IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 9, 2019, accepted April 12, 2019, date of publication May 8, 2019, date of current version May 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2913615

Identifying the Mode and Impact of Technological Substitutions

IAN P. MARR[1], CHRIS MCMAHON[2], MARK LOWENBERG[2], AND SANJIV SHARMA[1]
[1]Airbus Operations Ltd., Bristol BS34 7PA, U.K.
[2]Faculty of Engineering, University of Bristol, Bristol BS8 1TR, U.K.

Corresponding author: Ian P. Marr (ian.marr@airbus.com)

This work was supported in part by Airbus Operations Ltd., as a part of the Agile Wing Integration Project, and in part by the University of Bristol's Industrial Doctorate Centre in Systems under EPSRC Grant EP/G037353/1.

ABSTRACT Technological substitutions play a major role in the research and development efforts of most modern industries. If timed and provisioned well, successful technology substitutions can provide significant market advantages to firms that have anticipated the demand correctly for emergent technologies. Conversely, failure to commit to new technologies at the right time can have catastrophic consequences, making determining the likely substitution mode of critical strategic importance. With little available data, being able to identify at an early stage whether new technologies are appearing in response to the perceived stagnation in existing technical developments, or as a result of pioneering leaps of scientific foresight, poses a significant challenge. This paper combines bibliometric, pattern recognition, and statistical approaches to develop a technology classification model from historical datasets where literature evidence supports mode labeling. The resulting functional linear regression model demonstrates robust predictive capabilities for the technologies considered, supporting the literature-based substitution framework applied and providing evidence suggesting that substitution modes can be recognized through automated processing of patent data. Furthermore, preliminary evidence suggests that classification can be achieved based on partial time series, implying that future extensions to real-time classifications may be possible for decision-making in the early stages of research and development.

INDEX TERMS Adner's classification scheme, emergence, patent bibliometrics, pattern recognition, technological substitutions, technology life cycle.

I. INTRODUCTION

Technology substitutions occur when an incumbent technology is replaced by a radical innovation resulting in a new socio-technical regime [1]. The introduction of new technologies and replacement of incumbents in heavily regulated industries such as aerospace is often a very complex, time-consuming, and expensive challenge that requires significant levels of research and development to ensure a successful technology substitution. This challenge is exacerbated when new technologies represent a fundamental shift away from well-established principles, as the risk and uncertainties involved increase significantly. Simultaneously, the opportunities associated with these innovations may be sufficient to warrant decision-makers adopting new technological approaches. In some cases, new technologies arise even while existing technologies are still undergoing further developments, and have not yet reached the peak of their performance. This further complicates the decision for enterprises, as devoting significant resources to a new technological approach that may or may not out-perform the old one presents great commercial risk. The potential for high gains or equally high losses arising from the technology adoption choices made by a company reflects the importance of these substitution events for long-term planning, meaning they are often considered of critical strategic importance. It is therefore beneficial to be able to identify early whether a new technology is likely to have scope for development beyond that of the current dominant technology, and commercially when the tipping point might occur where the new candidate would become the industry 'mainstream' technology option.

This paper develops a new methodology for automatically classifying the dynamics observed in technological substitutions based on aligning scientific and technological

The associate editor coordinating the review of this manuscript and approving it for publication was Nikhil Padhi.

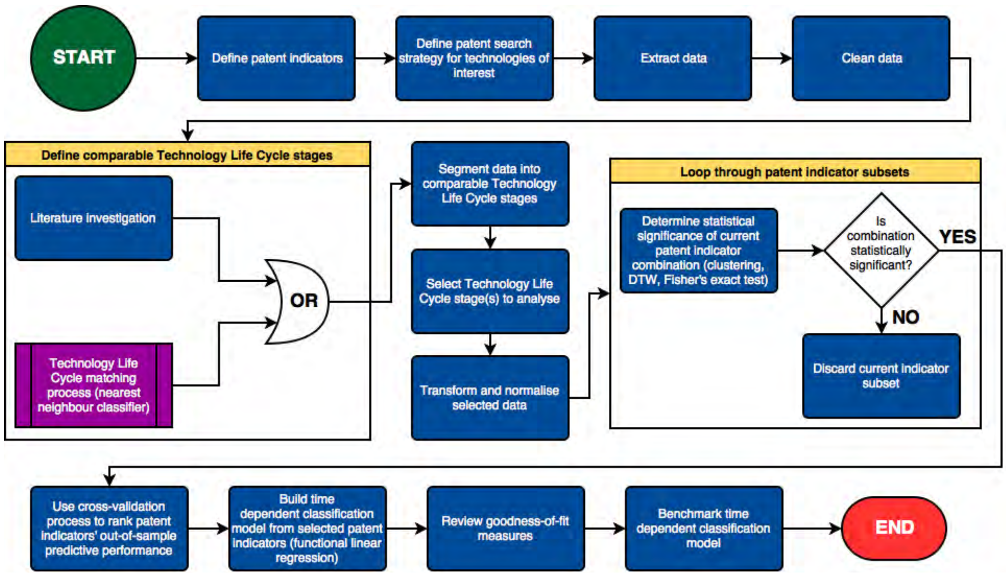


FIGURE 1. Overview of the analysis framework developed in this study.

development patterns recorded for a range of historical case studies against a recently proposed classification scheme. In particular, this paper looks to test a means of identifying the most suitable combination of bibliometric measures for indicating the likely substitution mode from patent data available during early stages of development. In doing so, this paper demonstrates how substitution modes may be recognized through automated processing of patent data. In addition, preliminary evidence from the technologies considered suggests that classification may be possible based on partially complete datasets (i.e. segmented time series), providing a potential route to real-time classifications in future extensions. Consequently, the methodology outlined in this paper moves towards a process that can be used in supporting technology strategy and innovation management by indicating the likely mode of adoption from early patent activity.

The multi-level regression methodology described in this paper combines bibliometric, pattern recognition, and statistical analysis techniques to patent data gathered for 20 representative technology substitutions. The substitutions considered in this analysis have been labeled based on coupling literature evidence with previously published conceptual models of substitution mechanisms. Statistical validation is presented for both the selection of the patent indicator set chosen for use in building the final classification model, and the suitability of the resulting functional linear regression model for in and out-of-sample predictions. The first of these validation stages is addressed by the application of exhaustive statistical significance testing and cross-validation processes to enable a complete ranking of all possible patent indicator combinations from the indicators considered. This in turn directly informs the selection of patent indicators used in the final time-based model. Subsequently it is found that the functional linear regression model developed performs well against the expected literature classifications. Further, benchmarking against other common regression models, coupled with permutation analysis, suggests that this result is not a chance occurrence and that the model should extend reasonably well to out-of-sample predictions (although based on a limited initial sample size). Lastly, the potential extension to real-time applications is supported by the successful classification of the technologies considered from the segmented time series data used in this analysis.

The paper begins by providing some background to technology substitutions and patent-based analysis techniques in section II, followed by an overview of bibliometric data sources, statistical analysis, and method selection in section III. Details of the derivation of the technology classification model using statistical ranking and functional data analysis are then provided in section IV, along with the corresponding results and discussions in section V. Finally, conclusions from the patent indicator ranking and classification model building exercises are then summarized in section VI. The methodological stages considered in this analysis are summarized in the framework shown in Fig. 1 to provide a more coherent picture of the methods adopted in the following sections.

II. BACKGROUND

Technological substitution often plays an important role in the fortunes of enterprises. As such, numerous studies have examined the many complex factors that influence technology development and adoption trends. An overview of the relationships between technological performance, perceived limits of science and technology, observed substitution patterns and behaviors, and patent-based forecasting techniques are provided here to explain the analysis that follows.

A. TECHNOLOGY FORECASTING, SUBSTITUTION PATTERNS, AND TECHNOLOGICAL FAILURE

Correctly predicting which emerging technologies are likely to be most influential can ensure that a company is best positioned to gain an advantage over its competitors when the new technology comes to fruition. Conversely, failure to anticipate the arrival of large technological shifts can leave businesses severely diminished. This is often illustrated by the dramatic impact on Kodak's business following the introduction of digital photography that rendered many of the company's existing film products obsolete, following an early lead in the digital field that was not fully capitalized upon [2]. Equally, investing heavily in a nascent technology too soon can have grave consequences, as Bertlesmann found from investing in Napster [3]. As such, forecasting techniques are commonly used to determine strategies in large organizations by providing an initial guide to future opportunities, risks, challenges, & areas of uncertainty [4].

In this field, considerable work has already been undertaken on modeling technology diffusion in these substitution events. This has included, amongst many other areas of study (see [5]), the influence of successive technology generations, and the impact of time delays on the perception of new technologies (see [6] and [7] respectively). Classically, the introduction of new technologies is often described as following an S-curve that assumes uptake is initially slow in the earliest stages, until performance and functional benefits of the new technology are seen to be greater than those of existing technologies, at which point uptake significantly accelerates [8], [9]. This model assumes that all technologies eventually arrive, driven by research and development efforts, at an ultimate limiting condition based on physical constraints, where performance improvements stagnate once again. However, in reality, periods of performance stagnation can also occur when challenging technical obstacles appear, or when market uptake slows (potentially due to market saturation, regulatory changes, or competition from new technologies), reducing investment in research and development [10], [11]. This results in substitutions to the next generation of technologies occurring either before or after arriving at a perceived performance limit, which may or may not be an actual, or ultimate, performance limit [12], [13].

This brings about the notion of continual technological (or functional) failure, at the point where a replacement technology is sought for a currently stalled technological paradigm [14]. However, the technological 'failures' that lead to this reactive type of substitution vary greatly, and cannot just assume a single simple definition. On this topic, previous work has examined what is meant by 'technological failure', and has broadly categorized these occurrences as outlined in the work of Gooday [15]. Beyond continually increasing human expectations of technology this work takes on board notions of non-linear development in the history of technologies (i.e. the stop-start nature of progress), the potential effects of social marginalization, as well as demographic and cultural influences that can lead to a divergence of opinions of whether a technology has 'succeeded' or 'failed'. More recently, the work of Edgerton has delved further into these concepts by introducing the idea of *Creole* technologies that can appear, disappear, and subsequently reappear throughout the course of history, whilst also highlighting the lag between technology development and widespread use [16]. In this regard, segmentation of technology life cycles into clearly defined sequential stages is not necessarily a straightforward task (as noted in section III-C.1). Additionally, Edgerton has contested the role of 'bleeding-edge' technologies, noting that conventional technologies have a remarkably long shelf-life, sustained impact, and are capable of resurgence [16]. Taking these notions of non-linear development into account, in the analysis that follows, this study focuses specifically on failures relating to the ever more demanding performance expectations that human users impose on their technologies. Specifically, the definition of technological failure used in this study is given as:

> "A point in time at which technology performance development stagnates/plateaus, with no further progressive trajectory improvements foreseen for a significant period of time in comparison to the overall technology lifecycle considered, which is subsequently followed by the substitution of a new technology/architecture that is on a progressive trajectory"

This means that a technology has been able to reach what could be observed to be a temporary performance limit in this condition before substitution to a new discontinuous technology occurs [17]. This definition also follows on from the work of Sood & Tellis which applied a sub-sampling approach to analyze different types of 'multiple S-curves', and subsequently concluded that technologies tend to follow more of a step-function, with long periods of static performance interspersed with abrupt jumps in performance, rather than a classical S shape. In this study, stagnation periods were recorded where technology performance during a given sub-sample had an upper plateau longer in duration than the immediately preceding growth phase, whilst the subsequent jump in performance in the year immediately after the plateau was almost double the performance gained during the entire plateau [14].

Up till now, only substitution patterns associated with technological failure have been discussed. However, previous studies have identified that technological substitutions are not just the result of the existing technology being deemed to have 'failed'. Edward Constant argued that a feature common to all technological revolutions is the emergence of 'technological anomalies', which can be traced to either scientific or technological crisis [18]. In the work of Constant the first, and most common, cause of these technological anomalies was attributed to functional failure. Conversely, technological anomalies were also identified as arising as a result of presumptive technological leaps. The mechanisms driving

TABLE 1. Identified examples of reactive and presumptive technological substitutions.

Examples of reactive substitutions	Examples of presumptive substitutions
Plug-compatible market (PCM) disk drives [20]	Transition from piston to jet engines [18, 21, 22]
Transition to fibre optic cables from Cu/Al wires for data transfer [14]	Transition to optical undersea cables from coaxial cables [21]
Transition to Low Pressure Sodium lights from Tungsten Filament Lamps [21]	Transition to water turbines from steam engines [18, 22]
Transition to Compact Fluorescent Lamps from Tungsten Filament Lamps [21]	Transition to early gas engines from steam engines [18]
Transition to White LED lighting from Low Pressure Sodium and Compact Fluorescent Lamps [21]	Transition to steam turbines from water turbines [18, 22]
Transition to hypersonic aircraft from supersonic [21]	Transition to catalytic petroleum cracking from thermal cracking [18]
Transition to coaxial undersea cables from single cable [21]	Transition to transistors from the vacuum tube [23]
Transition to T-carrier system from modem internet access [21]	Transition to atomic energy from fossil fuels [18, 24]
Transition to Synchronous Optical Networking (SONET) system from T-carrier internet access [21]	Renewable energy sources: transition to solar PV/thermal, wind, geothermal, hydropower, and marine energy from fossil fuels [22, 24]
Transition to ink jet and laser printers from dot matrix printers [14]	Transition to modern battery and plug-in hybrid electric vehicles from petrol and diesel vehicles [25]

technological substitutions are discussed in more detail in chapter 2 of [19].

B. MODES OF SUBSTITUTION

Building on the works of Constant, Schilling, and Sood, a conceptual framework for analyzing technology substitutions was published by Ron Adner that considers both the *emergence challenges* facing new technologies and the *extension opportunities* still available to existing technologies [12]. The relationships between *emergence challenges*, *extension opportunities* and the substitution regimes proposed by Adner are explored in greater detail in chapter 2 of [19], along with subsequent mapping to the more global classifications of reactive and presumptive substitution types used in this study.

Whilst Adner's framework provides a means of mapping observed substitutions to conceptually distinct patterns, the theoretical framework proposed by Adner does not go as far as developing a process for automatically recognizing substitution modes. The method outlined in this paper is therefore a first attempt at translating Adner's conceptual framework into a repeatable and generalizable methodology. As such, the current study only considers the *extension opportunity* dimension in its classification of substitution modes, to facilitate the development of the data-driven methodology presented here. It is worth noting that this analysis could be repeated and decomposed further into the higher fidelity regimes proposed by Adner, but this would require additional case studies to ensure a sufficient number of technologies are available in each category, whilst also requiring supplementary literature or expert evidence to support category assignments. For this reason, this study only considers the ability to distinguish between the two broader *extension opportunity* driven modes of substitution from analysis of historical scientific and technological data. More specifically, substitutions based on low extension opportunities for existing technologies are here termed *reactive*. Conversely, where there still appears to be high extension opportunities for existing technologies, substitutions are termed *presumptive*. In terms of performance trends this means that a reactive substitution corresponds to a period of performance stagnation prior to the new technology first appearing, whilst a presumptive substitution corresponds to the new technology first emerging as the existing technology continues to improve. The characteristics of these two broader substitution modes used in this paper are explored in greater depth in chapter 2 of [19]. The modes considered in this paper are illustrated in Fig. 2.

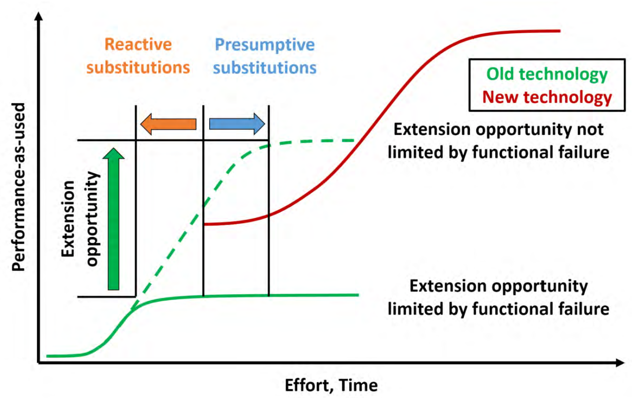


FIGURE 2. Illustration of reactive and presumptive substitution modes, based on Adner's framework.

Table 1 uses Adner's framework, alongside the definitions provided in section II-A and chapter 2 of [19], with performance evidence obtained from literature, to classify a sample set of technologies according to the broader modes of substitution observed.

In addition to the broader modes of substitution outlined in Table 1, other technologies have been identified as *non-starters*; these are marginalized technologies that were never mass commercialized (such as wire recorders or chain printers). In many cases these technologies could have been adapted for the target markets considered, but were either never used or failed to demonstrate the required features or performance and cost improvements necessary to warrant further development beyond initial trials. Non-starters are excluded in this study as there is often very little patent data pertaining to these technologies due to their very brief life-spans. However, as the analysis that follows is based on technologies that are known to have been successfully commercialized (falling into either the reactive or presumptive categories) it is not believed their inclusion would influence the results presented here. In reality, non-starters would need to be included for predicting the commercial success or failure of emerging technologies in the first instance [14], but this additional classifier dimension is left as an extension for future studies.

Based on Constant and Adner's classifications of substitution modes, this paper looks to test whether bibliometric measures of scientific and technological development can provide an indication of the mode of adoption likely to occur. Further classifier requirements based on these conceptual models of substitution mechanisms are discussed in chapter 2 of [19].

C. MEASURING PERCEPTIONS OF LIMITS OF SCIENCE AND TECHNOLOGY

Many indicators of science and technological progress have been developed in the fields of bibliometrics and scientometrics in recent decades. Whilst largely developed for the purposes of identifying and targeting gaps in existing knowledge, and determining the effectiveness of funding in specific fields of research, these indicators also provide a systematic approach to compare development trends across a broad range of scientific domains. When attempting to measure scientific and technological extension opportunities it is however important to ensure that any measurements taken are suitable indicators of the development characteristics that are being studied. In this regard conceptual distinctions exist between scientific activity, scientific production, and scientific progress [26]. In this study, the emphasis is not on assessing the performance or influence on technical direction of a specific set of papers, but rather to gauge the adoption trends of the field as a whole. As technology diffusion models also rely on non-invested parties being made aware of scientific and technological progress, communication and promotion of scientific research are important factors to include in adoption processes [6]. Adoption is equally dependent on perceptions of current scientific and technological rates of progress (shaped by social and political pressures, as well as technical [15], [16]), rather than the actual rates of progress (shaped by technical contributions to knowledge). Lastly, diffusion effects are population size, word-of-mouth, and time dependent [6]. As a result, measures of scientific production are felt to be a more relevant indication of likelihood to adopt than measures of scientific progress in this study.

D. PATENT-BASED TECHNOLOGY FORECASTING

The use of patents for forecasting technology development trends, and the close links to economic activity, has evolved considerably since the earliest literature was published on measuring innovation from patent statistics by the likes of Schmookler and Scherer in the 1960s [27], [28]. More recent publications have expanded these early concepts and demonstrated on numerous occasions how patterns in historic patent data can be used to build predictions of future development trends, including using partially complete or mined datasets when historical data is not yet available. Many of these studies attempt to assess the development maturity of a given technology (not to be confused with measures of commercial market adoption [12], [16]) against commonly recognized milestones and features in observed technology evolution patterns. Chief amongst these is comparison to Arthur Little's *Technology Life Cycle* (TLC) [29]. Comprising four stages (*emergence*, *growth*, *maturity*, and *saturation*), Little's framework describes a means of measuring technological development efforts relative to a technology's competitive impact and progress in transitioning from product to process-based innovation.

Classically, TLC studies have relied on simple counts of patent records to determine the maturity of technologies on this scale. However, contesting the accuracy and reliability of matching a single patent indicator against pre-determined growth curves, Watts, Porter, and Haupt advocated the use of multiple patent metrics in their technology evaluations [30], [31]. Building on this, Gao demonstrated the use of a trained nearest neighbor classifier, based on thirteen extracted patent data dimensions, to assess a technology's life cycle progress [32]. This was followed more recently by Lee's proposal for the use of a stochastic method based on multiple patent indicators and a hidden Markov model (i.e. an unsupervised machine learning technique) to estimate the probability of a technology being at a certain stage of its life cycle [33]. In parallel to these extensions to sets of indicators and pattern recognition techniques, the use of text-mining approaches to improve speed, relevance, and accuracy of patent analysis methods have been demonstrated by Ranaei's automatic retrieval of patent records for forecasting the development of electric and hydrogen vehicles [34]. Similarly, patent content clustering techniques for technology forecasting purposes have also been explored by the works of Daim *et al.* [4] and Trappey *et al.* [35]. Daim's analysis illustrated how technology forecasting results for emerging technologies can be improved by combining patent-based statistics with bibliometric clustering and citation analysis techniques for the purpose of data acquisition (as a proxy indicator for technology diffusion when historical data is unavailable). However, being able to determine the technical readiness of a new technology is only part of the technology forecasting problem. The other critical aspect that must be considered is the market adoption of the technology once it has been commercialized [12], [16]. Here, Daim's work subsequently coupled the patent-based and academic literature data-mining techniques employed with the use of system dynamics modeling as a means of exploring causal relationships and non-linear behaviors in technology diffusion. Based on these works, the current study looks to combine the recent advances made in pattern recognition applications with a simplified version of Adner's technology substitution framework.

III. METHODOLOGY

There is a range of possible techniques that can be used for measuring the progress of technological development and classifying substitution modes. Appendix A provides a summary of the methodologies considered in this analysis, their respective utility and limitations, and how these can potentially be adapted to the current study.

Based on the relative advantages of the different methodologies available, a multi-level regression approach is adopted here to enable the variation between different patent indicators to be viewed separately from the phase variation observed between different technologies. In this sense, bibliometric approaches enable patent indicators to be defined from the extracted patent datasets, pattern recognition techniques perform classification and time-based model building roles using these indicators, and statistical approaches enable significance testing, error checking, and ranking exercises to be carried out to verify the robustness of proposed models. The specific methods selected within each of these fields to achieve the desired research objective are discussed in further detail in section III-C.

Considering the sources of information available that chart technological development, growth trends, and substitutions, existing analysis suggests that patent data has been observed to account for 90 to 95% of the world's inventions [36], [37]. In this study, bibliometric data has therefore been extracted based on patent records as this has become a well-established means of assessment for both industry market comparisons and government policy setting purposes. An overview of the considerations taken into account in the selection of specific methods and model development techniques for analyzing this data are discussed below.

A. BIBLIOMETRIC DATA

Patent data has been sourced from the Questel-Orbit patent search platform in this analysis. More specifically, the full FamPat database was considered, which groups related invention-based patents filed in multiple international jurisdictions into 'families' of patents in accordance with EPO's strict family rules[1]. As such invention-based patent families are counted in this analysis, including both patent applications and granted patents to provide a complete reflection of associated technological development activities (bearing in mind the distinction between scientific production and progress discussed in section II-C). The data gathered covers all patent offices registered in the FamPat database[2]. Some of the core functionalities behind this search engine are outlined in [38]. This platform is accessed by subscribers via an online search engine that allows complex patent record searches to be structured, saved, and exported in a variety of formats. A selection of keywords, dates, and classification categories are used in this search engine to build relevant queries for each technology (this process is discussed in more detail in section IV-B). The provided search terms are then matched to the title, abstract, and key content of all family members included in a FamPat record, although unlike title and abstract searches, key contents searches (which include independent claims, advantages, drawbacks, and the main patent object) are limited to only English language publications.

B. STATISTICAL COMPARISONS OF TIME SERIES

This study considers 23 technologies, defined in Appendix B, where literature evidence has been identified to classify the particular mode of technology substitution observed. These technologies were selected based on four criteria:

1) Is there a historical narrative available?
2) Is there accompanying (and consistent) performance data available for both the preceding and replacement technologies? (i.e. to provide evidence of the mode of substitution)
3) Do a sufficient number of patent records exist for the replacement technology?
4) Is there accompanying adoption data present for the replacement technology for use in the subsequent technology diffusion studies? (not presented in this paper)

The evidence and process used in the subsequent categorization is outlined in detail in [19]. Using bibliometric analysis methods it is possible to extract a variety of historical trends for any technologies of interest, effectively generating a collection of time series data points associated with a given technology (these multidimensional time series datasets are referred to here as *technology profiles*). This raises the question of how best to compare dissimilar bibliometric technology profiles, in an unbiased manner, to investigate whether literature-based technology substitution groups can be determined using a classification system built on the assumptions given in section II-B. In particular, comparisons of technology time series can be subject to one or more areas of dissimilarity: they may be based on different number of observations (e.g. covering different time spans), out of phase with each other, subject to long-term and shorter term cyclic trends, at different stages through the Technology Life Cycle (or fluctuating between different stages) [29], or be representative of dissimilar industries. As such, a body of work already exists on the statistical comparison of time series, and in particular time series classification methods [39]. Most modern pattern recognition and classification techniques emerging from the machine learning and data science domains broadly fall within the categories of supervised, semi-supervised, or unsupervised learning approaches. Related to this, an overview of current preprocessing, statistical significance testing, classification, feature alignment, clustering, cross-validation, and functional data analysis techniques for time series is provided in Appendix C for further details of the considerations addressed in this study's methodology beyond those discussed directly in sections III and III-C.

C. METHOD SELECTION

Based on the technology classification problem considered, available methodologies, bibliometric data available, and specific methods discussed in Appendix C the following methods have been selected for use in this analysis:

[1]https://www.questel.com/wp-content/uploads/2016/04/FamPat-Rules.pdf

[2]http://static.orbit.com/orbit/help/1.9.6/en/index.html#!Documents/thefampatcollection.htm

1) TECHNOLOGY LIFE CYCLE STAGE MATCHING PROCESS

For those technologies where evidence for determining the transitions between different stages of the Technology Life Cycle has either not been found or is incomplete, a *nearest neighbor* pattern recognition approach has been employed based on the work of Gao [32] to locate the points where shifts between cycle stages occur. As noted in section II, technologies may in fact shift continually and non-sequentially between the different stages of the Technology life Cycle, however this is reflected in the outputs from the nearest neighbor pattern recognition approach (as illustrated in section 5.6 of [19]). This gives a measure of *progress* along the Technology Life Cycle S-curve, but does not compare the mode (i.e. *shape*) of the substitution observed to the typical classification patterns described by Adner. However, for the specific technologies considered in this paper, literature evidence has been identified for the transitions between stages, and so the *nearest neighbor* method for gauging progress is not discussed further here.

2) IDENTIFICATION OF SIGNIFICANT PATENT INDICATOR GROUPS

To identify bibliometric indicator groups that could form the basis of a data-driven technology classification model, a combination of Dynamic Time Warping and the 'Partitioning Around Medoids' (PAM) variant of K-Medoids clustering has been applied in this study. For the initial feature alignment and distance measurement stages of this process, Dynamic Time Warping is still widely recognized as the classification benchmark to beat (see Appendix C), and so this study does not attempt to advance the feature alignment processes used beyond this. Unlike the Technology Life Cycle stage matching process which is based on a well-established technology maturity model, this study is assuming that a classification system based on the modes of substitution outlined in section II-B is not intrinsically valid. For this reason, an unsupervised learning approach has been adopted here to eliminate human biases in determining whether a classification system based on reactive and presumptive technological substitutions is valid, before defining a classification rule system. This means that predicted clusters can be labeled, even if labels are only available for a small number of observed samples representative of the desired classes, or if none of the samples are absolutely defined. This is particularly useful if the technique is to be expanded to a wider population of technologies, as obtaining evidence of the applicable mode of substitution that gave rise to the current technology can be time-consuming, and in some cases the necessary evidence may not be publicly available (e.g. if dealing with commercially sensitive performance data). Clustering may therefore be able to provide an indication of the likely substitution mode of a given technology, without the need for prior training on classes of technologies. Under such circumstances this approach could be applied without the need for collecting performance data, providing that predicted groupings are broadly identifiable from inspection as being associated with the suspected modes of substitution. This is of course easier if some examples are known, but means it is no longer a hard requirement.

The 'PAM' variant of K-Medoids is selected here over hierarchical clustering since the expected number of clusters is known from literature (for the technologies considered), and keeping this number fixed enables easier testing of how frequently predicted clusters align with expected groupings. Additionally, a small sample of technologies is evaluated in this study, and as a result computational expense is unlikely to be significant in using the 'PAM' variant of K-Medoids over Hierarchical clustering approaches. It is also worth noting that by evaluating the predictive performance of each subset of patent indicator groupings independently it is possible to spot and rank commonly recurring patterns of subsets. This is not possible when using approaches such as Linear Discriminant Analysis, which can assess the impact of individual predictors but not rank the most suitable combinations of indicators.

3) RANKING OF SIGNIFICANT PATENT INDICATOR GROUPS

As the number of technologies considered in this study is relatively small, exhaustive cross-validation approaches provide a feasible means to rank the out-of-sample predictive capabilities of bibliometric indicator subsets that produce significant correlations to expected in-sample technology groupings. As such, 'leave-p-out' cross-validation approaches are applied for this purpose, whilst also reducing the risk of over-fitting in the following model building phases [40].

4) MODEL BUILDING

The misalignment in time between life cycle stages relative to other technologies can make it difficult to identify common features in time series. This is primarily because this phase variance risks artificially inflating data variance, skewing the driving principal components and often disguising underlying data structures [41]. Consequently, due to the importance of phase variance when comparing historical trends for different technologies, and the coupling that exists between adjacent points in growth and adoption curves, functional linear regression is selected here to build the time-based technology classification model developed in this study (see notes on Functional Data Analysis in Appendix C for further details). The prior clustering stages therefore test the suitability of Adner's classification scheme based on complete patent indicator profiles (testing variation and correlation in the patent indicator dimension), whilst the regression analysis builds time-dependent models for each patent indicator considered in the selected classification scheme.

IV. BUILDING A TECHNOLOGY CLASSIFICATION MODEL FROM TECHNOLOGY LIFE CYCLE FEATURES

A. PATENT INDICATOR DEFINITIONS

The work of Gao et al. identifies a range of studies that have been conducted previously based on using either single or multiple bibliometric indicators to investigate technological development and performance [32]. Their review of these methods concluded that multiple patent indicators are required to avoid generating potentially unreliable findings as a result of using a single indicator extracted from patent data. As such, the *nearest neighbor* matching process developed in Gao's study to assess progress through the Technology Life Cycle S-curve proposes thirteen separate patent indicators. The current study has accordingly reproduced these metrics where possible, resulting in a total of 10 patent indicators (i.e. producing time series for each technology with 10 dimensions). Indicators 11, 12, and 13 considered in [32] were specific to the Derwent Innovation Index[3], which was not used in this study due to the limited ability to bulk export the results from this database. Table 2 summarizes the bibliometric indicators extracted for each technology within this analysis. The dependencies between each of these indicators during different TLC stages is explored in the cross-correlation analysis presented in section 2.4 and Table 4 of [32]. Aside from indicator 1, all of the other patent counts considered in Table 2 are based on the earliest priority date of the collated patent family records.

TABLE 2. Bibliometric indicators used in this study (based on the work of Gao *et al.* [32]).

Indicator No.	Name	Description
1	Application	Number of patents in Questel-Orbit by application year
2	Priority	Number of patents in Questel-Orbit by priority year
3	Corporate	Number of corporates in Questel-Orbit by priority year
4	Non-corporate	Number of non-corporates in Questel-Orbit by priority year
5	Inventor	Number of groups of inventors in Questel-Orbit by priority year
6	Literature	Number of backward citations to literature in Questel-Orbit by priority year
7	Patent citation	Number of backward citations to patents in Questel-Orbit by priority year
8	IPC	Number of IPCs (4-digit) in Questel-Orbit by priority year
9	IPC top 5	Number of patents of top 5 IPCs in Questel-Orbit by priority year
10	IPC top 10	Number of patents of top 10 IPCs in Questel-Orbit by priority year

Apart from using the Questel-Orbit FamPat database instead of the Derwent Innovation Index, the indicator definitions and assumptions used in this study are consistent with those outlined in sections 2.1.1 to 2.1.5 of [32]. The only other notable difference is that the Questel-Orbit patent records are not automatically designated as corporate, non-corporate, or individual patent assignees. Consequently, counts of corporate and non-corporate indicators (which would otherwise be based on this assignee designation) are determined instead from the 'Family Normalized Assignee Name' field in the patent records, as records with entries in this field correspond to corporate designations.

B. SEARCH STRATEGY AND TERMS FOR IDENTIFYING RELEVANT PATENT PROFILES

Previous bibliometric studies have explored the different ways in which patent records can be correctly identified for a given field or topic [42]–[49]. Whilst filtering search results based on technology classification categories is generally preferred where possible to ensure a more rigorous search strategy [44], it is also advisable to keep the steps that supplement or remove patents from search queries to a minimum, to maintain data consistency and repeatability [49]. Accordingly, the search queries in this analysis are based primarily on filtering by International Patent Classification (IPC v2017.01) or Cooperative Patent Classification (CPC) labels. Where possible, IPC categories have been reused from previous studies to replicate existing search queries so as to extract comparative datasets, or based on expert defined groupings such as the European Patent Office's Y02 classification which specifically relates to climate change mitigation technologies. Otherwise, keyword search terms and IPC labels are combined that focus on matching closely adjoining instances of each search term (or their common synonyms). Using IPC technology category filters in this manner ensures that a higher level of relevance and repeatability is achieved. Based on these preprocessing steps, the final search queries are presented in Appendix B along with the number of records retrieved.

C. PATENT INDICATOR DATA EXTRACTION PROCESS

Using the technology classification categories, and where applicable the keywords in Appendix B, the results of these search queries were exported in batches of up to 10,000 records at a time in a tabulated HTML format. Exported records were based on only the representative family member for a given FamPat grouping in order to avoid duplication of records across multiple jurisdictions. Each record included key patent information and full details of both cited patent and non-patent literature references within the current record. As some searches generated very large numbers of records (i.e. hundreds of thousands), batch processing enabled large quantities of records to be handled in manageable formats, but required batches to be subsequently imported into a tool capable of processing the volumes of data considered. For this purpose, MATLAB was used, and a script (provided as a supplementary document upon request) was developed to convert each HTML batch file into a corresponding .MAT file (based on a pre-existing conversion script), ready for data cleaning processes.

D. PATENT INDICATOR DATA CLEANING PROCESS

Whilst the consistency of the Questel-Orbit patent data is of a high standard, several steps are required to extract patent indicator metrics from this data. This is based on WIPO preprocessing guidelines [48], to ensure that the datasets are translated into a tabulated format suitable for the automated analysis processes to follow, and to correct any easily rectifiable data-entry errors in the extracted data (such as the omission of application or priority dates from the relevant columns when these dates are available elsewhere). This allows a more accurate chronology of patent events to be established which is presented in chapter 5 of [19]. These chronologies when coupled with historical narratives in [19] provide evidence to suggest that patent data does in fact capture many of the real-life socio-economic, political, and

[3]https://doi.org/10.1108%2Fmi.2003.1235 21820cab.008

TABLE 3. Technology Life Cycle transition points based on literature evidence.

Case study	Last year of Emergence stage	Last year of Growth stage	Last year of Maturity stage	Technology Life Cycle transition point sources
Compact Fluorescent Lamps	1979	2011	–	[55, 56]
Electric vehicles	1997	2005	–	[57, 58]
Fiber optics (data transfer)	1970	1990	–	[59, 60]
Geothermal electricity	1958	–	–	[61]
Halogen lights	1959	–	–	[62, 63, 64]
Hydro electricity	1956	1975	–	[65]
Impact/Dot-matrix printers	1970	1984	1991	[66, 67, 68, 69, 70]
Incandescent lights	1882	1916	2008	[21, 64, 71]
Ink jet printer	1988	1996	2003	[69]
Internet	1982	2000	–	[72, 73, 74]
Landline telephones	1878	1945	2009	[75, 76]
Laser printer	1979	1993	–	[67, 77]
LED lights	2001	–	–	[55]
Linear Fluorescent Tube lights	1937	1990	2012	[62, 78, 79]
Nuclear electricity	1963	1981	–	[55]
Solar PV	1990	–	–	[55]
Solar thermal electricity	1968	–	–	[80, 81]
TFT-LCD	1990	2007	–	[32]
Thermal printers	1972	1985	2002	[67, 82, 83, 84, 85]
Tide-wave-ocean electricity	1966	–	–	[86, 87]
Turbojet	1939	1958	–	[88]
Wind electricity	1982	–	–	[55]
Wireless data transfer	1982	2002	–	[55]

organizational factors that influence the growth of a new technology beyond pure technological developments. As such, these profiles reflect the non-sequential nature of technology development observed by Gooday [15] and Edgerton [16]. This data cleaning process is not discussed in detail here, but is available as a supplementary document upon request.

E. TECHNOLOGY LIFE CYCLE STAGE MATCHING PROCESS

With bibliometric profiles extracted for each of the technologies considered, the first stage of analysis consists of identifying transition points between different stages of the Technology Life Cycle to establish time series segments for use in later comparative analysis. For the technologies considered in this study, evidence was identified from literature to suggest when these transitions had occurred, such as in the innovation timeline assessments prepared for a range of technologies by Hanna [50]. Full details of the transition points used in this study are provided in Table 3. These transition points define the time series segments each technology dataset was decomposed into relative to evidence presented by the complete historical development profiles and narratives.

Of the 23 technologies listed in Table 3, 20 had patent data pertaining to the emergence stage (i.e. excluding incandescent lights, landline telephones, and wireless data transfer). Therefore only those technologies with patent data available during the emergence stage are considered in the following analysis.

A nearest neighbor pattern matching process was also developed as outlined in section III-C.1 based on the work of Gao *et al.* [32]. This enables the analysis described in this paper to be expanded to additional technologies where evidence is not immediately apparent for the definition of TLC segments relative to the observed historical profile. This methodology is not discussed in further detail in this paper.

F. IDENTIFICATION OF SIGNIFICANT PATENT INDICATOR GROUPS

Having defined the time periods corresponding to each Technology Life Cycle stage for the technologies considered, it is now possible to segment the bibliometric time series into comparable phases of development. Significant predictors of substitution modes in each TLC stage are then identified by analyzing data from each TLC stage separately using the procedure outlined in Fig. 3.

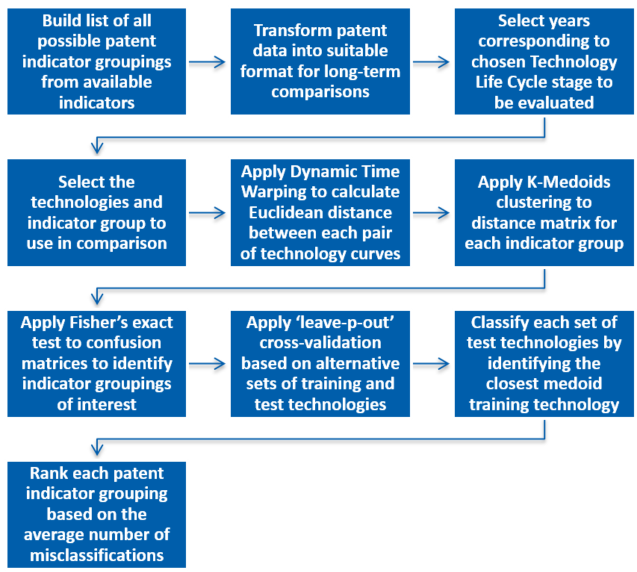


FIGURE 3. Overview of the process used to identify and rank significant patent indicator groups.

As discussed in sections III-C.2 and III-C.3 an unsupervised learning approach has been employed here based on applying Dynamic Time Warping (DTW) and the 'PAM' variant of K-Medoids clustering on the relative distance measures

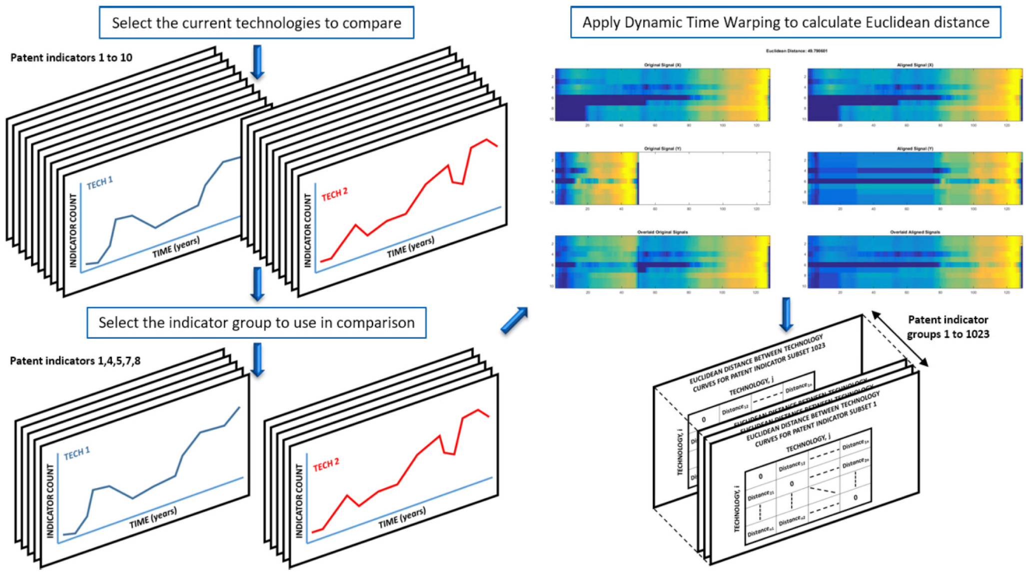


FIGURE 4. **Calculating the distance between each pair of technology time series for each indicator grouping (illustrative).**

calculated between time series. This is again implemented as a MATLAB script based on the DTW and K-Medoid functions made available by MathsWorks[4] [5], which is provided as a supplementary document upon request. The first step of this process involves generating a list of all unique subsets that can be created from the 10 patent indicator metrics considered. This produces 1,023 (i.e. $2^{10} - 1$) possible combinations of patent indicators to be tested.

Next, the raw patent data time series are transformed using an inverse hyperbolic sine function and normalized, to convert the data into a suitable format for long-term comparisons (see notes on preprocessing in Appendix C). The data points are then filtered based on the current Technology Life Cycle stage being considered, ensuring focus on comparable curve features.

After transforming the datasets and filtering based on the current Technology Life Cycle stage, Dynamic Time Warping is used to calculate the Euclidean distance between each pair of technology time series when compared using the time series dimensions specified by each patent indicator grouping in turn. This process is depicted visually in Fig. 4, illustrating the successive layers of filtering that are applied for each technology pairing and each patent indicator grouping considered. Fig. 4 also provides an illustration of how the DTW alignment process distorts technology profiles to reduce the dissimilarity between the multidimensional sets of features being compared (i.e. in this case aligning two ten-dimensional signals spanning different time periods). The output from this process is an $i \times j \times 1023$ distance matrix, where i and j specify the current technology pair, and the value quoted is the measured distance between multi-dimensional time series based on the current patent indicator subset. In parallel, the corresponding warping paths required to measure the distance between the N-dimensional curves in each condition are stored in two separate matrices for later use.

Using this distance matrix it is now possible to apply K-Medoids clustering to determine the technology groupings predicted when each patent indicator subset is used. By comparing the predicted technology groupings to those expected from the earlier literature classifications (see section II-B and Appendix B), a confusion matrix is created for each patent indicator subset that shows the alignment between predicted and target groupings. Fisher's exact test is then applied to each confusion matrix to calculate the probability of obtaining the observed clusters. In doing so, significant patent indicator subsets are identified based on those that have less than a 5% chance of natural occurrence.

G. RANKING OF GROUPED PATENT INDICATOR DIMENSIONS

As discussed in section III-C.3 and Appendix C *leave-p-out* cross-validation techniques provide a means to rank bibliometric indicator subsets that have been identified as producing a significant match to the expected technology groupings. More specifically, this form of permutation testing enables the ranking of these indicator subsets by providing an estimate of how accurately the current predictive model will perform in out-of-sample conditions (based on the results produced from using numerous reduced forms of the in-sample data sets). The first stage of this process consists of generating lists of all possible training technology and corresponding test technology combinations, when leaving one technology out at a time. Leave-one-out cross-validation was selected to ensure that a sufficient number of resampling points were present in each K-Medoids training set.

[4] https://uk.mathworks.com/help/stats/kmedoids.html

[5] https://stats.stackexchange.com/questions/131281/dynamic-time-warping-clustering

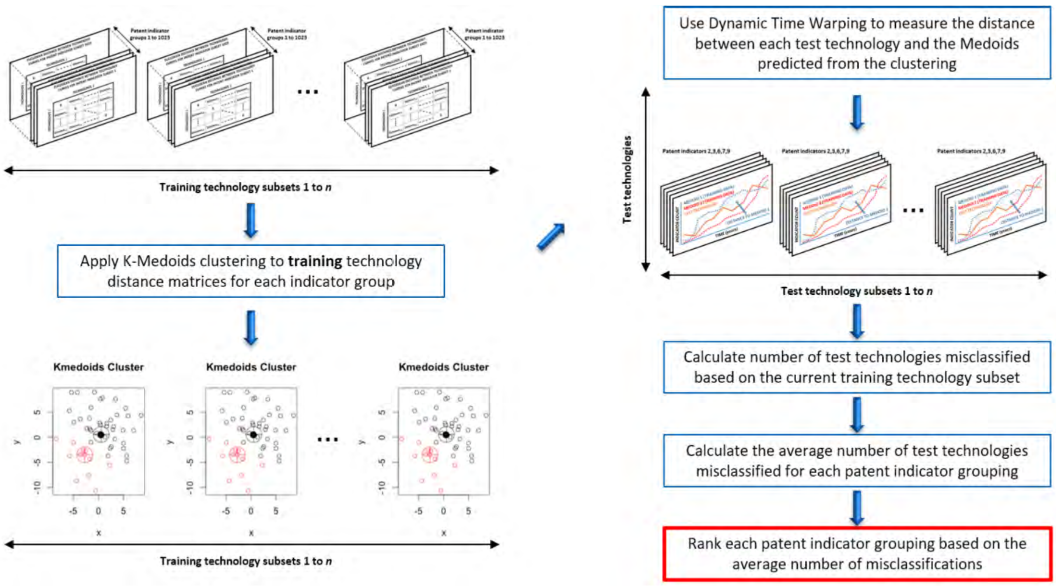


FIGURE 5. Ranking of grouped patent indicator dimensions (for illustration purposes only).

This enables meaningful clusters to be formed whilst still allowing a sufficient number of permutations to be tested. The procedure then progresses in a similar format to the initial calculation of distances between each pair of technology time series as shown in Fig. 4, except that this time distance measures are only calculated between pairs of training technologies, and the process is repeated for every possible combination of available training technologies. As such, the output from this process is now an i x j x 1023 x n distance matrix, where i and j now specify the current **training** technology pair considered, and n represents the number of training combinations that can be used.

K-Medoids clustering is once again applied to the resulting training technology distance matrices, from which two medoid technologies are identified for each patent indicator subset, in each training condition. The test technologies can now be evaluated individually against the two medoid curves identified in each training condition, to determine the closest medoid to the current test technology. This provides a classification for the test technologies based on each training condition and each patent indicator subset. Comparing predicted and expected technology classifications provides a count of the number of misclassified test technologies for the current combination of training technologies and patent indicators. This in turn is used to calculate the average number of test technologies misclassified for each patent indicator grouping across all of the training conditions considered. In this instance this means that each of the 1,023 possible patent indicator subsets is assessed for predictive performance, based on data only pertaining to the emergence stage, against 20 different training technology combinations. Consequently, an average misclassification value of *1* indicates that test technologies were incorrectly classified in all test conditions for the current patent indicator subset, whilst a value of *0* indicates no misclassifications in any test conditions. Using the average number of misclassifications ensures a symmetrical and unbiased treatment of all the patent indicator groupings considered. Finally, the results are sorted by the minimum average number of misclassifications, to rank the robustness of each patent indicator grouping. This procedure is illustrated in Fig. 5. From this ranked list of patent indicator subsets, the frequency of occurrence of individual patent indicators in the top ranked subsets can be observed. This is shown in Table 4 for the top performing 15% of subsets, with the average number of misclassifications against each subset. From this, the combination of indicators 4 and 6 is observed to appear in all of the best performing subsets (i.e. the four subsets that average a misclassification value of 0.1), whilst reappearing consistently in the majority of the remaining indicator subsets that achieve average misclassifications of less than 15%. This does not mean that all combinations with indicators 4 and 6 should automatically be used, as some sets containing additional indicators may have counter-acting effects. This is apparent since all combinations with indicators 4 and 6 were calculated in this analysis, with only those in Table 4 achieving the best levels of performance. It is normally advisable in model building to use as few parameters as possible (i.e. a parsimonious model), so Table 4 suggests that indicators 4 and 6 would be most appropriate for the classification scheme considered.

H. FUNCTIONAL MODEL BUILDING PROCESS

The ranking of different bibliometric indicator subsets provides a means to identify the time series dimensions that, when combined, are most likely to provide robust out-of-sample predictions of observed technological substitution modes. These indicators are therefore expected to form a reasonable basis for Adner's classification scheme.

TABLE 4. Frequency of individual patent indicators in the top ranked subsets.

Average number of misclassified test technologies	Subset	Subset indicators									
		1	2	3	4	5	6	7	8	9	10
0.1	[1,4,6]	X			X		X				
	[2,4,6]		X		X		X				
	[4,5,6]				X	X	X				
	[1,4,5,6]	X			X	X	X				
Frequency [<10% misclassified]		2	1	0	4	2	4	0	0	0	0
0.15	[2,4]		X		X						
	[4,5]				X	X					
	[4,6]				X		X				
	[4,7]				X			X			
	[1,2,4,6]	X	X		X		X				
	[2,4,5,6]		X		X	X	X				
	[1,2,3,4,6]	X	X	X	X		X				
	[1,2,4,5,6]	X	X		X	X	X				
	[2,4,5,6,7]		X		X	X	X	X			
	[2,4,6,7,9]		X		X		X	X		X	
	[2,4,6,7,10]		X		X		X	X			X
	[2,4,6,8,9]		X		X		X		X	X	
	[2,4,6,8,10]		X		X		X		X		X
	[1,2,3,4,5,6]	X	X	X	X	X	X				
	[2,3,4,6,8,9]		X	X	X		X		X	X	
	[2,3,4,6,8,10]		X	X	X		X		X		X
	[2,4,6,8,9,10]		X		X		X		X	X	X
	[2,3,4,6,8,9,10]		X	X	X		X		X	X	X
Frequency [<15% misclassified]		6	16	5	22	7	19	4	6	5	5

However, for subsequent causal exploration (not discussed in this paper), it is also necessary to trace the evolution of both observed technology profiles and corresponding mode predictions over time. For this, continuous time series are required. As such, a time-based regression model of mode prediction is also desirable, enabling technological development and substitution dynamics to be mapped directly against historical events, whilst accounting for the phase variation observed between different technologies. In this manner, variation in the patent measures defined in Table 2 is considered separately from phase variation between technologies, enabling model variation to be more accurately mapped to specific influences [84], [85]. This ensures that standard errors, confidence intervals, and significance tests are not misled by incorrectly aggregating distinct influences (i.e. overlaying influences specific to individual patent metrics with those linked to phase variance effects) [84]. Equally, the use of clustering means that this approach is less error prone and sensitive to outliers than using classical regression techniques in isolation [86]. Conversely, clustering provides limited insight into the residuals and variance associated with predictions of individual technologies, whereas the methods now applied enable further exploration of uncertainty in these predictions. Lastly, while the approach described in this section can generate a technology classification model without the preceding cross-validation and ranking exercises, doing so would not provide insight into how the chosen patent indicator subset may perform in comparison to other subsets, in terms of out-of-sample predictive capabilities. This means that in-sample classification results could potentially match those produced by other model variants, but when extended to new test cases the performance could vary drastically. The goodness-of-fit measures and permutation testing discussed in section V subsequently verify that the model developed in this section conforms to the predictive expectations inferred from the cross-validation exercises.

The preceding cross-validation exercise therefore acts as the first stage of the multi-level regression procedure discussed in section III, and provides a basis for an informed selection of the time series components to use in model building. Drawing on these findings, a time-dependent technology classification model is now developed using functional data analysis (see section III-C.4 and Appendix C) that is based on patent indicators 4 and 6 (i.e. the *number of non-corporate assignees* and the *number of cited references by priority year*).

Besides being present in all of the highest scoring sets of top ranked predictors, the chosen patent dimensions can potentially be associated with the rate of development in technology and science respectively. This is in the sense that *cited references* show a clear link to scientific production that is directly influencing technological development efforts, whilst the *number of non-corporates by priority year* (which counts the number of universities, academies, non-profit labs and technology research centers) is associated with the amount of lab work required to commercialize a technology. Considering the measure of non-corporates by priority year specifically, a large volume of lab work could indicate a lack of technological maturity, or the presence of considerable complexity in the emerging technology. By contrast, technologies with reduced non-corporate activity may represent simpler technologies that mature more rapidly or intuitively. *Non-corporates by priority year* could therefore equate to a measure of technological complexity, or effort required to mature.

However, it is also worth noting that there are other patent indicator subset couples/triples that perform nearly as well. It is possible that these other high-performing subsets may be in some way related to the chosen indicators (i.e. perfect orthogonality cannot be assumed between these metrics following on from the correlation analysis conducted by Gao *et al.* [32]). At this point it was decided to use the indicators specified as these have been seen to be the most statistically robust, whilst also being in good agreement with previous literature conclusions (as discussed further in chapter 2 of [19]).

Following on from the introduction to functional data analysis provided in Appendix C, and detailed methods presented in [87], the method outlined in Fig. 6 has been implemented in MATLAB for building a functional linear regression model for technology classification (the MATLAB script is available as a supplementary document upon request).

Taking the chosen time series dimensions as a starting point, a *functional data object* must first be created for each of the patent indicators (or *model components*) included in the chosen subset. This is necessary to combine all of the technology profiles considered into two regression terms: one representing the *number of non-corporates by priority year*, and a second representing the *number of cited references by priority year*. These terms, when multiplied by their

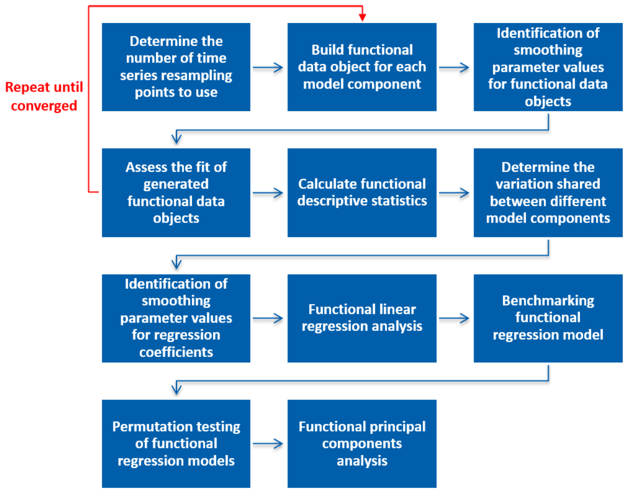


FIGURE 6. Functional model building process based on methods outlined in [87].

respective regression coefficients (calculated in the subsequent regression analysis), provide the relationship between the predicted mode of substitution and the two selected measures of science and technology. However, as the Technology Life Cycle segments being combined have a different number of observations for each case study technology, it is first necessary to resample the segmented time series based on a common number of resampling points. This ensures that even if a Technology Life Cycle stage spans 20 years in one time series, and spans 50 years in another, both time series will have 50 observations. This enables the two curves to be aligned relative to each other for the current Technology Life Cycle stage. Next, a B-spline basis system is created for each model component based on the common number of resampling points defined, and also for the regression coefficients (β_i) to be estimated by the functional linear regression analysis (see Eq. 1 and Eq. 3 in Appendix C, as well as sections 3.4.1, 3.4.2, 9.4.1 and 9.4.2 of [87]). Fig. 7 provides an illustrative example of how three B-spline basis systems are combined in this instance, corresponding to a single constant regression term in addition to two terms relating to the selected model components.

Before functional data objects can be generated from the B-spline basis systems, the degree of curve smoothing to be applied has to be determined (i.e. the tightness of fit). Following the process in [87] a *functional parameter object* that allows smoothness to be imposed on estimated functional parameters is now created (see section 5.2.4 of [87]). Functional parameter objects extend the existing datasets, by storing additional attributes relating to the smoothness constraints that need to be respected in any B-spline curve fit. A functional data object is then created for the current model component, using the new functional parameter object and an initial value of the smoothing parameter (λ). The degrees of freedom and generalized cross-validation criterion coefficient (see section 5.3 of [87]) can then be calculated for the current functional data object. By repeating this process for a range of λ values and plotting the results (not shown here) a suitable smoothing parameter can be identified to use in the final functional data object for each model component. Selection of a smoothing parameter in this fashion ensures that the functional data object generated will have the best chance of capturing dynamics present in the data, whilst being more likely to fit future out-of-sample technologies. An example of a smoothed functional data object generated for the *number of non-corporate assignees* associated with different technologies in a given priority year is shown in Fig. 8. This example illustrates how technology development profiles are realigned on to an equivalent time span, the duration of which is based on using either a) each technology's complete historical profile (as shown in this example), or b) specific comparable TLC stages, in the analysis. It is worth noting that although multiple technology profiles are shown in Fig. 8, as a functional data object, these curves are treated as a single data object when applied in the later functional regression analysis. In this regard, a single model component (i.e. each patent indicator) includes curves representative of all of the technologies considered.

Having created a functional data object representation of each model component from the selected bibliometric subset, the MATLAB script assesses the fit of each functional data object to the trend data. This is accomplished by calculating the residuals, variance, and standard deviations between the real and modeled values across the technology curves included, and across the time span of the Technology Life Cycle stage considered (see section 5.5 of [87]). Residuals are typically found to be within 10% of the actual data points, with RMSE values of less than 5%. As such, the distributions appear to show a good functional fit has been achieved on a technology-by-technology and life cycle basis (this is reviewed in more detail in chapter 5 of [19]). A related sanity check for the functional data objects generated for each model component (before they are used in the functional linear regression analysis) is the plotting of functional descriptive statistics (see section 6.1.1 of [87]). The functional mean and standard deviation of the data objects (i.e. solid and dashed lines) for the *number of non-corporates* and the *number of cited references* by priority year are shown in Fig. 9 and Fig. 10 respectively. These figures show that for both model components, variation from the mean generally increases towards the end of the emergence stage (as may be expected for a relatively diverse spread of technologies and industries). More specifically, for the two patent indicators plotted the standard deviation indicates that once these technologies begin to emerge, the rate of growth observed for these particular patent metrics varies significantly between technologies. In addition, mean functional data object values show that there is often an early surge, followed by a dip, in *non-corporates by priority year*, during the emergence phase before a technology achieves mainstream adoption. This potentially corresponds well to the hype cycle associated with new technologies in early development, when significant

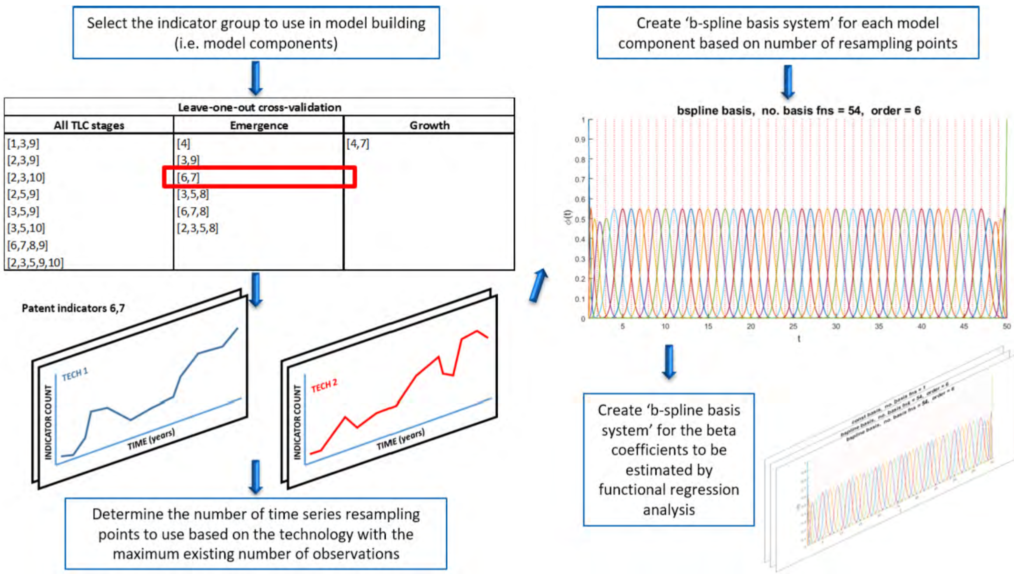


FIGURE 7. Building functional models of selected patent indicator groupings (for illustration purposes only).

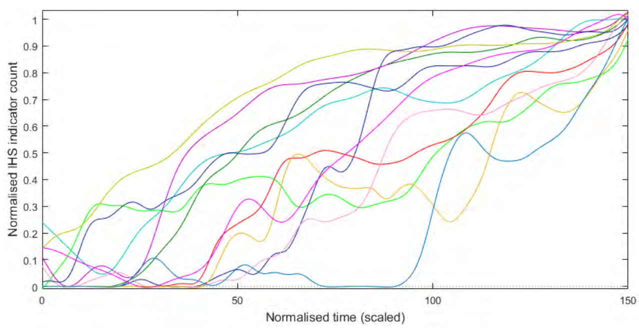


FIGURE 8. Functional Data Object for all technology profiles based on *non-corporates by priority year*.

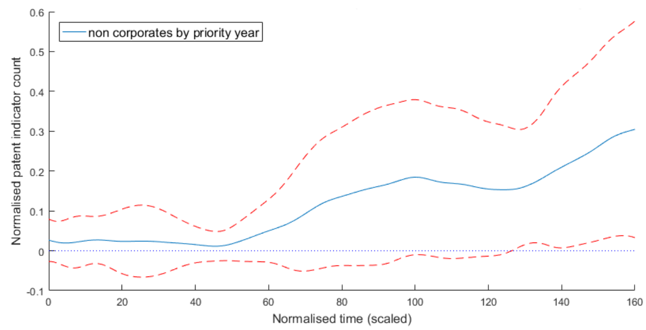


FIGURE 9. Mean and standard deviation of functional data object created for *non-corporates by priority year*.

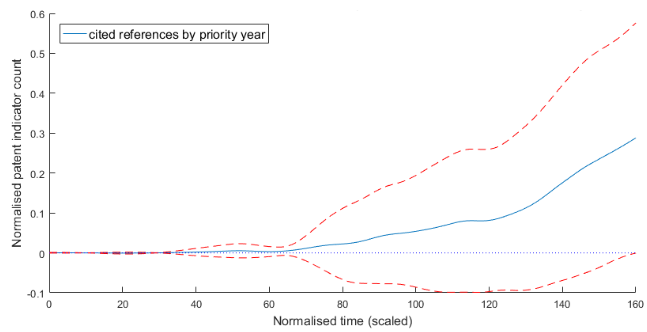


FIGURE 10. Mean and standard deviation of functional data object created for *cited references by priority year*.

levels of R&D may initially be committed to achieve commercialization, which can sometimes prove premature or short-lived [88]. By contrast, mean *cited references by priority year* values show a steadily accelerating growth during the emergence phase, without significant fluctuation, potentially implying that scientific development efforts are less sensitive to disturbances as they accumulate.

1) IDENTIFICATION OF SMOOTHING PARAMETER VALUES FOR REGRESSION COEFFICIENTS

With the functional data objects for each model component now ready, a cell array containing these components along with a constant predictor term (i.e. a cell array equal to 1 for all technology terms) is generated for use in the functional linear regression. Before running the final regression analysis, a smoothing parameter for the regression coefficient basis system has to be selected. This is separate from the earlier parameter for smoothing the technology profiles; this second parameter only addresses the roughness of regression coefficients. This is again necessary to try to prevent over-fitting, and ensure that functional linear regression converges on a model that has the best chance of performing well out-of-sample when extended to future datasets. This smoothing parameter is selected by calculating leave-one-out cross-validation scores (i.e. error sum of squares values) for functional responses using a range of smoothing parameter values, as per section 9.4.3 and 10.6.2 of [87]. The results of this selection process are presented in chapter 5 of [19]. The functional parameter object for the regression coefficient basis system is then redefined using this more optimized smoothing parameter value.

V. RESULTS AND DISCUSSION

The functional linear regression analysis is now run with the identified smoothing parameters and scalar response variables to identify the β_i coefficients and corresponding variance (used to define the 95% confidence bounds; see sections 9.4.3 and 9.4.4 of [87] respectively). Fig. 11 and Fig. 12 show the resulting β_i coefficients and confidence bounds (solid and dashed lines respectively) for the *number of non-corporates* and the *number of cited references* by priority year during the emergence phase when using a high-dimensional regression fit (i.e. when the beta basis system for each regression coefficient is made up of a large number of B-splines). In the high-dimensional model, the constant regression coefficient is found to have a value of 0.0071. Fig. 11 and Fig. 12 meanwhile show that values of the β_i coefficients and 95% confidence limits calculated for the two selected patent indicators change continuously with time during the emergence stage. Based on these coefficient functions the regression fit successfully identifies the correct mode of substitution, from patent data available in the emergence stage, for 19 of the 20 technologies considered, as summarized in Table 5. Therefore on preliminary inspection, this time-based classification model looks to provide a good degree of accuracy. However, further investigation is required to ensure the model is not over-fitted, and that the result is not simply a naturally occurring phenomenon.

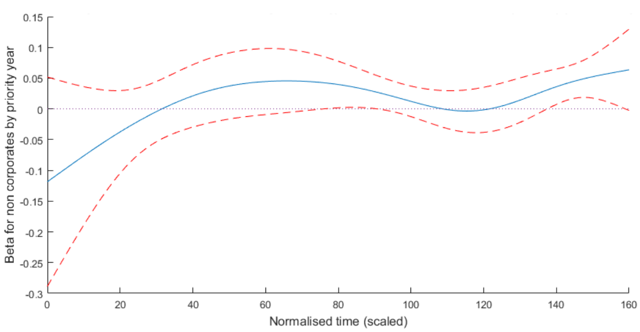


FIGURE 11. **Estimated regression coefficient for predicting technology cluster from *non-corporates by priority year* based on emergence stage data.**

From the confidence bounds on these plots it can be seen that for both the *number of non-corporates* and *cited references* indicator counts the variance across technology profiles is highest at the start of the emergence phase. This is typically when the least amount of data is available for comparing each technology, and also when development activity is most sporadic, which is unsurprising as this represents the point of greatest uncertainty. Consequently, the confidence intervals

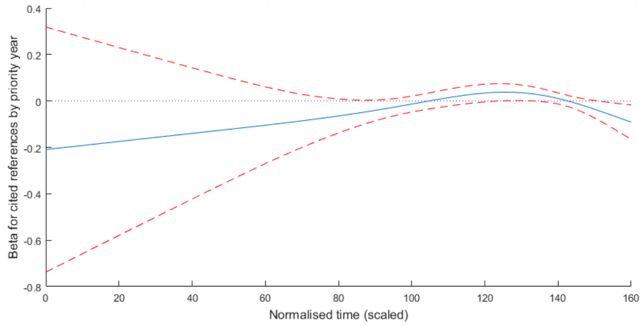


FIGURE 12. **Estimated regression coefficient for predicting technology cluster from *cited references by priority year* based on emergence stage data.**

suggest that the largest uncertainty around derived regression coefficients occurs at this point, particularly in the case of *cited references* (see Fig. 12). As time advances and more patent data becomes available, confidence bounds tighten around both of the calculated functional regression coefficients. By about 60% of the way through the emergence phase, the confidence bounds for the two model components have both narrowed to what appears to be near their minimum bandwidth. This possibly infers that any real-time classifications made after this point may be converging towards their final predicted label. This is observed in more detail for the technologies considered in chapter 5 of [19] by plotting the inner product of the patent indicator count values and regression coefficients over time. Taking time series segmentation a step further, these results and the successful use of segmentation in this analysis suggest that future extensions to real-time applications may be possible.

Fig. 11 and Fig. 12 also illustrate how the relative importance of each patent indicator in determining the predicted mode of substitution varies in time throughout the emergence phase (based on the datasets used). However, no causal explanation for why they receive these relative weightings is directly provided by these functions. Deviations from zero in these coefficient functions represent an increased positive or negative weighting for the associated patent indicator count at that time, within the determination of the predicted mode of substitution. For example, Fig. 11 suggests that any patents registered to non-corporate assignees at $t = 0$ (assuming these are present) will have a more significant influence on the predicted classification than at any other point in the emergence phase. The regression results also suggest that the impact of non-corporate activity next peaks around 40% of the way through the emergence phase (potentially corresponding to the hype effect suggested by Fig. 9), and again at the end of the emergence phase. For the *number of cited references*, this regression model suggests that the times of greatest impact on the mode of substitution are at the very beginning and end of the emergence stage. Whilst these coefficient plots gives some indication of relative patent indicator count weightings as time progresses, the cumulative nature of the inner products used in functional linear regression

TABLE 5. Results of high dimensional model fit.

Correct mode type	R^2	Adjusted R^2	Degrees of freedom 1	Degrees of freedom 2	F-ratio
19/20	0.7954	0.7713	7.7837	11.2163	5.6024

TABLE 6. Benchmarking results.

Model basis	Correct mode type	R^2	Adjusted R^2	Degrees of freedom 1	Degrees of freedom 2	F-ratio	p-value
Low dimension	19/20	0.8514	0.8340	10	9	5.1584	0.0107
Constant	18/20	0.6200	0.5753	2	17	13.8684	0.0003
Monomial	19/20	0.8139	0.7920	8	11	6.0139	0.0040

(see Eq. 3 in Appendix C) means it is not possible to visually infer which mode the test technology is converging towards from the coefficients alone. This requires the corresponding patent indicator counts that the coefficient terms are multiplied by for specific test technologies.

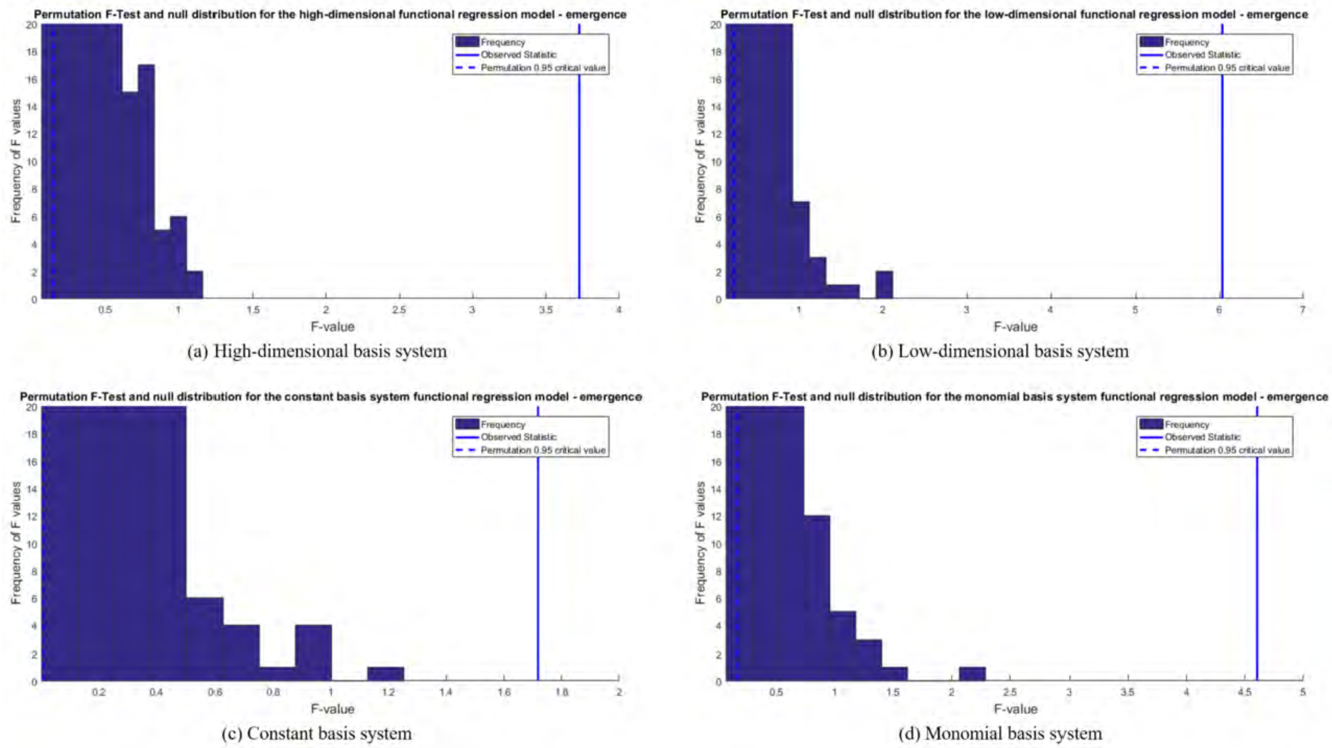
Regression coefficient plots help to provide a possible interpretation of relationships between each model component and predicted technology substitution modes. However, it is also necessary to check the *goodness-of-fit* measures associated with these results. These common statistical measures examine the amount of variability that is explained by the current model, and test the likelihood that the same result could be obtained by chance. As such, *R-Squared*, *adjusted R-Squared*, and *F-ratio* statistics are calculated (see section 9.4.1 and 9.4.2 of [87]) to assess the overall fit of the high-dimensional functional linear regression model. These are summarized in Table 5.

The R-squared and adjusted R-squared values in Table 5 suggest that a reasonable fit has been achieved with this model across the 20 technology profiles considered during the emergence phase. These values, which describe the proportion of variation that is predictable from the selected patent indicators, suggest a good level of accuracy based on the classification residuals. F-ratio values provide a measure of the variance observed between the two classification groups to the variance observed between individual technologies, taking into account the number of independent variables used in the model. In doing so, F-ratio values provide an indication of whether the classification grouping is significantly distinguishable from noise that might be otherwise observed between individual technologies. The degrees of freedom presented in Tables 5 and 6 are used to determine whether F-ratio values are above the critical F-ratio threshold or not, and are calculated based upon methods outlined for functional regression models in sections 9.4.1 and 9.4.2 of [87]. In this instance, the F-ratio of 5.60 with degrees of freedom 7.78 and 11.22 respectively implies that the relationship established has a p-value somewhere between 0.0041 and 0.0060. As such, this result appears to be significant at the 1% level, meaning that is unlikely that this classification label set would occur by chance. This compares well to the results of the cross-validation exercise outlined in section IV-G for ranking indicator sets based on likely predictive performance, and provides preliminary evidence to suggest that substitution classification based on a simplified version of Adner's framework is reasonable.

However, to ensure that it provides the most appropriate fit to available data, the original high-dimensional model was subsequently benchmarked against a low-dimensional model (i.e. a model where the beta basis system for each regression coefficient consists of a small number of B-splines), as well as constant and monomial based models. These variants use the same patent indicator terms as the high-dimensional model, ensuring that only the regression coefficients are changed (based on the alternative B-spline basis systems used). The corresponding 'goodness-of-fit' measures for the alternative functional linear regression variants are compiled in Table 6.

Whilst the R-squared and adjusted R-squared measures in Table 6 suggest that the low-dimensional model provides a better fit, the associated F-ratio score and corresponding p-value suggests a lower significance than the values observed for the high-dimensional model. Conversely, the constant basis model does not appear to provide as good a fit to the expected scalar responses from the R-squared and adjusted R-squared values, which is not surprising considering the more limited nature of models constructed from constant terms. Finally, the monomial basis system performs fractionally better on both the R-squared and adjusted R-squared measures, whilst also achieving a comparable level of significance to the high-dimensional model. Consequently, this benchmarking analysis suggests that the high-dimensional and monomial basis system models are the most suitable candidates. However, the performance of the models could possibly be further improved by sensitivity studies into the optimum number of B-splines to use in the regression fit.

To further validate the statistical significance of the four models considered here, permutation testing counts the proportion of generated F values that are larger than the F-statistic for each model (see section 9.5 of [87]).



(a) High-dimensional basis system

(b) Low-dimensional basis system

(c) Constant basis system

(d) Monomial basis system

FIGURE 13. Permutation F-Test and null distributions for functional regression models based on emergence stage data.

This involves repeatedly shuffling the expected mode classification labels versus the technology profiles being read (maintaining their original order) to see if it remains possible to fit the regression model to these reordered responses. This tests the sensitivity of the predicted classification labels to the order that the technology profiles appear in, to examine how the results would appear if there really was no relationship between the derived classification functions and original data. In doing so, this test also creates a null distribution versus the q^{th} quantile and observed F-statistic generated from the models themselves. The results of this analysis are shown in Fig. 13.

For statistical significance it is necessary that the observed test statistic (shown as a solid blue line) is in the tail of the distribution generated, implying that predicted classification responses would only occur very rarely (i.e. not by chance) if the data order was rearranged. Having generated classification models based on the most robust predictors from the earlier cross-validation exercise, all four models suggest that a significant relationship has been identified between expected substitution mode predictions and the two patent indicator dimensions used, that is specific to the data provided. However, as seen in Tables 5 and 6, the fit achieved varies depending on the model used. As such, these distributions appear to reinforce the significance of the patent indicators selected from the earlier clustering and ranking exercise. Additionally, the permutation testing in this last stage of analysis reveals that the high and low-dimensional model variants are likely to perform best out-of-sample as the observed F-statistics are furthest along each distribution's right tail, when compared to the distributions generated for the constant and monomial based models. This indicates that results from these two models have the lowest probability of occurring by chance, and are most likely to be generalizable to future datasets. A similar level of statistical significance is observed between the high and low-dimensional models, although as permutation testing was only based on 1,000 permutations, there is scope for the distributions to evolve further with more permutations. However, the constant basis system model appears not to perform as well out-of-sample here, with the observed F-statistic closest to the main body of the distribution. This, in combination with the other 'goodness-of-fit' measures in Tables 5 and 6, suggests that the high-dimensional functional linear regression model provides the best basis for a real-time technology substitution classification model, based on the selected patent indicators, from those tested in this analysis.

A. METHOD LIMITATIONS

Although precautions have been taken to ensure that the methods selected for this study address the problem posed of building a generalized technology classification model based on bibliometric data as rigorously as possible, there are known limitations that must be recognized. Many of these stem from the fact that technologies have been selected for which evidence is obtainable to indicate the mode of adoption followed. As such, the technologies considered here do not come from a truly representative cross-section of all industries, so it is possible that models generated will provide a better representation of those industries considered rather than a more generalizable result. This evidence-based approach also means that it is time-consuming to locate the necessary literature material to support classifying technology examples as arising from one mode of substitution or another, and to then compile the cleaned patent datasets for analysis. Consequently, a relatively limited number of technologies have been considered in this study, which should be expanded on to increase confidence in findings produced from this work. This also raises the risk that clustering techniques may struggle to produce consistent results for the small number of technologies considered. Furthermore, any statistical or quantitative methods used for modeling are unlikely to provide real depth of knowledge beyond the detection of correlations behind patent trends when used in isolation. Ultimately some degree of causal exploration, whether through case study descriptions, system dynamics modeling, or expert elicitation is required to shed more light on the underlying influences shaping technology substitution behaviors.

Other data-specific issues that could arise relate to the use of patent searches and the need to resample data based on variable length time series. The former relates to the fact that patent search results and records can vary to a large extent depending on the database and exact search terms used, although overall trends once normalized should remain consistent with other studies of this nature. The latter meanwhile refers to the fact that functional linear regression requires all technology case studies to be based on the same number of time samples. As such, as discussed in Appendix C, linear interpolation is used to ensure consistency between the number of observations, whilst possibly introducing some small errors which are not considered to be significant.

VI. CONCLUSIONS

Expanding on previous historical accounts of technological substitutions, this study has outlined a new methodology for automatically classifying the dynamics observed in substitutions based on matching scientific and technological development patterns against a recently proposed classification scheme. The conceptual framework outlined by Ron Adner defines technological substitutions in two dimensions based on the *emergence challenges* facing new technologies and the *extension opportunities* still available to existing technologies. The current study has focused on the *extension opportunity* dimension of this framework to facilitate a first attempt at translating Adner's work into a repeatable and generalizable method for automatically detecting substitution modes. From this, two high-level substitution classes appear to correspond to significantly different technology adoption characteristics (not discussed in this paper), with scientific foresight believed to play a crucial role in the identification of presumptive innovations, and performance stagnation leading to reactive transitions. The former class of substitution corresponds to situations where extension opportunities for existing technologies still appear to be high at the point when new technology emerges, whilst the latter relates to situations where the extension opportunities appear to be low (e.g. performance stagnation).

As such, this paper has considered 23 example technologies where literature evidence of performance development trends has been found, to test the ability to correctly identify associated adoption modes using bibliometric, pattern recognition, and statistical analysis techniques. This forms a multi-level regression methodology, where the patent indicators most likely to produce a reasonable basis for Adner's classification scheme are identified by an initial clustering and ranking analysis, before time dependent patent indicator models (for use in subsequent causal analysis) are constructed from functional linear regression. This allows variation specific to individual patent indicators to be considered separate from phase variation observed between technologies. The results obtained suggest that statistical analysis of patent indicator time series, segmented according to identified Technology Life Cycle features, provides a possible means for automated classification of technological substitutions using Adner's framework. Specifically, for the datasets considered, measures of the number of cited references and involvement of non-corporate entities by year during the emergence phase were found to provide a good indication of the expected mode of substitution when used as a basis for functional linear regression (correctly classifying 19 out of 20 technologies included in this stage), and performed consistently well in statistical ranking of both in and out-of-sample predictive capability. The selected patent data dimensions can also be associated with perceptions of scientific and technological production respectively.

Whilst these two patent metrics occur in all of the most robust predictor subsets (i.e. most reliable out-of-sample) when basing analysis on the emergence stage, this does not prove that these are the only indicators capable of predicting substitution modes. As discussed in section IV-H, the possibility of orthogonality has not been ruled out for the other patent indicators in Table 2. However, these two dimensions are also in good agreement with the technological anomaly arguments put forward by Constant in [18], so were felt to be reasonable for forming the basis of the time-based classification model that has been developed using functional linear regression. Subsequently, a regression fit made from beta coefficient functions with many B-spline elements was found to provide a viable means of correctly matching the mode of substitution to the technology profile being evaluated when considering multiple 'goodness of fit' measures.

Permutation testing of the time-based classification model further suggests that the regression fit is sensitive to the ordering of the expected mode labels, relative to the technology time series being considered. The relationship observed appears therefore to be based on the specifics of the individual technology curves considered, and does not appear to occur by chance. This implies that it may be possible to predict modes of substitution using Adner's framework from limited bibliometric data during the earliest stages of technology development, providing some evaluation of progress through the early stages of Technology Life Cycle is made (this can be obtained using a nearest neighbor matching process, not discussed in this paper). Equally, this suggests the functional regression corroborates the earlier statistical rankings produced using Dynamic Time Warping, K-Medoids clustering, and leave-one-out cross-validation leading to the selection of patent indicators, providing evidence of compatibility between the methods used in this analysis.

It is also important to remember the potential limitations of this study, which would need to be addressed for further confidence in the methodology. Firstly, only a relatively small number of technologies have been evaluated here due to the time-consuming process required for data extraction, preparation, and identification of supporting evidence from literature for the assignment of expected classification labels. Consequently, whilst precautions have been taken to minimize the risk of model over-fitting, the cross-validation procedures employed would benefit from further verification with a more diverse spread of technologies to ensure that out-of-sample errors are accurately captured. Regression models based on small sample sizes can be very fickle to the datasets they are calibrated to, so it cannot be ruled out that the results obtained are a better fit to the industries included, rather than a model that can be generalized to all technologies.

However, perhaps the most important note of caution relates to the quantitative approaches used here. Whilst statistical approaches are well-suited to detecting underlying correlations in historical and experimental datasets, this on its own does not provide a detailed understanding of the causation behind associated events. This is particularly relevant when considering the breadth of reasons for technological stagnations, 'failures', or presumptive leaps. Equally, statistical methods are not generally well suited to predicting disruptive events and complex interactions; other simulation techniques such as system dynamics and agent based modeling perform better in these areas. Accordingly, to identify causation and test the sensitivity of technological substitution patterns to variability arising from real-world socio-technical behaviors not captured in simple bibliometric indicators (such as the influence of competition, organizational, and economic effects), the fitted regression model also needs to be evaluated from a causal perspective.

Similarly, to demonstrate practical applicability, the mode of substitutions considered here based on Adner's classification scheme need to be related to observed adoption characteristics (not discussed in this paper). Consequently, a system dynamics model built on the regression functions identified in this study is proposed (although not discussed here), to calibrate these extracted technology profiles and mode predictions to empirical adoption data. This aims to more thoroughly explore the causal mechanisms relating early indicators of technological substitution to the eventual adoption patterns observed, and provide a means of applying greater reasoning to the relationships identified here. In doing so, this may enable businesses to recognize substitution patterns at an early stage, and subsequently determine the likelihood of an emerging technology out-performing and displacing the existing dominant technology in a given time frame.

ACKNOWLEDGMENT

The authors are grateful for assistance provided by the University of Bristol's statistics clinic.

REFERENCES

[1] F. W. Geels and J. Schot, "Typology of sociotechnical transition pathways," *Res. Policy*, vol. 36, no. 3, pp. 399–417, 2007.

[2] H. C. Lucas, Jr. and J. M. Goh, "Disruptive technology: How Kodak missed the digital photography revolution," *J. Strategic Inf. Syst.*, vol. 18, no. 1, pp. 46–55, Mar. 2009. doi: 10.1016/j.jsis.2009.01.002.

[3] J. Hall and P. Rosson, "The impact of technological turbulence on entrepreneurial behavior, social norms and ethics: Three Internet-based cases," *J. Bus. Ethics*, vol. 64, no. 3, pp. 231–248, Mar. 2006. doi: 10.1007/s10551-005-5354-z.

[4] T. U. Daim, G. Rueda, H. Martin, and P. Gerdsri, "Forecasting emerging technologies: Use of bibliometrics and patent analysis," *Technol. Forecasting Social Change*, vol. 73, no. 8, pp. 981–1012, Oct. 2006. doi: 10.1016/j.techfore.2006.04.004.

[5] R. Peres, E. Muller, and V. Mahajan, "Innovation diffusion and new product growth models: A critical review and research directions," *Int. J. Res. Marketing*, vol. 27, no. 2, pp. 91–106, Jun. 2010. doi: 10.1016/j.ijresmar.2009.12.012.

[6] F. M. Bass, "Comments on 'a new product growth for model consumer durables the bass model,'" *Manage. Sci.*, vol. 50, no. 12, pp. 1833–1840, Dec. 2004. doi: 10.1287/mnsc.1040.0300.

[7] B. Dattée and H. B. Weil, "Dynamics of social factors in technological substitutions," *Technol. Forecasting Social Change*, vol. 74, no. 5, pp. 579–607, Jun. 2007. doi: 10.1016/j.techfore.2007.03.003.

[8] R. N. Foster, *Innovation: The Attacker's Advantage*. New York, NY, USA: McKinsey, 1986.

[9] J. M. Utterback, *Mastering the Dynamics of Innovation: How Companies Can Seize Opportunities in the Face of Technological Change*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[10] S. Myers and D. G. Marquis, *Successful Industrial Innovations. A Study of Factors Underlying Innovation in Selected Firms*. Alexandria, VA, USA: National Sci. Found., 1969.

[11] J. Poolton and I. Barclay, "New product development from past research to future applications," *Ind. Marketing Manage.*, vol. 27, no. 3, pp. 197–212, May 1998. doi: 10.1016/S0019-8501(97)00047-3.

[12] R. Adner and R. Kapoor, "Innovation ecosystems and the pace of substitution: Re-examining technology S-curves," *Strategic Manage. J.*, vol. 37, no. 4, pp. 625–648, Apr. 2015. doi: 10.1002/smj.2363.

[13] T. P. Hughes, *Networks of Power: Electrification in Western Society, 1880-1930*. Baltimore, MD, USA: JHU Press, 1983.

[14] A. Sood and G. J. Tellis, "Technological evolution and radical innovation," *J. Marketing*, vol. 69, no. 3, pp. 152–168, Jul. 2005. doi: 10.1509/jmkg.69.3.152.66361.

[15] G. Gooday, "Re-writing the 'book of blots': Critical reflections on histories of technological 'failure'," *Hist. Technol.*, vol. 14, no. 4, pp. 265–291, Jul. 1998. doi: 10.1080/07341519808581934.

[16] D. Edgerton, *The Shock of the Old: Technology and Global History Since 1900*. London, U.K.: Profile Books, 2011.

[17] M. A. Schilling and M. Esmundo, "Technology S-curves in renewable energy alternatives: Analysis and implications for industry and government," *Energy Policy*, vol. 37, no. 5, pp. 1767–1781, May 2009. doi: 10.1016/j.enpol.2009.01.004.

[18] E. W. Constant, "A Model for Technological Change Applied to the Turbojet Revolution," *Technol. Culture*, vol. 14, no. 4, p. 553, Oct. 1973. doi: 10.2307/3102443.

[19] I. Marr, "Identifying the mode and impact of technological substitutions," Ph.D. dissertation, Fac. Eng., Univ. Bristol, Bristol, U.K., 2018.

[20] C. M. Christensen and R. S. Rosenbloom, "Explaining the attacker's advantage: Technological paradigms, organizational dynamics, and the value network," *Res. Policy*, vol. 24, no. 2, pp. 233–257, Mar. 1995. doi: 10.1016/0048-7333(93)00764-K.

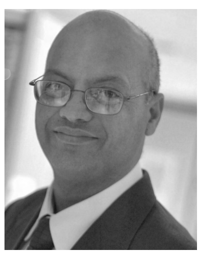
[21] Y. S. Chang and S. J. Baek, "Limit to improvement: Myth or reality?" *Technol. Forecasting Social Change*, vol. 77, no. 5, pp. 712–729, Jun. 2010. doi: 10.1007/978-3-319-91394-0_7.
[22] V. Smil, "World history and energy," in *Encyclopedia of Energy*. Amsterdam, The Netherlands: Elsevier, 2004, pp. 549–561. doi: 10.10162Fb0-12-176480-x2F00025-5.
[23] R. N. Foster, "Timing technological transitions," *Technol. Soc.*, vol. 7, nos. 2–3, pp. 127–141, Jan. 1985. doi: 10.1016/0160-791X(85)90022-3.
[24] W. Graus and E. Worrell, "Trend in efficiency and capacity of fossil power generation in the EU," *Energy Policy*, vol. 37, no. 6, pp. 2147–2160, Jun. 2009. doi: 10.1016/j.enpol.2009.01.034.
[25] T. Zachariadis, "On the baseline evolution of automobile fuel economy in Europe," *Energy Policy*, vol. 34, no. 14, pp. 1773–1785, Sep. 2006. doi: 10.1016/j.enpol.2005.01.002.
[26] B. R. Martin, "The use of multiple indicators in the assessment of basic research," *Scientometrics*, vol. 36, no. 3, pp. 343–362, Jul. 1996. doi: 10.1007/BF02129599.
[27] J. Schmookler, *Invention Economic Growth*. Cambridge, MA, USA: Harvard Univ. Press, Jan. 1966. doi: 10.41592Fharvard.9780674432833.
[28] F. M. Scherer, "Firm size, market structure, opportunity, and the output of patented inventions," *Amer. Econ. Rev.*, vol. 55, no. 5, pp. 1097–1125, 1965.
[29] A. D. Little, *The Strategic Management of Technology*. Cambridge, MA, USA: Arthur D. Little, 1981.
[30] R. J. Watts and A. L. Porter, "Innovation forecasting," *Technol. Forecasting Social Change*, vol. 56, no. 1, pp. 25–47, Sep. 1997. doi: 10.1016/S0040-1625(97)00050-4.
[31] R. Haupt, M. Kloyer, and M. Lange, "Patent indicators for the technology life cycle development," *Res. Policy*, vol. 36, no. 3, pp. 387–398, Apr. 2007. doi: 10.1016/j.respol.2006.12.004.
[32] L. Gao *et al.*, "Technology life cycle analysis method based on patent documents," *Technol. Forecasting Social Change*, vol. 80, no. 3, pp. 398–407, Mar. 2013. doi: 10.1016/j.techfore.2012.10.003Get.
[33] C. Lee, J. Kim, O. Kwon, and H.-G. Woo, "Stochastic technology life cycle analysis using multiple patent indicators," *Technol. Forecasting Social Change*, vol. 106, pp. 53–64, May 2016. doi: 10.1016/j.techfore.2016.01.024.
[34] S. Ranaei, M. Karvonen, A. Suominen, and T. Kässi, "Patent-based technology forecasting: case of electric and hydrogen vehicle," *Int. J. Energy Technol. Policy*, vol. 12, no. 1, p. 20, 2016. doi: 10.1504/IJETP.2016.074490.
[35] C. V. Trappey, H.-Y. Wu, F. Taghaboni-Dutta, and A. J. C. Trappey, "Using patent data for technology forecasting: China RFID patent analysis," *Adv. Eng. Informat.*, vol. 25, no. 1, pp. 53–64, Jan. 2011. doi: 10.1016/j.aei.2010.05.007.
[36] S.-J. Liu and J. Shyu, "Strategic planning for technology development with patent analysis," *Int. J. Technol. Manage.*, vol. 13, nos. 5–6, pp. 661–680, 1997.
[37] A. Chen and R. Chen, "Design patent map: An innovative measure for corporative design strategies," *Eng. Manage. J.*, vol. 19, no. 3, pp. 14–29, Apr. 2007.
[38] N. Lambert. (2008). *Orbit and Questel-Orbit: Farewell and Hail*. Accessed: Jun. 2017. [Online]. Available: http://www.infotoday.com/searcher/feb00/lambert.htm
[39] J. Lin, S. Williamson, K. Borne, and D. DeBarr, "Pattern recognition in time series," *Adv. Mach. Learn. Data Mining Astron.*, vol. 1, pp. 617–645, Mar. 2012.
[40] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statist. Surveys*, vol. 4, no. 0, pp. 40–79, 2010. doi: 10.12142F09-ss054.
[41] J. S. Marron, J. O. Ramsay, L. M. Sangalli, and A. Srivastava, "Functional data analysis of amplitude and phase variation," *Stat. Sci.*, vol. 30, no. 4, pp. 468–484, Nov. 2015. doi: 10.12142F15-sts524.
[42] A. Verbeek, K. Debackere, M. Luwel, and E. Zimmermann, "Measuring progress and evolution in science and technology—I: The multiple uses of bibliometric indicators," *Int. J. Manage. Rev.*, vol. 4, no. 2, pp. 179–211, Jun. 2002. doi: 10.1111/1468-2370.00083.
[43] U. Schmoch, "Indicators and the relations between science and technology," *Scientometrics*, vol. 38, no. 1, pp. 103–116, Jan. 1997. doi: 10.1007/BF02461126.
[44] V. Albino, L. Ardito, R. M. Dangelico, and A. M. Petruzzelli, "Understanding the development trends of low-carbon energy technologies: A patent analysis," *Appl. Energy*, vol. 135, pp. 836–854, Dec. 2014. doi: 10.1016/j.apenergy.2014.08.012.
[45] F. Rizzi, N. J. van Eck, and M. Frey, "The production of scientific knowledge on renewable energies: Worldwide trends, dynamics and challenges and implications for management," *Renew. Energy*, vol. 62, pp. 657–671, Feb. 2014. doi: 10.1016/j.renene.2013.08.030.
[46] G. Mao, X. Liu, H. Du, J. Zuo, and L. Wang, "Way forward for alternative energy research: A bibliometric analysis during 1994–2013," *Renew. Sustain. Energy Rev.*, vol. 48, pp. 276–286, Aug. 2015. doi: 10.1016/j.rser.2015.03.094.
[47] B. Dong, G. Xu, X. Luo, Y. Cai, and W. Gao, "A bibliometric analysis of solar power research from 1991 to 2010," *Scientometrics*, vol. 93, no. 3, pp. 1101–1117, May 2012. doi: 10.1007/s11192-012-0730-9.
[48] A. Trippe, "Guidelines for preparing patent landscape reports," WIPO, Geneva, Switzerland, WIPO Publication No. 946E, 2018.
[49] S. Helm, Q. Tannock, and I. Iliev, "Renewable energy technology: Evolution and policy implications - Evidence from patent literature," WIPO, Geneva, Switzerland, Tech. Rep., 2014. [Online]. Available: https://www.wipo.int/publications/en/details.jsp?id=3891&plang=EN
[50] R. Hanna, R. Gross, J. Speirs, P. Heptonstall, and A. Gambhir, "Innovation timelines from invention to maturity. A rapid review of the evidence on the time taken for new technologies to reach widespread commercialisation," UKERC Technol. Policy Assessment. UK Energy Res. Centre, London, U.K., Tech. Rep., 2015. [Online]. Available: http://www.ukerc.ac.uk/asset/ADA12E92-C1DC-4033-8CFA63AC9EA9FE59.6B438A7E-474E-437E-89025FD302A10A7C/
[51] M. Weiss, H. M. Junginger, and M. K. Patel, "Learning energy efficiency: Experience curves for household appliances and space heating," Copernicus Inst., Utrecht Univ., Utrecht, The Netherlands, Tech. Rep. NWS-E-2008-26, 2008. [Online]. Available: http://dspace.library.uu.nl/handle/1874/32937
[52] S. Ranaei, M. Karvonen, A. Suominen, and T. Kässi, "Forecasting emerging technologies of low emission vehicle," in *Proc. Conf. Portland Int. Center Manage. Eng. Technol. Infrastruct. Service Integr.*, Jul. 2014, pp. 2924–2937.
[53] F. Yuan and K. Miyazaki, "Understanding the dynamic nature of technological change using trajectory identification based on patent citation network in the electric vehicles industry," in *Proc. Conf. Portland Int. Center Manage. Eng. Technol. Infrastruct. Service Integr.*, Jul. 2014, pp. 2780–2790.
[54] G. Cattani, "Technological pre-adaptation, speciation, and emergence of new technologies: How Corning invented and developed fiber optics," *Ind. Corporate Change*, vol. 15, no. 2, pp. 285–318, Mar. 2006. doi: 10.1093/icc/dtj016.
[55] J. Hecht, *City of Light: The Story of Fiber Optics*. New York, NY, USA: Oxford Univ. Press, 2004.
[56] W. E. Glassley, *Geothermal Energy: Renewes Energy Environment*. Boca Raton, FL, USA: CRC Press, 2014.
[57] P. Waide, S. Tanishima, and P. Harrington, *Light's Labour's Lost: Policies for Energy-Efficient Lighting*. Paris, France: OECD, 2006.
[58] P. Menanteau and H. Lefebvre, "Competing technologies and the diffusion of innovations: The emergence of energy-efficient lamps in the residential sector," *Res. Policy*, vol. 29, no. 3, pp. 375–389, Mar. 2000. doi: 10.1016/S0048-7333(99)00038-4.
[59] Navigant Consulting Europe, "Life cycle assessment of ultra-efficient lamps," in *Department of the Environment and Energy*. London, U.K.: FARA, 2009.
[60] M. C. Connelly and J. Sekhar, "U. S. energy production activity and innovation," *Technol. Forecasting Social Change*, vol. 79, no. 1, pp. 30–46, Jan. 2012. doi: 10.1016/j.techfore.2011.05.001.
[61] A. F. Mayadas, R. C. Durbeck, W. D. Hinsberg, and J. M. McCrossin, "The evolution of printers and displays," *IBM Syst. J.*, vol. 25, nos. 3–4, pp. 399–416, 1986.
[62] E. Tomash. (2009). *The U.S. Computer Printer Industry*. Accessed: Aug. 2017. [Online]. Available: http://jacques-andre.fr/chi/chi90/tomash.html
[63] A. Agrawal and J. Dwoskin. (2013). *Timeline of Printers*. Accessed: Aug. 2017. [Online]. Available: http://www.jeffdwoskin.com/printers/menu.html
[64] N. Clymer and S. Asaba, "A new approach for understanding dominant design: The case of the ink-jet printer," *J. Eng. Technol. Manage.*, vol. 25, no. 3, pp. 137–156, Sep. 2008. doi: 10.1016/j.jengtecman.2008.06.003.
[65] H. Acee, "Disruptive technologies: an expanded view," Ph.D. dissertation, Sloan School Manage., MIT, Cambridge, MA, USA, 2001.

[66] M. F. Gendre. (2010). *Two Centuries of Electric Light Source Innovations*. Accessed: Aug. 2017. [Online]. Available: http://www.einlightred.tue.nl/lightsources/history/light_history.pdf
[67] W. Lemstra, *The Internet Bubble Impact Develop. Path Telecommunication Sector*. Eemnes, The Netherlands: Industry-Insights BV, 2006.
[68] R. Zakon, ''Hobbes' Internet timeline,'' Zakon Group LLC, USA, Nov. 1997. [Online]. Available: https://www.zakon.org/robert/internet/timeline/
[69] P. von Stackelberg, (2018). *Technological Eras*. Accessed: Aug. 2017. [Online]. Available: http://www.transmediadigest.com/wp-content/uploads/2012/05/TechEras-Interactive.pdf
[70] J. R. Ortt and J. P. Schoormans, ''The pattern of development and diffusion of breakthrough communication technologies,'' *Eur.J. Innov. Manage.*, vol. 7, no. 4, pp. 292–302, Dec. 2004. doi: 10.11082F14601060410565047.
[71] S. Teltscher *et al.*, *Measuring the Information Society Report*. Geneva, Switzerland: ITU, 2013.
[72] A. E. Grant *et al.*, *Communication Technology Update Fundamentals*. Abingdon, U.K.: Taylor Francis, 2013.
[73] J. Tidd, J. R. Bessant, and K. Pavitt, *Managing Innovation: Integrating Technological, Market Organizational Change*. Hoboken, NJ, USA: Wiley, 1997, vol. 4.
[74] A. R. Köhler, ''Anticipatory eco-design strategies for smart textiles: Perspectives on environmental risk prevention in the development of an emerging technology,'' doctoral thesis, Ind. Des. Eng., Dept. Des. Eng., Des. Sustainability, Sep. 2013. doi: 10.4233/uuid:850be7ae-1f9e-4b3f-b49f-242488bab216.
[75] EIA. (2018). *Energy timelines - Solar Thermal*. Accessed: Aug. 2017. [Online]. Available: https://www.eia.gov/kids/energy.?page=tl_solarthermal
[76] A. Grübler *et al.*, ''Policies for the energy technology innovation system (ETIS),'' in *Global Energy Assessment-Toward a Sustainable Future*. Cambridge, U.K.: Cambridge Univ. Press, 2012, ch. 24, pp. 1665–1744.
[77] I. McLoughlin and K. C. Keong. (2001). *Computer Peripherals*. [Online]. Available: www.lintech.org
[78] P. Gregory, *Chemistry and Technology of Printing and Imaging Systems*. Berlin, Germany: Springer, 1996.
[79] S. cientific. (2018). *United States Securities and Exchange Commission—Form 10-K*. Accessed: Aug. 2017. [Online]. Available: https://www.sec.gov/Archives/edgar/data/41133/000004113307000026/fy200710k.htm
[80] B. Cartridges. ; ''The Evolution Of Printing,''2017, accessed: Aug. 2017. [Online]. Available: https://www.redbuscartridges.com/printing/evolution-of-printing/
[81] J. W. Tester, E. M. Drake, M. J. Driscoll, M. W. Golay, and W. A. Peters, *Sustainable Energy: Choosing Among Options*. Cambridge, MA, USA: MIT Press, 2012.
[82] T. D. Corsatea, ''Increasing synergies between institutions and technology developers: Lessons from marine energy,'' *Energy Policy*, vol. 74, pp. 682–696, Nov. 2014. doi: 10.10162Fj.enpol.2014.07.006.
[83] F. Geels, ''Co-evolutionary and multi-level dynamics in transitions: The transformation of aviation systems and the shift from propeller to turbojet (1930–1970),'' *Technovation*, vol. 26, no. 9, pp. 999–1016, Sep. 2006. doi: 10.10162Fj.technovation.2005.08.010.
[84] H. Goldstein, *Multilevel Statistical Models*, vol. 922. Hoboken, NJ, USA: Wiley, 2011.
[85] A. Gelman and J. Hill, *Data Analis Using Regression Multilevel/Hierarchical Models*. Cambridge, U.K.: Cambridge Univ. Press, 2006.
[86] P. Joshi. (2017). *When to Use Linear Regression, Clustering, or Decision Trees*. Accessed: Jun. 2018. [Online]. Available: https://dzone.com/articles/decision-trees-vs-clustering-algorithms-vs-linear
[87] J. Ramsay, G. Hooker, and S. Graves, *Functional Data Analysis with R and MATLAB*. New York, NY, USA: Springer, 2009. doi: 10.10072F978-0-387-98185-7.
[88] A. Linden and J. Fenn, ''Understanding gartners hype cycles,'' Gartner, Stamford, CT, USA, Tech. Rep. NÂřR-20-1971, 2003.

IAN P. MARR received the M.Eng. (Hons.) degree in aeronautical engineering from Loughborough University in 2008, before moving to Bristol to work on Operational Research topics in Airbus' Future Projects Office. In this role he focuses primarily on disruptive transport system operations. He received his EngD in Systems from the University of Bristol in 2018, with a focus on the application of System of System methodologies to commercial aviation and technology development. His research interests include the modeling and simulation of disruptive innovations, technology diffusion processes, game theory, machine learning, network dynamics, behavioral modeling, and agent-based simulations.

CHRIS MCMAHON was a Professor of engineering design with the University of Bristol, until 2016, where he was the Director of the Engineering Design and Systems Programmes. He has worked as an academic since 1984 carrying out teaching and research in engineering design. He is currently a Visiting Researcher/Professor with the Technical University of Denmark and the University of Bristol. His research interest includes design informatics. He is on the Editorial Advisory Board of several journals. From 2011 to 2013, he was the President of the International Design Society.

MARK LOWENBERG is currently a Professor of flight mechanics with the University of Bristol. His research interests include nonlinear analysis of dynamical systems, principally in aircraft dynamics applications but also in other related fields, and innovation in aircraft design and development.



SANJIV SHARMA currently works at Airbus within the Department of Engineering Strategy, Processes Methods and Tools. He is also a member of the Airbus Experts Community, with expertise in modeling and simulation methods and tools. He has over 30 years of experience in engineering design analysis and integrated system engineering frameworks. He is also the Project Leader of the Airbus' Uncertainty Quantification and Management Project. His research interests include set-based design, decision-making under uncertainty, and nonlinear analyses. He is a Fellow of the Institute of Mathematics and its Applications.