Identifying the mode and impact of technological substitutions

Ian Marr[a], Chris McMahon[b], Mark Lowenberg[c], Sanjiv Sharma[a]

[a] *Airbus Operations Ltd*
[b] *Technical University of Denmark*
[c] *University of Bristol*

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 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, statistical, and data-driven approaches to develop a technology classification model from historical datasets where literature evidence supports mode labelling. 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 substitution modes can be recognised through automated processing of patent data. Further, 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.

Keywords:

Technological substitutions, Patent bibliometrics, Pattern recognition, Classification, Technology Life Cycle, Emergence

2. Introduction

The introduction of new technologies into heavily regulated industries such as aerospace is often a very complex, time-consuming and expensive challenge that requires significant levels of research and development in order to ensure a successful technology substitution. This challenge is exacerbated when new technology options represent a fundamental shift away from well-established principles, as the risk and uncertainties involved increase significantly. This is currently the case in the anticipated transition from conventional turbojet aircraft architectures to all new electric configurations, and equally for the adoption of technologies enabling mass manufacturing and customisation processes in aerospace production lines. At the same time, 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. In this regard it is beneficial to be able to identify early on 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 approach would become the industry 'mainstream' technology option.

This paper examines historical cases where emerging technologies have been presumed in advance to have development opportunities beyond those of pre-existing technologies, subsequently leading to transitions occurring before performance of the existing technology has stagnated. Based on conceptual models published previously that consider the mode of technological substitution and the relation to both scientific and technological developments, this paper looks to test whether separate bibliometric measures of scientific and technological development can be combined to provide an indication of the mode of adoption likely to occur from patent data available during the early stages of development. Bibliometric, pattern recognition, statistical and other data-driven analysis techniques are applied to technologies identified as having been adopted as a result of either prior technological stagnation (which we term technological failure with reactive substitution), or as a result of a presumptive leap being made, in order to identify early indicators of the mode of technological substitution. In the case of substitutions as a result of a presumptive leap, some forthcoming technical limit is recognised that prompts a transition before the current technology has stagnated. This historical classification has led to the development of a functional linear regression model that can be used in supporting technology strategy and innovation management by indicating the likely mode of adoption from key technology development indicators. In doing so, this paper has found good evidence in historical records to support the literature based categorisation into reactive and presumptive modes of substitution, and demonstrated that these modes can be recognised through automated processing of patent data. Preliminary evidence is also provided that suggests it may be possible to use partially complete datasets (i.e. segmented time series) to predict the end mode of substitution, potentially enabling future extensions to real-time applications. The paper begins by providing some background to technology substitutions and patent-based analysis techniques in section 3, followed by an overview of bibliometric data sources, statistical analysis, and method selection in section 4. Details of the derivation of the technology classification model using statistical ranking and functional data analysis are then provided in section 5, along with the corresponding results and discussions in section 6. Finally, conclusions from the patent indicator ranking and classification model building exercises are then summarised in section 7.

3. Background

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

3.1. Technology forecasting, substitution patterns, and technological failure

Correctly predicting which emerging technologies are likely to be most influential can ensure that a firm is best positioned to gain a large advance over their competitors when the new technology comes to fruition. Conversely, failure to anticipate the arrival of big technological shifts can leave firms severely diminished. This is illustrated by the dramatic impact on Kodak's business following the introduction of digital photography, that rendered many of the firm's existing film products obsolete following an early lead in the digital field that was not fully capitalised upon [47]. Equally, investing heavily in a nascent technology too soon can have grave consequences, as Bertlesmann found from investing in Napster [33]. As such, forecasting techniques are often used to determine strategies in large organisations by providing an initial guide to future opportunities, risks, challenges, & areas of uncertainty [17].

In this field, considerable work has already been undertaken on the modelling of technology diffusion as part of these substitution events. This has included, amongst many other areas of study (see [58]), the influence of successive technology generations, and the impact of time delays on the perception of new technologies (see [9] and [18] 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 [23, 75]. This model assumes that eventually all technologies then arrive, driven by research and development efforts, at an ultimate limiting condition that is 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 [56, 59]. 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 [5, 38].

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 [70]. However, the technological 'failures' that lead to this reactive type of substitution vary greatly, and cannot just assume a single simple definition. In this regard, previous work has examined what is meant by 'technological failure', and has broadly categorised these occurrences as outlined in the work of Gooday [28]. In the analysis that follows, this study focuses on failures relating to the ever more demanding 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 [65]. This definition also follows on from the work of Sood & Tellis which applied a sub-sampling approach to analyse 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 [70].

3.2. Anomalies associated with scientific and technological crisis

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'. In this sense Edward Constant argued that a feature common to all technological revolutions was the emergence of 'technological anomalies', which could be traced to either scientific or technological crisis [39]. 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 demarcation between functional-failure anomaly and presumptive anomaly is that presumptive anomaly is deduced from science before a new paradigm is formulated and that scientific deduction is the sole reason for the sole guide to new paradigm creation. No functional failure exists; an anomaly is presumed to exist, hence presumptive anomaly" [39]

The type of crisis that emerges is dependent on which type of anomaly precedes it. Scientific crisis can occur irrespective of whether an alternative theoretical framework exists or not when a persistent, unresolved, scientific anomaly successfully refutes an established theory. In this condition the crisis is directly linked to the anomaly. However, technological anomaly and crisis are rarely so logically driven, and can arise in conditions where existing technological paradigms are still performing favourably. This is illustrated by the turbojet revolution of the 1930s and 1940s, where piston-engine developments provided remarkable performance improvements and continuing success, but were superseded by scientific predictions of a performance limit arising from propeller compressibility effects. Consequently scientific foresight was directly responsible for the radical technological changes that followed. In addition, in order for a technological anomaly to provoke a technological crisis, a convincing alternative paradigm must exist, so that the relative functional failure of the conventional system is observable. As such, the alternative technological paradigm instigates the crisis, whilst the technological anomaly may only be seen as speculation or as a limiting condition to the normal technology [39].

3.3. *Modes of substitution*

Building on the works of Constant, Schilling, and Sood, a conceptual framework for analysing 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 [5]. In this, four substitution regimes are proposed considering low and high scenarios for both new technology emergence challenges and old technology extension opportunities, and are demonstrated in the context of developments in semiconductor lithography equipment. These regimes are characterised as 1) *Creative Destruction* (low extension opportunity and low emergence challenge), 2) *Robust Coexistence* (high extension opportunity and low emergence challenge), 3) *Resilience Illusion* (low extension opportunity and high emergence challenge), and 4) *Robust Resilience* (high extension opportunity and high emergence challenge). Based on the definitions of functional failure and presumptive anomaly described in sections 3.1 and 3.2, reactive technology substitutions correspond to quadrants 1 and 3 in Adner's substitution framework (i.e. substitutions based on low extension opportunities for existing technologies), whilst presumptive technology substitutions correspond to quadrants 2 and 4 (i.e. substitutions where there still appears to be high extension opportunities for existing technologies). Further details and examples of these technological substitution regimes are provided in [5] along with a review of the corresponding technology adoption S-curves.

The current study only considers the *extension opportunity* dimension in its classification of substitution modes in order 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 four 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 (i.e. reactive or presumptive) from analysis of historical scientific and technological data. Whilst the higher level modes considered here are characterised by the low and high *extension opportunity* scenarios respectively at the tail end of the existing technology's S-curve, variability in the *emergence challenge* dimension is assumed to slow the development of the new technology at the start of the subsequent S-curve. As such, this varies the initial curvature of the new technology's S-curve, rather than shifting in time the point of first emergence (which for this analysis is effectively treated as a static point). 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. This is illustrated in Fig. 1.

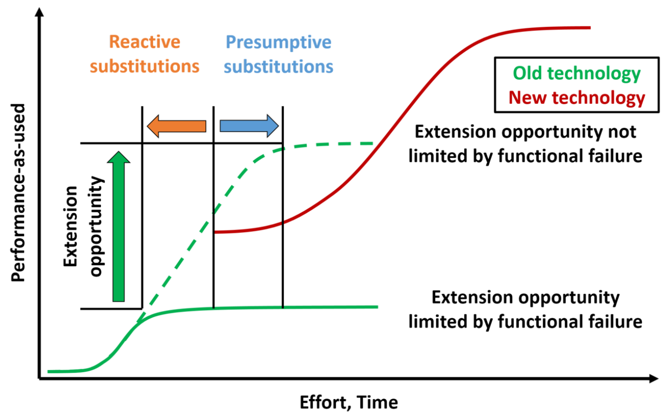


Figure 1: Illustration of reactive and presumptive substitution modes, based on Adner's framework

Table 1 uses Adner's framework, alongside the definitions provided in sections 3.1 and 3.2, and 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 marginalised technologies that were never mass commercialised (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 the analysis that follows classifies individual technologies based on training technologies that are known to have been successfully commercialised, and as such it is not believed their inclusion would influence the results presented here, although non-starters would need to be included for predicting the commercial success or failure of emerging technologies in the first instance [70].

Based on Constant's hypothesis regarding scientific and technological anomalies and their influence on the mode of technological substitution, 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. Constant's conceptual model theorises that presumptive technological anomalies emerge from scientific insights before a functional failure has occurred. Consequently, this study theorises that in order to identify cases of technological substitution arising from presumptive anomaly a classification scheme would need to be able to identify if a functional failure already exists, and if new scientific discoveries have preceded such a failure. As a result, the classification scheme needs to consider:

1. a population's perception of the current rate of scientific development in observed domains [39]
2. a population's perception of the current rate of technological development in observed domains [39]

3.4. 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 these have largely been developed for the purposes of identifying and targeting gaps in existing knowledge, as well as for determining the effectiveness of funding in specific fields of research, they also provide a systematic approach to compare development trends across a broad range of scientific domains. When attempting to measure science 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 [51]. 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 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 [9]. Adoption is equally dependent on perceptions of current scientific and technological rates of progress (shaped by social and political pressures, as well as technical), 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 [9]. 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.

3.5. 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 [67, 64]. More recent publications have expanded these early concepts and have demonstrated on numerous occasions how patterns in historic patent data can be used to build predictions of future development trends, including the use of 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) against commonly recognised milestones and features in observed technology evolution patterns. Chief amongst these is comparison to Arthur Little's Technology Life Cycle (TLC) [46]. 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 a simple count 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 [78, 35]. Building on this, Gao demonstrated the use of a trained nearest neighbour classifier, based on thirteen extracted patent data dimensions, to assess a technology's life cycle progress [24]. 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 [43]. In parallel to these extensions to sets of indicators and pattern recognition techniques, the use of text-mining approaches to improve the 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 [62]. Similarly, patent content clustering techniques for technology forecasting purposes have also been explored by the works of Trappey and Daim [74, 17]. 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

Examples of reactive substitutions	Examples of presumptive substitutions
Plug-compatible market (PCM) disk drives [13]	Transition from piston engine to jet engine [39, 12, 69]
Transition to fibre optic cables from Cu/Al wires for data transfer [70]	Transition to optical undersea cables from coaxial cables [12]
Transition to Low Pressure Sodium lights from Tungsten Filament Lamps [12]	Transition to water turbines from steam engines [39, 69]
Transition to Compact Fluorescent Lamps from Tungsten Filament Lamps [12]	Transition to early gas engines from steam engines [39]
Transition to White LED lighting from Low Pressure Sodium and Compact Fluorescent Lamps [12]	Transition to steam turbines from water turbines [39, 69]
Transition to hypersonic aircraft from supersonic [12]	Transition to catalytic petroleum cracking from thermal cracking [39]
Transition to coaxial undersea cables from single cable [12]	Transition to the transistor from the vacuum tube [22]
Transition to T-carrier system from modem internet access [12]	Transition to atomic energy from fossil fuels [39, 30]
Transition to Synchronous Optical Networking (SONET) system from T-carrier internet access [12]	Renewable energy sources: transition to solar PV/thermal, wind, geothermal, hydropower, and marine energy from fossil fuels [30, 69]
Transition to ink jet and laser printers from dot matrix printers [70]	Transition to modern battery and plug-in hybrid electric vehicles from petrol and diesel vehicles [82]

Table 1: Identified examples of reactive and presumptive technological substitutions

of data acquisition (as a proxy indicator for technology diffusion when historical data is not present). 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 then be considered is the market adoption of the technology once it has been commercialised. Here Daim's work subsequently coupled the patent-based and academic literature data-mining techniques employed with the use of system dynamics modelling as a means of exploring causal relationships and non-linear behaviours 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.

4. Methodology

There is a range of possible techniques that can be used for gauging the progress of technological development. In this study, bibliometric data has been used 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 in to account in method selection and development are discussed below.

4.1. Bibliometric data

Patent data has been sourced from the Questel-Orbit patent search platform in this analysis. More specifically, the full FamPat database was queried in this study, which groups related invention-based patents filed in multiple international jurisdictions into families of patents. Some of the core functionalities behind this search engine are outlined in [42]. 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, or classification categories are used in this search engine to build relevant queries for a given technology (this process is discussed in more detail in section 5.2). The provided search terms are then matched in 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.

4.2. Statistical comparisons of time series

This study considers 23 technologies, defined in Table 3, where literature evidence has been identified to classify the particular mode of technology substitution observed. The evidence and process used in this categorisation is outlined in detail in [49]. 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 in order to investigate whether literature based technology substitution groupings can be determined using a classification system built on the assumptions given in section 3.3. In particular, comparisons of technology time series can be subject to one or more areas of dissimilarity: time series may be based on different number of observations (e.g. covering different time spans), be out of phase with each other, may be subject to long-term and shorter term cyclic trends, be at different stages through the Technology Life Cycle (or be fluctuating between different stages) [46], 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 [45]. 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 A for further details of the considerations addressed in this study's methodology beyond those discussed directly in section 4.3.

4.3. Method selection

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

4.3.1. 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 neighbour pattern recognition approach has been employed based on the work of Gao [24] to locate the points where shifts between cycle stages occur. However, for the specific technologies considered in this paper, literature evidence has been identified for the transitions between stages, and so the nearest neighbour methodology is not discussed further here.

4.3.2. Identification of significant patent indicator groups

In order to identify those bibliometric indicator groupings 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 recognised as the classification benchmark to beat (see Appendix A), and so this study does not look 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 3.3 is not intrinsically valid. For this reason an unsupervised learning approach has been adopted here to enable human biases to be eliminated in determining whether a classification system based on presumptive technological substitution is valid or not, before subsequently defining a classification rule system. In doing so this additionally means that labelling of predicted clusters can be carried out even if labels are only available for a small number of observed samples representative of the desired classes, or potentially even if none of the observed samples are absolutely defined. This is of particular use if this 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 a time-consuming process, and in some cases the necessary evidence may not be publicly available (e.g. if dealing with commercially sensitive performance data). As such, clustering can provide an indication of the likely substitution mode of a given technology without the need for prior training on technologies that belong to any given class. Under such circumstances this approach could be applied without the need for collecting performance data, providing that the groupings produced by the analysis are broadly identifiable from inspection as being associated with the suspected modes of substitution (this is of course made easier if a handful of examples are known, but means that this 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 the literature, and keeping the number of clusters fixed allows for 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 not likely 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, which 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.

4.3.3. 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 those bibliometric indicator subsets that have been identified as producing 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 [8].

4.3.4. 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 [50]. 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 technology classification model developed in this study (see notes on Functional Data Analysis in Appendix A for further details).

5. Building a technology classification model from Technology Life Cycle features

5.1. Patent indicator definitions

The work of Gao et al. identifies a range of studies that have been conducted previously based on the principle of using either a single or multiple bibliometric indicators as a means of investigating technological development and performance [24]. Their review of these methods concluded that multiple patent indicators are required to avoid generating potentially unreliable results if just using a single indicator extracted from patent data. As such, the nearest neighbour classification process developed in Gao's study proposes the use of thirteen separate patent indicators. This current study has accordingly reproduced these metrics where possible, resulting in a total of ten patent indicators (i.e. producing time series for each technology with ten dimensions), as three of the previous list of indicators were specific to the Derwent Innovation Index [1] which was not used in this study due to the limited ability to bulk export the necessary results from this database. Table 2 summarises the bibliometric indicators extracted for each technology within this analysis.

With the main exception of the use of the Questel-Orbit FamPat database instead of the Derwent Innovation Index, the indicator definitions and assumptions used in this study are otherwise consistent with those outlined in sections 2.1.1 to 2.1.5 of [24]. The only other notable difference to be recorded is that the Questel-Orbit patent records are not automatically given a designation as being a corporate, non-corporate, or individual patent assignee. As such, the counts of corporate and non-corporate indicators (which would otherwise be based on this assignee designation) are determined instead based on the 'Family Normalized Assignee Name' field available in the patent records, as records with entries in this field correspond to corporate designations.

5.2. Search strategy and terms for identifying relevant patent profiles

Previous bibliometric studies have explored the many different ways in which patent records can be correctly identified for a given field or topic [76, 66, 7, 63, 48, 19, 80, 37]. Whilst filtering of search results based on technology classification categories is generally preferred where possible to ensure a more rigorous search strategy [7], 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 [37]. As such, the search queries used in this analysis are based primarily on filtering by International Patent Classification (IPC) or Cooperative Patent Classification (CPC) labels. Where possible the IPC categories applied have been reused from previous studies in order to replicate existing search queries so as to extract comparative datasets, or have been 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 the appearance of closely adjoining instances of the search terms (or of their common synonyms) to be matched. The use of 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 used for the technologies to be considered are presented in Table 3.

5.3. Patent indicator data extraction process

Using the technology classification categories, and where applicable the keywords specified in Table 3, 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. Additionally, each exported record included the key patent information along with full details of both cited patent and non-patent literature references made in the current record. As some searches could generate very large numbers of records (i.e. hundreds of thousands), the use of batch processing enabled large quantities of records to be handled in manageable formats, but required that the batches were subsequently imported into a tool capable of processing the volumes of data considered. For this purpose, MATLAB was used, and a script (provided in Appendix B) 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.

5.4. Patent indicator data cleaning process

Whilst the consistency of the Questel-Orbit patent data is of a high standard, several steps are still required to be able to extract patent indicator metrics from this data. This is done 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 that may be present in the extracted data (such as the omission of application or priority dates from the relevant columns when these dates are available elsewhere). In doing so, this allows a more accurate chronology of patent events to be established. This process is not discussed in detail here, but is available in Appendix C for more information.

5.5. Technology Life Cycle stage matching process

With bibliometric profiles extracted for each of the technologies considered in this study, the first stage of analysis consists of identifying the transition points between different stages of the Technology Life Cycle in order to establish time series segments for use in subsequent 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 [34]. Full details of the transition points used in this study are provided in Table 4.

Of the 23 technologies listed in Table 4, 20 were found to have patent data pertaining to the emergence stage (i.e. excluding incandescent lights, landline telephones, and wireless data transfer). As such only those technologies with patent data

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 citation	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

Table 2: Bibliometric indicators used in this study (based on the work of Gao et al. [Gao 2013])

available during the emergence stage are considered in the analysis that follows.

For subsequent expansion of this analysis to additional technologies where evidence is not immediately apparent for the definition of these segments, a nearest neighbour pattern matching process was also developed as outlined in section 4.3.1 based on the work of Gao et al. [24]. This methodology is not discussed in further detail in this paper.

5.6. 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 Technology Life Cycle stage are then identified using the procedure outlined in Fig. 2.

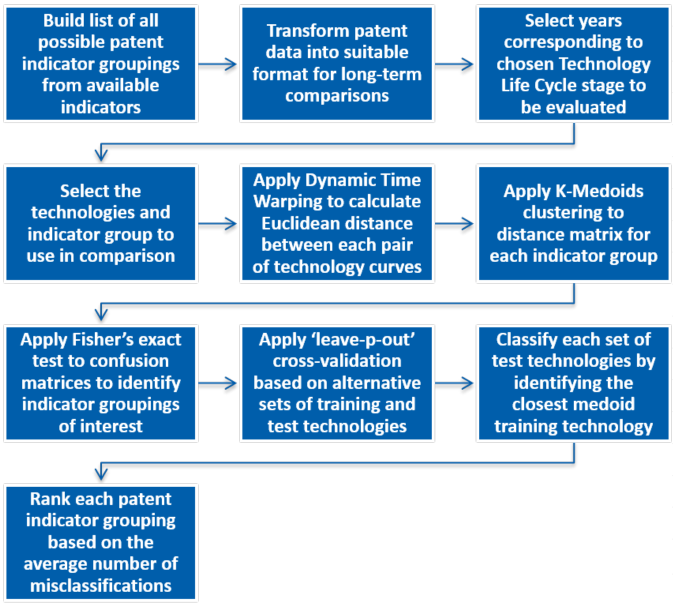


Figure 2: Overview of the process used to identify and rank significant patent indicator groups

As discussed in sections 4.3.2 and 4.3.3 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 calculated between time series. This is again implemented as a MATLAB script based on the DTW and K-Medoid functions made available by MathsWorks [52, 3], which is provided in Appendix B. The first step of this process involves generating a list of all the unique subsets that can be created from the ten patent indicator metrics considered in this study. This produces 1,023 (i.e. $2^{10} - 1$) possible combinations of the ten patent indicators to be tested, as illustrated by Fig. 3.

Next, the raw patent data time series are transformed by using an inverse hyperbolic sine function and normalised to convert the data into a suitable format for long-term comparisons (see notes on preprocessing in Appendix A). Once in this format, the data points are filtered based on the current Technology Life Cycle stage being considered, as illustrated by Fig. 4, ensuring comparable curve features are considered.

After the datasets have been transformed and filtered based on the current Technology Life Cycle stage, Dynamic Time Warping is then 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. 5, illustrating the successive layers of filtering that are applied for each technology pairing and each patent indicator grouping considered. The output from this process is an i x j x 1023 distance matrix, where i and j specify the current technology pairing being considered, and the value quoted is the measured distance between multi-dimensional time series based on the current patent indicator subset being used. In parallel to this 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 specific patent indicator subset is used. By comparing the predicted technology groupings to those expected from the earlier literature classifications (see section 3.3), a confusion matrix is created for each patent indicator

Case study	Class	Orbit patent search keywords	IPC or CPC categories	No. of patent families
Compact Fluorescent Lamp	R	(compact+ or CFL+ or (energ+ s (sav+ or low+))) AND fluores+	CPC: Y02B-020/16+ OR Y02B-020/18+ OR Y02B-020/19+	1,169 (21/07/2017)
Electric vehicles	P	–	CPC: Y02T-010/62+ OR Y02T-010/64+ OR Y02T-010/70+ OR Y02T-010/72+ OR Y02T-090/1+	100,870 (24/07/2017)
Fiber optics (data transfer)	R	((fiber+ or fibre+) 3d optic+)	IPC: G02B OR H04B OR C03B OR C03C OR D01C OR D04H OR D06L OR G02F OR G06E OR G06K OR G11B OR G11C OR H02G OR H03K OR H04J OR H04N OR G01P	176,299 (20/07/2017)
Geothermal electricity	P	–	CPC: Y02E-010/1+	5,272 (24/07/2017)
Halogen lights	R	–	CPC: Y02B-020/12+	645 (24/07/2017)
Hydro electricity	P	–	CPC: Y02E-010/2+	46,125 (24/07/2017)
Impact/Dot-matrix printers	R	((impact+ or (dot+ or matri+) or (daisy 1w wheel+)) 3d print+)	IPC: G03G OR B41J OR G06F OR G06K OR H04N OR G06T OR G02B OR H04L OR G01R OR G03C OR B41M OR G03B OR B65H	24,993 (24/07/2017)
Incandescent lights	P	Incandescen+ or filament+	IPC: F21H OR F21L OR F21S OR F21V OR F21W OR F21Y	17,597 (03/08/2017)
Ink jet printer	R	(ink+ 3d jet+ 3d print+)	IPC: B41J-002/01 OR G03G OR B41J OR G06F OR G06K OR H04N OR G06T OR G02B OR H04L OR G01R OR G03C OR B41M OR G03B OR B65H	46,135 (24/07/2017)
Internet	R	(internet+ 3d protocol+ 3d suite+) OR (computer+ 1w network+)	IPC: G06F OR H04L OR G06N OR H04K OR G09F	42,861 (24/07/2017)
Landline telephones	P	(((land_line+ or main_line+ or home or fixed_line+ or wire_line+) 3d (+phone)) OR (speaking telegraph+) OR (telephon+)) NOT (mobil+ or (cell+ 3d (+phon+ or communi+)) or smart_phon+ or port+)	IPC: H04B OR H01Q OR H01P OR H04J OR G01R OR H04Q OR H01H OR H04M OR H04R OR G10L	139,895 (03/08/2017)
Laser printer	R	(laser+ 3d print+)	IPC: G03G OR B41J OR G06F OR G06K OR H04N OR G06T OR G02B OR H04L OR G01R OR G03C OR B41M OR G03B OR B65H	17,827 (24/07/2017)
LED lights	R	–	CPC: Y02B-020/3+	8,596 (24/07/2017)
Linear Fluorescent Tube lights	R	((fluores+ 3d (lamp+ or light+ or tube+))) NOT (compact or (energ+ 3d sav+))	IPC: F21K OR F21L OR F21S OR F21V OR F21W OR F21Y	25,126 (24/07/2017)
Nuclear energy	P	–	CPC: Y02E-030+	60,017 (24/07/2017)
Solar PV	P	–	CPC: Y02E-010/5+ OR Y02E-010/6+	112,068 (24/07/2017)
Solar thermal electricity	P	–	CPC: Y02E-010/4+ OR Y02E-010/6+	91,553 (24/07/2017)
TFT-LCD	R	((((thin film+) 1w transistor+) or TFT+) AND (((liquid crystal+) 1w display+) or LCD)) or TFT_LCD	IPC: G02F-001/13	5,181 (24/07/2017)
Thermal printers	R	(thermal+ 2d print+)	IPC: G03G OR B41J OR G06F OR G06K OR H04N OR G06T OR G02B OR H04L OR G01R OR G03C OR B41M OR G03B OR B65H	23,388 (24/07/2017)
Tide-wave-ocean electricity	P	–	CPC: Y02E-010/28+ OR Y02E-010/3+	19,224 (24/07/2017)
Turbojet	P	((Gas w turbin+) or (jet+ w engine+) or turbo_fan+ or turbo_prop+ or turbo_jet+ or turbo_shaft+ or prop_fan+ or ((open w rotor+) 3d (engine+ or technolog+ or counter_rotat+)))	IPC: B60K OR B60L OR B60P OR B60V OR B61B OR B61C OR B62D OR B63B OR B63H OR B64C OR B64D OR B64F OR B64G OR F01D OR F02B OR F02C OR F02K	71,024 (24/07/2017)
Wind electricity	P	–	CPC: Y02E-010/7+	67,035 (24/07/2017)
Wireless data transfer	R	(Wireless 3d data 3d trans+)	IPC: H03K OR H04H OR H04W OR G06K OR G06T	17,188 (24/07/2017)

Table 3: Technologies considered in study, classification, and patent data search terms

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	–	[34, 79]
Electric vehicles	1997	2005	–	[61, 81]
Fiber optics (data transfer)	1970	1990	–	[11, 36]
Geothermal electricity	1958	–	–	[27]
Halogen lights	1959	–	–	[2, 55, 21]
Hydro electricity	1956	1975	–	[15]
Impact/Dot-matrix printers	1970	1984	1991	[53, 73, 6, 14, 4]
Incandescent lights	1882	1916	2008	[12, 26, 21]
Ink jet printer	1988	1996	2003	[14]
Internet	1982	2000	–	[44, 83, 77]
Landline telephones	1878	1945	2009	[57, 40]
Laser printer	1979	1993	–	[29, 73]
LED lights	2001	–	–	[34]
Linear Fluorescent Tube lights	1937	1990	2012	[2, 72, 41]
Nuclear electricity	1963	1981	–	[34]
Solar PV	1990	–	–	[34]
Solar thermal electricity	1968	–	–	[20, 32]
TFT-LCD	1990	2007	–	[24]
Thermal printers	1972	1985	2002	[54, 31, 73, 68, 10]
Tide-wave-ocean electricity	1966	–	–	[71, 16]
Turbojet	1939	1958	–	[25]
Wind electricity	1982	–	–	[34]
Wireless data transfer	1982	2002	–	[34]

Table 4: Technology Life Cycle transition points based on literature evidence

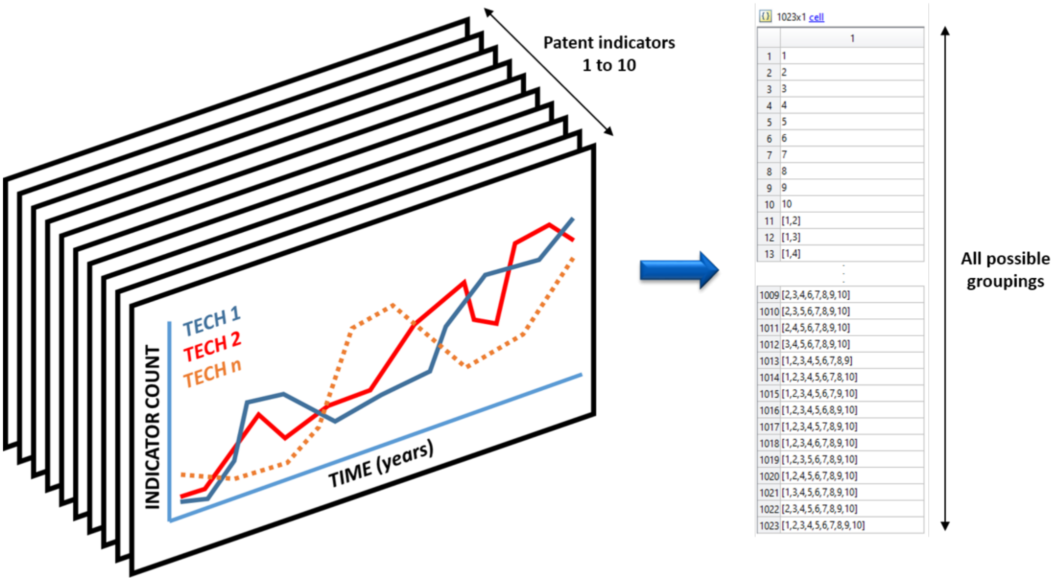


Figure 3: Generating list of all possible patent indicator groupings from time series dimensions considered

subset that shows the alignment between predicted and target groupings as shown in Fig. 6. 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.

5.7. Ranking of grouped patent indicator dimensions

As discussed in section 4.3.3 and Appendix A leave-p-out cross-validation techniques provide a means to rank those bibliometric indicator subsets that have been identified as producing a significant match to the expected technology groupings. The first stage of this process consists of generating lists of all possible training technology combinations and corresponding test technology combinations based on leaving one technology out at a time. 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. 5, except that this time distance measures are only calculated between pairs of training technologies, and that this process is repeated for every possi-

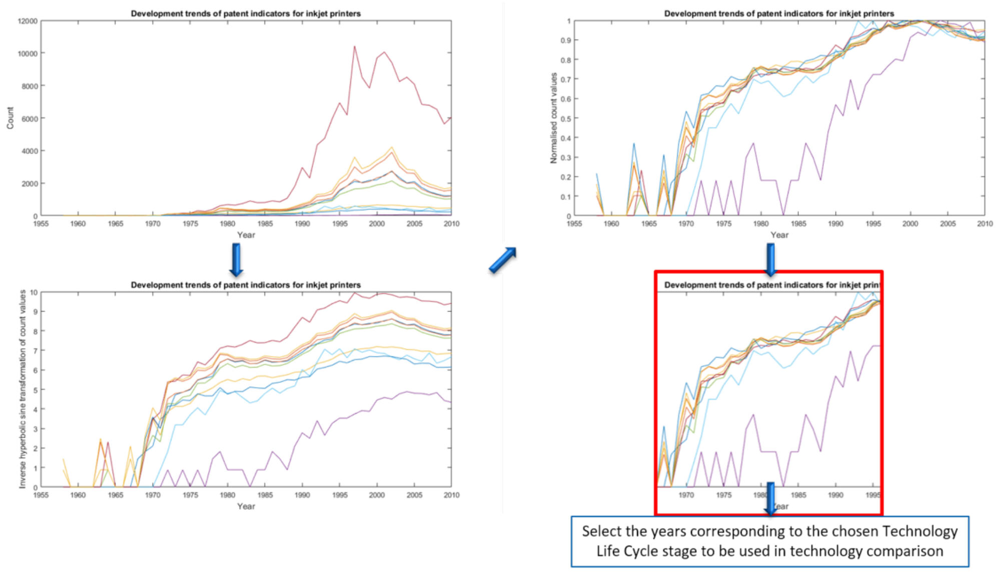


Figure 4: Transforming extracted patent data time series into a suitable format for long-term comparisons

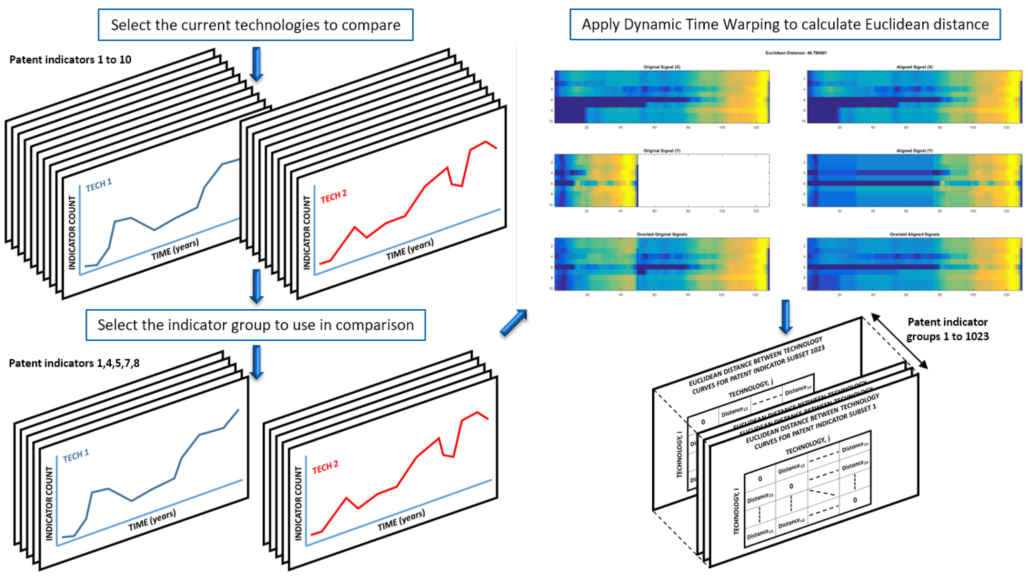


Figure 5: Calculating the distance between each pair of technology time series for each indicator grouping

ble combination of training technologies that are available. 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 pairing being considered, and n represents the number of training combinations that can be used. This is illustrated in Fig. 7.

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. At this point the test technologies can now be evaluated individually against the two medoid curves identified in each training condition, in order 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. From this the number of test technologies misclassified based on the current training condition can be determined. This in turn is then used to calculate the average number of test technologies misclassified for each patent indicator grouping across all of the training conditions considered. Finally, the results are sorted in terms of the minimum average number of misclassifications in order to rank the robustness of each patent indicator grouping. This procedure is illustrated in Fig. 8.

5.8. *Functional model building process*

The ranking of different bibliometric indicator subsets provides a means to identify the time series dimensions that, when

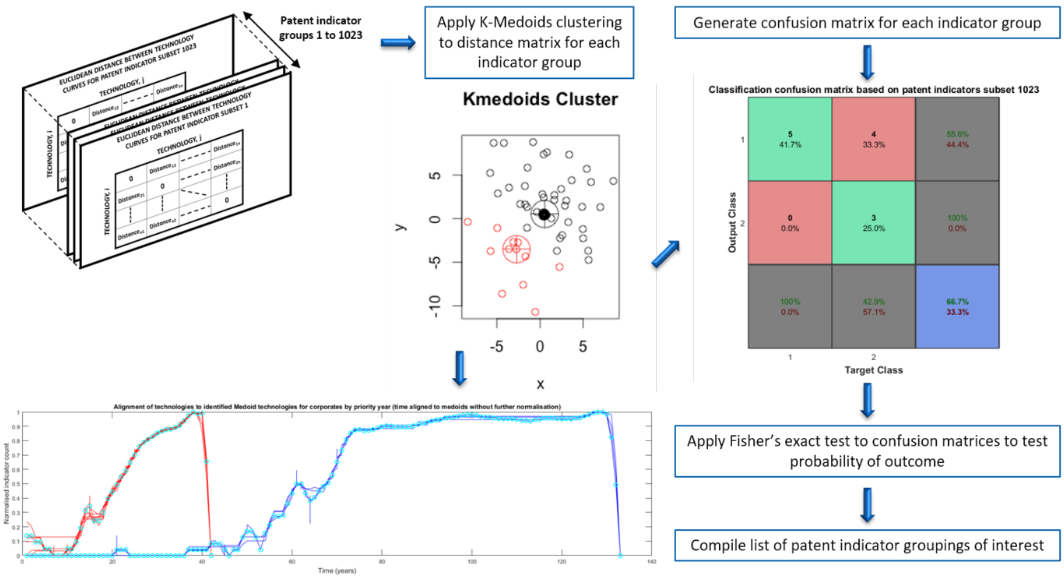


Figure 6: Identifying patent indicator groups of interest

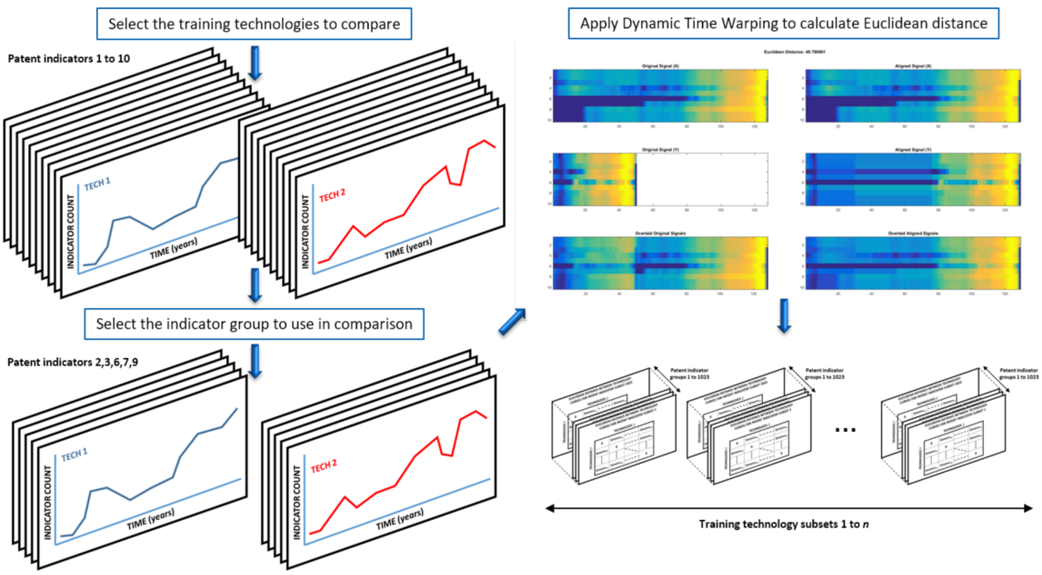


Figure 7: Calculating the distance between each pair of training technologies for each indicator grouping

combined, are most likely to provide robust out-of-sample predictions of the observed technological modes of substitution. The preceding cross-validation exercise therefore provides a basis for an informed selection of the time series components to use in model building. As a result, a technology classification model is now developed using functional data analysis (see section 4.3.4 and Appendix A) that is based on patent indicators 4 and 6 (i.e. the number of non-corporates and the number of cited references by priority year). Besides being present in all of the highest scoring sets of top ranked predictors, these 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 shows 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 centres) is associated with the amount of lab work required to commercialise 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 technology being developed. By contrast, those technologies with reduced non-corporates by priority year 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

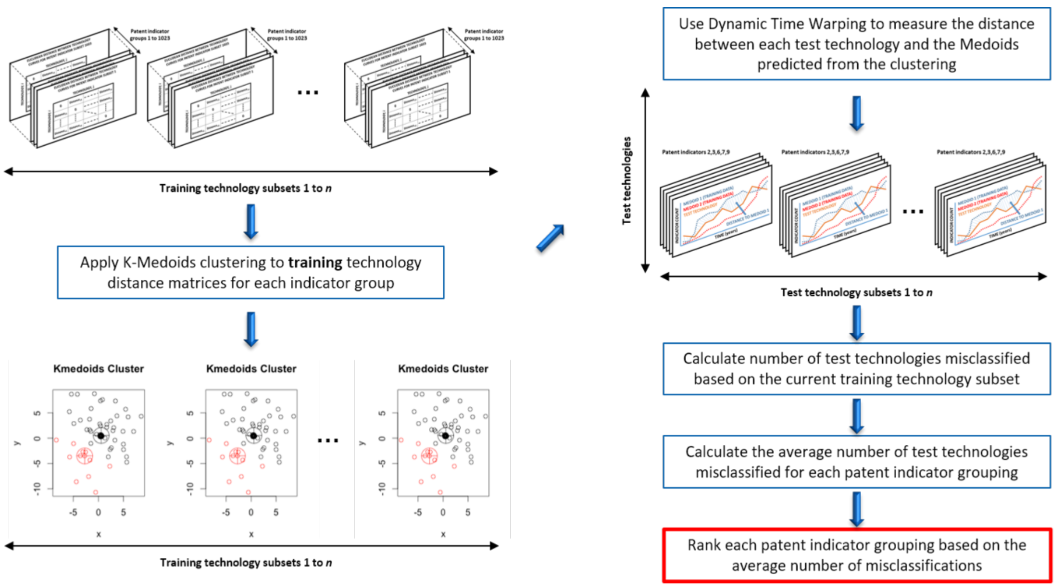


Figure 8: Ranking of grouped patent indicator dimensions

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 patent indicators (i.e. perfect orthogonality can not necessarily be assumed between these metrics), and so at this point the choice has been taken 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.

Following on from the initial introduction to functional data analysis provided in Appendix A, and more detailed methods presented in [60], the method outlined in Fig. 9 has been implemented in MATLAB for building a functional linear regression model for the purposes of technology classification (the MATLAB script is available in Appendix B for further details).

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 in order to combine all of the different technology profiles being evaluated into two regression terms: one representing the number of non-corporates by priority year, and a second term representing the number of cited references by priority year. These terms, when multiplied by their respective regression coefficients (which are 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 will 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 one Technology Life Cycle stage spans 20 years in one time series, and spans 50 years in another, both time series will have

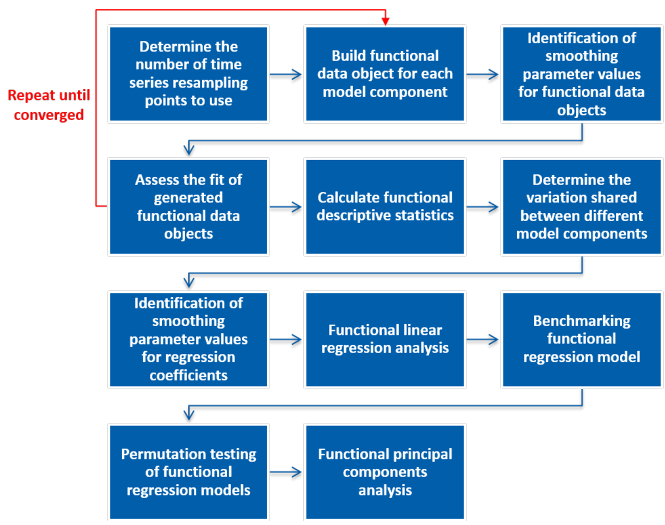


Figure 9: Functional model building process based on methods outlined in [60]

50 observations, which 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 at the same time for the regression coefficients (β_i) to be estimated by the functional linear regression analysis (see Eq. 1 and Eq. 3 in Appendix A, as well as sections 3.4.1, 3.4.2, 9.4.1 and 9.4.2 of (Ramsay 2009)), as illustrated in Fig. 10.

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 outlined in [60] a 'functional parameter object' that allows smoothness to be imposed on estimated functional

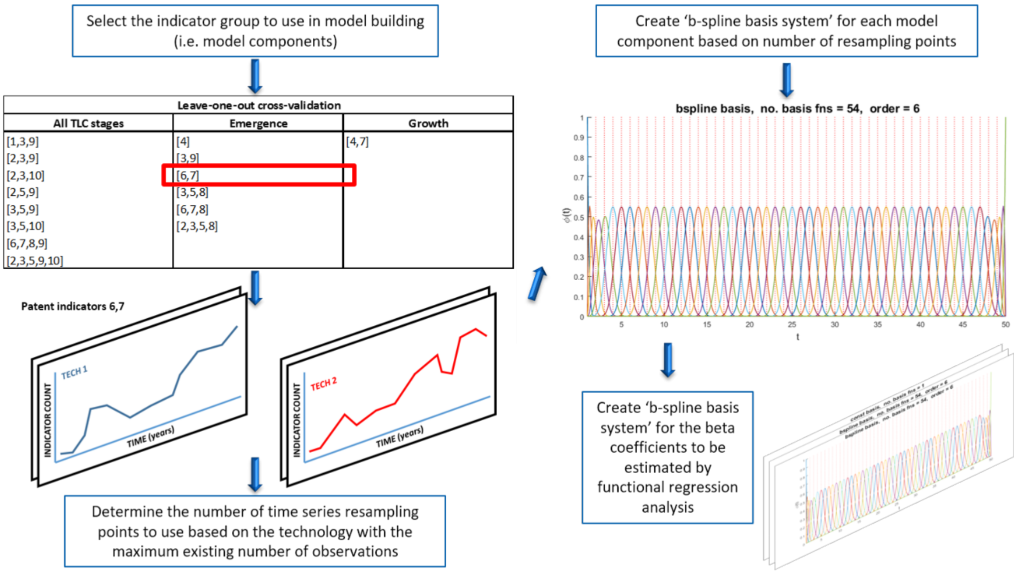


Figure 10: Building functional models of selected patent indicator groupings

parameters is now created (see section 5.2.4 of [60]). 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, along with an initial value of the smoothing parameter (λ). The degrees of freedom and generalised cross-validation criterion coefficient (see section 5.3 of [60]) 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 that will be used 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 the dynamics present in the current datasets, whilst also being more likely to be adaptable to future out-of-sample technologies. An example of a smoothed functional data object generated for the number of corporations associated with different technologies in a given priority year is illustrated in Fig. 11.

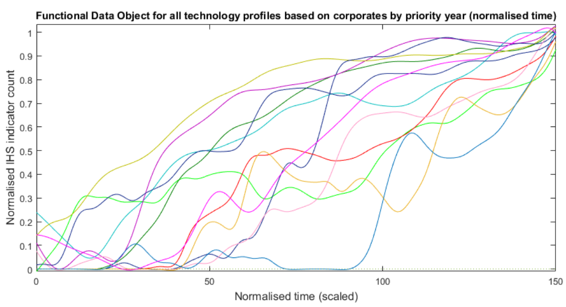


Figure 11: Functional Data Object for all technology profiles based on corporates by priority year

Having created a functional data object representation of each model component from the selected bibliometric subset, the MATLAB script then 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 modelled values across the different technology curves included, but also across the time span of the Technology Life Cycle stage considered (see section 5.5 of [60]). 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 [60]). The functional mean and standard deviation of the data objects for the number of non-corporates and the number of cited references by priority year are shown in Fig. 12 and Fig. 13 respectively, and show that for both model components variability from the mean generally increases as time progresses (as would be expected for an increasingly divergent spread of technology trajectories). In addition the mean functional data object values show that there tends to be a notable early surge followed by a dip in non-corporates by priority year during the emergence phase before a technology achieves mainstream adoption. This corresponds well to the hype cycle associated with new technologies during early development when significant levels of R&D are first launched in a race to achieve commercialisation, which can often prove premature or short-lived. By contrast, the mean cited references by priority year measure shows that a steadily accelerating growth is observed during the emergence phase, without significant undulation, potentially implying that scientific development efforts are less phased by disturbances as they begin to accumulate.

5.8.1. *Identification of smoothing parameter values for regression coefficients*

With the functional data objects for each model component now ready, a cell array containing each model component along

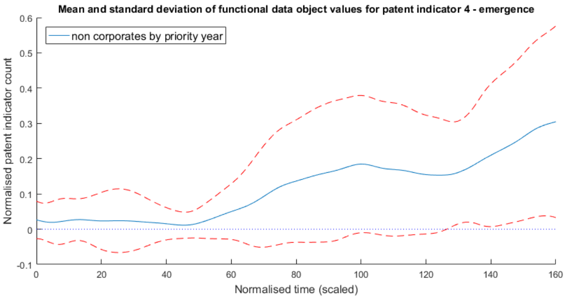


Figure 12: Mean and standard deviation of functional data object created for non-corporates by priority year

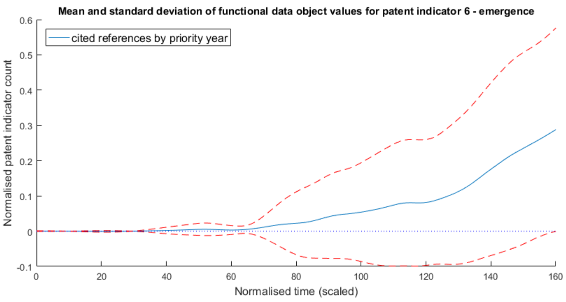


Figure 13: Mean and standard deviation of functional data object created for cited references by priority year

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 the final regression analysis can be run, a smoothing parameter for the regression coefficient basis system has to be selected. This is separate from the earlier smoothing parameter selected for smoothing the technology profiles; this second smoothing parameter only addresses the roughness of the regression coefficients. This is again necessary to try to prevent over-fitting, and ensure that the model converged on by the subsequent functional linear regression analysis has the best chance of performing well out-of-sample when extended to future datasets. In this instance, the selection of smoothing parameter is achieved by calculating leave-one-out cross-validation scores (i.e. error sum of squares values) for functional responses using a range of different smoothing parameter values, as per section 9.4.3 and 10.6.2 of [60]. The functional parameter object used in the regression coefficient basis system is then redefined using this more optimised smoothing parameter value.

6. 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 the corresponding variance, used to define the 95% confidence bounds (see sections 9.4.3 and 9.4.4 of [60] respectively). Fig. 14 to Fig. 16 show the resulting β_i coefficients and confidence bounds for the number of non-corporates and the number of cited references by priority year during the emergence phase of development 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). This 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 such, from a preliminary inspection, this classification model looks to provide a good degree of accuracy, but further investigation is required to ensure the model is not over-fitted, and that the result is not simply a naturally occurring phenomenon.

From the confidence bounds on these plots it can be seen that for both the number of non-corporates and the number of cited references by priority year indicator counts the variance across technology profiles is highest at the start of the emergence phase: this is often when the least amount of data is available for comparing each technology, and also when development activity is most haphazard and sporadic, so this is not entirely surprising as this represents the point of greatest uncertainty. However, Fig. 15 and Fig. 16 also illustrate how the relative importance of the chosen science (Fig. 16) and technology (Fig. 15) patent indicators in determining the predicted mode of substitution varies with time during the emergence phase (based on the datasets used), although no causal explanation as to why they have this relative weighting is directly provided by these functions. More specifically, deviations away from zero in these coefficient functions equate to an increased positive or negative weighting for the associated patent indicator count at that moment in time, within the determination of the predicted mode of substitution. As such it can be seen from Fig. 15 that any patent indicator counts at $t = 0$ for the number of non-corporates by priority year (assuming these are present) will have a more significant influence on the predicted classification than at any other point in the emergence phase. Equally, Fig. 15 would suggest that the impact of non-corporates activity next peaks around 40% of the way through the emergence phase (potentially corresponding to the hype effect suggested by Fig. 12), and again at the end of the emergence phase. For the number of cited references by priority year, this regression model suggests that the times of greatest impact on the mode of substitution are at the very beginning and at the very end of the emergence stage. Whilst these coefficient plots gives some indication of the relative weighting applied to patent indicator counts as time progresses, the cumulative nature of the inner products used in functional linear regression means it is not possible to visually infer from these plots alone which mode the technology under evaluation is currently converging towards. For this it is also necessary to include the corresponding patent indicator count values that these coefficient terms are multiplied by for the specific technology being assessed.

Whilst the regression coefficient plots help to provide a possible interpretation of the relationship between the different model components and the predicted technology substitution classifications, 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, as well as testing the likelihood

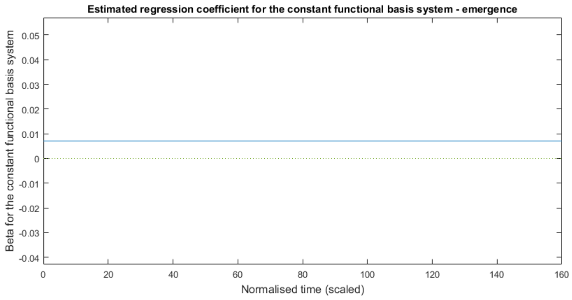


Figure 14: Estimated regression coefficient for the constant functional basis system - emergence

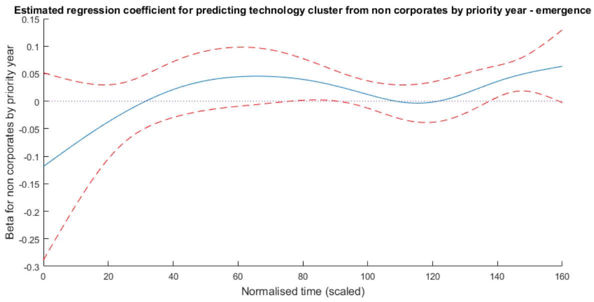


Figure 15: Estimated regression coefficient for predicting technology cluster from non-corporates by priority year - emergence

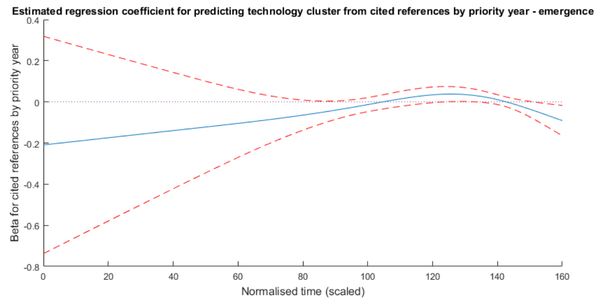


Figure 16: Estimated regression coefficient for predicting technology cluster from cited references by priority year - emergence

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 [60]) to assess the overall fit of the high-dimensional functional linear regression model, and are summarised in Table 5.

The R-squared and adjusted R-squared values shown in Table 5 would suggest that a reasonable classification fit has been achieved with this model across the 20 technology profiles considered during the emergence phase. Specifically, this suggests a good level of accuracy based on the classification residuals, whilst 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.

However, to ensure that this is the most appropriate fit to the data presented, the high-dimensional model initially developed was subsequently benchmarked against a low-dimensional model (i.e. when the beta basis system for each regression coefficient is made up of a small number of B-splines), as well as a constant and a monomial based model. The corresponding 'goodness-of-fit' measures for the alternative functional linear regression models are compiled in Table 6.

Whilst the R-squared and adjusted R-squared measures observed in Table 6 would suggest that the low-dimensional model provides a better fit, the associated F-ratio score and corresponding p-value suggests a lower significance than those 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, but this is not surprising considering the more limited nature of models built on 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, from this benchmarking analysis it would appear that the high-dimensional and monomial basis system models are the most suitable candidates, but it is possible that the overall performance of all of the models could 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 is applied to count the proportion of generated F values that are larger than the F-statistic for each model (see section 9.5 of [60]). This involves repeatedly shuffling the expected mode classification labels versus the technology profiles being read (maintaining their original order) to see if it is still 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 what the results would look like if there really was no relationship between the classification functions derived and the original data. In so doing, 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. 17.

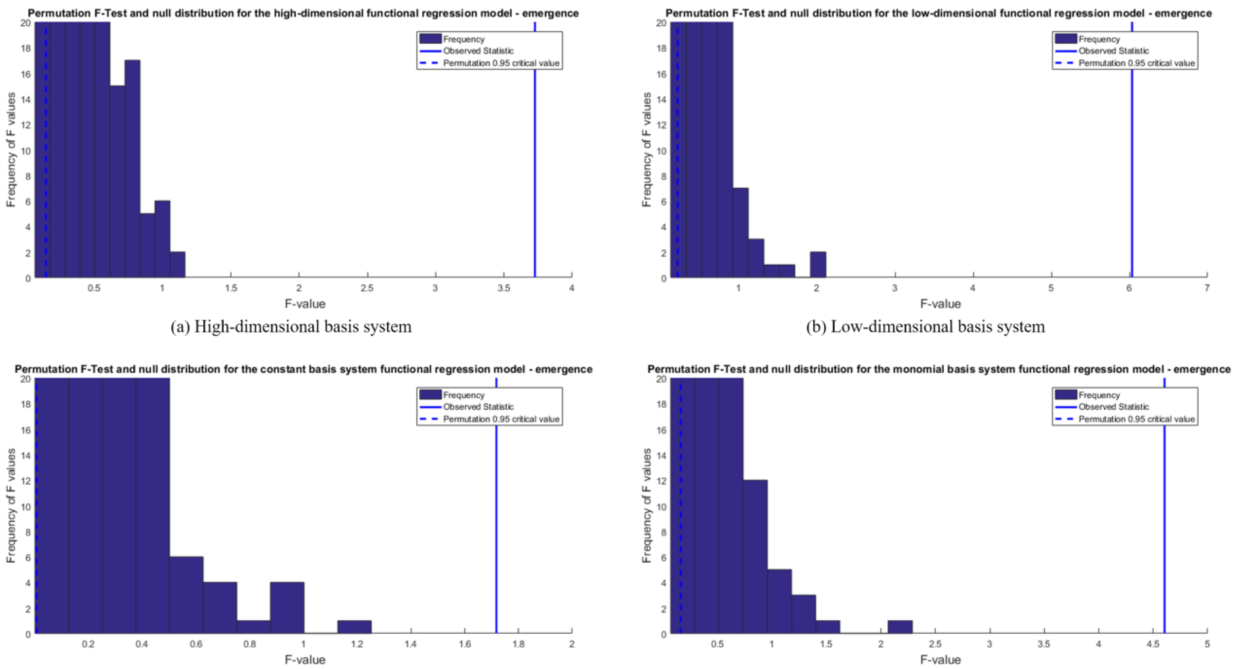
For statistical significance it is necessary that the observed test statistic is found in the tail of the distribution generated, implying that the classification responses predicted would only occur very rarely (i.e. not by chance) if the data order was rearranged. Having generated classification models based on the most robust predictors from the earlier cross-validation exercise, all four models imply that some relationship has been identified between the substitution mode predictions expected and the two patent indicator dimensions used that is specific to the data provided, although as seen in Tables 5 and 6 the fit achieved varies depending on the model used. In this last stage of the analysis the permutation testing now reveals that the high and low-dimensional models are likely to perform best out-of-sample as the observed F-statistics are furthest along each distribution's right tail in relative terms in comparison to the other distributions generated for the constant and monomial

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

Table 5: Results of high dimensional model fit

Model basis	Correct mode type	R-squared	Adjusted R-squared	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

Table 6: Benchmarking results



(a) High-dimensional basis system

(b) Low-dimensional basis system

(c) Constant basis system

(d) Monomial basis system

Figure 17: Permutation F-Test and null distributions for functional regression models - emergence

based models. This shows these two models have the lowest probability of occurring by chance, and are most likely to be generalisable to future datasets. A similar level of statistical significance is observed between the high and low-dimensional models, although as this permutation testing was only based on 1,000 permutations, the distributions could still evolve further with a greater number of permutations. However, the constant basis system model is more clearly seen here not to perform as well out-of-sample, with the observed F-statistic closest to the main body of the distribution. This, in combination with the other 'goodness-of-fit' measures shown In Tables 5 and 6, would therefore suggest that the high-dimensional functional linear regression model provides the best basis for a technology substitution classification model from those tested in this analysis.

6.1. Method limitations

Although precautions have been taken where available to ensure that the methods selected for this study address the problem posed of building a generalised technology classification model based on bibliometric data in as rigorous a fashion as possible, there are some known limitations to the methods used in this work that must be recognised. Many of the current limitations stem from the fact that in this analysis technologies have been selected based on where 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 generalisable result. This evidence-based approach also means that it is still a time-consuming process to locate the necessary literature material to be able to support classifying technology examples as arising based on one mode of substitution or another, and to then compile the relevant cleaned patent datasets for analysis. As a result only a relatively limited number of technologies have been considered in this study, which should be expanded on to increase confidence in the findings produced from this work. This also raises the risk that clustering techniques may struggle to produce consistent results based on the small number of technologies considered. Furthermore, any statistical or quantitative methods used for modelling 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 modelling, or expert elicitation will be required to shed more light on the underlying influences shaping technology substitution behaviours.

Other data-specific issues that could arise relate to the use of patent searches in this analysis 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 based on the database and exact search terms used, however overall trends once normalised 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 A, linear interpolation is used as required to ensure consistency on the number of observations whilst possibly introducing some small errors which are not felt to be significant.

7. Conclusions from statistical ranking and functional data analysis

Expanding on previous historical accounts of technological substitutions this study has examined the premise that two principal modes are often observed when considering transitions between successive commercially prevalent technologies: reactive and presumptive technological substitutions. These two modes are believed 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. In both cases, technological anomalies are believed to arise, either as a result of scientific or technological crisis, that subsequently trigger the eventual shift to the next technological paradigm. As such, this paper has considered 23 example technologies where literature evidence of performance development trends has been found in order to test the ability to correctly identify observed adoption modes using bibliometric, pattern recognition, and statistical analysis techniques. The results obtained from this analysis suggest that statistical analysis of patent indicator time series, segmented based on identified Technology Life Cycle features, provides a possible means for classification of technological substitutions. Specifically, for the datasets considered measures of the number of cited references and the 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 predictive capability. These selected patent data dimensions can be associated with perceptions of scientific and technological production respectively, consistent with the basic prerequisites listed in section 3.3 for a classification scheme that can identify presumptive technological substitutions.

Whilst these two patent dimensions occur in all of the most robust predictor subsets (i.e. in terms of out-of-sample reliability) when basing analysis on the emergence stage, this does not prove that these are the only indicators capable of predicting modes of technological substitution. As discussed in section 5.8, the possibility of orthogonality has not been ruled out with regards to the other patent indicators shown in Table 2. However, these two dimensions are in good agreement with the technological anomaly arguments put forward by Constant in sections 3.2 and 3.3, and so were felt to be reasonable for forming the basis of the technology classification model that has been developed using functional linear regression. In particular, a regression fit made up of 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 derived technology 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, so this relationship would appear to be based on the specifics of the individual technology curves considered, and does not appear to be occurring by chance. This implies that it may be possible to predict modes of substitution from limited bibliometric data during the earliest stages of technology development, providing some evaluation of the progress through the early stages of Technology Life Cycle is made (this can be obtained using a nearest neighbour matching process, not discussed in this paper). Equally this shows that the functional data approach employed corroborates well the earlier statistical rankings produced using Dynamic Time Warping, K-Medoids clustering, and leave-one-out cross-validation of the selected patent indicators, suggesting that these two methods are compatible for this type of analysis.

It is also important to remember the potential limitations of this study that would need to be addressed for further confidence in the methodology used. Firstly, only a relatively small number of technologies have been evaluated in this study 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 minimise 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 here. 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 presented here are a better fit to the industries included in this analysis, rather than a model that can be necessarily generalised to all technologies.

However, perhaps the most important note of caution regarding this work 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, particularly in this case when considering the breadth of reasons for technological stagnations, 'failures', or presumptive leaps to occur. Equally, statistical methods are not generally well suited to predicting disruptive events and complex interactions, with other simulation techniques such as System Dynamics and Agent Based Modelling performing better in these areas. Accordingly, to identify causation effects and test the sensitivity of technological substitution patterns to variability arising from real-world sociotechnical behaviours not captured in simple bibliometric indicators (such as the influence of competition, organisational, and economic effects), the fitted regression model presented here also needs to be evaluated in a causal environment.

Similarly, in order to demonstrate practical applicability the mode of substitutions considered here 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) in order 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.

8. Acknowledgements

This work was supported by Airbus Operations Ltd, as part of the Agile Wing Integration project, and by the University of Bristol's Industrial Doctorate Centre in Systems [EPSRC grant number EP/G037353/1].

[1] Derwent innovations index version 4.0 offers expanded functionality. *Microelectronics International*, 20(3), dec 2003. doi:10.1108/mi.2003.21820cab.008. URL https://doi.org/10.1108%2Fmi.2003.21820cab.008.

[2] *Light's labour's lost author/editor = IEA*. OECD Publishing, jun 2006. doi:10.1787/9789264109520-en. URL https://doi.org/10.1787%2F9789264109520-en.

[3] Dynamic Time Warping Clustering, 2015. URL https://stats.stackexchange.com/questions/131281/dynamic-time-warping-clustering. Accessed: March 2017.

[4] Hap Acee. *Disruptive technologies: an expanded view*. PhD thesis, Massachusetts Institute of Technology, 2001.

[5] Ron Adner and Rahul Kapoor. Innovation ecosystems and the pace of substitution: Re-examining technology S-curves. *Strategic Management Journal*, 37(4):625–648, mar 2015. doi:10.1002/smj.2363. URL https://doi.org/10.1002%2Fsmj.2363.

[6] Abhinav Agrawal and Jeffrey Dwoskin. Timeline of Printers, 2003. URL http://www.jeffdwoskin.com/printers/menu.html. Accessed: August 2017.

[7] Vito Albino, Lorenzo Ardito, Rosa Maria Dangelico, and Antonio Messeni Petruzzelli. Understanding the development trends of low-carbon energy technologies: A patent analysis. *Applied Energy*, 135:836–854, dec 2014. doi:10.1016/j.apenergy.2014.08.012. URL https://doi.org/10.1016%2Fj.apenergy.2014.08.012.

[8] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4(0):40–79, 2010. doi:10.1214/09-ss054. URL https://doi.org/10.1214%2F09-ss054.

[9] Frank M. Bass. Comments on "A New Product Growth for Model Consumer Durables The Bass Model". *Management Science*, 50(12_supplement):1833–1840, dec 2004. doi:10.1287/mnsc.1040.0300. URL https://doi.org/10.1287%2Fmnsc.1040.0300.

[10] Red Bus Cartridges. The Evolution Of Printing, 2017. URL https://www.redbuscartridges.com/printing/evolution-of-printing/. Accessed: August 2017.

[11] G. Cattani. Technological pre-adaptation speciation, and emergence of new technologies: how Corning invented and developed fiber optics. *Industrial and Corporate Change*, 15(2):285–318, mar 2006. doi:10.1093/icc/dtj016. URL https://doi.org/10.1093%2Ficc%2Fdtj016.

[12] Yu Sang Chang and Seung Jin Baek. Limit to improvement: Myth or reality? *Technological Forecasting and Social Change*, 77(5):712–729, jun 2010. doi:10.1016/j.techfore.2010.02.010. URL https://doi.org/10.1016%2Fj.techfore.2010.02.010.

[13] Clayton M. Christensen and Richard S. Rosenbloom. Explaining the attacker's advantage: Technological paradigms organizational dynamics and the value network. *Research Policy*, 24(2):233–257, mar 1995. doi:10.1016/0048-7333(93)00764-k. URL https://doi.org/10.1016%2F0048-7333%2893%2900764-k.

[14] Neil Clymer and Shigeru Asaba. A new approach for understanding dominant design: The case of the ink-jet printer. *Journal of Engineering and Technology Management*, 25(3):137–156, sep 2008. doi:10.1016/j.jengtecman.2008.06.003. URL https://doi.org/10.1016%2Fj.jengtecman.2008.06.003.

[15] Michael C. Connelly and J.A. Sekhar. U. S. energy production activity and innovation. *Technological Forecasting and Social Change*, 79(1):30–46, jan 2012. doi:10.1016/j.techfore.2011.05.001. URL https://doi.org/10.1016%2Fj.techfore.2011.05.001.

[16] Teodora Diana Corsatea. Increasing synergies between institutions and technology developers: Lessons from marine energy. *Energy Policy*, 74:682–696, nov 2014. doi:10.1016/j.enpol.2014.07.006. URL https://doi.org/10.1016%2Fj.enpol.2014.07.006.

[17] Tugrul U. Daim, Guillermo Rueda, Hilary Martin, and Pisek Gerdsri. Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change*, 73(8):981–1012, oct 2006. doi:10.1016/j.techfore.2006.04.004. URL https://doi.org/10.1016%2Fj.techfore.2006.04.004.

[18] Brice Dattée and Henry Birdseye Weil. Dynamics of social factors in technological substitutions. *Technological Forecasting and Social Change*, 74(5):579–607, jun 2007. doi:10.1016/j.techfore.2007.03.003. URL https://doi.org/10.1016%2Fj.techfore.2007.03.003.

[19] Bensi Dong, Guoqiang Xu, Xiang Luo, Yi Cai, and Wei Gao. A bibliometric analysis of solar power research from 1991 to 2010. *Scientometrics*, 93(3):1101–1117, may 2012. doi:10.1007/s11192-012-0730-9. URL https://doi.org/10.1007%2Fs11192-012-0730-9.

[20] EIA. Energy timelines - Solar Thermal, 2008. URL https://www.eia.gov/kids/energy.cfm?page=tl_solarthermal. Accessed: August 2017.

[21] Navigant Consulting Europe et al. Life cycle assessment of ultra-efficient lamps. In *Department of Environment*. FARA UK, 2009.

[22] Richard N. Foster. Timing technological transitions. *Technology in Society*, 7(2-3):127–141, jan 1985. doi:10.1016/0160-791x(85)90022-3. URL https://doi.org/10.1016%2F0160-791x%2885%2990022-3.

[23] Richard N Foster. *Innovation: The Attacker's Advantage*. 1986.

[24] Lidan Gao, Alan L. Porter, Jing Wang, Shu Fang, Xian Zhang, Tingting Ma, Wenping Wang, and Lu Huang. Technology life cycle analysis method based on patent documents. *Technological Forecasting and Social Change*, 80(3):398–407, mar 2013. doi:10.1016/j.techfore.2012.10.003. URL https://doi.org/10.1016%2Fj.techfore.2012.10.003.

[25] F.W. Geels. Co-evolutionary and multi-level dynamics in transitions: The transformation of aviation systems and the shift from propeller to turbojet (1930–1970). *Technovation*, 26(9):999–1016, sep 2006. doi:10.1016/j.technovation.2005.08.010. URL https://doi.org/10.1016%2Fj.technovation.2005.08.010.

[26] Maxime F Gendre. Two centuries of electric light source innovations. 143, 2003. URL http://ufowaves.org/rendlesham/Rendlesham%20Incident%20Forum/www.einlightred.tue.nl/lightsources/history/light_history.pdf.

[27] William E Glassley. *Geothermal energy: renewable energy and the environment*. CRC Press, 2014.

[28] Graeme Gooday. Re-writing the 'book of blots': Critical reflections on histories of technological 'failure'. *History and Technology*, 14(4):265–291, jul 1998. doi:10.1080/07341519808581934. URL https://doi.org/10.1080%2F07341519808581934.

[29] August E Grant, Jennifer H Meadows, et al. *Communication technology update and fundamentals*. Taylor & Francis, 2013.

[30] Wina Graus and Ernst Worrell. Trend in efficiency and capacity of fossil power generation in the EU. *Energy Policy*, 37(6):2147–2160, jun 2009. doi:10.1016/j.enpol.2009.01.034. URL https://doi.org/10.1016%2Fj.enpol.2009.01.034.

[31] Peter Gregory. *Chemistry and technology of printing and imaging systems*. Springer, 1996.

[32] Arnulf Grubler, Francisco Aguayo, KS Gallagher, Marko Hekkert, Kejun Jiang, Lynn Mytelka, Lena Neij, GF Nemet, and Charlie Wilson. Policies for the energy technology innovation system (ETIS). 2012.

[33] Jeremy Hall and Philip Rosson. The Impact of Technological Turbulence on Entrepreneurial Behavior Social Norms and Ethics: Three Internet-based Cases. *Journal of Business Ethics*, 64(3):231–248, mar 2006. doi:10.1007/s10551-005-5354-z. URL https://doi.org/10.1007%2Fs10551-005-5354-z.

[34] Richard Hanna, Robert Gross, Jamie Speirs, Philip Heptonstall, and Ajay Gambhir. Innovation timelines from invention to maturity. *UK Energy Research Centre*, 2015.

[35] Reinhard Haupt, Martin Kloyer, and Marcus Lange. Patent indicators for the technology life cycle development. *Research Policy*, 36(3):387–398, apr 2007. doi:10.1016/j.respol.2006.12.004. URL https://doi.org/10.1016%2Fj.respol.2006.12.004.

[36] Jeff Hecht. *City of light: the story of fiber optics*. Oxford University Press on Demand, 2004.

[37] S Helm, Q Tannock, and I Iliev. Renewable energy technology: Evolution and policy implications–evidence from patent literature (Global Challenges Report). Geneva: WIPO. *World Intellectual Property Organization, online at: http://www. wipo. int/export/sites/www/policy/en/climate_change/pdf/ccmt_report. pdf*, 2014.

[38] Thomas Parke Hughes. *Networks of power: electrification in Western society, 1880-1930*. JHU Press, 1983.

[39] Edward W. Constant II. A Model for Technological Change Applied to the Turbojet Revolution. *Technology and Culture*, 14(4):553, oct 1973. doi:10.2307/3102443. URL https://doi.org/10.2307%2F3102443.

[40] ITU ITU. Measuring the information society, 2013.

[41] Andreas Ralf Köhler. Anticipatory eco-design strategies for smart textiles. *Perspectives on environmental risk prevention in the development of an emerging technology*, 2013.

[42] Nancy Lambert. Orbit and Questel-Orbit: Farewell and Hail, 2000. URL

http://www.infotoday.com/searcher/feb00/lambert.htm. Accessed: June 2017.

[43] Changyong Lee, Juram Kim, Ohjin Kwon, and Han-Gyun Woo. Stochastic technology life cycle analysis using multiple patent indicators. *Technological Forecasting and Social Change*, 106:53–64, may 2016. doi:10.1016/j.techfore.2016.01.024. URL https://doi.org/10.1016%2Fj.techfore.2016.01.024.

[44] Wolter Lemstra. The Internet Bubble.

[45] Jessica Lin, Sheri Williamson, Kirk Borne, and David DeBarr. Pattern recognition in time series. *Advances in Machine Learning and Data Mining for Astronomy*, 1:617–645, 2012.

[46] Arthur D Little. *The strategic management of technology*. Arthur D Little, 1981.

[47] Henry C. Lucas and Jie Mein Goh. Disruptive technology: How Kodak missed the digital photography revolution. *The Journal of Strategic Information Systems*, 18(1):46–55, mar 2009. doi:10.1016/j.jsis.2009.01.002. URL https://doi.org/10.1016%2Fj.jsis.2009.01.002.

[48] Guozhu Mao, Xi Liu, Huibin Du, Jian Zuo, and Linyuan Wang. Way forward for alternative energy research: A bibliometric analysis during 1994–2013. *Renewable and Sustainable Energy Reviews*, 48:276–286, aug 2015. doi:10.1016/j.rser.2015.03.094. URL https://doi.org/10.1016%2Fj.rser.2015.03.094.

[49] Ian Marr. *Identifying the mode and impact of technological substitutions*. PhD thesis, University of Bristol, 2018.

[50] J. S. Marron, James O. Ramsay, Laura M. Sangalli, and Anuj Srivastava. Functional Data Analysis of Amplitude and Phase Variation. *Statistical Science*, 30(4):468–484, nov 2015. doi:10.1214/15-sts524. URL https://doi.org/10.1214%2F15-sts524.

[51] B. R. Martin. The use of multiple indicators in the assessment of basic research. *Scientometrics*, 36(3):343–362, jul 1996. doi:10.1007/bf02129599. URL https://doi.org/10.1007%2Fbf02129599.

[52] MathWorks. k-medoids clustering, 2016. URL https://uk.mathworks.com/help/stats/kmedoids.html. Accessed: March 2017.

[53] A. F. Mayadas, R. C. Durbeck, W. D. Hinsberg, and J. M. McCrossin. The evolution of printers and displays. *IBM Systems Journal*, 25(3.4):399–416, 1986. doi:10.1147/sj.253.0399. URL https://doi.org/10.1147%2Fsj.253.0399.

[54] Ian McLoughlin. Computer Peripherals.

[55] Philippe Menanteau and Hervé Lefebvre. Competing technologies and the diffusion of innovations: the emergence of energy-efficient lamps in the residential sector. *Research Policy*, 29(3):375–389, mar 2000. doi:10.1016/s0048-7333(99)00038-4. URL https://doi.org/10.1016%2Fs0048-7333%2899%2900038-4.

[56] Sumner Myers and Donald George Marquis. Successful industrial innovations. A study of factors underlying innovation in selected firms. 1969.

[57] J. Roland Ortt and Jan P.L. Schoormans. The pattern of development and diffusion of breakthrough communication technologies. *European Journal of Innovation Management*, 7(4):292–302, dec 2004. doi:10.1108/14601060410565047. URL https://doi.org/10.1108%2F14601060410565047.

[58] Renana Peres, Eitan Muller, and Vijay Mahajan. Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing*, 27(2):91–106, jun 2010. doi:10.1016/j.ijresmar.2009.12.012. URL https://doi.org/10.1016%2Fj.ijresmar.2009.12.012.

[59] Jenny Poolton and Ian Barclay. New Product Development From Past Research to Future Applications. *Industrial Marketing Management*, 27(3):197–212, may 1998. doi:10.1016/s0019-8501(97)00047-3. URL https://doi.org/10.1016%2Fs0019-8501%2897%2900047-3.

[60] James Ramsay, Giles Hooker, and Spencer Graves. *Functional Data Analysis with R and MATLAB*. Springer New York, 2009. doi:10.1007/978-0-387-98185-7. URL https://doi.org/10.1007%2F978-0-387-98185-7.

[61] Samira Ranaei, Matti Karvonen, Arho Suominen, and Tuomo Kässi. Forecasting emerging technologies of low emission vehicle. In *Management of Engineering & Technology (PICMET), 2014 Portland International Conference on*, pages 2924–2937. IEEE, 2014.

[62] Samira Ranaei, Matti Karvonen, Arho Suominen, and Tuomo Kässi. Patent-based technology forecasting: case of electric and hydrogen vehicle. *International Journal of Energy Technology and Policy*, 12(1):20, 2016. doi:10.1504/ijetp.2016.074490. URL https://doi.org/10.1504%2Fijetp.2016.074490.

[63] Francesco Rizzi, Nees Jan van Eck, and Marco Frey. The production of scientific knowledge on renewable energies: Worldwide trends dynamics and challenges and implications for management. *Renewable Energy*, 62:657–671, feb 2014. doi:10.1016/j.renene.2013.08.030. URL https://doi.org/10.1016%2Fj.renene.2013.08.030.

[64] Frederic M Scherer. Firm size, market structure, opportunity, and the output of patented inventions. *The American Economic Review*, 55(5):1097–1125, 1965.

[65] Melissa A. Schilling and Melissa Esmundo. Technology S-curves in renewable energy alternatives: Analysis and implications for industry and government. *Energy Policy*, 37(5):1767–1781, may 2009. doi:10.1016/j.enpol.2009.01.004. URL https://doi.org/10.1016%2Fj.enpol.2009.01.004.

[66] U. Schmoch. Indicators and the relations between science and technology. *Scientometrics*, 38(1):103–116, jan 1997. doi:10.1007/bf02461126. URL https://doi.org/10.1007%2Fbf02461126.

[67] Jacob Schmookler. *Invention and Economic Growth*. Harvard University Press, jan 1966. doi:10.4159/harvard.9780674432833. URL https://doi.org/10.4159%2Fharvard.9780674432833.

[68] Gerber Scientific. UNITED STATES SECURITIES AND EXCHANGE COMMISSION - FORM 10-K, 2007. URL https://www.sec.gov/Archives/edgar/data/41133/000004113307000026/fy200710k.htm. Accessed: August 2017.

[69] Vaclav Smil. World History and Energy. In *Encyclopedia of Energy*, pages 549–561. Elsevier, 2004. doi:10.1016/b0-12-176480-x/00025-5. URL https://doi.org/10.1016%2Fb0-12-176480-x%2F00025-5.

[70] Ashish Sood and Gerard J. Tellis. Technological Evolution and Radical Innovation. *Journal of Marketing*, 69(3):152–168, jul 2005. doi:10.1509/jmkg.69.3.152.66361. URL https://doi.org/10.1509%2Fjmkg.69.3.152.66361.

[71] Jefferson W Tester, Elisabeth M Drake, Michael J Driscoll, Michael W Golay, and William A Peters. *Sustainable energy: choosing among options*. MIT press, 2012.

[72] Joseph Tidd, John R Bessant, and Keith Pavitt. *Managing innovation: integrating technological, market and organizational change*, volume 4. Wiley Chichester, 1997.

[73] Erwin Tomash. The U.S. Computer Printer Industry, 1990. URL http://jacques-andre.fr/chi/chi90/tomash.html. Accessed: August 2017.

[74] Charles V. Trappey, Hsin-Ying Wu, Fataneh Taghaboni-Dutta, and Amy J.C. Trappey. Using patent data for technology forecasting: China RFID patent analysis. *Advanced Engineering Informatics*, 25(1):53–64, jan 2011. doi:10.1016/j.aei.2010.05.007. URL https://doi.org/10.1016%2Fj.aei.2010.05.007.

[75] James M Utterback. Mastering the dynamics of innovation: how companies can seize opportunities in the face of technological change Harvard Business School Press. *Boston, MA*, 1994.

[76] Arnold Verbeek, Koenraad Debackere, Marc Luwel, and Edwin Zimmermann. Measuring progress and evolution in science and technology - I: The multiple uses of bibliometric indicators. *International Journal of Management Reviews*, 4(2):179–211, jun 2002. doi:10.1111/1468-2370.00083. URL https://doi.org/10.1111%2F1468-2370.00083.

[77] Peter von Stackelberg. Technological Eras, 2011. URL http://www.transmediadigest.com/wp-content/uploads/2012/05/TechEras-Interactive.pdf. Accessed: August 2017.

[78] Robert J. Watts and Alan L. Porter. Innovation forecasting. *Technological Forecasting and Social Change*, 56(1):25–47, sep 1997. doi:10.1016/s0040-1625(97)00050-4. URL https://doi.org/10.1016%2Fs0040-1625%2897%2900050-4.

[79] Martin Weiss, H Martin Junginger, and Martin K Patel. Learning energy efficiency: experience curves for household appliances and space heating, cooling, and lighting technologies, 2008.

[80] WIPO. Patent-based Technology Analysis Report - Alternative Energy Technology. Technical report, 2009.

[81] Fei Yuan and Kumiko Miyazaki. Understanding the dynamic nature of technological change using trajectory identification based on patent citation network in the Electric Vehicles industry. In *Management of Engineering & Technology (PICMET), 2014 Portland International Confer-*

ence on, pages 2780–2790. IEEE, 2014.

[82] Theodoros Zachariadis. On the baseline evolution of automobile fuel economy in Europe. *Energy Policy*, 34(14):1773–1785, sep 2006. doi: 10.1016/j.enpol.2005.01.002. URL https://doi.org/10.1016%2Fj.enpol.2005.01.002.

[83] R. Zakon. Hobbes' Internet Timeline. Technical report, nov 1997. URL https://doi.org/10.17487%2Frfc2235.