2014 Proceedings of PICMET '14: Infrastructure and Service Integration.

Technological Frontiers and Embeddings: A Visualization Approach

Scott W. Cunningham, Jan H. Kwakkel
Policy Analysis Section, Faculty of Technology, Policy and Management
Delft University of Technology, Delft, The Netherlands

***Abstract*--The paper concerns the measurement and forecasting of technological change, a topic relevant to many high-tech organizations and their customers. We revisit recent and classic data sets from technology forecasting data envelopment analysis (TFDEA) research and technometrics in light of a new visualization technique known as t-Distributed Stochastic Neighbor Embedding (t-SNE). The technique is a non-linear visualization technique for preserving local structure in high-dimensional spaces of data. The technique may be classified as a form of topological data analysis. Specifically each point in the space represents a potential technological design or implementation, and each line segment in the space represents a local measure of technological improvement or degradation. We hypothesize six distinct kinds of performance development in technology within this space including the frontier, the fold, the salient, the soliton, and the lock-in. Then we examine the spaces to determine which kinds of development are the best explanations for observed development. The technique is not extrapolative, and therefore cannot supplant existing technometric methods. Nonetheless the approach offers a useful diagnostic to existing technometric methods, and may help advance theories of technological development.**

I. INTRODUCTION

How can we forecast the speed and direction of specific change for new technologies? Despite the scope and scale of technological changes since the industrial revolution, we still know relatively little about this question. Empirical successes are few – there is for instance the eponymous “Moore’s Law” which in fact reflects performance change across a range of computing technology. There are a variety of theories of technological change, but few look narrowly and specifically at performance changes. Still fewer efforts attempt to corroborate theories with empirical evidence to corroborate sources of change.

Effective measures for tracking, modelling and anticipating the direction of technological performance will have considerable payoff in a variety of fields of technology and innovation management. The first, and perhaps initial inspiration, is in systems acquisition. Technological systems which require a long lead time in research and development require a correspondingly long window of foresight for planning. Without such a window, the system is outdated before it is even developed. With such a window, reasonable functional requirements and systems integration can proceed as technologies develop apace.

A second opportunity area is in the area of research coordination. Many technologies today are distributed and staged across multiple organizations. So for instance in the automobile industry, parts of the drive chain may be developed and produced by different companies. Other newer technologies, such as high performance electric batteries, may still be in research or development. A further complication is that these research and development organizations may be public sector organizations with correspondingly different incentive structures and intellectual property regimes. Tracking technological performance, even today, enables these diverse organizations to coordinate their activities across alliance networks.

A third opportunity area lies in managing and planning for technological disruption. New technologies are part of a wave of creative destruction which reshapes political and economic life. Understanding new technological capabilities is an important part of military planning, and an essential element of maintaining security in an era of unconventional warfare. Technological change brings new public infrastructures, and the public monies are best spent with a planful and coordinated transition towards new technologies. In the private sphere many well-established companies have been unmade because of a lack of attention to their critical assumptions. Too often these critical assumptions are undermined by new innovation. Improved tools for tracking technological performance could also assist with managing these transitions.

How then to muster improved models and measures of technological performance? Evidence on technological performance is plentiful, albeit often of a proprietary character. So it’s a surprise that conventional statistics or operations research techniques have not been more extensively applied to the question of modelling technological performance. In previous research (below) we briefly survey these efforts. There are at least two obstacles to modelling performance change. The first is that technologies are inherently multivariate, and therefore simple input-output models are unable to capture this fundamental characteristic. The second challenge is that technological change is omnipresent, causing rises in performance across all measures, and confounding the ability to separate the distinct measures of change.

Technologies have a systemic character which must be captured in an effective model of technology performance. In Sahal’s fundamental conceptualization of technology [1] he offers three different alternatives. The first conceptualization is adopted from economics. Technologies are functions which transform inputs into productive output. The problem with such a conceptualization is that technological change itself is too often left unexplained – the so-called endogeneity problem. A second conceptualization sees technologies as waypoints along a trajectory which can be easily parameterized. This conceptualization, while useful, lacks material to understand the source of performance change. The third and final conceptualization by Sahal conceives of technologies as systems. This conceptualization usefully encompasses the prior ideas of technology, and has been considerably expanded by others [c.f. 2].

A systems conceptualization of technology is characteristic of technometrics. Technometrics is a discipline which "measures and evaluates technological change with important policy implications [3]." We therefore root our question within an applied tradition of technometrics. This paper presents a new technometric approach for tracking technological change which builds on a new renaissance in the visualization of complex data sets. We argue that these new multivariate visualization techniques, drawn from machine learning and data science, are ideal for monitoring the complex datasets derived from technological performance databases. A second advantage of these visualization techniques are their inductive, non-parametric character. We may know too little about the character of technological change to impose restrictive assumptions, and to foreclose some of the knowledge that can be gleaned directly from the data.

The paper is organized as follows. The second section reviews previous research in the field of technometrics, with a particular attention to the technique known as technological forecasting data envelopment analysis (TFDEA) [4]. This is followed by a literature review examining structural ideas about how technologies progress over time. We then present a method new to technometrics. The method, called stochastic network embedding (t-SNE) [5], is outlined with algorithmic and implementation details. We describe the case, which involves tracking hybrid and electrical vehicle performance. The case is ideal for comparative purposes since it has been previously investigated and analysed using TFDEA. The data and database of vehicles used in the case are described, and then the method is applied in the analysis section. We visualize the resulting dimensions of change in the electrical vehicle sector, and reflect on which of the hypothesized structures of technological change are best revealed by the data set. The paper then concludes with new avenues for research.

II. METHODOLOGICAL REVIEW

The purpose of this section is to review prior research in the field of technometrics. The review pays particular attention to the TFDEA technique since as noted, the case later presented permits a multi-methodological comparison of TFDEA using a common data set. The section begins with a broad survey of the various schools of technometrics, focusing particularly on a few critical assumptions in the field. These critical assumptions enable a comparison with TFDEA, which is broadly reviewed, and a new application of an existing technique. This new technique, stochastic network embedding, is more broadly described in the subsequent methodological section.

Our review of technometrics is beholden to the comprehensive efforts of Coccia [3]. Coccia describes three major schools and multiple subschools in the discipline, with an intellectual history reaching back more than sixty years. Rather than repeat the review here, the following paragraphs will highlight a number of critical assumptions which underpin the wide variety of technometric methods. Two assumptions are of critical importance for the methods in this paper; for interested readers we also highlight other differences across technometric schools and methods (table 1).

The two technometric assumptions which are of critical importance for this paper are parameterized versus non-parameterized approaches, and compositional versus decompositional techniques. Parametric approaches create explicit metrics, whether of technological performance, or of the variance in designs around market leaders, or both. Non-parametric techniques attempt neither. The advantage of parametric approaches are explicit guidance regarding the direction of technological change, and a metric to evaluate the uncertainties in future forecasts. The disadvantages of parametric modelling is the need to develop explicit understanding: of the system, the principal drivers of change, and the sources of diversity in design.

Compositional approaches attempt to model the technology as a whole, recognizing that performance may be an emergent property of multiple, unmodeled technological functions. The compositional approach is related to what Coccia [3] calls "summative" modelling. Decompositional approaches explicitly represent these technological functions. The advantage of the compositional approach is the ability to understand emergent performance characteristics of technologies. A disadvantage is the lack of resolution; without a model of the underlying technological functionality, important engineering and economic constraints may be omitted.

This covers two significant differences in technometric modelling. At least four other differences are worth noting (table 1). The measures used in technometric studies differ greatly. Some studies are based on actual performance data, while others are based on proxies. And, even among those studies which use actual data, some are attempting objective measures of performance while others are evaluating the user utility or satisfaction with enhanced performance. Two other differences lie in the level and locus of analysis. Many studies, particularly in the broader literature on the economics of technological change, study the impact of technology on whole industries. Correspondingly, many studies also investigate the impact of new technologies on users and groups. In this paper we are primarily interested in changes in specific studies; the technology (and not its user community) is the principle object of study.

TABLE 1. ASSUMPTIONS IN TECHNOMETRIC MODELLING

Assumption	Explanation
Parameterized	Does the model attempt an explicit metric of either technological change, or design variance?
Compositional	Is the model summative in character, or does it attempt a multivariate perspective?
Representation	Is the model based upon actual data, or are proxies used when assessing technological performance?
Objectivist	Is the model measuring objective outcomes of technological interest, or does it attempt to measure increases in utility?
Level of Analysis	Is the technological context an entire industry, or a specific technological component?
Locus of Analysis	Is the model concerned with the object of technological change, or its knock-on effects on those who might adopt the technology?

We now turn to a particular approach within technometrics known as technology forecasting data envelopment analysis (TFDEA). TFDEA is based on a foundation of linear programming, a technique for optimizing linear or linearized systems. This versatile technique has been used for a variety of applications in operations management. The application of interest here is data envelopment analysis (DEA). Consider a number of different organizations or bureaus, each charged with a certain amount of input, and transforming this into gainful output. There are multiple inputs and outputs, and also multiple production technologies. The question underlying DEA is "What is the most credible claim that each bureau can make that it has been a credible steward of the resource it has been given?" Credible answers to this question are cast as optimal approaches to a frontier of peak performance. Bureaus can then be assessed as being best in class, or below best in class. Below best in class organizations can then be given benchmark values for potential improvement. The summary here is vastly simplified; a more thorough-going account is offered by Zhu [6]. The original work in the field is by Charnes et al. [7].The summary here is vastly simplified; a more thorough-going account is offered by Zhu [6]. The original work in the DEA field is by Charnes et al. [7].

The translation to technological forecasting begins with a view of technology as a function which transforms productive inputs to outputs. Over time technologies mature and develop, potentially delivering more output, or requiring less input to deliver the same performance. The technological frontier advances over time, with some technologies in a given year being market leaders, and others having fallen behind the performance frontier.

The technique is non-statistical, and so requires no functional specification of the variances in design. Furthermore the technique does not require a functional form for the technological frontier, instead creating a piecewise linear manifold of best-in-class technological performances. However the technique does deliver a parameterized function describing an optimum approach of a given technology towards the high-performance frontier. This function may differ from technology to technology.

The original TFDEA work is by Inman [4]. This versatile technique has been widely applied. The approach has been used to investigate jet fighters, database technologies, display technologies, microprocessors, and wireless communications -- among other technologies [8-12]. The approach is parameterized since it describes a desirable direction of technological change. It is also decompositional since it delivers an explicit model of a single technological function. In this paper a radically different approach is presented, an approach which is compositional and non-parameterized. First however, it is necessary to take a deeper look at theories describing the nature and direction of technological change.

III. THEORY AND FRAMEWORK

While there is a wealth of theories concerning the role and impact of technology in industries and the marketplace, there has been relatively lesser attention paid to the nature and direction of technology itself. In this section we survey three major schools of thought concerning the nature and direction of technological change. Significant schools of thought include reverse salients, radical innovation, and technological lock-in. Synthesizing these schools of thought permit the creation of a general framework concerning the spread and diffusion of technologies on a manifold or landscape of technological possibilities.

Hughes [13] developed the concept of the reverse salient as part of a history of electrification in the United States. The reverse salient is originally a military term, describing the point in front of troops where morale is lost. Hesitation at that point means that momentum has been lost, and the battle may potentially be turned to the opponents. In a technological sense the reverse salient indicates an obstacle, whether social, economic or technological, which presents the full expression of potential technological designs. Proponents of the school are quick to describe the difference between a bottleneck and a salient. A bottleneck indicates a winnowing of possibility, while the salient completely shuts out possible directions of growth. Mulder and Knot [14] present a case study of reverse salients in the development of plastic. Dedehayir and Mäkinen [15] present an analytical measure of performance gaps, thereby directly operationalizing the concept of a reverse salient in terms of observable measures of technology.

There are diverse perspectives on radical innovation; we highlight three. The three share a common argument that radical change occurs as individual systems and subsystems are strained and grow obsolescent. A famous model due to Henderson and Clark [16] argues that technological change occurs at both the system and the component level. Therefore there can be both novel components as well as novelty in technological architecture. Arthur [2] continues this theme, arguing that performance constraints can eventually limit future technological development until whole subsystems are designed and redeveloped. Corroborating Arthur, Suh [17] argues that technological change may be driven from the top down, or may emerge endemically from constraints in individual system components. He presents a design methodology which simultaneously addresses both sources of change.

A third school of thought describes technological lock-ins. The school characteristically questions whether it is necessarily true that technology progresses in a unilateral manner, displaying objectively better measures of technological performance over time. David [18] presented a famous case study which argued that the modern keyboard shut out an objectively better design – the Dvorak keyboard. The contention that the Dvorak keyboard is objectively better is subject to some controversy. Arthur [19] presents a simple model which purports some of the driving forces leading to a technological lock-in. According to Arthur positive network externalities can outweigh initial gains in technological performance, causing superior designs to remain unexploited.

Figure 1 presents a general framework synthesizing major perspectives on the direction of technological change. Technological change occurs in a landscape, a design space or manifold of potential technological opportunities. The design space is multivariate, representing a variety of different performance measures for the technology. The most general conceptualization of this manifold is that it is non-linear in the underlying performance measures, and that change is multi-directional. All changes or trade-offs between performance measures are possible, including absolute reductions. A much more specific conceptualization of technology, often linear and directional, is the technology frontier. (See the *soliton* and the *frontier* in the diagram below.)

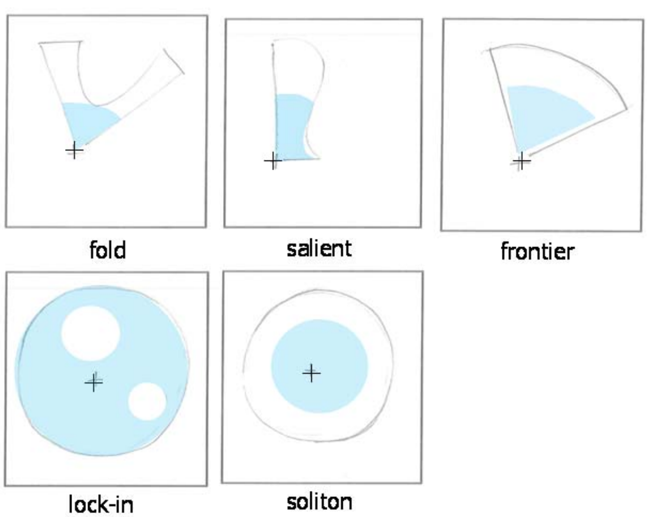

Figure 1. A Framework for Describing the Varied Direction of Technological Change

The manifold is presented by enclosed shapes in the figure above. The initial design is represented by a cross, and represents a particular expression of technological possibilities. Over time the design is adapted to different purposes, and changes in its expression of performance. It may or may not fill out the full space of technological possibilities. Those design possibilities which are filled are shown above in blue. Theories address the highly varied nature of these technological changes. Radical change is visualized using a *fold.* The landscape permits only two increasingly divergent forms of the technology. When change occurs technological changes cause increasingly divergent expressions of technological possibilities. The salient shows a landscape of technological possibility which is pinched or constrained in some manner (whether economically, technologically, or socially). The result is a limited, directional expression of possibilities. When the space of possibility expands, but fails to explore certain possibilities, the manifold describes a *lock-in.*

These manifolds, and the expression of technological possibilities over time, can be visualized using technology performance data. In the sections to follow this conceptual framework is confronted with data from an electrical vehicle case. After describing and analyzing the case we return to this framework to determine which if any of the surveyed theories of technological change best describe recent directions of change in the electrical vehicle sector.

IV. METHOD

This method section justifies the use of a novel and versatile data visualization technique for tracking technological change. The section first surveys relevant linear and non-linear data visualization and data compression techniques. The discussion then turns to the particular challenges of visualizing high-dimensional data. One specific non-linear data visualization technique is described in terms of its underlying principles and mathematics. Implementation details and outputs from the technique conclude the section.

The conventional approach for data visualization is based on linear algebra. The fundamental technique is known as singular value decomposition and entails computing the leading eigenvalues and eigenvectors of the system. These leading eigenvectors encompass a set of orthogonal, maximally informative vectors, which can be used to approximate the rest of the data. A range of techniques is built of this versatile core. Related techniques include principal components analysis, factor analysis, correspondence analysis, and multidimensional scaling. In the case of multidimensional scaling in two dimensions, the technique amounts a geometric approximation. The geometric approximation entails finding the best fitting plane through a multidimensional cloud. Best in this sense involves finding the plane that minimizes the orthogonal distance from the data points to the surface of the plane.

A variety of different non-linear techniques have also been proposed. Like multidimensional scaling, these techniques seek finding a subspace of data which preserves the structure in the data. Many real world datasets are very high-dimensional, so it can be a substantial challenge to either visualize the data set in its native dimensions, or to find an appropriate reduction in the data so that a sense of the data is appropriately conveyed. Techniques in this space include self-organizing maps, Sammon mapping, isomaps, generative topographical maps, locally linear embedding, Laplacian eigenmaps, and maximum variance unfolding. In this paper we examine the utility of one particular non-linear visualization technique, t-Distributed Stochastic Network Embedding (t-SNE). The technique has received attention in the literature because of its effectiveness in visualizing a wide variety of difficult, high-dimensional data sets.

The goal of the t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm is to take a full-dimensioned data set (n cases by q dimensions) and to embed this data in a lower dimensioned space (n cases by r dimensions, where r < q). Unlike many linear models, the algorithm only reduces dimensionality but not cases. A reduction in cases would involve finding exemplary cases and representing the remainder of the data set as neighbors of these cases. A full description of elements of the algorithm is offered in appendix A.

The algorithm requires the user to specify the perplexity of the model. Perplexity is a measure of how much of the local structure of the data must be considered when fitting the model. A low perplexity (for instance one) means that only the relationships between the nearest neighbors are considered, while a high perplexity (for instance a value equal to the total number of variables, n, the total number of cases in the model) indicates that the model will attempt to model the relationships between all of the points in the data. This measure smoothly interpolates between low and high dimensional data, and also between regularly spaced and irregularly spaced data. Lower perplexity values will result in a high-fidelity representation of the high density regions of the space of data.

Affinities are non-reflexive. For instance, case i may be more similar to case j than vice versa. In this model this occurs because the data points occur in regions which vary in density. When the embedding fits a data point in a high density region, it must consider a potentially greater variety of information in adjusting the fit then when it fits a low-density region. The model changes these affinities to be reflexive (and thus comparable to Euclidean distances) by taking an average value of the two. The model also ignores self-similarity in its calculations.

The algorithm takes as givens (P, T, η, $\alpha(t)$, r), which are respectively the perplexity, number of steps in the algorithm, the learning parameter, the momentum rate, and the desired dimensionality of the embedding. The data (X) is given as an input. As part of the initialization the algorithm generates an initial low dimensional embedding (Y). The algorithm can be initialized with a random solution or the outputs of a linear algorithm such as multidimensional scaling. Then the affinities between the data points are computed (p_{ij}). The algorithm itself minimizes a cost function C, with respect to the low-dimensional embedding (Y), and constrained by the actual data (X). The cost function is based on the Kullback-Leibler divergence, a general-purpose measure of the differences between two probability distributions (table 2).

TABLE 2. SIMPLIFIED DESCRIPTION OF THE T-SNE ALGORITHM

Given:	P, T, η, $\alpha(t)$, r
Inputs:	X
Initialization:	Y, p_{ij}
Objectives:	Minimize C with respect to Y, subject to X
Iterate:	for t = 1 to T do
	compute low dimensional affinities q_{ij}
	compute a gradient of the cost function $\partial C/\partial y$
	update Y(t)
	end
Output:	Y(T)

The heart of the algorithm is a non-linear optimization procedure. The algorithm iterates, making sequential reductions in the cost function. The learning parameter and the momentum function ensure that the algorithm smoothly improves over time, without being prematurely trapped in local optima. The algorithm is similar to a number of other non-linear optimization procedures, including hill-climbing algorithms and simulated annealing. Van der Maaten and Hinton [5] make explicit comparisons between their algorithm and a spring-graph layout. In short[5] make explicit comparisons between their algorithm and a spring-graph layout. In short, the stiffness of the springs is proportional to the degree of mismatch between the low-dimensional and high-dimensional affinities, while the force exerted on the springs is proportional to the placement of the cases in the low-dimensional embedding.

There are advantages as well as disadvantages to the choice of a non-linear visualization technique. The advantage of a non-linear visualization is the possibility of vastly reducing the needed dimensions to create an effective visualization. Another potential advantage is that the technique maintains only the local structure of the data. Similar objects in the high-dimensional space remain close in the reduced space, while the sparse parts of the data are only loosely modeled. This may be the most principled way of modeling sparse and under-sampled data.

The disadvantages of a non-linear visualization are three-fold. The problems are that the model: is no longer idempotent, must overcome the crowding problem, and must overcome the curse of dimensionality. Furthermore the non-linear procedure scales less effectively with large data set. These problems are discussed below.

The disadvantage is that the relationship between individual records and their encoding in the high-dimensional data is essentially lost. The relevant property of linear systems that maintains this relationship is known as idempotency. Further, very dissimilar objects may be conflated in the visualization, by being placed together in a spurious fashion. Another peculiar challenge of high dimensional data is the fact that there are many ways objects can be proximal in a high dimensional space – this is the so called crowding problem. Furthermore data in a high-dimensional space is increasingly sparse as the dimensionality increases. This makes the empirical estimation of distances between very distant objects in the data set increasingly more difficult.

The t-SNE algorithm makes further modifications beyond the pseudo-code in figure 2. These modifications simplify and speed the optimization process, and reduce the crowding problem. In the analysis which follows we use a matlab implementation of t-SNE [20] running on a standard laptop. In the next section we describe the input data (X). The input data includes different makes and models of hybrid and electric vehicles as cases, and incorporates a number of different performance measures for these vehicles as the high-dimensional data.

The output results are a two-dimensional embedding (Y). Additional calculation and visualization is applied to the outputs. The visualization shows the time of market entry. Even if the design space is itself highly non-linear, it can be useful to describe locally linear vectors in the space. These locally linear vectors are useful in determining the rate and direction of technological change, and in understanding the boundaries between the distinct regimes of design revealed in the data.

V. DATA, ANALYSIS AND RESULTS

In this section the t-SNE method is applied to a hybrid electric vehicle and battery electric vehicle case. The case has been previously discussed in the literature [21, 22], and thereby presents an opportunity to compare the t-SNE method with an established technometric approach (TFDEA). Such a comparison enables the work to examine the underlying assumptions of t-SNE and TFDEA which might otherwise go unexamined. This paper will not attempt a cross-validation of the methods. Ultimately such validation may require a retrospective comparison of the forecast in light of the new makes and models in the coming ten years. An additional scientific merit of the case lies in the design complexity of the vehicles themselves; the case is an opportunity to explore in a quantified manner the configurational learning concept endorsed by previous researchers [23].

The case offers considerable managerial and social relevance as well. The emergence of electrical vehicles presents a considerable challenge for research coordination among the various automotive supply chains. There are serious technological uncertainties in the case, although the uncertainties may be more configurational in character, than in component [24]. The social relevance of the case is also high. The penetration of electrical vehicles may dramatically change the transportation landscape, further altering the trajectory of carbon emissions and economic growth. Depending on policies, many first time drivers in the emerging world may soon select hybrid or battery vehicles rather than conventional fuel vehicles [25].

Table 3 clarifies the essential elements of the electric vehicle. The elements are first described, and then their role in the respective design configurations is discussed. The *battery* stores electrical change, and is used in some configurations to power the vehicle. The *charger* enables the engine to be recharged from various sources of electrical power. The *electrical motor* converts electricity to motive power. The *engine* burns liquid fuel also to power the vehicle. In those vehicles running on conventional fuel the *fuel system* delivers the fuel from tanks to engine. The *generator* converts kinetic energy into electrical change for storage and later use. The *plug* enables the car to be directly charged from the electricity mains where desired. In all designs the *transaxle* utilizes kinetic energy from the engine or electric motor to propel the vehicle forward.

TABLE 3. ESSENTIAL COMPONENTS OF THE ELECTRIC VEHICLE

	ICV	HEV	PHEV	BEV
Engine	✓	✓	✓	
Fuel System	✓	✓	✓	
Charger		✓	✓	✓
Electric Motor		✓	✓	✓
Battery		✓	✓	✓
Generator			✓	
Plug			✓	✓
Transaxle	✓	✓	✓	✓

The *internal combustion vehicle* (ICV) uses three of these elements – the engine, fuel system and transaxle. A *hybrid electrical vehicle* (HEC) adds three additional elements – the charger, the battery and the electric motor. A *plug-in hybrid* (PHEV) adds in addition a generator and a plug so that the user can choose to recharge from gasoline or the mains. *The battery electrical vehicle* (BEV) actually removes components so that the car becomes entirely reliant on electricity. The BEV removes engine and fuel system from the configuration. See also [21] , appendix A, for a more fuller discussion.

The data, collected by [21], involves 106 HEV, PHEV and BEV vehicles introduced in the market from 1997-2012. The data set is believed to be comprehensive for all modern makes and models up to 2012. The data was sourced from archives at SKF corporation and supplemented with freely available information on the internet. The data was further supplemented by [22] , but this additional information is not used here.

Table 4 describes the variables which are used in the analysis. Type is a logical variable encoding whether the vehicle is an HEV or PHEV, or a BEV. Weight is the total weight of the vehicle in kilograms. Output power is the power of the vehicle, measured in Watts. This is contrasted with battery capacity, which is the capacity for sustained power output (kilowatt-hours). The acceleration of the vehicles are noted, in units equivalent to meters per second square. The CO^2 emissions of the vehicles are recorded. Units for this are in grams of carbon per kilometer. As a convention we set the emissions of a battery electric vehicle to zero. In reality the sourcing of the electricity in the conventional generation system will also create carbon outputs. The resultant outputs affect all three vehicle types. Gas mileage is in terms of kilometers per liter, while battery range is kilometers per fully charged vehicle. These two variables are transformed into total driving range (kilometers) by assuming a standard tank size, and setting charge to one hundred percent on a new battery. This eases comparison across designs.

TABLE 4. VARIABLES USED IN THE ANALYSIS

Variable	Units / Explanation
Type	Logical variable; zero for HEV and PHEV, one for BEV
Weight	Kilograms
Output Power	Watts
Battery Capacity	kWatt-hours
Acceleration	Kilometer/hr/s
CO^2	Grams / kilometer; set to zero for BEV
Gas Mileage	Kilometers / liter
Electric Range	Kilometers / charge

The variables are in varied units ranging from kilometers to kilograms. We therefore choose to rescale the units. This is accomplished by rank-sorting the various makes and models, replacing the observed value with its rank. Identical values are coded at the average rank across all vehicles with the same performance score. The algorithm is robust to differences in unit scaling, but the scaling alleviates a potential threat to validity in the output. The algorithm is also dealing with mixed data types – both rank ordered as well as logical. Our hypothesis is that the algorithm can gracefully manage a variety of different data types. The local structure of the data enables an appropriate reweighting of the logical and ordinal variables. The data used in the case an opportunity to test this hypothesis.

The data is submitted to a t-SNE algorithm in matlab [20]. The resultant output is rapid and robust to multiple re-initialization. The raw output of the data is a matrix indexed by cases (on the rows), and embedded in a lower dimensional space (on the columns). The embedding is a ratio-scaled, two-dimensional vector for each make and model in the sample. The resultant output is readily visualized as a scatter plot (figure 2). As usual the axes of such plots are not readily interpretable given the varied data measures used as input.

The scatter plot is supplemented with a third dimension, the date of market availability of the vehicle. This data was not used in the analysis, since we believe the date of release is not a relevant variable for comparison of the vehicles. Nonetheless the overlay of the date of release adds an additional element in understanding the dynamics of change in this case. Cooler colors are used to represent the earliest vehicles (c. 1998) while warm colors are used to represent the newest vehicles (c. 2012). Interpreting the plot, similar designs are placed similarly together in the plot. Dissimilar designs are placed far apart. Unusual designs are shown as isolated in the resultant plot. There are noticeable groupings of designs over time, despite that time is exogeneous to the model. This suggests that there are technological dynamics at work in the case.

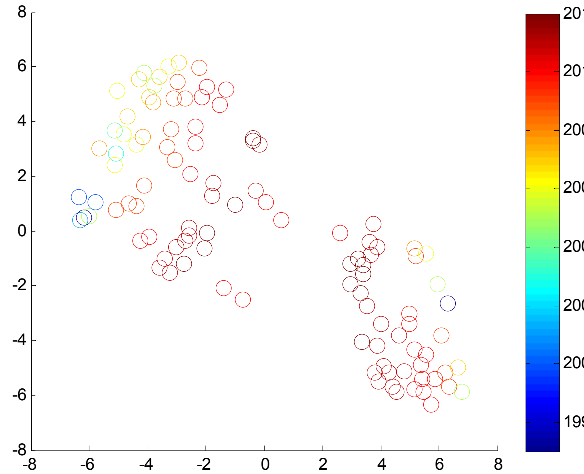

Figure 2. Placement of the Make and Models in a Scatter Plot

Figure 3 further labels the points by their design configuration – HEV, PHEV and EV. Each of these vehicle configurations occupies a different part of the space. Hybrid vehicles are located at the upper left, plug-in hybrids in the middle of the space, and pure electrical vehicles at the bottom right. No normative interpretation can be attached to the axes. Figure 3 provides additional confirmation that the different design configurations are captured in the reduced embedding.

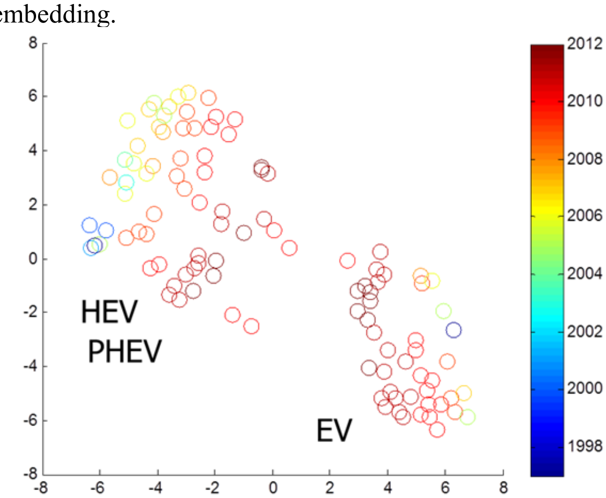

Figure 3. Different Configurations in the Design Space

We now take a closer look at the technological frontiers and outliers in the data, further annotating the figure to make these clear (figure 4). Two fronts are readily apparent – one in the space of the hybrid vehicles, and another in the space of the battery vehicles. The plug-in hybrids occupy an isolated space in the center. It is clear that new entries to the market are crowding along both fronts. It is further clear that there are many vehicles along the front, suggesting a range of viable designs in the marketplace.

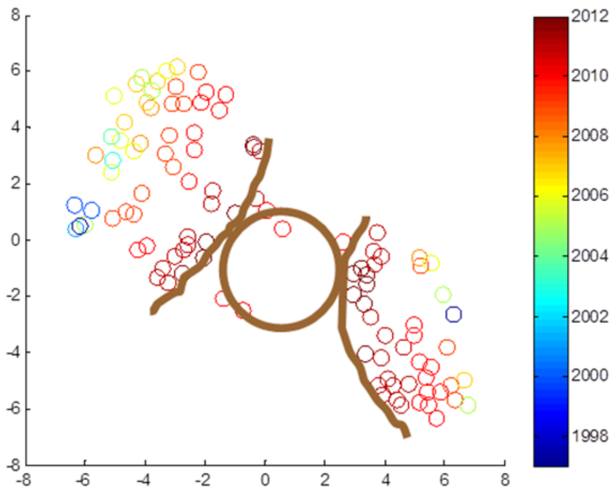

Figure 4. Frontiers in the Design Space

Given these results we now take a closer look at the implied dynamics of the space. These trajectories (figure 5) are not explicitly modelled. Nonetheless a clear direction of change can be seen for both vehicle configurations. On the hybrid vehicle side the design initially headed north or north east from a set of initial designs in 1998-199. A dramatic shift in technological emphasis happened around 2006 (incidently when battery electrical vehicles became readily available in the market. Design configurations then systematically turned towards the south-east, where a set of plug-in hybrid entries were already available.

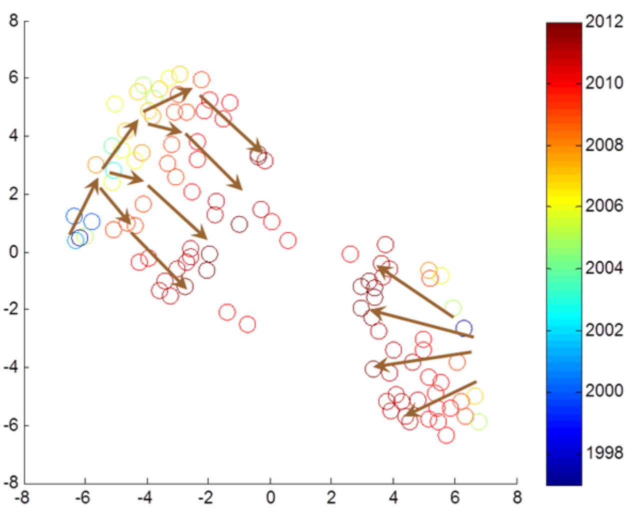

Figure 5. Trajectories in the Design Space

On the battery electric vehicle side, designs have steadily radiated away from an initial set of designs pioneered in 2004. The battery electrical designs are growing closer over time to the plug-in hybrids, and the fleet of hybrid electrical vehicles currently on the market. The plug-in hybrid designs are very distinctive. Given the current rate of change the rest of the market is likely to converge on comparable designs, but not until the 2020 time frames. Given the unusual design characteristics of these vehicles, and the apparent convergence of the market, we now take a closer look at two of these PHEV vehicles.

Both of the vehicles are Chinese makes. The F3DM (center isolate, upper right) is produced by the BYD Auto Company. The Besturn B50 (center isolate, lower left) is manufactured by the FAW Group ("First Automotive Works"). These companies are respectively the sixth and fourth largest automotive manufacturers in China. Both are attempting to aggressively expand into electrical vehicles given a strong domestic market and a perceived gap created by other western and Asian entrants. Note that neither the F3DM nor the Besturn B50 are championed as high-performance or high-cost-performance vehicles, despite their unusual characteristics.

We now turn to a discussion of performance trade-offs and of gradients of performance improvement. As discussed previously the local embedding is non-linear. Local linearization of the embedding is possible by drawing line segments of sufficiently limited length. Each line segment in the graph potentially represents a different weighting on the underlying variables. (The underlying variables are described in table 4). Of interest are technological trade-offs, which show a range of potential design swaps which are currently viable both technologically and economically in the market place. Two design trade-offs are apparent. One is the trade-off of HEV vehicles, shown at left. The second is the trade-off of electrical vehicles, shown at right. The fact that along each trade-off there are a variety of makes and models suggest widespread market viability for a range of designs.

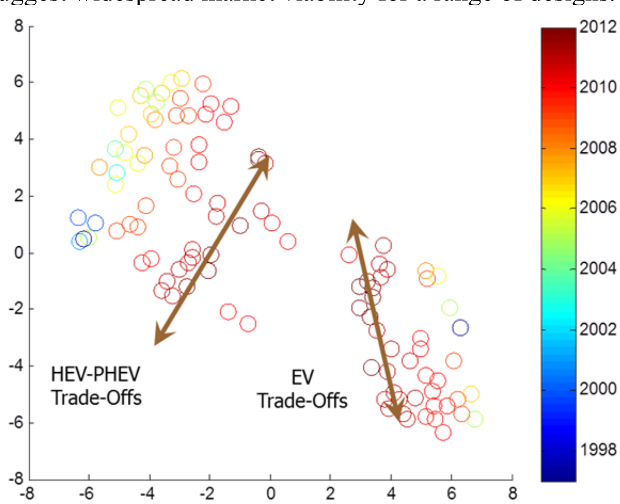

Figure 6. Performance Trade-offs

Also of interest are performance gradients. Performance gradients are the leading direction of technical change. The change is anticipated, but not yet expressed in many actual designs. As noted the t-SNE technique is not extrapolative, so we cannot with precision describe the exact gradient. It is nonetheless very interesting to examine this gradient, at least as a thought experiment (figure 7).

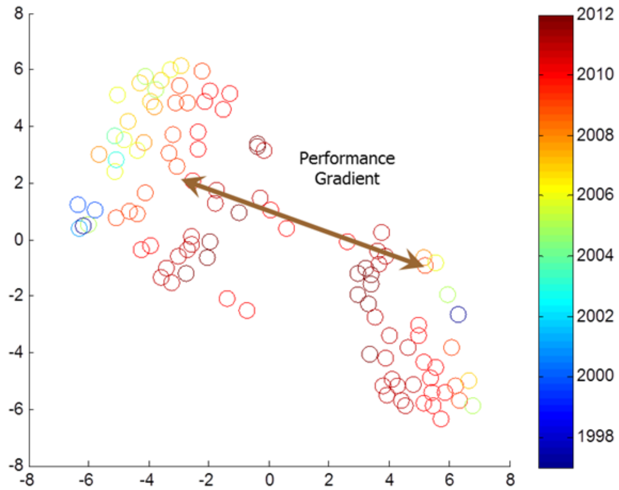

Figure 7. Performance Gradient

One gradient of change is shown in figure 7. This particular gradient is already populated with multiple novel designs, and is therefore selected for further examination. The approximation process involves selecting several makes and models on or near the gradient. Each model, if not already on the gradient, is mapped in its closest approach to the gradient. The gradient itself parameterized in terms of distance in the embedding. One make or model may be set arbitrarily to zero for this purpose, and other points scored according to their distances. Then a regression on technological performance by distance on the gradient may be performed. Similarly, the rate of technological change can be approximated by regressing the entry date of each of the selected makes and models with their location along the gradient.

Table 5 shows the result of one such analysis of the gradient. The procedure as described affords an estimate of where the average performances of vehicles will sit at an arbitrary date in time. A date of 2018 is selected for reference purposes. Since the two design configurations are converging, increasing gains in performance relative to one configuration are decreasing gains when compared to the other configuration. So for instance the forecasted vehicle in 2018 will be heavier, and have a greater battery capacity than the selected current design. On the other hand, relative to modern electrical vehicles the weight will be lighter, and the battery capacity will be lesser. The performance parameters of the 2018 vehicle are shown, and the respective units described.

Table 5 shows the resultant conclusions for the other four performance parameters considered. It is important to note that EV vehicles cannot, in their purely electric configuration, actually increase in CO^2 output and so the resultant performance change is marked as stable. This raises the issue that there may be regions in the embedding which cannot be reached without either logical contradictions, or potential violations of physical law. Table 5 also shows the rate of change (whether positive or negative). Of particular note are the rapid reductions in CO2 output relative to the standard HEV designs. Also interesting are the rapid changes in acceleration – increases relative to the selected HEV norm, but decreases relative to the selected EV norm.

A possible market scenario for 2018 is shown (figure 8). Given the previous changes in the design trajectories it seems entirely likely that the HEV and EV technologies will merge by 2018. The current range of market designs are already moderately constrained – whether by technology or economic realities is uncertain. We would expect this natural variation of market designs to continue into 2018.

Nonetheless the underlying technologies will continue to evolve, and there will still be a broad range of different possible market niches possible. This is expressed by the straight line shown on figure 8, which represents a gradient of technological performance. The gradient itself is not necessarily linear although it can be represented as linear in this graph. A range of possible vehicle designs are shown along this gradient, expressed as open circles.

The gradient might be further segmented – for instance into two niches. The upper niche is almost already occupied by the F3DM vehicle as discussed earlier. It may represent a low-cost vehicle most suitable for mass production. The lower niche is almost already occupied by the B50, although rather more technological change towards electrical vehicles might be expected here. This segment might represent more expensive or elite vehicles for the upper class, given the current customer segment for Besturn.

TABLE 5. LOCAL LINEAR APPROXIMATION OF THE GRADIENT

Performance Indicator	HEV	EV	Rate of Change	2018 Estimates	Units
Weight	Increasing	Decreasing	0.3% / year	1669	Kg
Output Power	Decreasing	Increasing	0.8% / year	121	kW
Battery Capacity	Increasing	Decreasing	1.8% / year	31	kW-hr
Acceleration Rate	Increasing	Decreasing	2.1% / year	11	km/hr/s
CO^2 Output	Decreasing	Stable	3.7% / year	60	g/km
Functional Range	Decreasing	Increasing	1.4% / year	59	km

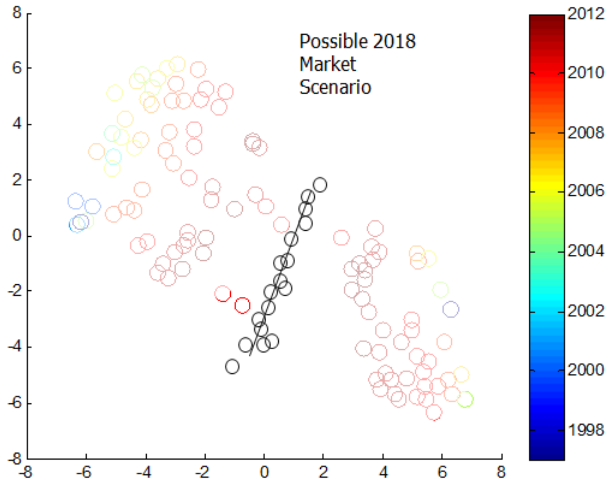

Figure 8. Possible Future Market Scenarios for 2018

In summary several interesting conclusions emerge from the analysis. The t-SNE successfully addresses distinct data types producing a coherent embedding of the technology. The embedding shows face validity – separating distinct design configurations in the embedding, and demonstrating the presence of rapid technological change in the time period studied. The diagram revealed fronts and trade-offs, and a gradient of rapid technological change which may presage continuing reductions in CO^2 output. In the final section of the paper we reflect further on the method and the framework. Limitations are acknowledge, and opportunities for future research are noted.

VI. REFLECTION AND CONCLUSION

The principal result of the visualization exercise is that plug-in hybrid electrical vehicles (PHEV) are becoming the dominant design in the marketplace. Existing hybrid electrical vehicles are adding increasingly powerful batteries, and slowly shedding their reliance on internal combustion engines. The F3DM is particularly noteworthy in the analysis as an isolated design which will increasingly be duplicated by others. PHEVs are not new of course, but the trend towards these designs has been obscured by the complexity of the space of vehicle attributes.

Another result of the visualization exercise is the presence of eroding technological goals. In this technology segment, a race towards uniformly superior technological attributes does not seem to be present. Electrical vehicles are becoming less able to accelerate over time. In return however, the vehicles are growing lighter weight, more powerful, and able to cruise for longer distances.

The implications of the model are that world policies for carbon reduction are succeeding. One of the most apparent directions of vectors of change is in CO_2 reduction; the newest designs are rapidly shedding their output of carbon. This suggests that current regulations for fleet reduction have been very successful, and should be continued as long as CO_2 reduction is an important goal for policy. More nuance could be applied to the model by accounting for the intended market for each make and model. The differential effects of polices by region might thereby be better evaluated.

The t-SNE model contrasts with prior analyses using TFDEA.Tudorie [21] Tudorie [21] analyzed the current data set using TFDEA. The author concluded that there would be no absolute reductions in overall performance improvement. This contrasts with the current model where, as noted, there are eroding technological performance goals. The model by Tudorie [21] also is far less optimistic about the potential for future reductions in carbon emission. A similar comparison can be made with Jahromi et al. [22] who also analyzed a modified version of this data set using TFDEA. The corresponding TFDEA model anticipates much higher rates of technological change over-all, while being relatively conservative about the prospects of improvements in fuel economy.

Both sets of model differences, we suggest, stem from the assumption of technological frontiers used in the TFDEA model. Further exploration is needed to determine which of the two methods is most valid for this data. The current model shows relatively many makes and models on a front of new activity. The front however, is not a frontier, since it does not depict an increasing gradient of technological performance.

The patterns of technological change in this example were more varied than expected. The framework used suggested only five patterns of change. These patterns were derived from the theory of reverse salients, technological lock-ins, and radical innovation. This model instead revealed two unexpected patterns – technological convergence, and a phase change. Both patterns have their antecedents in the literature. Technological convergence is reflected by the fronts of activity in the HEV and BEV industry, and the apparent momentum towards a middle ground. The existing literature on technological convergence has largely spoken about convergence in industry sectors, rather than convergence in technological design.Lee [26] Lee [26] describes the convergence in the telecommunications and computer networking industry. Gambardella and Torrisi [27] explicitly contrast convergence in technological design with convergence in the marketplace, using the microelectronics sector as a case.

The data suggest that the trajectory of hybrid electrical vehicles dramatically changed with the advent of battery electrical vehicles. The presence of an alternative paradigm on the market apparently caused a dramatic shift in design priorities on the part of the hybrid electric manufacturers. Speculatively, this shift may have occurred because of the ease of sourcing new vehicle parts, and because of a shift in consumer expectations regarding price and performance. There has been relatively little research on the matter, at least under the label of "phase change." Nonetheless Peine [23] describes "configurational learning," a process by which developers change and adapt technological architectures.

Murmann and Frenken [28] describe industrial changes ultimately leading to dominant designs.

Two of these articles in particular illustrate the complex relationship between technological dominance and industrial predominance [27, 28].[27, 28]. As previously noted the technometric literature distinguishes studies at the level of industries, as well as individual technologies. This enables us to take a closer look at the appropriate interpretation and use of these forecasts. Is the intention to describe the best-in-class technologies in a given year? Or instead, is it more appropriate to describe the technology is which is likely to dominate the market? This answer might differ according to the purposes of the study. Studies for systems acquisition purposes, or studies for analyzing technological disruption, will likely have the need to understand the high performance frontier. Studies for the purpose of technological coordination will be more concerned with the dominant design in an industry.

The case demonstrated several advantages for using an explicitly non-linear visualization technique. The technique afforded a local metric of change. A given line segment on the plot is structurally different from others on the same axis. This provides a needed degree of freedom in modeling the variety of directions of change in the data. For instance, the non-linearities in the model permit easy visualization of sudden bursts of growth and change.

There are limitations to the work as well, both inherent in the method as well as in the case design. The method is inherently inductive. Arguably more structuring of the input data is needed. The comparison of different designs in a single design space may have been naïve and may invite false comparisons. Acknowledging limitations in the case design, this was only a single case comparison. More extended application across multiple data sets are needed, and more rigorous cross-validation effort with TFDEA and other methods is needed. The desire to catalog patterns of technological change will undoubtedly require many more examples, with extended corroboration from the literature, to produce a workable framework.

A topic for future research would be modeling the swarming of new designs. The t-SNE model invites speculation about the future direction of technological change. Nonetheless it is not an extrapolative model. An approach using stochastic differential equations may permit a rigorous estimation of the numbers and varieties of designs likely to be manifested across the design space. The equations could possibly be corroborated with a maximum likelihood approach to calibrate the location of previous designs. Then, these patterns of growth could be used as an extrapolative exercise to determine likely future scenarios.

The principal contribution of this paper is to apply a novel non-linear visualization technique to technometrics. More broadly the technique may have relevant for a range of technology and innovation management questions including systems acquisition, transitions management, and research coordination. The paper also isolated a particular strand of literature on the dynamics of technological change and design, and corroborated the literature with a framework which can be tested using techniques of modelling and design. Finally, the paper contributed to the monitoring of advances in the hybrid electrical vehicle industry, pointing out a potential convergence not previously noted in technometrics.

ACKNOWLEDGEMENTS

An early version of this paper was presented without publication to the Higher School of Economics, Moscow, 28 November 2013.

REFERENCES

[1] Sahal, "Alternative conceptualizations of technology," *Research Policy,* vol. 10, pp. 2-24, 1981.

[2] B. W. Arthur, *The Nature of Technology: What It Is and How It Evolves* New York: Free Press, 2010.

[3] M. Coccia, "Technometrics: Origins, Historical Evolution and New Directions," *Technological Forecasting and Social Change,* vol. 72, pp. 944-979, 2005.

[4] O. L. Inman, "Technology forecasting using data envelopment analysis," Portland State University, 2004.

[5] L. van der Maaten and G. Hinton, "Visualizing Data Using t-SNE," *Journal of Machine Learning Research,* vol. 9, pp. 2579-2605, 2008.

[6] J. Zhu, *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheet*. New York: Springer, 2009.

[7] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision-making units," *European Journal of Operational Research,* vol. 2, pp. 429-444, 1978.

[8] D. J. Lim, T. R. Anderson, and J. Kim, "Forecast of Wireless Communication Technology: A Comparative Study of Regression and TFDEA Model," presented at the Portland International Conference on Management of Engineering and Technology, 2012.

[9] T. Anderson, T. Färe, S. Grosskopf, L. Inman, and X. Song, "Further Examination of Moore's Law with Data Envelopment Analysis," *Technological Forecasting and Social Change,* vol. 69, pp. 465-477, 2002.

[10] D. J. Lim, N. Runde, and T. Anderson, "Applying Technology Forecasting to New Product Development Target Setting of LCD Panels," *Advances in Business and Management Forecasting,* vol. 9, pp. 134-152, 2013.

[11] L. Inman, T. Anderson, and R. Harmon, "Predicting U.S. Jet Fighter Aircraft Introductions from 1944 to 1982: A Dogfight between Regression and TFDEA," *Technological Forecasting and Social Change,* vol. 73, pp. 1178-1187, 2006.

[12] T. Anderson, K. Hollingsworth, and L. Inman, "Assessing the Rate of Change in the Enterprise Database System Market Over Time Using DEA," in *Portland International Conference on Management of Engineering and Technology*, 2001.

[13] T. P. Hughes, *Networks of power: Electrification in western society, 1880-1930*. Baltimore: Johns Hopkins University Press, 1983.

[14] K. Mulder and M. Knot, "PVC plastic: A history of systems development and entrenchment," *Technology in Society,* vol. 23, pp. 265-286, 2001.

[15] O. Dedehayir and S. J. Mäkinen, "Dynamics of Reverse Salience as Technological Performance Gap: An Empirical Study of the Personal Computer Technology System," *Journal of Technology Management and Innovation,* vol. 3, pp. 55-66, 2008.

[16] R. M. Henderson and K. B. Clark, "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms," *Administrative Science Quarterly,* vol. 35, pp. 9-30, 1990.

[17] N. P. Suh, *Axiomatic Design: Advances and Applications* Oxford: Oxford University Press, 2001.
[18] P. A. David, "Clio and the Economics of QWERTY," *The American Economic Review,* vol. 75, pp. 332-337, 1985.
[19] B. W. Arthur, "Competing Technologies, Increasing Returns, and Lock-In by Historical Events," *The Economic Journal,* vol. 99, pp. 116-131, 1989.
[20] L. van der Maaten. (2014). *t-Distributed Stochastic Neighbor Embedding*. Available: http://homepage.tudelft.nl/19j49/t-SNE.html
[21] A. A. Tudorie, "Technology Forecasting of Electric Vehicles Using Data Envelopment Analysis," Delft University of Technology, 2012.
[22] S. R. Jahromi, A. A. Tudorie, and T. Anderson, "Forecasting Hybrid Electric Vehicles," presented at the Portland International Conference on Management of Engineering and Technology, Portland, OR, 2013.
[23] A. Peine, "Understanding the dynamics of technological configurations: A conceptual framework and the case of Smart Homes," *Technological Forecasting and Social Change,* vol. 76, pp. 396-409, 2009.
[24] N. Wang, "The exploration of the reverse salient of electric vehicle systems for urban mobility," Delft University of Technology, 2013.
[25] A. L. Porter, S. W. Cunningham, and A. Sanz, "Advancing the Forecasting Innovation Pathways Approach: Hybrid and Electric Vehicles Case," *International Journal of Technology Management,* in press.
[26] G. K. Lee, "The significance of network resources in the race to enter emerging product markets: The convergence of telephony communications and computer networking, 1989-2001," *Strategic Management Journal,* vol. 28, pp. 17-37, 2008.
[27] A. Gambardella and S. Torrisi, "Does technological convergence imply convergence in markets? Evidence from the electronics industry," *Research Policy,* vol. 27, pp. 445-463, 1998.
[28] J. P. Murmann and K. Frenken, "Towards a systematic framework for research on dominant designs, technological innovations, and industrial change," *Research Policy,* vol. 35, pp. 925-952, 2006.

APPENDIX A. ELEMENTS OF THE ALGORITHM

Scalars	
i,j	iterators over elements in a vector, n
t	iterator, algorithm step
n	data points
q	dimensionality of the full data
r	reduced dimensionality of the embedding
P	algorithm, perplexity
T	algorithm, number of iterations
η	algorithm, learning
$p_{i\|j}$	high-dimensional affinity between case i and j, non-reflexive
p_{ij}	high-dimensional affinity between case i and j, reflexive
q_{ij}	low-dimensional affinity between case i and j, reflexive
Functions	
α (t)	momentum, as a function of time
C(p,q)	cost, as a function of affinities
Vectors	
$x_i = \{x_{i1}, x_{i2}, \dots, x_{iq}\}$	data point for a given case i
$y_i = \{y_{i1}, y_{i2}, \dots, y_{ir}\}$	data point for a given case i
Matrices	
$X = \{x_1, x_2, \dots, x_n\}$	data set, dimensioned n cases and q variables
Tensors	
$Y(T) = \{y_1, y_2, \dots, y_n\}$	reduced representation of the data set as a function of iteration