21ST INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED17
21-25 AUGUST 2017, THE UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER, CANADA

INTERDISCIPLINARY LIFE CYCLE DATA ANALYSIS WITHIN A KNOWLEDGE-BASED SYSTEM FOR PRODUCT COST ESTIMATION

Altavilla, Stefania (1); Montagna, Francesca (1); Newnes, Linda (2)
1: Politecnico di Torino, Italy; 2: University of Bath, United Kingdom

Abstract

Product lifecycle has become one of the most relevant cost drivers for many manufacturing companies. Companies, in fact, are forced to reduce not only the costs that directly determine the price of the product, but they must also evaluate those costs that affect the entire lifecycle since customers are now considering after-sale services in their purchasing decision. However, implementing a product lifecycle perspective is challenging for an organisation if the uncertainty related to the duration of the life cycle process as well as the collection of a large amount of data are taken into account. Data, if correctly collected, contain valuable information that can be extracted and integrated into the analysis, to improve better design decisions and enhance product quality. For this purpose, this research proposed a knowledge-based system that uses empirical data, and information available across phases of product's lifecycle, and suggests how different qualitative and quantitative analyses can be performed, to transform the results in valuable feedback for future product development, while also providing a cost estimation.

Keywords: Design costing, Design for X (DfX), Conceptual design, Knowledge management

Contact:
Stefania Altavilla
Politecnico di Torino
Department of Management and Production Engineering
Italy
stefania.altavilla@polito.it

Please cite this paper as:
Surnames, Initials: *Title of paper*. In: Proceedings of the 21st International Conference on Engineering Design (ICED17), Vol. 5: Design for X, Design to X, Vancouver, Canada, 21.-25.08.2017.

1 INTRODUCTION

Nowadays, companies are constantly looking for new ways of providing additional value to customers and of gaining competitive advantage. If ten years ago, manufacturing companies invested into product quality, cost and time to market (Asiedu, 1998), today they focus on developing products that differentiate themselves from the others, as well as also being affordable, reliable and sustainable (Mévellec and Perry, 2006). Product Life Cycle hence has emerged as a critical area for investment. In particular, the focus on post-sales and, more generally, on customer is becoming so relevant that companies take into account not only the most obvious aspects related to services (e.g., shorter delivery times), but even they must consider the associated costs and the ones that a customer sustain during the utilization phase of a product. These costs, in fact, are becoming extremely important in customers purchasing decision (Dunk Alan S., 2004). Therefore, manufacturers are forced to reduce not only the costs, which directly will determine the price of the product, but they must also evaluate those costs that impact on users, as well as on society, extending their considerations on all the product life cycle costs (Altavilla and Montagna, 2015).

However, effectively implementing a Product Life Cycle perspective is challenging for an organization because it requires a long-term cost management, when the horizon and the environment are uncertain (Lindholm and Suomala, 2007), and the collection and usage of big amount of data from different phases of the product life (Ma et al., 2014). Available data contain valuable information and knowledge that can be integrated into the life cycle analysis to improve better design decisions and enhance product quality.

The horizon represents a challenge because of life cycle cost analysis changes in scope and kind of evaluation at a different stage of maturity of the development process, resulting even more demanding when it is adopted at the early stages of product development. Data gathering, instead, is challenging since it aims at collecting comprehensive information.

Data must be collected from different company departments, and even from several companies within a supply chain (Liu et al., 2008). Hence, the process of extracting and evaluating the data from different sources results time-consuming and can affect the quality and accuracy of the analysis. The most of the time, moreover, the majority of information available are related to the phases of product development and manufacturing. In particular, information coming from the later stages of product life cycle are difficult to reach, mostly because companies often do not have good and comprehensive IT systems, so to cover in particular processes outside the company. Another issue is related to the presence of missing data in the historical information accessible. Over the years, organisations have changed and evolved, as well as their Information System (IS) infrastructure: IT systems have been updated or modified, and this creates misalignments in the way information have been stored, especially for long-lasting products. However, an enormous amount of data is daily generated, and companies must be put in the condition of benefiting from them. Product designs, bills of materials, production processes, monitoring reports and market forecasting, etc. are collected and stored easily and quickly in databases at various stages and levels (Choudhary et al., 2009). Nevertheless, databases developed and then archived, lead to obtaining poor information, although data richness. How to enable designers and managers to extract and effectively use data available remains a problem (Wang and McGreavy, 1998). Knowledge can be extracted by data relationships, for finding the better configuration of a product architecture, or for instance discovering the connection between a particular design decision at a component level and the number of failures during the product utilisation, but manually this task results highly complicated.

For this reason, the support of computer-based and mining approaches is gaining importance in providing more in-depth and timely analyses (Germani et al., 2011). The goal is not only enhancing life cycle cost estimation but also creating a procedure to be systematically employed during the development stages, which represent for designers the enabling opportunity to learn from past experience and to suggest directions for future improvements, while also reducing costs.

For this purpose, we hence propose a knowledge-based system that uses empirical data, and information available across phases of product's life cycle, and suggests how different qualitative and quantitative analyses can be performed to transform the results in valuable feedbacks for future product development, while also providing a cost estimation.

A brief review of literature related to various approaches for cost estimation, focusing on the employment of knowledge reuse and mining techniques is presented in Section 2. The new system is proposed in Section 3. Section 4 illustrates the application case in an automotive company that designs, produces, and commercialises measurement devices for engine development while in Section 5 some conclusions are provided.

2 COST ESTIMATION TECHNIQUES AND THEIR DATA USAGE

A wide variety of product cost estimation methods is available in the literature. However, the time required for collecting the data, the availability of the information, and the stage of the design process impose significant constraint in choosing the most appropriate approach. Mainly cost estimation techniques can be divided into qualitative and quantitative methods (Niazi et al., 2006). Qualitative cost estimation methods are usually more appropriate in the early stages of design, as they are based on the analogy of a new product with the previous ones. Often, these techniques require historical cost data, past design experience and cost engineering experts to identify the similarity between products. The methods included in this category are implementations of a case-based methodology, a decision support system, and analogical cost estimation. The quantitative approach instead aims to provide evaluations that are more accurate, detailed, and hence their usage is often restricted to the final stages of the development process (Roy et al., 2001). These methods, mainly, decompose product processes into cost factors based on their features, geometrical characteristics, quality requirement and supporting activities. The costs are calculated using parametric or analytical functions, which aim to assess product and process variables as cost drivers.

However, a new trend in cost estimation is combining different qualitative and quantitative methods (Altavilla and Montagna, 2015). The proposed combining models try to solve the drawbacks within each single cost estimation technique, achieved mainly by enabling quantitative methods to be used from the conceptual design stage, which also improves the accuracy of the analysis. Qualitative techniques instead are meant to structure and collect the input information, mainly based on past experience, expert knowledge or theoretical information, employed then by more quantitative cost estimation methods. In particular, knowledge-based system approach of artificial intelligence has been extensively used in the process of extraction from databases implicit, unknown and potentially useful information, provided in the form of rules, trends and patterns. Germani et al. (2011) developed a knowledge-based tool that analyses CAD models to extract product structure (e.g. geometrical features, components, assemblies, etc.), and use the information to obtain the estimation of manufacturing costs, automatically. Naranje et al., (2014) report a similar example for the evaluation of sheet metal parts. In this case, the knowledge required is extracted in from of interviews with experienced designers or referring to industry research articles and manuals, and consequently represented in the form of production rules.

When extensive, and mostly qualitative, historical product configurations and cost data are available, artificial intelligent algorithms have been used to acquire, discover and classify the knowledge. Seo and Ahn (2006) used learning algorithms for the estimation of maintenance cost of electronics products. Che (2010) instead applied a particle swarm optimisation based approach for training artificial neural network in the evaluation of plastic injection moulding, while Deng and Yeh (2011) presented a least squares support vector machine method for the estimation of manufacturing cost of the airframe.

Alternatively, the data mining based cost estimate has emerged as a new approach to improve the accuracy and consistency of cost estimation. Data mining techniques have been recently used in the process of discovering patterns, associations and trends from large amounts of data stored in a data warehouse or other information repositories (Choudhary et al., 2009). They have been employed in a variety of fields (Gröger et al., 2012). In cost estimation, data mining have frequently been applied in particular for software cost estimation (Khalifelu and Farhad, 2012). However, a recent example is proposed by Sajadfar and Ma (2015) for the cost estimation of welding operations. The authors suggest a computer-aided tool for more accurate cost estimation, by using manufacturing process data, commonly available through IS within an organisation. Similar examples of projects cost estimation can also be found in Chou and Tseng (2011), or Gunduz et al. (2011).

Although a number of attempts have been made to develop advanced and automated knowledge-driven costing approaches within the early design process, most of them suffer from the following issues: i)the cost estimation analysis is mainly dedicated to one single phase of the product lifecycle, neglecting, in particular, the product in-use stage; ii) only a few attempts have been made in explaining how efficiently data can be extracted from companies repository; iii) less have been reported in the kind and amount of data that it is necessary to acquire for a life cycle cost analysis; iv) even fewer examples of how the knowledge can be interpreted and adapted by the development team for future cost estimation purposes has been found.

In response to the shortcomings above, this paper introduces a general comprehensive approach for product life cycle cost estimation, based on the analysis of a large amount of empirical data coming from different phases of product's life cycle. The proposed method is meant to guide, by a series of steps, the development team in the process of data collection, extraction, and interpretation for the purpose of product cost estimation analysis. Data mining techniques have been proposed to discover potential trend and pattern in data, while also identify the most influent cost drivers. An integration of quantitative techniques, namely parametric and Activity-Based-Costing approaches, have been introduced for building the cost estimation equation. This integrated approach has been previously proposed by the authors (Altavilla and Montagna, 2015). The structure of the model is finally bundled in the form of a generally applicable and automated knowledge-based system, based on three units: the Data Collection Module, the Data Mining module, and the Cost Estimation Module.

3 THE PROPOSED SYSTEM AND ITS MAIN ELEMENTS

The system that we are proposing is based on three main modules explain in the following, as well as depicted in Figure1:

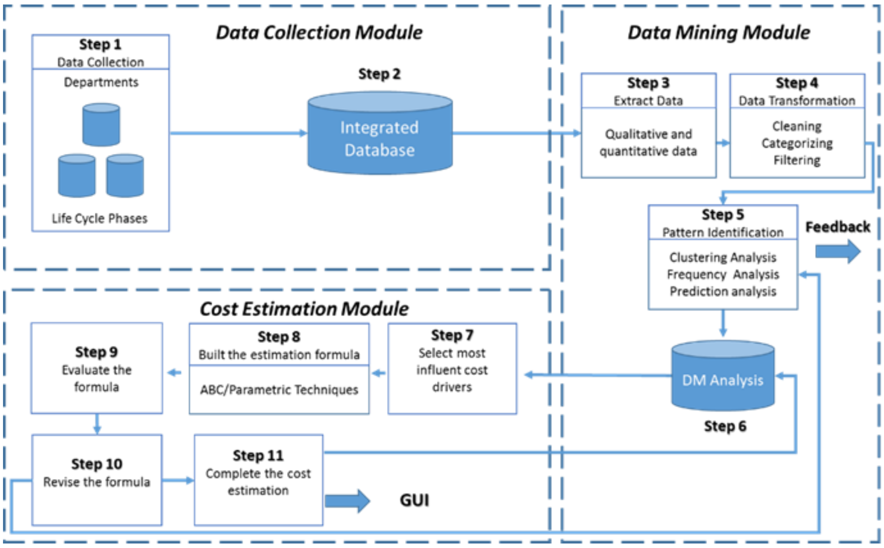


Figure 1. The knowledge-based system and its modules

A Data Collection Module: Product lifecycle data are usually dispersed and spread among different departments in a company. These departments may also have different databases and different formats for storing the data. Moreover, the amount and granularity of available information are also diverse among the various products, as well as throughout the different life cycle stages. Different kind of data must be collected (step 1): they are both tangible and intangible, usually consisting of product families and their structure (as also physical relationships among components and parts), activity subdivision and required resources, product and process cost data, as well as unstructured knowledge collected by interviews and questionnaires with the central figures in the organization. Therefore, creating a central and unique database (step 2) is the most important task in conducting a life cycle cost analysis, so that the different sources are collected and the data successively efficiently extracted. Accordingly, the database allows estimators and managers to transform unprocessed data into explicit knowledge, by providing a solid foundation for the application of Data Mining Module.

A Data Ming Module: The Data Mining Module is fed by the previous element to extract quantitative and qualitative information related to product and processes, and eventually analyse the patterns on product lifecycle (step 3). Obviously, not all the data are fundamental. Hence, for cost estimation purposes, the most relevant information has to be recognised and selected. Data need then to be transformed (step 4) and filtered to reduce redundancy and improve the quality of the database. On the new set of collected and cleaned data, different extraction techniques can be applied (step 5), according to the purposes of the analysis. Clustering algorithms can be employed to group similar products, components or parts. Frequency analysis is useful to get the most frequent patterns, such as development/manufacturing/maintenance activities, standard shared part and components, customised orders, defective parts or recalled products, etc. Sequence analysis can be used to recognize successive activities in the design and development, or in manufacturing and testing the customers' orders, or subsequent actions in repairing and maintaining a product. Normally, all the knowledge acquired at the end of this process is more than the one required for a cost estimation. Some of the analyses hence performed will constitute a valuable feedback and knowledge as "lesson learned" for product managers and designers. The patterns instead that are inherent with the cost estimation activity will be stored (step 6). They represent not only the physical and geometrical features of products but also refer to all the activities in the lifecycle stages. These patterns, namely cost drivers, in this way denote the most influent cost factors for each single product and will constitute the cost variables in the estimation model.

A Cost Estimation Module: In the Cost Estimation Module, the cost drivers are recalled (step 7) to be activity by activity included in the cost model. An integration of two cost estimation techniques is here performed: a Parametric method and an Activity Based Costing (ABC) approach. In particular, the ABC provide the structure of the method, focusing the analysis on each lifecycle activity, while the cost drivers have been parametrically derived by previous investigations carried out on the historically company's data. In cases where historical data are not available, or the product is entirely new and not at all similar to previously developed ones, a bottom-up approach, based again on the ABC technique, can be employed. In this case, the cost drivers can be discovered by referring to the judgement of experts or to the company's accounting system, where the activity drivers and the relative drive rates are already stored. In both cases, the cost equation is eventually derived, by manipulating the cost drivers based on the magnitude of their impact (associated polynomial order) or in respect to the cross effects (step 8). The total lifecycle cost is then expressed by the following equation:

$$LCC = f(P1, P2, P3, P4 \dots Pn) \quad (1)$$

Where LCC is the final product cost, which is a function of different variables (P1, P2, P3, ….Pn), each of which represents a single product lifecycle phase to be included in the cost analysis. Not all the phases have to be considered simultaneously. Based on the intended scope of the investigation, a single lifecycle phase can also be evaluated in the formula and separated in the cost estimation. The formula can be finally validated (step 9), considering that all the cost drivers and a set of coefficients should minimise the estimation error. In cases in wich, the accuracy of the estimation is not adequate, the cost model must be updated (step 10). Eventually the evaluation of coefficients, the equation can be finally used for the estimation of a new product (step 11), replacing the variables with the pool of existing data. It is clear that the information is mostly related to previous products at the beginning of a new product development process. However, the cost estimation can continuously be changed, by updating the formula as soon as new detailed data are available. The evaluation of the new product is stored for future analysis, and the final estimated lifecycle cost is showed by a Graphical User Interface (GUI).

4 THE APPLICATION CASE

The above-described system has been validated at a company that designs, produces, and commercialises measurement devices for engine development, for which it has four competitors in the world. Among the company's departments, the case study was placed in the segment that provides devices for combustion measurement. The choice on this particular segment was due to the planning of an upgrade of the technology of a particular instrument (a device for data acquisition), on which the proposed cost estimation approach could have been tested. The complete segment's portfolio was included in the analysis, which contains 9 product families (ranging from acquisition devices, amplifiers and angle calculator, to sensors and cables), 16 product lines (for different applications e.g. non-complex and first time use, or light/heavy-duty, racing, and large engines development and calibration) and 37 product versions (different in e.g. number of input/analog channels, speed of the analysis, dimensions, precision, and versions of post processing software).

4.1 Data collection and extraction: Step 1 to 4

To identify the main lifecycle phases and the source of data from which gather the necessary information, a first business process analysis was performed. The company operates in an assembly-to-order mode: product models are modularly pre-designed, and components are assembled once products are sold. Hence, after a preliminary market analysis, the research and development of products start. The production will only begin when an order is confirmed. During this period, sales activities, order fulfilment and customer service of the product follow one after another, until the ownership of the product shift from the company to customers.

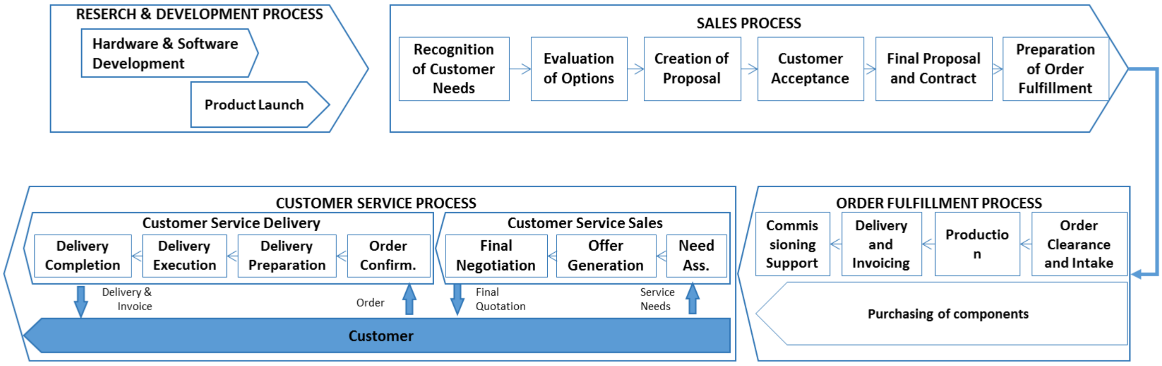


Figure 2. Company's Life Cycle Process

Accordingly, data were collected from each department involved in the lifecycle process. Research and Development data were obtained from the Accounting and Engineering departments, containing planned and actual cost information, requirements list, technical descriptions, drawings, BOMs, etc. Twenty years of Research and Development activities were collected at the end, divided in 50 different projects. Production data were extracted from the Accounting, Purchasing, Engineering and Manufacturing departments, including assembling and testing instruction, material requirements, suppliers' reports, etc. Data, in this case, were filtered for a time span of 5 years (2010 to 2015), contained a total of 1000 customer orders. For the Operation and Support Phase data were obtained instead mainly by the company's Affiliates, the Customer Service and Manufacturing departments. In this case, warranty reports, invoices, customer complaint, repair instruction and costs, were the set of information gained (around 600 records collected). All the data were available in different formats, and the specificity of the costing information ranged a lot among the various datasets. Moreover, there was a discrepancy between the various sources of data. These inconsistencies have shown the crucial importance of creating a centralised database, as a way to better aggregate information and efficiently understand product performance. For this reason, an integration of all the information was performed, and a subdivision of the data on a single product basis was provided. For the purpose of the future analysis, a representation of a product at different levels of lifecycle stage was hence created.

4.2 Data analysis and identification of cost drivers: Step 4 to 7

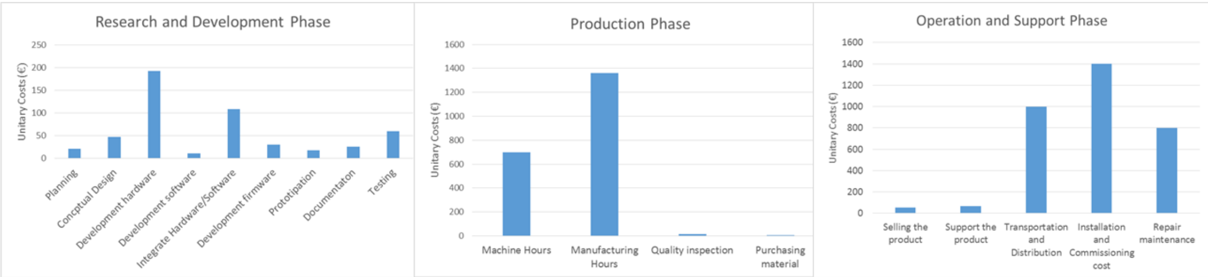
A first step in finding potential patterns in data was the application of clustering algorithms, to group products by analysing the similarity of their characteristics. The hypothesis is that products belonging to the same cluster, share the same characteristics, and as a consequence the same costs. Therefore, apply data mining analysis in a group with the appropriate size, can increase the accuracy of the results due to the existence of common relations among the components with similar data patterns. For this purpose, some distinctive variables describing the different company devices were chosen, namely the design effort (expressed by the number of design hours), the product complexity (expressed by functions decompositions), the component commonality (expressed by the number of parts) as well as the market price. The analysis resulted in the identification of five clusters, described in Table 1.

For each cluster, a separate analysis was hence conducted, aiming to search the most influential cost factors and to create the cost estimation model. For a non-comprehensive illustrative purpose, the following analysis will show the results obtained on a single product family, namely the Data Acquisition Devices (fourth cluster).

Table 1. Result of the Cluster Analysis

Cluster	Size	Design Effort (avg.)	Product Complexity (avg.)	Common ality (avg.)	Market Price (avg.)
C1: Simple Amplifiers	16	4152	Low	499	3200
C2: Sensors	8	2800	Low	130	1300
C3: Angle Calculators & Advance Amplifiers	2	6500	Medium	241	1700
C4:Data Acquisition Devices (DaD)	6	10230	High	545	6200
C5: Integrated DaD & Amplifiers	5	8200	Medium	530	5600

On the selected family of products, each life cycle phase was at first analysed, looking at the most influent activities in all process. Figure 3 shows the analysis performed and describes the activities based on the four main categories of costs (i.e. research and development, production, operation and support) on a single product basis.



(a) b) (c)

Figure 3. Cost of life cycle activities divided by R&D (a), Production (b), Operation (c)

Comparing graphs above, some considerations can be derived. In particular, during the Research and Development Phase, the cost are mainly driven by the development activities (Hardware Development and Hardware/Software Integration). Due to the nature of the products, the Testing activity represents a second large part of the costs. Regarding the Production phase, the major cost component is represented by the cost of the materials. However, looking at the activity costs, particular attention should be paid to the assembling and testing costs. Some additional costs are accounted as repair and maintenance, whose value consequently raise in the last phase of product operation. These costs are usually due to some additional repair and maintenance required during the warranty period by the customer. They often result from saving choices in the design phase that impact on the quality and feature of components.

Therefore, the cost factors that best described the costs, activity by activity, have been searched. The experiences of the designers and product managers involved in the process were also taken into account, to guarantee a broad spectrum of possible cost factors to be selected. At the beginning, a total number of nineteen possible variables were identified. These activity drivers reflected the characteristic of the single product (e.g. number of components, volume, width, depth), of the product family (e.g. number of product versions, number of standard parts, number of hardware and software options), and of the process (e.g. activity's duration, number of work orders, number of projects).

For each activity, a simple linear regression analysis was performed to investigate and model the one to one relationship between the identified cost factors and each activity cost. The best set of activity drivers was eventually recognised for each life cycle phase, as shown in Table 2.

The complete list of variables and their definitions is reported in Table 3. The process of cost drivers identification is still ongoing for the selected cluster of products. In fact, the results are available for the Research & Development and Production phase.

4.3 Cost estimation model formulation and validation: Step 8 to 11

The above-identified cost factors were quantitatively used in the construction of the parametric estimation model. Different models were created, to account for each life cycle phase individually, having in this way a better view on how the total cost, phase by phase, were influenced by the cost drivers. In particular, in this paragraph, the model created for the Research and Development costs and the Production costs are reported.

Table 1. Activity cost driver divided by life cycle phase

Life Cycle Phase	Activity	Activity Driver
Research and Development	Planning	Duration
	Conceptual Design	Number of Modules
	Development Hardware	Number of Parts
	Development Software	Number of Modules
	Development Firmware	Number of product Versions
	Integration Hardware/Software	Number of Parts
	Prototyping	Number of product Versions
	Documentation	Number of Parts
	Testing	Duration
Production	Purchase Material	Quantity of unit sold
	Assembling	Number of Modules, Number of Hardware and Software Options
	Testing	
	Quality inspection	Number of Parts

Table 2. Activity drivers and their definitions

Activity Driver #	Activity Driver Name	Description
X1	Duration (h)	It is the amount of projects' hours, directly assigned from projects to a single product.
X2	Quantity (pc.)	The variable expresses the number of units of product sold in a single order.
X3	Number of SW Options (pc.)	The number of software option sold in a single order.
X4	Number of HW Options (pc.)	The number of hardware options sold in a single order.
X5	Number of Parts (pc.)	The variable expresses the number of parts in the product. In particular, corresponds to the average of parts between different product variants.
X6	Number of Module (pc.)	The variable represents the number of the module that is a proxy of product complexity.
X7	Number of Product Version (pc.)	The variable accounts for the different number of product versions available for the product family.

The regression analysis was performed by using SPSS statistical software. With the inclusion of critical cost factors and associated known values of the response variable (namely the total cost of Research and Development and Production phases), the multivariate regression model determined the coefficients that produce the best fit for historical data with the largest variance explanation. Since all variables were continuous in the data set, the least square procedure was applied while analysing the data. The correlations of the independent variables were also investigated to eliminate multicollinearity. Table 4 reports the resulting best models, including for each of them the cost factors that presented significant levels of p-value < 0.05. Moreover, the Anova distribution of both the two models showed an overall p-value less than 0.05. This indicates that the regression model contains an effective set of attributes that affect the response quality.

Table 3. Linear parametric models with regression coefficients

Life Cycle Phase	**Response Variable**	**Parametric equation**	**Adj-R^2**
Research and Development	Total Research and Development Costs [€] (to be divided by the amount of produced products)	196359.8+(X1*54.9)+(X5*175.7)+(X7*(-55133))+(X13*(-3339.8))	0.95
Production	Production Cost [€/unit]	-18310.56+(X2 * (-52.9))+(X3*6.8)+ +(X6*177.9)+(X*44.9)	0.925

For model validation, previously separated data were used. The predicted cost was calculated by the previously defined statistically significant variables and final regression equations, resulting in an average absolute error of around 5.34%. This performance level could be further improved by extending the historical database as new data become available and by improving the quality of the data in the database through a more precise accounting of actual cost data.
Usually, even if these models do not represent the actual costs, but rather a proxy for the cost estimation, obtain such level of detail is challenging at the beginning of the design stage, though being valuable information for the products to be designed. Starting from the validated coefficients obtained, actual data will be used to predict the cost of a new product, by replacing the cost estimation models with the set of existing available information. A further sensitivity analysis on these cost drivers will allow assessing the main effects of their variations on the total cost profile. This step of the proposed approach is still ongoing within the company, as well as the creation of a unified cost estimation model that accounts of all the life cycle phases simultaneously. Given the absolute novelty of the method for the company and the natural resistance for the application, the proposed system is in use only for within one company segment, at the moment. However, introducing the system at the all company level will help in supporting real-time evaluation of new product design, providing the users with an automatic way to perform lifecycle cost estimation. The output can be used either to evaluate the impact of different design alternatives or to derive the actual costs of the new product once the design choices are made.

5 CONCLUSIONS

Companies nowadays are competing on global markets and hence the product quality, cost and time to market have become crucial strategic elements to survive, together with its affordability, reliability and sustainability. Moreover, companies are required to be flexible in providing product variants, to develop products looking at the entire lifecycle and to deploy these products considering service implications. However, decisions on costs are made during the early stages of development, but these considerably constrain the entire product life and its technical end economic performance.
Techniques for product life cycle cost estimation are increasing in importance to help companies to deal with these issues. However, implementing a Product Life Cycle perspective is challenging for an organisation if the uncertainty related to the duration of the life cycle process as well as the collection of a large amount of data are taken into account. In fact, life cycle cost analysis is highly dynamic and change in scope and kind of evaluation with the changing of the maturity of the life cycle process. Moreover, data gathering is even more complicated, since information must be collected from different departments, and even from various companies within a supply chain.
Data, if correctly collected, contain valuable information and knowledge that can be extracted and integrated into the life cycle analysis, to improve better design decisions and enhance product quality, while also reducing costs. For this purpose, this research proposed a knowledge-based system that uses empirical data, and information available across phases of product's life cycle, and suggests how different qualitative and quantitative analyses can be performed, to transform the results in valuable feedback for future product development, while also providing a cost estimation. The proposed approach has been developed for an automotive company that designs, produces and internationally markets measurement devices for engine development. Among the company's departments, the case study was placed in the segment that provides devices for combustion measurement. The entire segment's portfolio was investigated, which contains 9 product families, 16 product lines, and 37 product versions. The

complete application of the approach, as well as its optimisation and evaluation using different datasets, is still ongoing. This will be part of future works for the authors.

REFERENCES

Altavilla, S. and Montagna F. "When Costs From Being A Constraint Become A Driver For Concept Generation." *In DS 80-5 Proceedings of the 20th International Conference on Engineering Design (ICED 15)*, Vol 5: Design Methods and Tools-Part 1, Milan, Italy, 27-30.07. 15. 2015.

Asiedu, Y. and Gu, P. (1998), "Product Life Cycle Cost Analysis: State of the Art Review", *International Journal of Production Research*, Vol. 36, No. 4, pp. 883- 908.

Che, Z.H. (2010), "PSO-based back-propagation artificial neural network for product and mold cost estimation of plastic injection molding", *Computers & Industrial Engineering*, Vol. 58(4), pp.625-637.

Chou, J.S. and Tseng, H.C. (2011), "Establishing expert system for prediction based on the project-oriented data warehouse", *Expert Systems with Applications*, Vol. 38(1), pp.640-651.

Choudhary, A.K., Harding, J.A. and Tiwari, M.K. (2009), "Data mining in manufacturing: a review based on the kind of knowledge", *Journal of Intelligent Manufacturing*, Vol. 20(5), pp.501-521.

Deng, S., and Tsung-Han, Y. (2011), "Using least squares support vector machines for the airframe structures manufacturing cost estimation", *International Journal of Production Economics*, Vol. 131, No. 2, pp. 701-708.

Dunk, A.S. (2004), "Product life cost analysis: the impact of customer profiling, competitive advantage, and quality of IS information", *Management Accounting Research*, Vol. 15, pp. 401-414.

Germani, M., Cicconi, P. and Mandolini, M. (2011), "Manufacturing cost estimation during early phases of machine design", *In DS 68-5: Proceedings of the 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design*, Vol. 5: Design for X/Design to X, Lyngby/Copenhagen, Denmark, 15.-19.08. 2011.

Gröger, C., Niedermann F. and Mitschang B. (2012), "Data mining-driven manufacturing process optimization", *In Proceedings of the World Congress on Engineering*, Vol. 3, pp. 4-6.

Gunduz, M., Ugur, L. O. and Ozturk, E. (2011), "Parametric cost estimation system for light rail transit and metro trackworks", *Expert Systems with Applications*, Vol. 38, No. 3, pp. 2873-2877.

Khalifelu, Z.A. and Gharehchopogh F.S. (2012), "Comparison and evaluation of data mining techniques with algorithmic models in software cost estimation", *Procedia Technology*, Vol. 1, pp. 65-71.

Lindholm, A. and Suomala, P. (2007), " Learning by costing: Sharpening cost image through life cycle costing?", *International journal of productivity and performance management*, Vol. 56(8), pp.651-672.

Liu, H., Gopalkrishnan, V., Ng, W.K., Song, B. and Li, X. (2008), "An intelligent system for estimating full product Life Cycle Cost at the early design stage", *International Journal of Product Lifecycle Management*, Vol. 3(2-3), pp.96-113.

Ma, Y-S., Sajadfar, N. and Campos, L. (2014), "A feature-based semantic model for automatic product cost estimation", *International Journal of Engineering and Technology*, Vol. 6, No. 2, pp. 109.

Mévellec, P. and Perry, N. (2006), "Whole life-cycle costs: a new approach", *International Journal of Product Lifecycle Management*, Vol. 1(4), pp.400-414.

Naranje, V., Kumar, S. and Hussein, H.M.A. (2014), "A Knowledge Based System for Cost Estimation of Deep Drawn Parts", *Procedia Engineering*, Vol. 97, pp.2313-2322.

Niazi, A., Dai, S.J., Balabani, S., and Seneviratne, L. (2006), "Product Cost Estimation: Technique Classification and Methodology Review", *Journal of Manufacturing Science and Engineering*, Vol. 128, pp. 563-575.

Roy, R., Kelvesjo, S., Forsberg, S. and Rush C. (2001), "Quantitative and qualitative cost estimating for engineering design", *Journal of Engineering Design,*Vol. 12, No. 2, pp. 147-162.

Sajadfar, N. and Ma, Y. (2015), "A hybrid cost estimation framework based on feature-oriented data mining approach", *Advanced Engineering Informatics*, Vol. 29, No. 3, pp. 633-647.

Seo, K.K. and Ahn, B.J. (2006), "A learning algorithm based estimation method for maintenance cost of product concepts", *Computers & Industrial Engineering*, Vol. 50(1), pp.66-75.

Wang, X. Z., and McGreavy, C. (1998), "Automatic classification for mining process operational data", *Industrial & Engineering Chemistry Research*, Vol. 37, No. 6, pp. 2215-2222.