology is capable of extracting valuable information from semi-structured sources of data. While this study is still preliminary, this information is already useful in helping to improve our understanding about trends and patterns in research, and would already be of great interest to a researcher in the field of DG. As discussed in Sections 4 and 5, the trends highlighted by the bibliometric analysis certainly appear to be in agreement with developments in the field of DG, and could support the formulation of well informed and effective research policies. Some recommendations are:

1. Smart-Grids and Micro-Grids are considered amongst the hottest research topics in the field of DG. This is also evident from the increasing number of conferences that have panels focusing on this topic.
2. Among all DG interface types, inverter based DG seems to be the dominant type with many interesting and challenging problems.
3. Topics such as protection and stability in distributed system design are now becoming of major importance due to the high penetration of DG.

6.2. Outlook

As with any computational framework which exploits semi-structured data, there were some issues which need to be highlighted. Firstly, it is important to note that successful use of this approach is contingent upon our ability to correctly identify relevant publications. So, for example, the frequency of the term "wind" is used to estimate the level of activity in wind power research. While this successfully identifies many relevant papers, there will certainly be some false positives or negatives.

A further problem is with inconsistent database capabilities. To access a larger body of documents, queries could be submitted to a number of different academic search engines (for example, the "Scirus" search engine, or Google's Scholar search engine), or full text searches could be performed. Unfortunately, many search engines do not allow full-text searches, or exporting of search results for use in bibliometric analysis.

To help counter these problems and to increase the quality and applicability of this approach, we propose the following avenues for future work:

1. Natural Language Processing (NLP) — NLP based techniques could be used to enhance the existing methodology. One possibility would be to extract noun phrases from the Abstract, Title, and Keyword fields. Relevant phrases could then be combined with the existing terms to give a richer set of search terms.
2. Intelligent feature extraction — A variety of techniques from the machine learning and semantic technology communities could be brought to bear. In particular, it would be interesting to see the value of incorporating semantically-enabled features into the search process — i.e. instead of using manually generated keyword searches, computational techniques could be used to group together

terms which are either synonymous, or which are observed to co-occur frequently, and to combine these terms appropriately when conducting the searches.
3. Information fusion from multiple, heterogeneous data sources — as noted above, different databases provide different capabilities, and cover different subsets of the academic literature. Instead of accessing individual databases in isolation, information extracted from different sources could be combined, and in a way which allows for the heterogeneity. So for example, a weighting mechanism could be devised to allow results of full-text searches from one database to be combined with a title-only search from another.

Acknowledgments

The work described in this paper was carried out with the financial support of the Masdar Institute of Science and Technology The authors would like to thank the Masdar Institute for its support of this research.

Appendix A. Term extraction

The following regular expressions were used to detect term occurrences in order to calculate *TF*. All regular expressions were compiled using case insensitive mode.

- solar\spv|photovoltaic — “Solar PV”
- micro[-\s]{0,1}turbine — “Micro-turbines”
- wind\s(?:power|farm|turbine) — “Wind”
- fuel\scell — “Fuel cells”
- inverted — “Inverter”
- synchronous — “Synchronous”
- induction — “Induction”
- Control — “Control”
- Reliability — “Reliability”
- Islanding\s+detection — “Islanding Detection”
- Planning — “Planning”
- (?:voltage|frequency)\sstability — “Stability”
- power\squality — “Power Quality”
- electricity markets|economic — “Electricity markets/Economics”
- protection — “Protection”
- forecasting — “Forecasting”
- micro[-\s]{0,1}grid|smart[-\s]{0,1}grid — “Micro-/Smart-Grids”

References

[1] A. Porter, Tech mining, Compet. Int. Mag. 8 (1) (2005) 30–36.
[2] A. Porter, How “tech mining” can enhance r&d management, Res. Technol. Manage 50 (2) (2007) 15–20.
[3] P. Losiewicz, D. Oard, R. Kostoff, Textual data mining to support science and technology management, J. Int. Inf. Syst. 15 (2) (2000) 99–119.
[4] J. Martino, Technological Forecasting for Decision Making, McGraw-Hill Engineering and Technology Management Series, 1993.
[5] H. Small, Tracking and predicting growth areas in science, Scientometrics 68 (3) (2006) 595–610.
[6] D. Thorleuchter, D. Van den Poel, A. Prinzie, A compared R&D-based and patent-based cross impact analysis for identifying relationships between technologies, Technol. Forecast. Soc. Change 77 (7) (2010) 1037–1050.
[7] R.N. Kostoff, Text mining using database tomography and bibliometrics: a review, Technol. Forecast. Soc. 68 (2001) 223–253.
[8] M. de Miranda Santo, G. Coelho, D. dos Santos, Text mining as a valuable tool in foresight exercises: a study on nanotechnology, Technol. Forecast. Soc. 73 (8) (2006) 1013–1027.
[9] M. Kim, A bibliometric analysis of the effectiveness of Korea's biotechnology stimulation plans, with a comparison with four other Asian nations, Scientometrics 72 (3) (2007) 371–388.
[10] K. Anuradha, S. Urs, Bibliometric indicators of Indian research collaboration patterns: a correspondence analysis, Scientometrics 71 (2) (2007) 179–189.
[11] W.-T. Chiu, Y.-S. Ho, Bibliometric analysis of tsunami research, Scientometrics 73 (1) (2007) 3–17.
[12] T. Braun, A.P. Schubert, R.N. Kostoff, Growth and trends of fullerene research as reflected in its journal literature, Chem. Rev. 100 (1) (2000) 23–38.
[13] R.J. Watts, A.L. Porter, Innovation forecasting, Technol. Forecast. Soc. Change 56 (1) (1997) 25–47.
[14] D. Thorleuchter, D. Van den Poel, A. Prinzie, Mining innovative ideas to support new product research and development, in: H. Locarek-Junge, C. Weihs (Eds.), Classification as a Tool for Research, Springer, Berlin-Heidelberg-New York, 2010.
[15] N.R. Smalheiser, Predicting emerging technologies with the aid of text-based data mining: the micro approach, Technovation 21 (10) (2001) 689–693.
[16] T.U. Daim, G.R. Rueda, H.T. Martin, Technology forecasting using bibliometric analysis and system dynamics, Technology Management: A Unifying Discipline for Melting the Boundaries, 2005, pp. 112–122.
[17] T.U. Daim, G. Rueda, H. Martin, P. Gerdsri, Forecasting emerging technologies: use of bibliometrics and patent analysis, Technol. Forecast. Soc. 73 (8) (2006) 981–1012.
[18] T. Tran-Quoc, L. Le Thanh, C. Andrieu, N. Hadjsaid, C. Kieny, J. Sabonnadiere, K. Le, O. Devaux, O. Chilard, Stability Analysis for the Distribution Networks with Distributed Generation, Transmission and Distribution Conference and Exhibition, 2005/2006 IEEE PES, 2005, pp. 289–294.
[19] N. Kroutikova, C. Hernandez-Aramburo, T. Green, State-space model of grid-connected inverters under current control mode, IET Electr. Power Appl. 1 (2007) 329–338.
[20] A. Girgis, S. Brahma, Effect of distributed generation on protective device coordination in distribution system, Power Engineering, 2001. LESCOPE'01. 2001 Large Engineering Systems Conference on, 2001, pp. 115–119.

[21] A. Chowdhury, S. Agarwal, D. Koval, M. Co, I. Davenport, Reliability modeling of distributed generation in conventional distribution systems planning and analysis, IEEE Trans. Ind. Appl. 39 (5) (2003) 1493–1498.
[22] R. Dugan, T. McDermott, G. Ball, E. Inc, T. Knoxville, Distribution planning for distributed generation, Rural Electric Power Conference, 2000, 2000, p. C4.
[23] G. Celli, F. Pilo, MV network planning under uncertainties on distributed generation penetration, IEEE Power Engineering Society Summer Meeting, 2001, vol. 1, 2001.
[24] P. Jahangiri, M. Fotuhi-Firuzabad, Reliability assessment of distribution system with distributed generation, IEEE 2nd International Power and Energy Conference, 2008. PECon 2008, 2008, pp. 1551–1556.
[25] R. Brown, Impact of Smart Grid on distribution system design, 2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008, pp. 1–4.
[26] D. Novosel, Emerging technologies in support of smart grids, 2008 IEEE Power and Energy Society General Meeting — Conversion and Delivery of Electrical Energy in the 21st Century, 2008, pp. 1–2.
[27] H. Zeineldin, J. Kirtley, A simple technique for islanding detection with negligible nondetection zone, IEEE Trans. Power Delivery 24 (2009) 779–786.
[28] Y. Kajikawa, Y. Takeda, K. Matsushima, Computer-assisted roadmapping: a case study in energy research, Management of Engineering & Technology, 2008. PICMET 2008. Portland International Conference on, 2008, pp. 2159–2164.
[29] R. Kostoff, R. Tshiteya, K. Pfeil, J. Humenik, G. Karypis, Power source roadmaps using bibliometrics and database tomography, Energy 30 (5) (2005) 709–730.
[30] Y. Kajikawa, Y. Takeda, Structure of research on biomass and bio-fuels: a citation-based approach, Technol. Forecast. Soc. Change 75 (9) (2008) 1349–1359.
[31] W. El-Khattam, M. Salama, Distributed generation technologies, definitions and benefits, Electr. Power Syst. Res. 71 (2) (2004) 119–128.
[32] W. Xu, K. Mauch, S. Martel, An Assessment of DG Islanding Detection Methods and Issues for Canada, report# CETC-Varennes 2004-074 (TR), CANMET Energy Technology Centre–Varennes, Natural Resources Canada, 2004.
[33] M. Bengisu, R. Nekhili, Forecasting emerging technologies with the aid of science and technology databases, Technol. Forecast. Soc. 73 (7) (2006) 835–844.
[34] S.W. Jordan, Technology forecasting with science indicators: the case of laptop battery futures, Management of Engineering and Technology, Portland International Center for, 2007, pp. 1643–1650.
[35] Y. Kajikawa, J. Yoshikawa, Y. Takeda, K. Matsushima, Tracking emerging technologies in energy research: toward a roadmap for sustainable energy, Technol. Forecast. Soc. Change 75 (6) (2008) 771–782.
[36] D. Thorleuchter, D. Van den Poel, A. Prinzie, Mining ideas from textual information, Expert Syst. Appl. 37 (10) (2010) 7182–7188.
[37] G. Vidican, W. Woon, S. Madnick, Measuring innovation using bibliometrics: the case of solar photovoltaic industry, Advancing the Study of Innovation and Globalization in Organizations (ASIGO), Nuremberg, Germany, 2009.
[38] M. Patel, Wind and solar power systems: design, analysis, and operation, CRC Press, 2006.
[39] J. Wong, P. Baroutis, R. Chadha, R. Iravani, M. Graovac, X. Wang, A methodology for evaluation of permissible depth of penetration of distributed generation in urban distribution systems, 2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008, pp. 1–8.
[40] H. Zeineldin, E. El-Saadany, M. Salama, Distributed generation micro-grid operation: control and protection, Power Systems Conference: Advanced Metering, Protection, Control, Communication, and Distributed Resources, 2006. PS'06, 2006, pp. 105–111.
[41] C. Sao, P. Lehn, Control and power management of converter fed microgrids, IEEE Trans. Power Syst. 23 (3) (2008) 1088–1098.
[42] B. Kroposki, T. Basso, R. DeBlasio, Microgrid standards and technologies, 2008 IEEE Power and Energy Society General Meeting — Conversion and Delivery of Electrical Energy in the 21st Century, 2008, pp. 1–4.
[43] W. Feero, D. Dawson, J. Stevens, White paper on protection issues of the microgrid concept, Consortium for Electric Reliability Technology Solutions, 2002.
[44] A. Girgis, S. Brahma, Effect of distributed generation on protective device coordination in distribution system, Power Engineering, 2001. LESCOPE'01. 2001 Large Engineering Systems Conference on, 2001, pp. 115–119.
[45] K. Kauhaniemi, L. Kumpulainen, Impact of distributed generation on the protection of distribution networks, Eighth IEE International Conference on Developments in Power System Protection, 2004, vol. 1, 2004.
[46] G. Kaur, M. Vaziri, Effects of distributed generation (DG) interconnections on protection of distribution feeders, IEEE Power Engineering Society General Meeting, 2006, 2006, p. 8.
[47] S. Chaitusaney, A. Yokoyama, Impact of protection coordination on sizes of several distributed generation sources, Power Engineering Conference, 2005. IPEC 2005. The 7th International, 2005, pp. 669–674.
[48] M. Doyle, Reviewing the impacts of distributed generation on distribution system protection, 2002 IEEE Power Engineering Society Summer Meeting, vol. 1, 2002.

Dr. Wei Lee Woon received a B.Eng. in Electronic Engineering with first class honours from UMIST, UK (now merged with the University of Manchester) in 1997, and his PhD in 2002 from the Neural Computing Research Group at Aston University, Birmingham, UK. Upon graduation he joined the Malaysia University of Science and Technology (MUST) as an Assistant Professor, where he served until 2007. He subsequently joined the faculty of the Masdar Institute of Science and Technology (Abu Dhabi, UAE). Dr. Woon has also worked as a Visiting Researcher at the Massachusetts Institute of Technology and at the RIKEN Brain Science Institute in Tokyo. Dr. Woon's research interests span a variety of topics, all of which are centered around the analysis of large data sets. Recent examples include technology mining, text analysis and EEG signal analysis. He has authored/co-authored over 25 academic papers.

Dr. Hatem Zeineldin received the B.Sc. and M.Sc. degrees in electrical engineering from Cairo University, Cairo, Egypt, in 1999 and 2002, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2006. He was with Smith and Andersen Electrical Engineering, Inc., where he was engaged in projects involving distribution system design, and protection and distributed generation. He was then a Visiting Professor at the Massachusetts Institute of Technology (MIT), Cambridge. He is currently an Assistant Professor at the Masdar Institute of Science and Technology (Abu Dhabi, UAE). His current research interests include power system protection, distributed generation, and deregulation.

Prof. Stuart Madnick is the John Norris Maguire Professor of Information Technologies in the MIT Sloan School of Management and Professor of Engineering Systems in the MIT school of Engineering (Cambridge, USA). He is presently co-Director of the PROductivity From Information Technology (PROFIT) Initiative and co-Heads the Total Data Quality Management (TDQM) research program. His current research includes integrating disparate information systems, database technology, and strategic use of information technology. He is co-author of over 300 books, articles, or reports including Operating Systems and The Dynamics of Software Development. He was a key designer of IBM's VM/370 operating system and Lockheed's DIALOG information retrieval system and has been the co-founder of high-tech firms, including Intercomp, Mitrol, iAggregate, and currently operates Langley Castle Hotel in England. Dr. Madnick has degrees in Electrical Engineering, Management, and Computer Science from MIT. He has been a Visiting Professor at Harvard University, Nanyang Technological University, University of Newcastle, Technion, and Victoria University.