



Open Archive Toulouse Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of some Toulouse researchers and makes it freely available over the web where possible.

This is an author's version published in: <https://oatao.univ-toulouse.fr/20073>

To cite this version :

Yuskevich, Ilya and Vingerhoeds, Rob A. and Golkar, Alessandro Two-dimensional Pareto frontier forecasting for technology planning and roadmapping. (2018) In: 2018 SysCon, 24 April 2018 - 26 April 2018 (Vancouver, Canada).

Any correspondence concerning this service should be sent to the repository administrator:

tech-oatao@listes-diff.inp-toulouse.fr

Two-dimensional Pareto frontier forecasting for technology planning and roadmapping

Ilya Yuskevich
Skoltech Space Center
Skolkovo Institute of Science
and Technology
Moscow, Russia
Email: ilya.yuskevich@skolkovotech.ru

Rob Vingerhoeds
Dept. Engineering Complex Systems
ISAE SUPAERO
University of Toulouse
Toulouse, France
Email: rob.vingerhoeds@isae.fr

Alessandro Golkar
Skoltech Space Center
Skolkovo Institute of Science
and Technology
Moscow, Russia
Email: a.golkar@skoltech.ru

Abstract—Technology evolution forecasting based on historical data processing is a useful tool for quantitative analysis in technology planning and roadmapping. While previous efforts focused mainly on one-dimensional forecasting, real technical systems require the evaluation of multiple and conflicting figures of merit at the same time, such as cost and performance. This paper presents a methodology for technology forecasting based on Pareto (efficient) frontier estimation algorithms and multiple regressions in presence of at least two conflicting figures of merits. A tool was developed on the basis of the approach presented in this paper. The methodology is illustrated with a case study from the automotive industry. The paper also shows the validation of the methodology and the estimation of the forecast accuracy adopting a backward testing procedure.

Index Terms—Technology forecasting, technology planning, trend extrapolation, Pareto frontiers.

I. INTRODUCTION

Forecasting technology evolution and developing research investment plans are at the core of the mission of a technology planning and roadmapping function of any engineering organizations. Companies need to identify and capture those trends, in order to remain ahead of their competition. Evolution forecasts are often done based on the intuition of experts and senior executives, and are complemented by methods such as technology scouting and other approaches adopted by strategy departments. Quantitative methods based on historical data analysis are decision-making support methods that help decision makers identifying patterns and defining evolution trends from data. In particular, trend extrapolation methods stand out from other technology forecasting approaches by being simple, fully quantitative and free from subjectivity biases. Trend extrapolation can be mathematically modelled using quantitative inputs and its accuracy can be likewise estimated. The disadvantage of this type of methods lies in its simplicity. First of all, there is always a risk of the emergence of unknown events which can break evolution trend (breakthrough innovations, economic crisis, and so on). This kind of events are ignored during the numerical and unsupervised process and should be taken into account by human intervention into the mathematical model (what at the same time introduce new biases). Secondly, classical growth curves trend extrapolation [1] address one-dimensional forecasting ignoring the fact that reality is more complex in a sense that a technology or a system can rarely be evaluated by just one figure of merit (FoM). In reality, for any given year a technology needs to be evaluated under multiple FoMs, thus defining an n-dimensional surface of Pareto efficient design points.

This paper proposes a methodology to quantify technology evolution using a Pareto frontier forecasting approach assuming evaluations based on two competing Figures of Merit. A similar approach to solve this problem was proposed by researchers at Portland State

University, under the name of Technology Forecasting Using DEA (TFDEA). With TFDEA, they adopted the data envelopment analysis (DEA) approach for technology forecasting to estimate the relative efficiency of each point and use this information to estimate the rate of change of technology evolution to verify Moore's law [2], predict US Jet Fighter Aircraft introductions [3], forecast the performances of wireless communication systems [4] and estimate performance of the hybrid electric vehicle [5]. Latest methodological advances in TFDEA development proposed frontiers segmentation and local rate of change calculation to improve accuracy in technology evolution forecasts [6].

We propose an alternative procedure for technology evolution forecasting targeting two main goals. The first goal is to forecast future nonlinear Pareto frontiers using a direct approach, differing from the ambition of TFDEA of accurately estimating technology rate of change. The second goal is to incorporate growth curve models (such as the S-curve model) into the analysis, as proven patterns describing technological progress.

Our proposed approach, being also visual, is suited for use in concurrent design studies, such as for technology roadmapping for example, where multiple experts convene and integrate their knowledge using models. The resulting framework is therefore suited for strategic planning and roadmapping and facilitates its future integration in a concurrent design data exchange platform such as CEDESK [7].

Furthermore, our approach accounts for physical limits in estimating technology evolution, which allows for additional validation of the results. By providing engineering teams tools for forecast adjustments, we allow experts to decide between different evolution scenarios and avoid fallacies in conclusion due to reliance on a purely automated approach. The approach furthermore informs users on the value of residual errors in evolution estimates, as well as its determination coefficients, giving them a chance to validate the accuracy of forecasts.

The remainder of the paper is structured as follows. Section II provides background and theory. Section III describes the mathematical approach here proposed for technology evolution forecasting. Successively, the results of the case study are discussed along with validation. The paper ends with conclusions and identification of avenues for future work.

II. BACKGROUND AND THEORY

The input for the proposed forecasting procedure is a time-sampled dataset of related technologies, systems or services with at least two conflicting figures of merit (that is, for which a tradeoff has been identified). It is assumed that each FoM has a monotonous behaviour over the evolution of the system or service of interest,

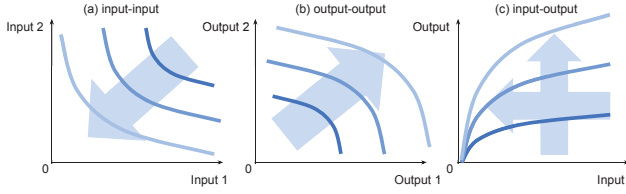


Fig. 1: Trade space classification by FoMs direction.

such as increasing quality or decreasing cost and development time. In measuring of productive efficiency theory each FoM is classified either as input or as output [8].

During system development we aim to maximize outputs (e.g. quality, power, performance) and minimize inputs (e.g. cost, mass, consumption). Technological progress occurs in such a way that inputs in average tend to decrease while outputs increase at the same time (which means, in turn, that average systems become more efficient over time). This is the case for example in data storage technology, where devices such as hard drives becomes cheaper and more capable at the same time over the last three decades.

Out of all possibilities, we can identify three mathematically different trade spaces that we define as “input-input”, “input-output” and “output-output”, respectively, as shown in Fig. 1. Inputs are all FoMs we wish to minimize (as they are proxy for resources, hence cost). Outputs are FoMs we wish to maximize (as they represent production, or other beneficial attribute contributing to value delivery). For the first type of trade space (Fig. 1a), the origin is an utopia point. Utopia is defined as the ideal optimum, which however cannot be achieved due to the opposing nature of the two FoMs. For the second type of trade space (Fig. 1b) the $[\inf, \inf]$ point is an utopia. Finally, for the third type of trade space (Fig. 1c) we have an utopia line (asymptote) coinciding with the y-axis.

To simplify the explanations and visualization this paper is concerned with two-dimensional Pareto frontiers only. However, we foresee to extend this approach to analyze n-dimensional Pareto frontiers, where each FoM again may be either input or output. With an increasing number of FoMs the number of possible trade space types is increasing as well. So to have a single unified mathematical process for all types of trade spaces, we are proposing to convert all inputs to outputs by taking outputs as the reciprocals of the inputs. For example, for cars we have two reciprocal metrics of the fuel consumption - miles per gallon (MPG) and liters per 100 Kilometers where first one is an output and second one is an input. We can easily switch between these metrics during pre- or postprocessing while core process works exclusively with outputs.

Year-by-year frontier movements can be mathematically expressed by means of multiple growth curves (one growth curve for each radial direction on a tradespace). The type of the fitting curves is the second attribute for the classification. The simplest case (Fig. 2) is the translational progress with constant rate of change in absence of physical limits. The growth curves in this case are straight lines.

In the second case, (Fig. 3) technology evolution rate of change slows down with time and asymptotically approach the physical limit. Such situations occur more often in applications and commonly described as the S-curve growth model.

In the most general case we have several growth curve models in a single trade space. This means that the shape of frontier is changed in time.

For instance, the double S-curve like the one depicted on Fig. 4 may occur due to the emergence of disruptive innovations and

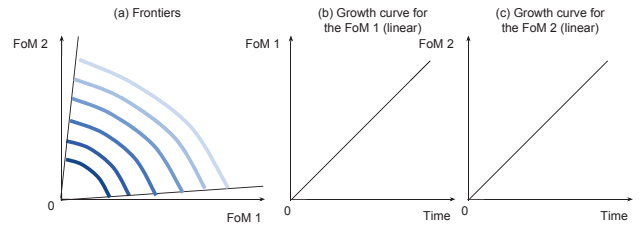


Fig. 2: Constant rate of change progress.

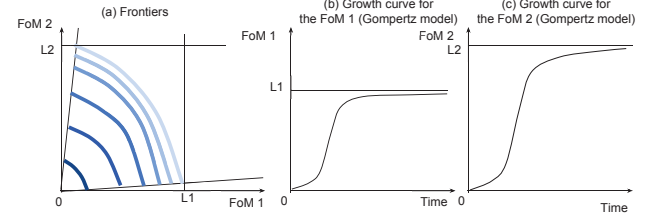


Fig. 3: Slowing rate of change progress.

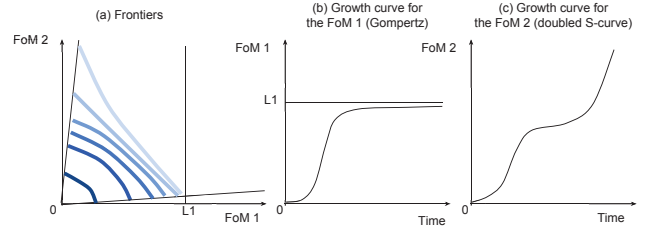


Fig. 4: Progress with two different growth curve models (e.g. in case of disruptive innovation along FoM 2).

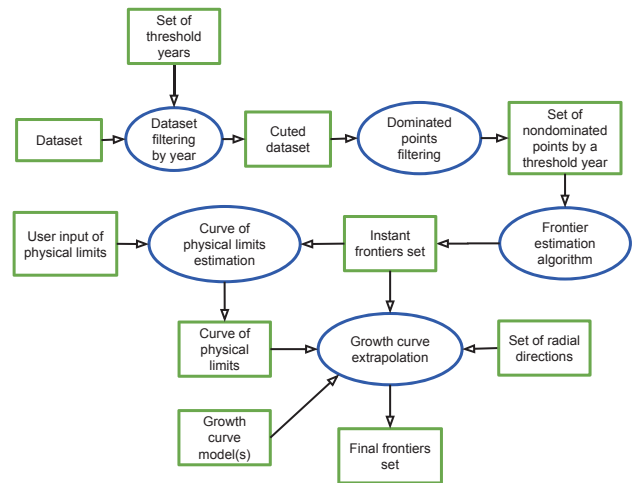


Fig. 5: OPM diagram of the proposed procedure.

paradigm shifts. Because of such influential events the convex frontier may become concave which can be mathematically expressed as a superposition of two or more growth curves.

III. MATHEMATICAL PROCEDURE

The proposed approach of frontiers estimating and forecasting is illustrated in the Object Process Model (OPM) [9] shown in Fig. 5.

After the dataset is obtained and all FoMs are converted to outputs we construct efficient frontiers by identifying Pareto efficient solutions for each year. It is impossible to use directly the standard Pareto frontier definition in our estimation algorithm. Instead, we require the Pareto frontier to satisfy the following properties:

- 1) No point shall be ahead of the frontier (e.g. in the unfeasible region);
- 2) The frontier is a monotonically decreasing function in a FoM space (first derivative is greater than zero);
- 3) The frontier can be either convex or concave on the entire domain, but its first derivative must be monotonic (no concavity changes).

The latter statement is worth discussing in depth because the concavity of the Pareto frontier in the output-output space is not obvious, and traditional econometrics approaches such as Data Envelopment Analysis (DEA) fail at estimating results for non-convex Pareto frontiers. Pareto front convexity is not always a realistic assumption in technology trade space analysis, due to the nonlinear physical processes underlying engineering tradeoffs. For instance, the set of frontiers of thermal versus propulsive efficiencies for commercial aircraft engines have been shown to be non-convex [10].

The problem of the Pareto optimal solutions determination in the non-convex regions can be solved by applying the adaptive weighted sum (AWS) method [11]. However, in our case this method is not applicable due to the inverse nature of the considered problem. AWS is the Pareto frontier generator which means that the model of the system is known and mathematically defined. This situation is common for the model-based systems engineering. In contrast, in the considered problem the model of the system is unknown, the set of solutions is limited and given as algorithm input. Thus, instead of Pareto frontier generation (multi-objective optimization) we deal with Pareto frontier estimation problem (curve fitting).

Further the exact mathematical routine is discussed.

For each dataset year we sample a set of nondominated points. By definition point x_i is nondominated if there is no any other point x_j for which $F_k(x_i) < F_k(x_j)$ for $\forall k, j$, where F is the vector function of multiple FoMs (which corresponds to dominated points filtering step on Fig. 5).

The orange circles on Fig. 6 represent the set of nondominated points of some given year for our case study. The frontier consisted of this set is not smooth or, more precisely, doesn't meet the formulated above conditions for valid Pareto frontier. This means that due to the finite number of existing products on the market (lack of data) some points are not lying on true Pareto frontier which is current cutting-edge for the given technology. To find this true Pareto frontier we need to apply a curve-fitting procedure (referred to as frontier estimation algorithm on OPM diagram on Fig. 5).

The problem of finding this true frontier can be formulated as standard linear optimization problem of minimizing distances between true frontier f_{true} and nondominated set f_{nd} under the condition of dominance f_{true} over f_{nd} and f_{true} monotony:

$$\begin{aligned} \min \|f_{true} - f_{nd}\|_{\infty} \\ f_{true} \geq f_{nd} \\ f''_{true} \geq 0 \text{ or } f''_{true} \leq 0 \end{aligned} \quad (1)$$

After we've calculated set of frontiers for all years (we call it instance frontiers set in a sense that we've not applied interpolation

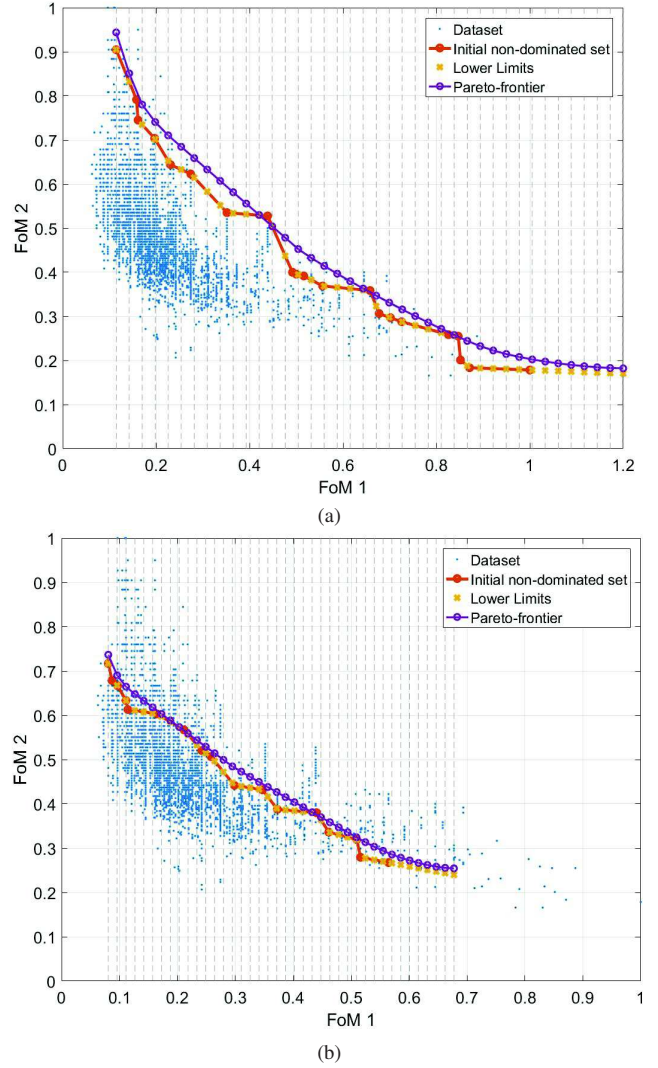


Fig. 6: Two results of frontier estimation algorithm (for two different cut-off years).

in a time domain yet). The result of solving this problem for two different years is shown on Fig. 6.

We perform the time domain interpolation along set of radial directions (the concept of this step is illustrated on Fig. 7).

Each direction has its own combination of FoM weights and represents very important concept of market segmentation (e.g. if FoM 1 is engine power and FoM 2 is MPG the direction A stands for the powerful cars whereas C stands for the economical).

Points $A', A'', B', B'', C', C''$ are inputs to our approach (described as follows). Their location is found as intersection of piecewise interpolated efficient frontiers and radial directions A, B and C . The resulting point series could be fitted by any growth curve. In this paper the Gompertz curve was used in straight line form [12]:

$$\log \log \frac{L}{y} = a - bx \quad (2)$$

Before fitting, the physical limit L or its allowed interval has to be set or estimated as regression model parameter if dataset is full

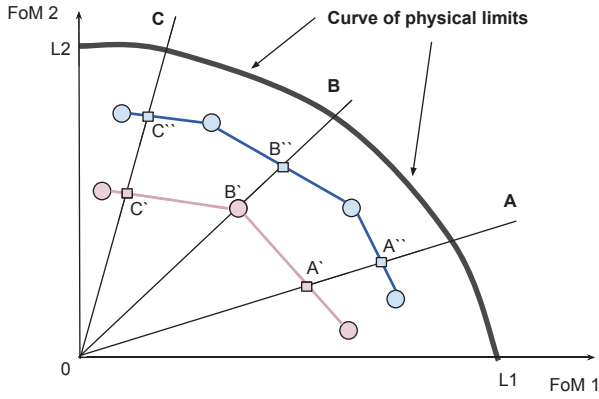


Fig. 7: Time domain interpolation concept.

enough to make a reliable prediction. In the two-dimensional space set of all physical limits in all possible radial directions forms a limiting curve to which frontiers approach asymptotically. There are several possibilities to define this limiting curve. One can define it directly, but this is redundant and impractical because this curve has the same shape as most recent frontiers. Therefore, it is enough to set the estimates of some parameters of this curve e.g. absolute or relative maximal value of each FoM (L_1 and L_2 on Fig. 7). Then, the averaged shape of a several the most recent frontiers is estimated and scaled to L_1 and L_2 (curve of physical limits estimation step on Fig. 5).

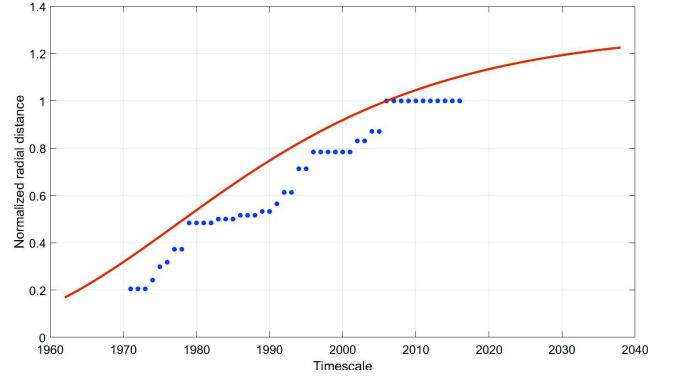
In our methodology curve fitting procedure is applied to all radial directions simultaneously with couple of additional conditions to preserve the right shape curve transformation in time (see Fig. 2-4). The resulting problem is a nonlinear convex optimization problem (NLP). To give an example let's formulate NLP for the Gompertz curve-case model and also require the frontier to preserve concave behaviour in time:

$$\begin{aligned}
 & \min_{a,b} \left\| \log \log \frac{L}{f(t)} - (a - bt) \right\|_2 \\
 & \log \log \frac{L}{f(t)} - (a - bt) \leq 0 \\
 & b_{i-1} - 2b_i + b_{i+1} \leq 0 \\
 & 0.2(a_i - a_{i-1}) \leq a_{i+1} - a_i \leq a_i - a_{i-1}
 \end{aligned} \quad (3)$$

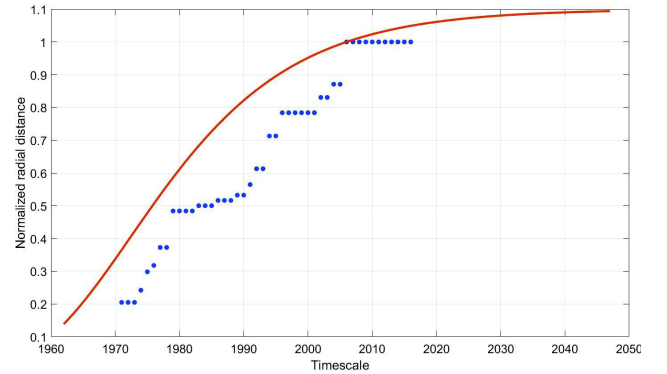
Where $i = 1 \dots n$ is the indexes of the radial directions set, t is the timescale, $f(t)$ is the instant frontiers set, a_i, b_i are the parameters of Gompertz curves for each radial direction, L_i is the physical limit estimation for each radial direction.

First equation determines the least-squares minimization of the fitting residual errors. Second equation require the growth curves to dominate the instance frontiers set in all points. Third and fourth equations constrain the optimization parameters and mutually link them in the FoMs space to preserve the right shape of the frontiers.

Here we would like to draw the attention to the following. After time domain interpolation with growth curves most of the products that were the best at their time actually no longer lie on cutting edge. Consequently, frontiers estimated independently from time (blue dots on Fig. 8) do not always coincide with true frontier set which depends also on time trends of trade-off evolution. This corresponds to real world experience; a technological frontier always goes ahead of mass



(a) $L = 1.4$



(b) $L = 1.1$

Fig. 8: Fitting the efficient frontiers in time domain along one of the radial directions with Gompertz curves.

TABLE I: Cars dataset sample

Model	Cost, €	Model year	MPG	Engine Power, Hp
Volkswagen Golf Sportsvan 1.4	34 970	2016	43.5	150
Smart forfour	11 195	2015	56.0	71
Porsche 911 Targa 4	148 776	2015	19.0	370
Toyota Celica Liftback 2000 ST	10 596	1980	24.0	89
Hyundai Atos Multi 1.0i GLS	9 525	1998	37.3	55

market products. From this perspective it is important that proposed analysis allows to calculate two kinds of slacks separately: delays of best practise products and delays of all products from a general technological frontier.

IV. CASE STUDY

We applied our approach to the estimation of Pareto frontiers in the the automotive industry using a publicly available dataset (as obtained from <http://www.cars-data.com>). The data set used for this case study contains more than 5 000 car models with petrol engine and manual transmission and covers over 40 years. The tradeoff is the engine power vs fuel consumption in miles per gallon (output-output type). An example of the data included in the data set is shown in Table 1.

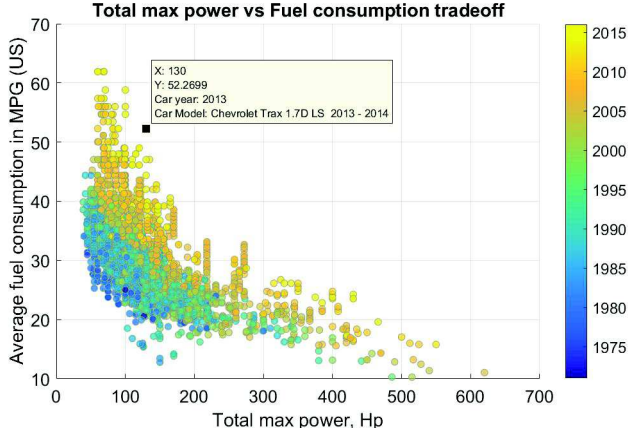


Fig. 9: Cars case study.

Fig. 9 shows the data sets of cars plotted on two FoMs of interest: maximum engine power (as measured in horsepower) and average fuel consumption in miles-per-gallon (MPG). This tradeoff is representative of a typical compromise a customer makes between performance and cost efficiency when purchasing a new vehicle. This of course does not mean that there are no other factors which could affect the decision. Among such factors are the level of comfort, price, body type, trunk capacity, and so on. However, the chosen tradeoff is probably a main technology-driven one in the automobile industry. The analysis proposed here can (and has to) be extended to account multiple figures of merit at the same time. However, for illustration purposes in this paper, we limit the analysis here to two FoMs, with no loss of generality, and leave multi-dimensional analysis for future work.

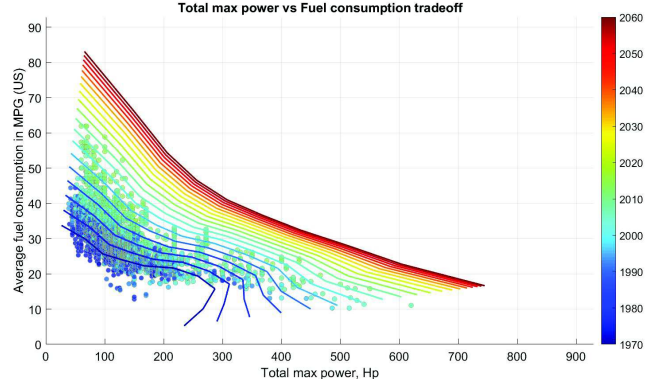
As we account for two FoMs only in this case study, we focused on data points different from each other in engine power and fuel consumption, and neglected car variants distinguished by other characteristics. For example, we did not consider in the data set those cars that have the same engine power and fuel consumption values, and different between each other under other characteristics (for instance automatic versus manual transmission, and different engine types such as diesel and hybrid).

Fig. 10 show the generated families of Pareto frontiers exhibiting different physical limits. Suppose we have two estimates of physical limits for our tradeoff - pessimistic and optimistic one. The historical data are the same for both cases.

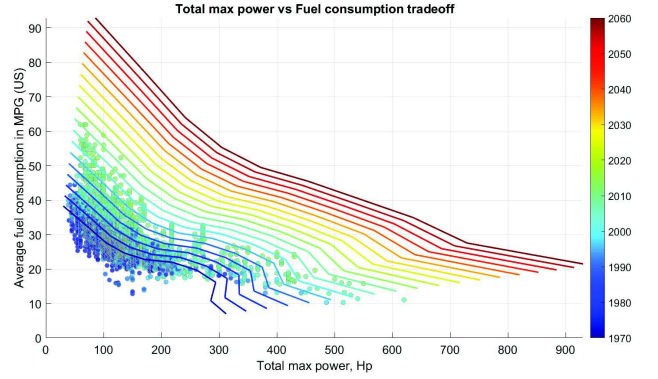
V. VALIDATION

We estimated the accuracy of our proposed Pareto forecasting method using a backward testing approach. That is, we partition our data set into training and validation subsets considering a given threshold year, and verify the accuracy of our forecasts. The error of forecasting is the difference between the real frontier (using known data) and the estimated one (using only data before the given threshold year).

The error itself could be measured either in years (as time difference between two corresponding frontiers) or as normalized distance in FoM space (normalized with respect to maximal variance of FoMs in the dataset). We used normalized distance since the estimated models could have different performance in different regions of the Pareto curve, due to varying rate of change as in the case of S-curve models. Fig. 11 illustrates the logic of the accuracy estimation.



(a) Pessimistic estimate with physical limits 992 Hp, 99 MPG



(b) Optimistic estimate with physical limits 1860 Hp, 186 MPG

Fig. 10: Cars case study frontiers.

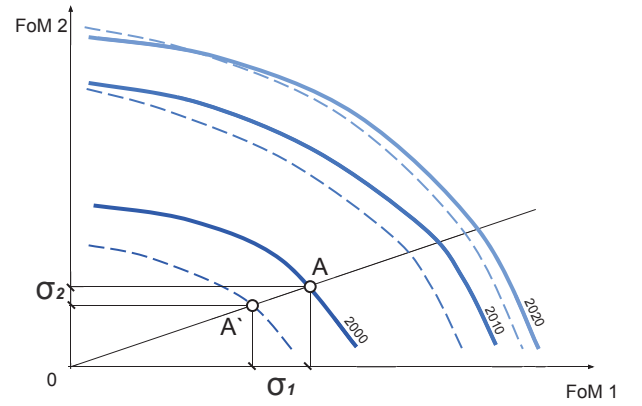
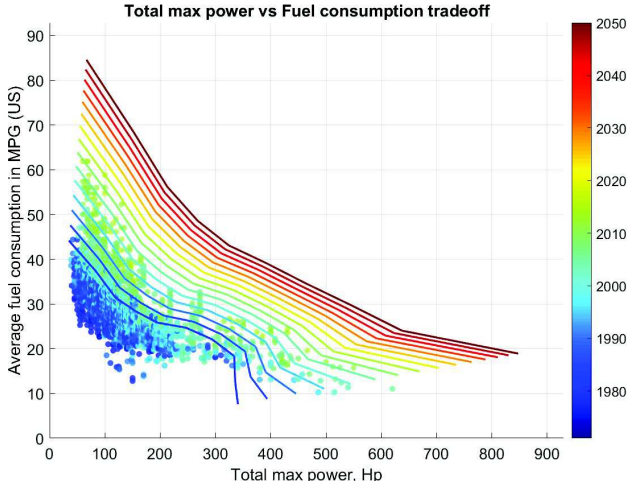


Fig. 11: The error calculation.

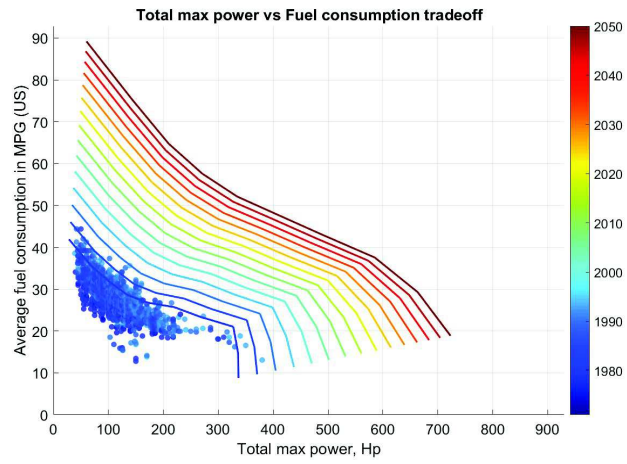
The normalized distance between points A and A' is the accuracy estimation for point A by proposed backward (or any other) test:

$$\begin{aligned} \sigma(A) &= \sqrt{\sigma_1^2 + \sigma_2^2} = \\ &= \sqrt{\left(\frac{f_1(A) - f_1(A')}{\max(f_1) - \min(f_1)}\right)^2 + \left(\frac{f_2(A) - f_2(A')}{\max(f_2) - \min(f_2)}\right)^2} \quad (4) \end{aligned}$$

In a given formula $f_1(A)$ and $f_2(A)$ are the frontier values estimated with full dataset and $f_1(A')$ and $f_2(A')$ are the frontier values estimated with cutted dataset for the same radial direction.



(a) 1972-2017 timeframe dataset



(b) 1972-1994 timeframe dataset

Fig. 12: Frontiers generated during backward test.

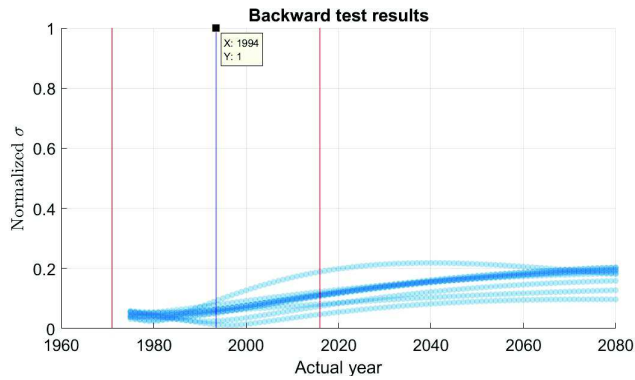


Fig. 13: Normalized error of 1972-1994 timeframe dataset with respect to 1972-2017.

Fig. 12 show the Pareto frontiers of the case study calculated using the full data set versus the reduced data set (threshold year 1994).

The normalized error chart of the backward test shown on Fig. 13.

A forecast error is a function of a dataset sufficiency and of a chosen approximation curve shape. It is clear from the figures that

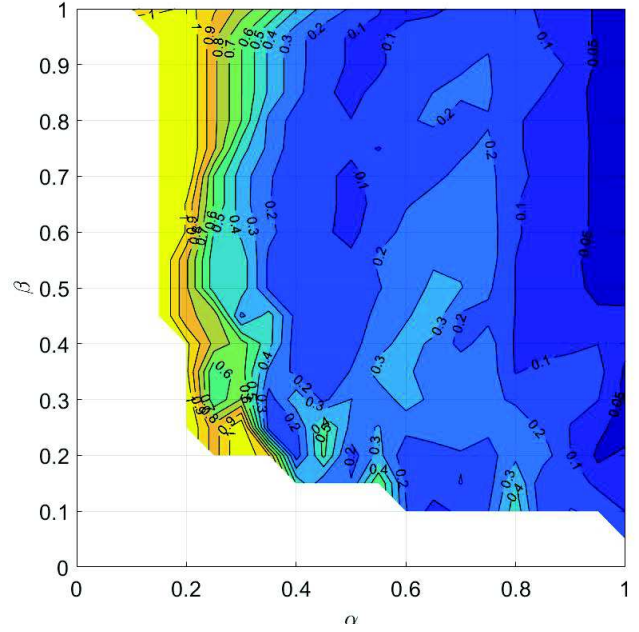


Fig. 14: Accuracy contour plot of α and β .

forecasting errors are greater in the segment of more powerful cars where the dataset is sparser. The forecast errors upon next 23 years is inferior then 20%. In segment of low-cost cars, which segment is more densely populated, the error is inferior then 10% which is quite a good result.

To see how the threshold years choice affects the forecast accuracy we performed parametrized version of the backward test. We introduce two parameters and representing the relative values of upper and lower threshold years:

$$\alpha = \frac{L_{upper} - \min(T)}{\max(T) - \min(T)}; \beta = \frac{L_{upper} - L_{lower}}{L_{upper} - \min(T)}. \quad (5)$$

Where L_{upper} and L_{lower} are the upper and lower threshold years respectively. The average error is calculated for frontier 2027 (10 years prediction).

Parametric contour plot of and is shown on Fig. 14. As can be seen from this figure accuracy is more correlated with α than with β which means that recent data (cars from 90s, 00s and 10s) have more influence on the prediction (recent data is more sufficient). This is not necessarily means that this is true for any dataset. We are not using any time weighting in our algorithm. Such situation occurs because this time period is characterized with more competitive and diverse market. This basically means more data which makes trends more evident.

The last test is intended to confirm the sufficiency of a dataset. Our case study is quite big (2160 data points). Obviously most of the points are not Pareto optimal and, because of that, do not determine the trend. The process of iterative exclusion of random points from the full dataset and comparison of the results of frontier forecasting gives us the estimation of the dataset sufficiency. The output of this test is shown on Fig. 15. If the threshold accuracy for us is 10% then the sufficient dataset should contain at least 400-500 data points.

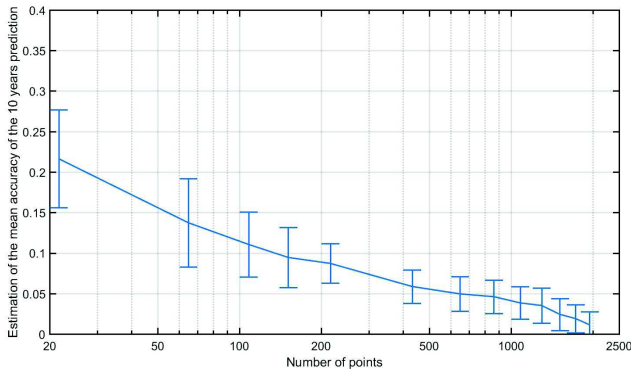


Fig. 15: Dataset sufficiency estimation.

VI. CONCLUSION

We propose a Pareto forecasting approach to estimate technology evolution over time. This approach helps users to visualize technology evolution and facilitates the identification of efficient and non-efficient technologies and supports decision-making in technology investment by setting realistic target performances and forecasting a development time for future technology. At the same time, to be effective, the proposed approach requires proper calibration in terms of growth curve and physical limits selection in characterizing technologies.

Future work will extend this approach to n-dimensional Pareto frontier forecasting, and development of additional case studies in industrial applications. We also foresee experimentation of the approach in concurrent design studies for technology planning and roadmapping.

REFERENCES

- [1] N. Meade, "The use of growth curves in forecasting market development — a review and appraisal", *Journal of Forecasting*, vol. 3, no. 4, pp. 429-451, 1984.
- [2] T. Anderson, R. 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, no. 5, pp. 465-477, 2002.
- [3] O. 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, no. 9, pp. 1178-1187, 2006.
- [4] T. Anderson, T. Daim and J. Kim, "Technology forecasting for wireless communication", *Technovation*, vol. 28, no. 9, pp. 602-614, 2008.
- [5] S. Jahromi, A. Tudori and T. Anderson, "Forecasting hybrid electric vehicles using TFDEA," 2013 Proceedings of PICMET '13: Technology Management in the IT-Driven Services (PICMET), San Jose, CA, 2013, pp. 2098-2107.
- [6] D. Lim and T. Anderson, "Improving forecast accuracy by a segmented rate of change in technology forecasting using data envelopment analysis (TFDEA)," Proceedings of PICMET '14 Conference: Portland International Center for Management of Engineering and Technology; Infrastructure and Service Integration, Kanazawa, 2014, pp. 2903-2907.
- [7] D. Knoll and A. Golkar, "A coordination method for concurrent design and a collaboration tool for parametric system models", *Concurrent Engineering*, 2017.
- [8] M. Farrell, "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society. Series A (General)*, vol. 120, no. 3, p. 253, 1957.
- [9] D. Dori, *Object-Process Methodology*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002.
- [10] A. Epstein, "Aeropropulsion for Commercial Aviation in the Twenty-First Century and Research Directions Needed", *AIAA Journal*, vol. 52, no. 5, pp. 901-911, 2014.

- [11] Y. Kim and O. L. D. Weck, "Adaptive weighted sum method for multiobjective optimization: a new method for Pareto front generation," *Structural and Multidisciplinary Optimization*, vol. 31, no. 2, pp. 105-116, 2005.
- [12] C. Winsor, "The Gompertz Curve as a Growth Curve", *Proceedings of the National Academy of Sciences*, vol. 18, no. 1, pp. 1-8, 1932.