Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Corso di Dottorato in Ingegneria Industriale

Cost estimation and product value optimization during conceptual design of gas turbine

Ph.D. Dissertation of:

Irene Martinelli

Supervisor:

Prof. Michele Germani

Ph.D. Course coordinator:

Prof. Giovanni Di Nicola

XXXIII edition - new series

Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Corso di Dottorato in Ingegneria Industriale

Cost Estimation and Product Value Optimization during Conceptual Design of Gas Turbine

Ph.D. Dissertation of:

Irene Martinelli

Supervisor:

Prof. Michele Germani

Ph.D. Course coordinator:

Prof. Giovanni Di Nicola

XXXIII edition - new series

Università Politecnica delle Marche
Dipartimento di (nome del dipartimento dove la tesi e' stata sviluppata)
Via Brecce Bianche — 60131 - Ancona, Italy

To Diana

Acknowledgments

I wish to express my most sincere gratitude to Prof. *Michele Germani*, and Prof. *Marco Mandolini* for their guidance through the Ph.D. years. You helped me in my research growth and in the realization of this milestone.

Thanks to Prof. *Claudio Favi* and to colleague of doctoral school, Engineer *Federico Campi,* for all the important and valuable opportunities of learning and development.

During these years, a great result has been learning a way of working and this has been possible thanks to an excellent and harmonious working group within Baker Hughes team. Thanks to all my colleagues, especially *Emanuele Checcacci, Giancarlo Magni, Francesco Pescatori, Antonio Pumo,* and my manager *Giulio Lo Presti*.

Last but not least, a special thanks to my husband Federico to support me every single day, and to my parents for their unconditional selflessness, their unwavering support, their countless sacrifices that have opened many doors for me to grow and succeed, and for all of the little (and big) things they have taught me along the way.

Abstract

Traditional and consolidated design process mainly focuses on achieving technical performances and requirements, keeping attention on quality, cost reduction, and safety. For these reasons, several design tools have been developed to assist designers in this activity.

Nowadays, the development of a new product cannot leave aside value assessment and product cost management.

Value becomes more and more important and represents a key and successful market factor. Therefore, the need for design tools to support the designers in considering both performance indicators (including cost) and value aspects in the design process arises.

The research goal of this work is the definition of a systematic approach, based on parametric cost modelling and value analysis, for the optimization of design solutions during the early design stage of a gas turbine.

The proposed approach enables design engineers to model components to achieve target cost, leveraging conceptual cost modelling and Value Analysis Value Engineering (VAVE) approach. Conceptual cost modelling allows to predict the overall cost of the module/product under analysis, and several methodologies have been compared and tested (regressions, artificial neural network, random forest, etc.)

On the other hand, VAVE is used to generate disruptive ideas whenever the target cost is not matched by cost estimated through cost modelling.

Two case studies are provided to address the advantages and limitations of the developed methodology. The first one deals with gas turbine blades, which have been analyzed from a parametric cost modelling point of view and partially redesigned, leveraging the methodology proposed, to achieve a cost reduction of 25% and increasing product value up to 33%. The second one focuses in particular on axial compressor's discs and spacers and has been used to perform a comparison among different cost modelling methodologies, describing advantages and drawbacks of regression, neural network, deep learning and random forest, demonstrating the potential of machine learning techniques and their improved reliability.

This research work is a step toward the development of methodology helpful to anticipate the cost of products since the early design stages, reaching an acceptable accuracy level.

The target foreseen is a strong increase in product value since the conceptual design stage and a reduction of iterations and changes in design in the following design stages.

Next steps and future development have been identified and described with a proposed software platform able to manage both conceptual costing and VAVE analysis, for which system architecture, needed modules, use scenarios and requirements have been defined.

Contents

List of Figures

List of Tables

Chapter 1.

Introduction

1.1. Overall context of the research work

One of the most challenging aspects in the development of turbomachinery, and in particular of a gas turbine, is the Product Cost Model Management.

Gas turbines are the fundamental assets for power plants to produce electric current. A gas turbine is a combustion engine that is able to convert natural gas or other liquid fuels into mechanical energy. With this energy, it is possible to drive generators to produce electrical energy or to drive a centrifugal compressor that may be used in pipeline (to move the gas from the production to customer site) or to support processes, such as reinjection of high-pressure natural gas to improve oil recovery.

A gas turbine is composed of the followings main modules or sub-systems, see Figure 1 for reference:

- Axial Compressor: an upstream rotating gas compressor
- Combustion Chamber
- Turbine: a downstream turbine. It is often possible to define two different subsystems, the High-Pressure Turbine and Low-pressure Turbine.

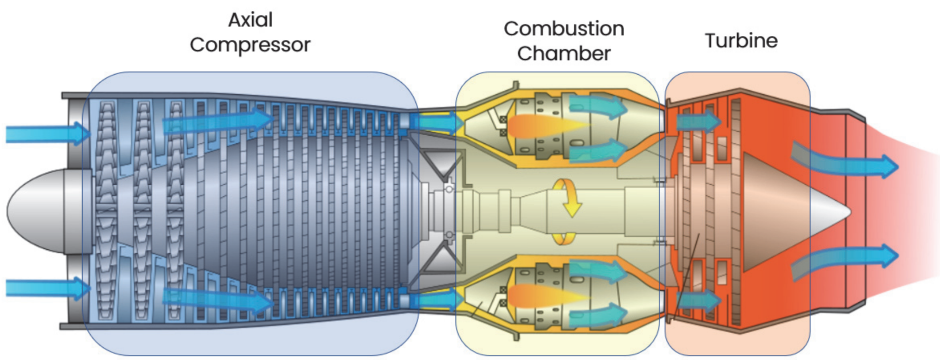


Figure 1. Modules or sub-systems of a Gas Turbine. Picture from web, for reference only

All those subsystems constitute the so called "flange to flange" portion of the gas turbine product. In order to operate a gas turbine, flange to flange systems needs to be supported by many other auxiliary subsystems. All the accessory systems are needed to guarantee gas turbine correct functionality, for example:

- Lube oil system for bearings
- Filter house to purify inlet air
- Fuel gas system to provide the correct amount of fuel to the combustion chamber

Gas turbine functionality is based on Brayton cycle with air acting as the working fluid. Atmospheric air is compressed through axial compressor stages, arrives in the combustion chamber where it is sprayed with fuel and ignited so that the combustion creates a high-temperature flow. High-temperature pressurized gas enters the downstream turbine, generating shaft work output in the process that can be used to drive an electrical generator or a centrifugal compressor.
This thesis work focuses on the engineering design of gas turbines with a specific interest in axial compressors and turbine subsystems.

The industrial framework in which this thesis work has been performed is the Gas Turbine New Product Development, in Baker Hughes company. Within this framework, the product design development process is represented in three steps of Figure 2 below. During stage 1, the problem statement is deeply defined, with relative boundary conditions statements. During Stage 2 preliminary and conceptual design stage is performed, and in this stage cost estimation is approached as well as gap versus target monitored, as shown in following Figure 2. Stage 3 is the development of a detailed design, during which also risk mitigation is analyzed and monitored. To complete each stage activities, a dedicated formal review shall be performed to receive approval from senior engineers and controlled title holders.

Figure 2 - Product development process – © 2020 Baker Hughes, LLC - All rights reserved

Program cost management aims to reach product cost targets aligned with profit goals, starting from the very beginning of production rather than reducing the costs only during industrialization. Companies working in turbomachinery design and manufacturing often use parametric cost estimation to be sure that the expected cost will be in line with the target cost.

Target cost is set by the market price. During those checks, if the gap between parametric cost estimation and target cost is sufficiently small, in particular, lower than 10%, it will be possible to continue with design steps. During each review dedicated to a given architecture, the cost will be managed in the same way as all the other performance parameters, to reach the target cost.

The result of the conceptual design for a gas turbine is the definition of a "product architecture" which summarizes all the technical characteristics implemented to address the design constraints. The choice of product architecture is a trade-off between product performance and cost. The analysis of the gap between the cost model and target cost must start from the pre-conceptual stage when the decision impact on the total cost of the product is greatest and there is room for a design change. Ability to lower costs decreases while cost accuracy rises over time in the design process.

The need observed both at an academic and industrial level is to define an engineering method that could anticipate as much as possible cost evaluation. The requirement is to be able to start cost evaluations since the conceptual design phase, with an acceptable level of accuracy.

Cost estimation enables to discard or modify unsuitable projects as quickly as possible before significant economic resources have been invested in their implementation. Cost estimation is usually evaluated using modelling techniques during the preliminary stages of the project when detailed costs cannot yet be provided.

The need for a cost estimation method and tool within this framework is even more important in the early stages of product development, because, as known and consolidated by literature studies, there is a "cost paradox" that states that “the initial conceptual design phase takes about 20% of the budget allocated for the development of a new product but the choices made during this phase affect 80% of production costs”.

It is therefore clear that costs must be estimated and optimized as soon as possible, where the degrees of freedom are higher since any variations during the production phase could be very expensive. For the above-mentioned reasons, the research work proposed within this thesis has been defined to pursue this goal.

Parametric estimation techniques have been applied and, in particular, there are two compatible approaches regarding cost estimation: traditional approach and innovative approach. In the first case, the cost was expressed with linear analytical functions called "Cost estimation relationships" (CER - Cost estimation relationship) and constructed through parametric methodology applications. In the second case, the cost was estimated using machine learning techniques which can be neural networks or random forests.

Different models arise from these prediction methods and it is useful to compare them for the same project to determine which of these is more convenient to use; in this way, both the predictive capabilities and the shortcomings of each model are highlighted. Cost estimates are always approximate as they are used when it is not possible to determine a precise cost value yet. It is expected that the value identified in this phase for the assessment of economic feasibility may fluctuate within a range depending on the specific knowledge of the project and the references previously accumulated over time.
Leveraging this method, the overall objective is to evaluate the order of magnitude of the costs to be faced: not being able to carry out this assessment means encountering serious problems in the production phases of a project, such as lack of funds in the case of underestimation, or, in the opposite case, the immobilization of capital that could be used for other purposes. Furthermore, an incorrect cost estimate does not allow to meet the initial project expectations. Therefore, these methods and the models that will be applied are extremely useful tools to decide whether to proceed in one direction or another in the production of a project; or even if it is convenient to carry out the project.

To face all those aspects, the methodology presented will leverage product cost management techniques and in particular, two following main topics will be deeply discussed:

- Parametric Costing Estimation
- VAVE Techniques (Value Analysis Value Engineering)

A systematic procedure to estimate costs of different product architectures (leveraging module analysis) in the early stages of product design (conceptual design) will be described, combining functional decomposition, Value Analysis Value Engineering (VAVE) approach and conceptual costs modelling.
The main aim is to compare different design options taking into account cost parameters since the very conceptual design stage in order to choose the best solution among cost trade-offs and performance requirements.
Functional analysis is used to find out product features and relative aims, VAVE methodology is used to uncover disruptive ideas and design modification proposals, and cost modeling in the conceptual stage is used as a parameter for design decisions to optimize design options even if within a certain uncertainty.

This design methodology enables product cost management (PCM) for gas turbines, helping to reach the configuration cost target.
A software platform will be described, with its architecture and completely integrated with both design and costing tools, to propose a guided process towards the optimized cost and product value design.
The software platform and all the integrated tools enable designers to compare different design alternatives easily and dynamically during the conceptual phase, evaluating several design choices based on the different scenarios.

1.2. Technical and scientific research objectives

One of the major goals of this thesis work is to define a systematic approach to evaluate different architecture options and estimate their cost implications leveraging parametric cost models defined.

The idea is to select parts cost drivers to generate quick cost assessments for each gas turbine's main component. In order to estimate total gas turbine cost, this approach shall be followed for all the components included in the gas turbine bill of material. Moreover, it will be necessary to add the assembly cost contribution, to take into account relative hours' impact on cost.

To do that, it will be important to estimate both components cost and processes cost. In Case Study Section, case studies of the axial compressor and turbine components will be described (blades and discs/spacers).

The goal is to use cost drivers of Gas Turbine parts to generate rapid cost estimations, doing this accurately but also quickly in the very early stages of design when lack of information is strictly related to those design stages.

Parametric Cost Estimates are a traditional cost estimation approach that determines cost by means of statistical relationships between variables (CERs). The main advantage of using a parametric methodology is that the estimation can usually be done quickly and can be easily replicated.

The first objective of this thesis work is the definition of conceptual cost models. Those models are needed to evaluate parts and product costs in the preliminary and detailed stages of design. The starting point of this activity is the definition of the set of data on which models will be built. The dataset could be constituted by historical data of purchased orders placed, if we want to estimate a prediction of real cost, based on the current supply chain, or could be constituted by should cost values if we want to estimate a possible optimized cost for the component on the market. The important aspect is to focus on a dataset that is uniform and referred to a specific timeframe (for example last three years purchase orders). Considering this aspect, it is easy to understand that a crucial role is played by cost model maintenance, as far as the dataset is a dynamic list of values that change over time. In order to define conceptual cost models, regression, neural network, and random forest methodologies have been applied to specific components of gas turbines in the framework of these research activities. To challenge and improve these techniques, also neural network has been taken into account, defining possible improvement in terms of accuracy and also of self-learning.

The second objective of this thesis work is the integration of VAVE (Value Analysis, Value Engineering) technique in the process of product development with the aim of improving the value of the solution proposed increasing ratio between function and cost. This methodology is used especially when target cost is not matched and a deepen analysis is needed in order to reach the desired product margin. This improvement can be reached by both increasing function perceived by the customer (that is the functionality for which the customer is willing to pay) or reducing production cost, in terms of resource expenses, or both the approach mutually.

1.3. Progress beyond the state of the art

In turbomachinery environment it is very hard to find common and universally used approaches to derive product cost models by reading the available literature in this field. The state-of-art is particularly fragmented in this field due

to the product complexity and due to lack of data-driven by confidential reasons.
This research thesis wants to make up for this lack and develops product cost model techniques, such as parametric cost modelling, VAVE techniques applied to Gas Turbine Components, and a related software design tool able to predict the cost of an Architecture configuration starting from the preliminary study in the conceptual stage.
The main novelty of this thesis work is the integration of the method described in the gas turbine development environment, analyzing the cost of the product in the very preliminary design stages and using the limited design information available. In particular, the approach leads to product architecture definition, guided by the cost indicators. The proposed approach has been applied to components and subsystems of a gas turbine, as well as to two complete gas turbine configurations, following all the steps that will be integrated into the tool that enable designers to evaluate different design alternatives easily and dynamically during the early product development phase.
Novelties concerning the current state of art, as anticipated, are the definition of cost models specifically tailored for gas turbine products, as well as the application of VAVE methodology to gas turbine components. Cost model definition allows increasing cost evaluation accuracy while anticipating the timing of the estimation as well. VAVE approach leads to product value optimization since the very early stage of design when usually it is not considered as a design parameter yet.

1.4. Thesis Overview

Following the methodology presented and integrated platform usage, two case studies are described, showing the application of the approach to two different components during the conceptual design of a gas turbine.
The case study section will help to understand the benefits of the methodology with respect to the product cost. Two components are analyzed in terms of conceptual cost: blades and discs of the gas turbine. A conceptual costing model has been built for both elements, which addresses all cost factors and assesses the confidence level in the methodology developed.
Component cost reduction leads to product value optimization. In fact, guaranteeing the same functionality of component, a reduction of cost will lower only the denominator portion of value equation (value = function/cost), and considering that, it will help to increase product value. For sure there are other ways to increase product value, for example, maintaining the same cost but increasing functions absolved by the component, in the next paragraphs all the possible ways to increase product value will be presented.

This thesis is organized as follows:

Section 1 illustrates a brief introduction of the research work and the main developed topics. Context of Gas Turbine Product Development is presented, and the theme of product cost management is introduced. In the following sections, the current state of art will be described more in detail and together with all the lack and limitations that this thesis aims to overcome. Then all the methodologies and methods applied in this thesis work to overcome those limitations will be presented.

Section 2 presents a detailed literature analysis highlighting the limits of the current state of the art, dealing with both parametric cost estimation methods and VAVE (Value Analysis Value Engineering) techniques.

Section 3 In chapter 3 Method is described and in particular Product Cost Management is discussed with a particular focus on Gas Turbines components and processes. The parametric Cost Estimation approach for a gas turbine is presented and all the steps of the VAVE methodology are described and discussed.

Section 4 presents a proposed software platform for both conceptual costing and VAVE analysis, showing a scheme of suggested system architecture, modules, use scenarios, and requirements. For main modules, several tools necessary for methods application are presented, focusing both on Parametric Cost Model Tools (Leancost, Rapid Miner, SPSS) and on VAVE methodology tools.

Section 5 describes, after a brief introduction on Baker Hughes environmental contest of gas turbine, two main cases study. The first one deals with Gas Turbine blades and, because in this case, after the cost evaluation process, the component cost was not aligned to target cost, VAVE methodology has been applied to optimize product value and costs. The second one focuses on axial compressor discs and spacers: for these components all the parametric cost model methodologies have been applied (Regressions, Neural Network, Backpropagation Algorithm, Cross-validation, Random forest, etc.) comparing them in terms of complexity and accuracy of results obtained.

Section 6 discusses the Results of both the Product Cost model and VAVE Methodology and Machine Learning Techniques applied to the two cases study presented. In particular, the first case study is more focused on the entire methodology application, leveraging VAVE analysis technique to optimize product value. The second case study deepens machine learning technique comparison, to understand which one gives better results in terms of predictability and accuracy.

Section 7 presents the Discussion and Concluding remarks of both the Product Cost model and VAVE Methodology and Machin Learning Techniques, underlying the advantages and drawbacks of the proposed methodologies.

Chapter 2.

State of Art

As described in the Introduction Chapter, product cost management, in the wider environment of engineering design, is the main topic of this thesis work. State of art for the following related topics will be deeply analyzed in dedicated paragraphs of this chapter:

- Engineering Design
- Cost Estimation methodologies in engineering design
- VAVE Techniques (Value Analysis Value Engineering)

2.1. Engineering Design

Engineering design is the process to follow to identify and solve problems. It has been described in the literature in many ways, but all descriptions include some common attributes: it shall include plans of action, to link working steps and design phases according to content and organization.
It shall also include strategies, rules, and principles to achieve the most general and specific goals as well as methods to solve specific design problems or sub-tasks.
This powerful approach to problem-solving is flexible enough to work in most situations.

Defining a systematic process to structure the main problem and every single task makes it easier to both follow the process and generalize results.
In literature, one of the most important and detailed studies on design methods of industrial products is proposed by [1]. In [1] four main phases for the product development process are identified:

1. **Planning and Task Clarification:** information collection of all the requirements that must be satisfied by the product, and about the existing constraints and their importance. The output of Task Clarification is the definition of design specifications, collected in a document that is called "product technical specification".

2. **Conceptual Design**: abstraction of the essential problem, definition of function structures, and first proposal of suitable working principles. During conceptual design, a preliminary specification of a principle solution is performed (concept).

3. **Embodiment Design:** definition of the construction structure (overall layout) of a technical system in line with technical and economic criteria. Embodiment Design output is the specification of the product layout.

4. **Detail Design:** final arrangement, geometry, shape, dimensions, and surface properties definition. During this phase all the materials are specified, the production possibilities compared, detailed costs estimated, and all the drawings and other production documents produced. The Detail Design output is the specification of the information in the form of construction drawings.

The main goal of the *Planning and Task Clarification* step is to create new ideas and new concepts both from a technical and non-technical point of view. The design engineering team concentrates on developing new product concepts to face more and more strategic business [2]. Design engineers inject new technology innovations that can be included in product development.
The product concept is defined based on tasks and functional requirements to be met. Design proposals generated at this stage may or may not be based on a historical solution already performed. The product specification list is a crucial starting point for all the next working steps. Product specification shall include both quantitative and qualitative information and is a basic document for any further product development.

In the *Conceptual Design* phase, the design process is recursive and interactive. Industrial engineers often start with solutions drawn from their stored tacit knowledge, heritage from previous design projects, and design education. Ideas are generated individually or during brainstorming group sessions. During this design stage, the designers go through the production of sketches, drawings, mock-ups, or models to test preliminary technical feasibility and to assess proposed production methods, refers to [3] and [4].

The third step in the design process is the *Embodiment Design* stage during which the design is developed. During this step, the main role of core design is to be found. Design engineers shall investigate competitor's products by taking them apart to see how they work and how they are manufactured.
Concepts have been embodied leveraging the use of 2D sketches, layout drawings, CAD models, schematics, and mock-ups. Prototypes are often used to test basic technical principles such as customers' needs, component configurations, and manufacturing process capabilities[3][4]. During this stage, all the needed calculations are performed and material selection finalized.

Moreover, one important technical analysis performed during this step is the refinement of cost-effectiveness.

The last step is the *Detail Design* phase, during which industrial designers and design engineers use manufacturing processes knowledge and material knowledge to make sure that proposed designs are efficient and suitable to be produced, guaranteeing all the requirements such as safety and usability.
One of the goals of this step is to issue all the working drawings, to provide information on the materials selected, tolerances, and manufacturing processes.
Last but not least, the designed product starts the manufacturing process, where design engineers end off with the production team.

As described in the introduction, the industrial approach (the one followed also by BH) mainly follows those steps, tailoring them to the specific products and business needs. Referring to Figure 2 it is possible to see how Planning and Task clarification correspond to Step 1 of Requirement Definition, Conceptual and Embodiment Design are developed in Step 2 of Preliminary and Conceptual design and Detailed Design is performed in Step 3. It is worth noting that these preliminary phases, from task clarification to conceptual design are the boundaries of this research work.

Figure 3 here below describes the main four steps of the discussed engineering design process and summarizes the main activities for each phase [1].

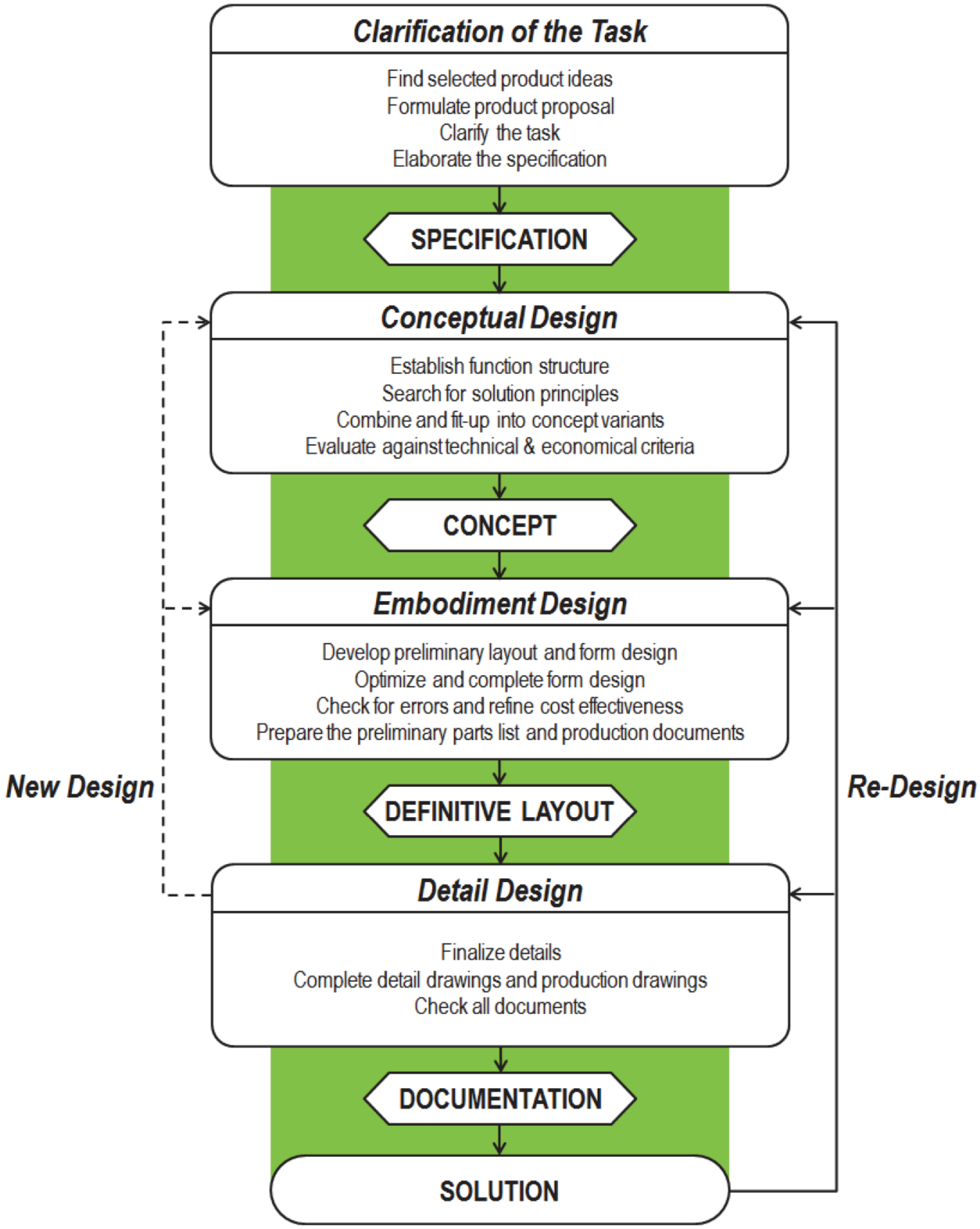


Figure 3. Schematic representation of the main step of product design development.

2.2. Cost Estimation in engineering design

Cost reduction impact is most effective when done in the preliminary phase, with the larger room for design changes and the chance to investigate several alternatives (product architectures). Product architecture definition starts from modules analysis, taking into account preliminary manufacturing evaluation for each item [5]. It is well known that 70% of product cost is committed during the preliminary design stage [6], and consequently, it is very important to evaluate and optimize costs as soon as possible. All changes in later times will have an impact in terms of re-design and cost of procurement. Literature analysis shows that two main aspects were addressed for answering the issue of cost control on each product production step: (i) functional and modular analysis linked to architectures definition [7][8], and (ii) cost evaluation models with proper accuracy level estimation[9][10].

From the literature, it is important to underline that ability to reduce cost decreases while cost accuracy increases with ongoing time in the design process as described in Figure 4.

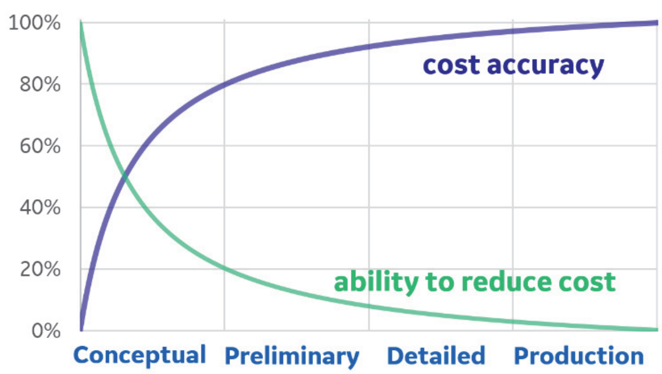


Figure 4. Impact on Cost Vs. NPI program stage - © 2020 Baker Hughes, LLC - All rights reserved.

It is important to anticipate as much as possible cost evaluation in the early stages of product development, because, as known and consolidated by literature studies, "cost paradox" affirms that “the initial conceptual design phase takes about 20% of the budget allocated for the development of a new product but the choices made during this phase affect 80% of production costs”. (see figure below)[11].

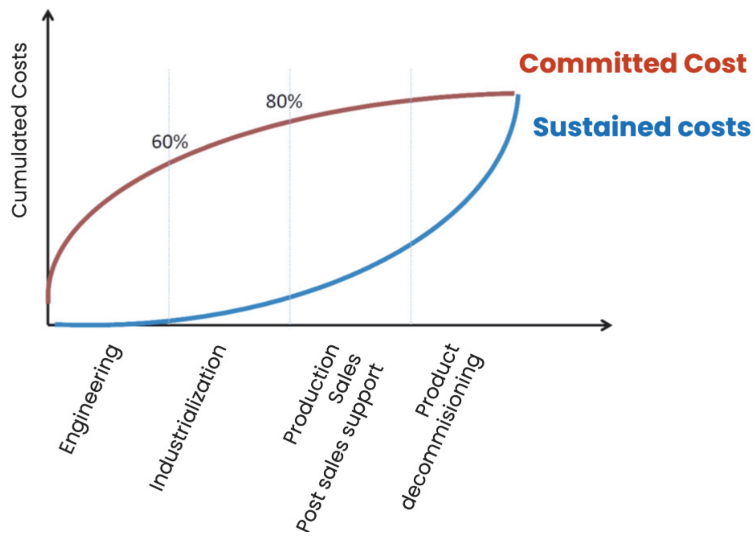


Figure 5. Cost Paradox -

Many design methodologies have been discussed during the last years to manage cost evaluation since the very preliminary design phases. All those methodologies take into account the adoption of functional analysis as part of Value Analysis Value Engineering (VAVE)[8], TRIZ (Theory of Inventive Problem Solving) [12], Quality Function Deployment (QFD) [13], System Engineering [7] and a few others. The mentioned methodologies present both advantages and drawbacks depending on the context of use and product development complexity. In the field of the gas turbine, being a complex product composed of thousands of parts and involving different manufacturing processes, many trials have been performed to apply the above methodologies, but there are some limitations to face.

The complexity of the product, the management of the supply chain, the intellectual property of its design, and many other aspects are just some of the factors that influence the analysis of the product at this early stage of technical design. As stated in these research papers [14][15][16][17], a systematic approach combining different design methods (concurrent design) would be useful to support the decision-making process involving both the entire gas turbine and all its sub-components.

An aspect of modular decomposition based on functional analysis is linked to several design methods, developed in the last years to facilitate engineering module definition and analysis. Functional analysis helps to define a general product framework to guide design choices. Functional analysis [1] is used for function identification and decomposition for the product since the conceptual level and generally, it is used together with the module heuristics approach to describe functional modules [18][19]. Functional modules help to understand product architecture and its relative interfaces, to meet project requirements while reducing total cost [20][14]. Functional analysis is proficiently used in several areas, for products and processes, with a systematic and guided process, with a multidisciplinary team involved. Theory of Inventive Problem Solving (TRIZ) helps in idea generation, breaking conventions, and introducing novelties [12]. On one hand, TRIZ helps to break conventional designers' mindsets [21], but on the other hand, it is a not repeatable procedure and so has some limitations such as lack of repeatability. It is worth to underline that TRIZ, as described in [21], presents another limitation: it is very useful while focusing on products, but has lower applicability of processes and services.

QFD methodology is customer-oriented and involves many aspects (such as costs, quality, reliability) describing them qualitatively and consequently without a detailed numerical analysis of cost [13]. QFD assured quality in the production

phase [22] but has more focus on planning than on design [22] and it does not deal with cost aspects [13]. System engineering [7] focuses on requirements and it is a rigorous approach that needs a not negligible time frame to be completed and a certified expert to be sure it is correctly applied. System engineer permits to vary the point of view to improve understanding [7] and to take into account not only short-term but also long-term, consequences of decisions [23].
CTOC methodology (Converter-Transmitter-Operator-Control) is a systematic process that aims to assure the correct translation from function to physical architecture [24]. It uses mathematical equations to characterize physical events as described in [25]. A drawback of CTOC methods consists in the fact that has no applicability for systems not described by energy flows [24][25].
VAVE method has system analysis in common with CTOC method; but helps to address design solutions and cost estimation to the specific requirement of the product, similar to the System Engineering approach [26][27][28][29]. VAVE methodology presents advantages of a multidisciplinary team involvement but as a drawback, it needs certified experts to be sure it is correctly executed [23][7]. VAVE approach is a combination of two separate aspects: Value Engineering and Value Analysis [8].

Dealing with Cost Estimation Techniques, literature presents a lot of design methodologies and tools to solve the cost estimation problem [30][31][32][33][34]. Cost evaluation is the design process focused on product cost estimation even before the design is completed and production is started [15].
There are several methods for estimating costs and they can be grouped into two main families [35]:

1. Qualitative methods: they are mainly based on a comparative analysis of a new product and an existing one;

2. Quantitative methods: they are based on a detailed analysis of the design of a product, including its characteristics and the corresponding production processes.

Among the existing methods for cost estimation, quantitative methods are the most suitable choice for product costing during the design phase, as they provide better results in terms of reliability (Figure 6).

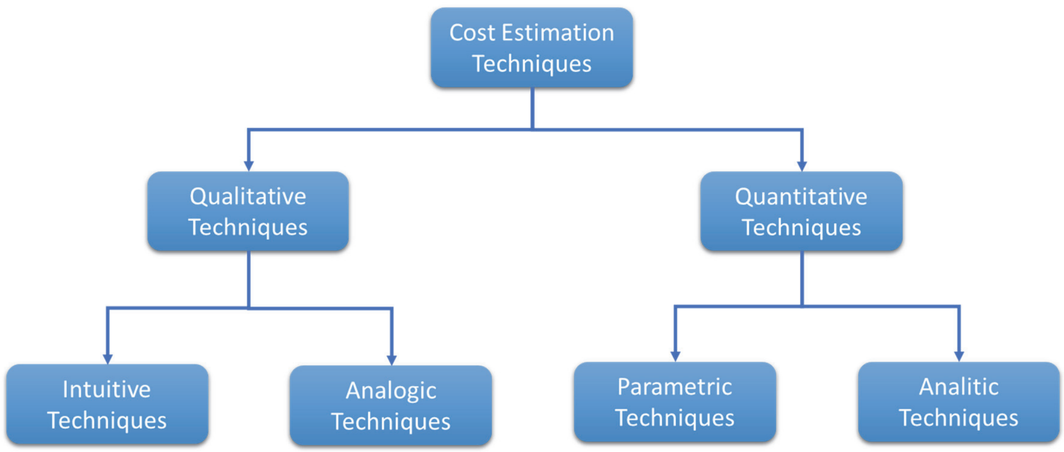


Figure 6. Cost estimation methods -

Cost evaluation starts with the definition of a mathematical model that takes into account all the aspects of cost [36]. Strategies utilized in literature for cost assessment can be recognized: (a) qualitative methods, which computes item cost as for already developed design and, (b) quantitative techniques, which gauges item based on design properties and production process phases [37][38][39]. Quantitative techniques are the favored decision for item costs assessment during the conceptual stage regarding precision and accuracy [15]. Fundamental points of interest of qualitative techniques are the chance to apply them from the preliminary phase of the project and the chance to apply them also in case of not linear problems, on the other side, the main drawback is the lower accuracy with respect to quantitative methods. Quantitative methods need detailed data and present a more elevated level of complexity, being developed to reach a higher accuracy level and to recognize all the cost drivers.

Qualitative methodologies present lower accuracy and thus they are not preferable to quantitative methods. Moreover, it is almost impossible to satisfy the need of reaching an estimation gap lower than 10% with a method that presents an accuracy of 30%. On the contrary, quantitative methods could eventually satisfy the accuracy requirements, but only in case, they receive proper inputs starting from the preliminary phase. Unfortunately, not all the inputs are available since the preliminary stage of design, and consequently, it is not possible to apply those methods without proper customization. Quantitative methodologies do not permit to satisfy the goal of the work: the need to identify a minimum set of cost drivers to guarantee cost prediction with a good level of accuracy and verify the cost estimation and the target cost is aligned with a maximum gap of 10%. Moreover, cost models available have general-purpose and cannot be used for the specific field of gas turbine items: in fact, gas turbine design had its own specific characteristics and technologies involved, and this causes the need for a tuned costing model.
For what concerns the previously described topics, few proposals to couple functional analysis with cost evaluation of modules have been presented. Saravi et al. [34] leveraged the Design of Experiments (DoE) technique to manage information in an efficient way to reach product cost estimation goals. DoE permits reach optimized design alternative with a food confidence level in cost evaluation. Also manufacturing cycle cost is analyzed with a particular focus on specific manufacturing technology, as described in [40][41][42][43][44], and total life cycle impact is more and more frequently taken into account in the whole analysis, as explained by [45][46]. Dealing with all these methodologies,

parametric cost modeling seems to be a robust practice for the definition of cost centers for specific manufacturing technologies, being also able to reach an acceptable gap between estimated and current cost, in the order of 10-15% [47][48][49].

2.2.1. Parametric Techniques

The primary method by which parametric cost estimation is developed is regression. This method attempts to establish the nature of the relationship between variables by providing a prediction mechanism. There are two different types of variables: dependent designated by the symbol y and independent designated by the symbol x. The variable y represents a type of cost, while the variables x represents various parameters of the system.
Regression, therefore, is a branch of applied statistics that allows to quantify the relationship between the dependent variable and one or more independent variables and to describe the accuracy of this relationship. Although it may be non-linear, only linear forms of regression (simple and multiple) have been analyzed. While staying in the linearity field may seem like an oversimplification of the problem, there are several good reasons to take this approach.
Costs should logically, and often do, vary linearly with most physical and performance characteristics. Additionally, many curvilinear and exponential functions can be transformed into a linear form, thus lending themselves to linear analysis.

2.2.1.1. Regressions

Simple linear regression occurs when only one independent variable is needed to estimate the value of y. The variable y is related to the variable x by the following expression:

$$y = \beta_0 + \beta_1 x \quad (1)$$

Where:

β_0 = Intercept

β_1 = Angular coefficient

X= independent variable

Y= dependent variable

Coefficients $\beta 1$ and $\beta 2$ can be calculated graphically by interpolation (Figure 7).

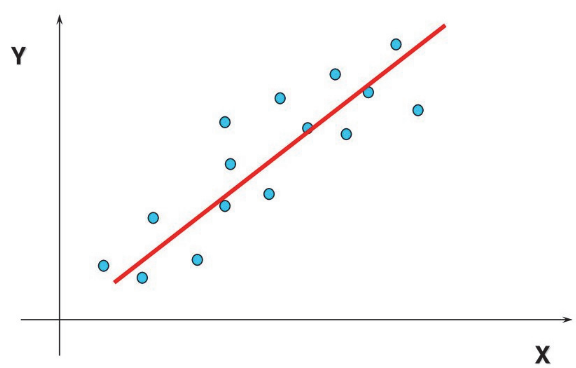


Figure 7 – Simple Linear Regression

Multiple linear regression is characterized by n dependent variables. The equation will take the following form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (2)$$

Where

β_0 = Intercept

β_1 = Angular coefficient for x_1

β_2 = Angular coefficient for x_2

X= independent variable

Y= dependent variable

Coefficients β_1 and β_2 can be calculated mathematically.

For both simple linear and multiple linear regression, the coefficient of determination R^2 allows measurement of the accuracy of equations. It can have a value between -1 and 1. In Figure 8 it is clearly seen that the higher the number, the better the fit of the regression line to the actual data, refer also to [50].

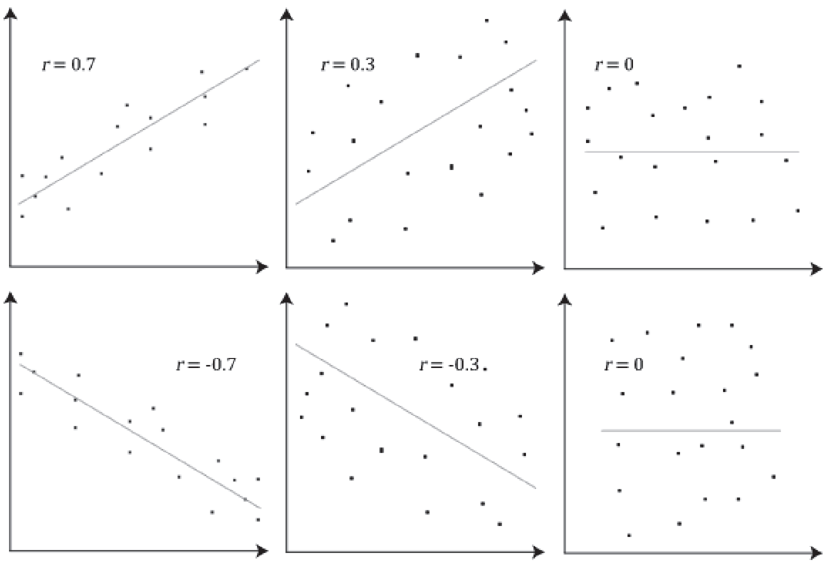


Figure 8 - R^2 Coefficient behaviour

2.2.1.2. Neural Network

Neural networks, defined by the acronym ANN (Artificial Neural Networks), are an innovative machine learning approach that allows you to develop complex computational problems quickly. They are inspired by the functionality of the human brain and its structure, which can be represented as a network of interconnected neurons and able to accumulate knowledge over time in a "distributed" way: the information is encoded through electrical impulses in neurons and is stored by modifying the molecular and physical structure of the connections. Similarly, the ANN is made up of many "units" called neurons. Considering the most common network architecture ("Multilayer Perceptron") (Figure 9), neurons are arranged in successive layers: each neuron is typically connected to all neurons of the next layer via weighted connections or synapses. A connection is a numerical value (the "weight" in fact). Three main layers can be identified:

- Input layer: consisting of neurons that represent the source of the data;
- Output layer: consisting of neurons that produce the final response of the network;
- Hidden layer: It can be more than one and contains hidden neurons.

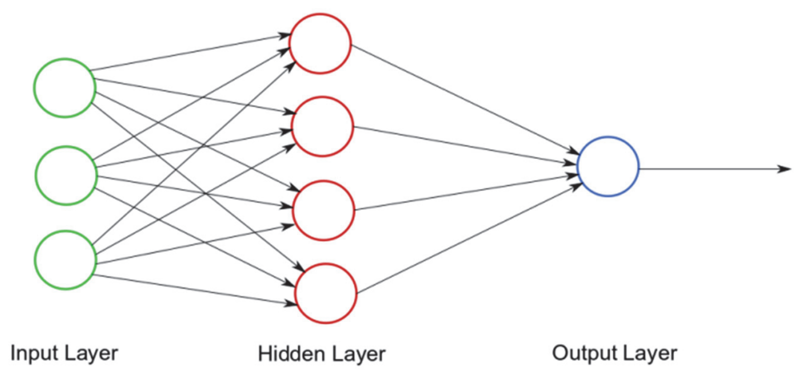


Figure 9 – Neural Network Architecture

Each neuron in the middle layer adds the weighted values of all neurons connected to it and adds a bias value. An activation function is applied to this result which mathematically transforms the value before moving on to the next layer. In this way, the input values are propagated through the network until the output is returned. Beyond the structure, another similarity between the neural

network and the human brain is the ability to learn. An ANN needs to be trained to function properly, that is, the user is required to store an increasing amount of detailed data and information that can be trained by the neural network (these data are called "patterns"). This operation allows to determine the final structure of the network and therefore to determine the values of the characteristic parameters (weights and bias).
From what has been said, a network can be seen as a closed box system capable of providing an output for a given input.

Three main aspects describe and characterize an ANN:

1. Network architecture: defined by the number of neurons for each layer, the number of intermediate layers, and the connections between neurons;
2. Type of activation function: typically, non-linear (sigmoidal function or hyperbolic tangent function)
3. Learning mode: it can be of the supervised type where the weights are modified based on the error made by the network concerning the real output data (this is possible if input-output pairs are known) or it can be of the unsupervised type where the weights vary during the learning process according to a rule defined a priori that does not use the error with respect to the actual data.

Among the various supervised learning techniques, the most used is the **Backpropagation algorithm**, BP [51]. The goal of this technique is to determine the intensity of the connections between the nodes and therefore the values of the characteristic parameters (weights and bias). This algorithm consists of two phases: forward (forward) and backward (backward). In particular, in the forward phase, the learning patterns (pairs of input-output data) are presented to the input nodes. The response of the network develops along with the intermediate levels up to the output nodes. During this phase, the weights and biases remain unchanged. In the backward phase, the error existing between the real process and the network response is calculated and propagated backward through the nodes. Through appropriate updating formulas, weights and biases are modified up to the first level of the network (the input level).

The updating of the characteristic parameters during the training can be seen as an optimization problem of an objective function aimed at minimizing the average error made on the set of learning patterns. A typical problem of this technique is the problem of overfitting (also called overlearning): it occurs when the network learns excellently the response to the input-output patterns, however losing the ability to generalize and respond to input data not yet experienced.

The learning process ends with a final validation phase where, using different patterns with respect to the learning ones, it is checked whether the algorithm provides satisfactory results.

Cross-validation is a statistical technique used in machine learning to eliminate the problem of overfitting in training sets. In particular, the so-called k-fold cross-validation is developed as follows:

1. Separate input data into k subsets of data (also known as folds);
2. Train the model on all but one subset (k-1) and then evaluate the model on the subset that was not used for training. This means that each time one of the subsets k is used as a test set, while the other subsets k-1 are put together to form a training set (Figure 10);
3. Process is repeated k times, each time with a different subset reserved for evaluation (and excluded from training).

Running a k-fold generates k machine learning models, k training data sources and k testing data sources. For each model, a performance parameter is generated and averaging the overall performance.

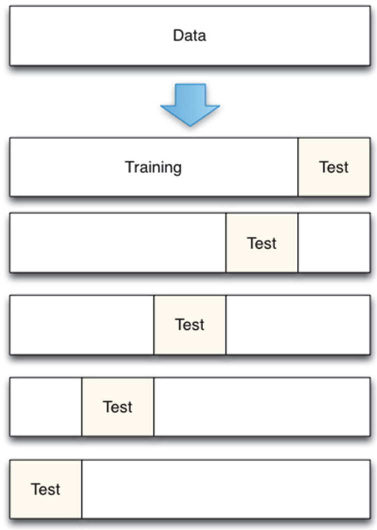


Figure 10 – Cross Validation

2.2.1.3. Random Forest

Random Forest is a very popular machine learning algorithm for its simplicity, ease of use, and interpretability; is a supervised learning methodology for ensembles that belongs to the decision tree family of algorithms. A decision tree represents a classification or regression model in a tree structure. Each node in the tree structure represents a particular "question" about a feature, each branch a decision, and each leaf at the end of a branch the corresponding output value (Figure 11).

To obtain a result, starting from a specific input, the decision process starts from the root node (at the top) and runs through the tree until it reaches a leaf that contains the result. In each node, the path to follow depends on the values assumed by the various features. Similar to neural networks, the tree is created through a learning process using training data.

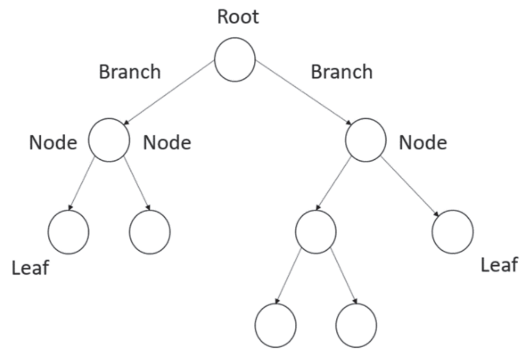


Figure 11 – Decision Tree

Random Forest consists of a set of different decision trees but with the same source of training data. Each decision tree is created from a different and random subset of the training data. This subdivision allows to solve a typical problem: if a single decision tree were created for the entire set of training data, one would encounter a prediction model with poor reliability due to the presence of over fittings on the training-set and high variance.
Instead, the Random Forest is made up of n weak trees but which on the whole build a set of "decision-makers" in which the collective result can be created with the majority vote in the case of classification or average in the case of regression (Figure 12).

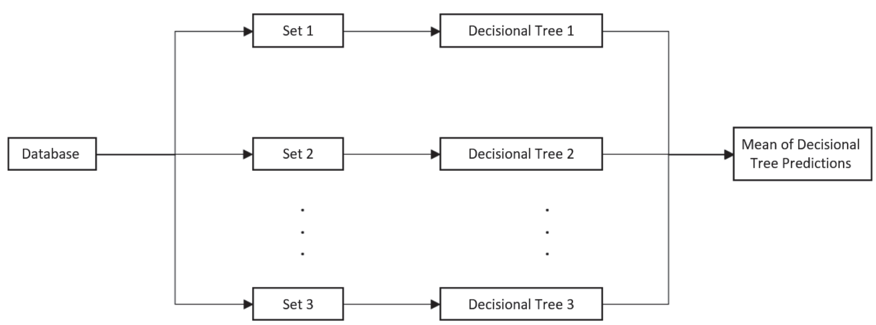


Figure 12 – Random Forest

This compensates for potential errors of individual trees so that variance is reduced, and the model is more accurate.

2.3. VAVE Techniques (Value Analysis Value Engineering)

Value engineering started at General Electric Co. during the second world war. Because of the war, there were shortages of skilled labor, raw materials, and component parts. Lawrence Miles, Jerry Leftow, and Harry Erlicher looked for acceptable substitutes and noticed that these alternatives often reduced costs, improved the product value, or both. What started as a necessity was turned into a systematic process, and they called their technique "value analysis" or “VAVE”. To focus on the component functions, they used a two-word statement composed of an active verb and a measurable noun to describe the goal that a part’s function provides.

A strategy in developing new products is to think about the specific functions the product should deliver to fulfill customer needs. For example, the function of an electric motor is to “generate torque” while the function of a gearbox is “convert torque”. Functional approach and modelling are key steps in the product design process, both in the case of a new design of or redesigned products. The big advantage of the functional approach is the chance to assess the new product behavior in the *Conceptual Design* phase when the products are conceived by designers.

Moreover, functional analysis is a systematic and repeatable approach that can be applied to product design and analysis.

The product function model helps to correlate design decisions to the customer’s needs. In particular, the functional approach to reach customer’s needs and to define functional specifications is called the “function-based modular design” [52].

After function identifications, the aim is to search for other ways or methods to achieve the same benefit of that intended function. Starting from their research, function analysis, the key foundation of value methodologies, has been developed and is now an important tool to help engineers and working teams to manage the way a concept is understood.

To identify value opportunities and develop innovation, both private industry and government agencies have been using a process known as the Vale Analysis Value Methodology (VAVE), refer also to [52].

VAVE process helps optimize projects, processes and product development in many different and significant ways. It aims to reach a decrease in costs, an increase in profits, a general improvement of quality and performance, and as a consequence of all these aspects, results in enhancing customer satisfaction.

Value Engineering is function-based method useful to identify all the functions for which the component is designed for and it permits correctly allocate the cost of the part to its functions.

Value Analysis is useful to receive feedback on unnecessary costs, those linked to secondary and indirect functions, that can be removed. It also helps to identify necessary costs, that are the ones linked to primary functions, and that needs to be increased as soon as a draft design is available.

As per [7] and also [8], according to European Standard EN16271:2013 [53] and EN1325:2014 [54], value is defined as the "reliable performance which a product or process must achieve to make it work and sell at the least practical cost", and product value can be defined as follows:

$$Product\ Value = \frac{Requirements\ Satisfaction}{Product\ Cost} = \frac{Function}{Cost} \tag{3}$$

Where “Requirement Satisfaction” refers to performance requirements of the customer and “Product Cost” to all the cost related to raw materials, labor, time, etc, as described in [23].

To numerically estimate product value as explained in [23], it is possible to further analyze the above equation as follows:

$$Product\ Value = \frac{Requirements\ Satisfaction}{Product\ Cost} = \frac{\prod_i 1 - \frac{|Requirement\ Value_i - Design\ Value_i|}{Requirement\ Value_i}}{\frac{Estimated\ Cost}{Target}} \tag{4}$$

Where difference from each product requirement is measured through the multiplication of distance between each single design value “i” from its associated requirement value “i” and product cost is compared to target cost. Product Value calculated in the equation above can be evaluated at different steps of project design to measure improvements on value-driven by redesign proposals.

As described in [23] “The Value Methodology (VAVE) is a systematic and structured approach for improving projects, products and processes. VAVE, which is also known as value engineering, is used to analyze and improve manufacturing products and processes, design and construction projects, and business and administrative processes.”

VAVE helps achieve an optimized balance between function, performance, quality, safety and cost. The proper balance results in the maximum value for the project.

VAVE follows a standard job plan, which consists of the following phases:

- **Information phase:** Gather information to better understand the project. The team analyzes and reviews the current design conditions of the project and identifies the goals of the redesign.
- **Function Analysis phase:** Analyze the project to understand and clarify the required functions. The team identifies all the product functions using a two-word active verb and measurable noun. The team studies and analyzes those functions to establish which one needs improvement, which others can be eliminated or created to guarantee product requirements satisfaction.
- **Creative phase:** Generate ideas on all the possible ways to accomplish the required functions. The team leverages creative techniques to identify and propose alternative ways to perform the product's function(s).
- **Evaluation phase:** Synthesize ideas and concepts and select those that are feasible for development into specific value improvements. The team performs a structured evaluation process, leveraging engineering best judgment, to select the most promising ideas with the potential for value increase maintaining the product's function(s) requirement absolved.
- **Development phase:** Select and prepare the 'best' alternative(s) for improving value. The team generates from the selected ideas one or more than one alternative proposal, collecting a sufficient level of documentation to allow management to decide if proceed with the implementation of proposals or not.

The team shall produce a report and a presentation to document the adequacy of the proposals developed as well as an estimation of the associated value improvement opportunity.

To qualify as a VAVE Study, the following conditions must be met:

- The VAVE Study Team follows a systematic and organized Job Plan that includes the five phases identified in this standard plus presentation phase. Function Analysis shall be performed during the study, by the team members.

- The VAVE Study Team is composed of a multidisciplinary group of experienced professionals and project stakeholders. Team members are selected based on their expertise and experience with the product. Individuals who have relevant expertise; but are not directly involved with the product or project can be invited as well, to provide a different and innovative point of view.

- The VAVE Team Leader shall be trained in VAVE methodology techniques and shall be qualified to lead a study team using the Job Plan.

As anticipated, the best results are achieved by a multi-disciplined team with experience and expertise relevant to the project being studied. A Certified Value Specialist (CVS) usually leads the team to ensure the VAVE methodology is properly followed.

VAVE methodologies can be followed at any stage of a project's development cycle, although the greatest benefit and resource savings can be achieved in case of the application early in development, during conceptual design stages.
In fact, during the conceptual stage, the basic product information is available, but major design and development have not yet been completed.
This constitutes the best time to apply VAVE methodology because how the basic function of the product is performed has not been established yet. This for sure gives room for improvement and alternative and better ways may be identified and considered.
VAVE methodologies can be applied also more than once during the life of the product. Early application of a VAVE methodology helps to get the product design initiated in the right direction, while repeating applications leads to refinement of product design, based on new or changing information.
Anyhow, in the industries framework, due to time constraints, VAVE application is often limited to one single time. The later a VAVE Study is performed during product development, the higher implementation costs will be.
VAVE methodology may be used as a quick response study to assess a specific problem or can be seen as an integral part of an overall organizational effort to improve performance characteristics leveraging on innovation stimulation.
VAVE methodologies can be applied to enhance not only product design but also processes, services and quality programs.

2.4. Advantages and drawbacks of each design method

Table 1 below represents a summary of all advantages and drawbacks of each design method. In particular VAVE methodology leverages a multidisciplinary team involvement but as a drawback, it is necessary to have certified experts to guarantee it is correctly executed [23][7]. To overcome the need for certified experts, BH trained their own experts to follow VAVE workshops.

VAVE methodology turned out to be the most promising method to optimize the product value of gas turbine component's design.

Table 1: Advantages and Drawbacks of design methods related to functional analysis - © 2020 Baker Hughes, LLC - All rights reserved.

Functional decomposition methodologies	**Advantages**	**Drawbacks**
Theory of Inventive Problem Solving	Allows Creative/Disruptive idea generation and provides intuitive means to clarify solution space [10] Breaks designers' conventional mind-set [25]	No repeatable process/solution [10] Lower focus on service application with respect to tangible context. [25]
Quality Function Deployment (QFD)	Customer Oriented [11] Guarantees quality assurance during production phase [26]	Not specifically focused on cost [11] More focused on planning than on design [26]
CTOC	Systematic procedure that guarantees transition from function to physical architecture [28] Applies mathematical formulas to describe physical phenomena [29]	Not optimized for modelling systems without energy flows [28][29]
System Engineering	Requirements oriented [9] Change perspective to increase understanding [12] Consider both short-term and long-term consequences of actions [27]	Need System certified experts to be executed [12] Long time to be developed [9]
VAVE	Cost included in analysis Creative/Disruptive idea generation [9] Multidisciplinary approach [27] Systematic process [12]	Long time to be developed [9] Need VAVE certified experts to be executed[12]

Focal points and Drawbacks of Cost Estimation Techniques are summed up in Table 2.

Table 2: Advantages and Drawbacks of design methods for cost estimation - © 2020 Baker Hughes, LLC - All rights reserved.

Cost Estimation methodologies		**Advantages**	**Drawbacks**
Qualitative Methods	Intuitive techniques	Applicable from preliminary stages, Innovative design approach [1]	Complex development [27] Accuracy not evaluable cause the results are always dependent on the estimator's knowledge [41]
	Analogical techniques	Applicable from conceptual design stage, fits uncertain and non-linear problems [1]	Accuracy lower than 30% [41][42]
Quantitative Methods	Parametric techniques	Cost driver identification [1]	Complex development [1]
	Analytic techniques	Easier method [1]	Require detailed design information [1]

All the described methods permit completion of the cost assessments in the preliminary phases of product design, but no one of those finds a practical and complete integration in gas turbine design development. Find a proper integration is complex in the case of a product of a high level of technology such as a gas turbine, which is made up of so many components with very demanding requirements in terms of safety, quality, reliability, etc. All those aspects contribute to build the product "value" that needs to be optimized.

Parametric cost modeling is the most robust practice for cost estimation of manufactured components and it is able to reach an acceptable accuracy of estimated cost versus current cost, in the order of 10-15%, that is why it has been selected to perform cost estimation of gas turbine components.

Chapter 3.

Materials and Methods

3.1. Product Cost Management for gas turbine components and processes

In this thesis work, product cost management will be presented, and two following topics will be deeply discussed:

Conceptual Costing Evaluation: through conceptual cost models definition it is possible to estimate the product cost in the preliminary and detailed stage of design.

VAVE Techniques: To improve the value of the solution proposed increasing the ratio between function and cost. It is used especially when target cost is not matched and a deepen analysis is needed.

In particular, those two topics are strictly connected by function analysis. In fact, during conceptual costing, not only parametric cost evaluation is performed, but also function definitions. Conceptual costing estimation has not a single value output, but a detailed cost split by functions.

This constitutes a sort of pre-work for VAVE analysis, for which functional analysis is one of the main tasks to be performed and that aims to assign to each function the correct percentage of total component cost associated.

As shown in figure "Product development process", gas turbine product development follows three main steps: Requirements definition, Preliminary and Conceptual Design, Detailed Design. Those steps are aligned with the Engineering Design Process described by Haik and Shahin in [55] and with the systematic approach to engineering design described by Beitz and Feldhusen in [56]. Both of them underline the importance of the requirement definition phase at the very beginning of the design process and defines conceptual, embodiment and detailed phases. In particular, the embodiment phase is called the “preliminary” phase in this thesis work, according to company common practice. [55] also discuss value theme and definition, while [55], during the phase of “developing concepts” is near to VAVE methodology process and in particular to Steps 2.3, 2.4 and 2.5 of following “flow chart” scheme (Figure 13).

The described method is strongly linked to the standard product development process (PDP) described in [1] and it concentrates on the starting design phase. The whole picture of the methodology is described in Figure 13 (flowchart).

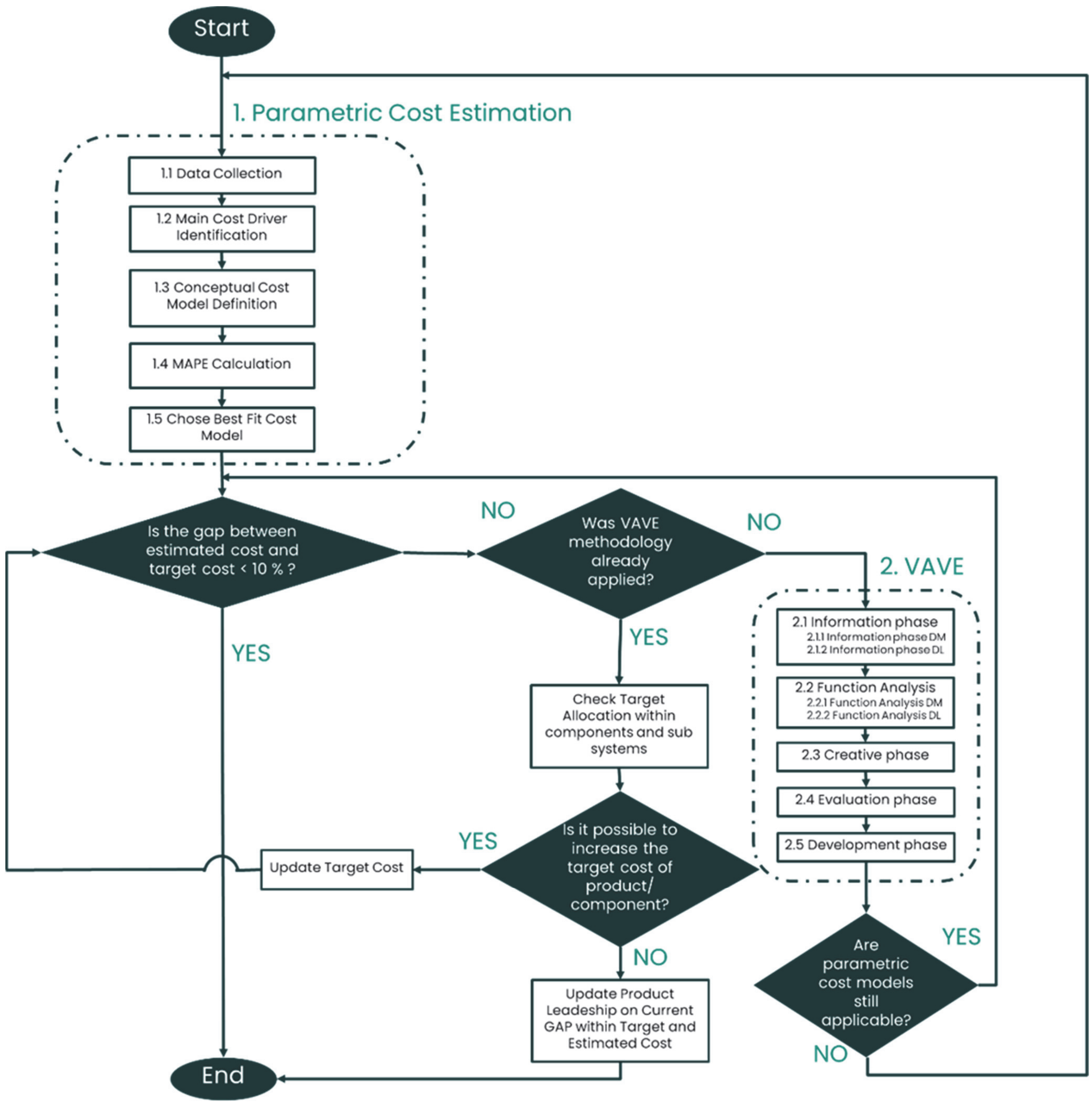


Figure 13 - Flow chart decision process – © 2020 Baker Hughes, LLC - All rights reserved

This Flowchart is composed of two main processes: one is "Parametric Cost Estimation" while the other "VAVE methodology".

It is important to underline that "VAVE methodology" stream is followed only in case "Parametric Cost Estimation" shows an important gap between estimated and target cost (in particular if the gap is higher than 10%, refers also to [56]and [57]).

"Parametric Cost Estimation" is characterized by five steps explained in the following paragraphs. This approach is useful to obtain a cost evaluation using cost models (based on regressions or artificial intelligence and neural networks) and leveraging knowledge and data from company expertise.

After parametric cost models are defined, the fitting of data is measured through MAPE value calculation.

As said, once product cost evaluation is completed, the estimated value is compared to target cost, defined on a market requirements basis. In case cost evaluation presents a gap lower than 10% with respect to target cost, it is possible to continue design stages and discuss the proposed design during a formal design review, where senior engineers of the company analyze both performances and cost parameters of the proposed product architecture.

On the contrary, in case the gap is higher than 10%, the proposed design shall be deeper analyzed to find a balance between performances and product cost.

In this specific case, the second process, "VAVE methodology", is used to help the design team optimizing product value. The aim is to find alternative disruptive solutions to the current design, through design to cost.

In particular, in this thesis work, a slightly different version of VAVE steps, with respect to current standard methodology, is presented, with the aim to analyze gas turbine components leveraging functional analysis. VAVE process is composed of five steps, applying which it will be possible to find new and alternative concepts to the current solution and introduce new technologies with the main scope of reducing cost and optimizing product value.

After completing VAVE steps, some alternative designs are proposed to the whole team. Before proceeding, it is important to validate which one is the most promising in terms of both performances and cost. To do that, a new cost estimation using conceptual cost models is used, and, in case of need, it could be also necessary to define new parametric costing models (in particular in case the new design is very far from the current solution).

In case the new design is so disruptive with respect to the current design, new dedicated costing models will be defined following steps of Process 1 "Parametric Cost Estimation" and associated cost will be estimated. If the VAVE application is successful, the new estimated cost will be within a range of 10% with respect to target cost, and in this case, it is possible to proceed with all the formal steps needed to approve the proposed design and finalize it. On the contrary, if despite the application of VAVE methodology, the gap is still bigger than 10%, it can be necessary to control and eventually update target cost. Target allocation is a System Engineer activity and, under their guidance, it is possible to check if there is room to increase specific target cost, for sure maintaining total target cost that, as already explained, is set by the market and not by the design team nor the company itself, refer also to [58]. If it is not possible to increase component target cost (in case there are no other components that can absorb a target decrease to maintain total target cost flat), the team shall notify immediately to product leader that the target is not reached on the specific component, explaining which is the gap with target foreseen.

Considering Figure 13 and the flowchart here represented, a detailed presentation of the parametric cost estimation model will be discussed in section 3.2 and VAVE methodology will be presented in section 3.3, describing both standard and modified methodologies, with the particular focus on gas turbine products.

3.2. Parametric Cost Estimation

In the case of complex components, and for sure it is the scenario of a gas turbine, production, manufacturing, and assembly is performed both inside the company but also in collaboration with many partners (suppliers) that provide one, or more than one, semi-finished component or subsystem, as described also in [59].

Cost driver definition is completed by VAVE engineer, which leverages both internal know-how and the supply chain capability and expertise, to understand all those parameters able to impact on analyzed components cost. During this activity, parametric cost models are widely applied together with a breakdown analysis of the current costs of already available configurations.

As per the current procedure, the aim is to select the smallest set of cost drivers needed to evaluate cost with a good level of accuracy (an acceptable level is to have an accuracy lower than 10%). With those drives, VAVE engineer will be able to evaluate total cost at each stage of the program and consequently also in very preliminary stages.

The aim of the "parametric cost estimation" process is to use cost drivers of gas turbine items to quickly complete a VAVE analysis, evaluating with acceptable accuracy the total cost, despite the lower number of information.

It is well known that lack of information is the main roadblock to cost estimation at the early stage of the project, as described also in [60][61][62]. To face this difficulty, the parametric cost estimation method is applied to any part or sub-system of a gas turbine, enabling evaluation of total product cost.

To estimate the cost of a component, more than one contribution shall be considered. In particular, raw material cost, the so-called direct material (DM) is one of the major contributors to cost, together with all the cost associated with production steps needed to obtain final geometry, starting from raw material one, the so-called direct labor (DL). On top of that, considering system cost, also assembly hours shall be counted in total cost estimation, refer also to [63].

Parametric costing estimation shall be completed for each main component and each module present inside a gas turbine, leveraging an internal database for input parameters.

The database will support cost estimation of different design alternatives, comparing different architectures in terms of cost. All the inputs needed are taken from geometrical features or manually identified by the design engineer and VAVE engineer. Filling conceptual costing models with all those inputs it is possible to estimate part cost and value. All the process steps of the "parametric cost estimation" process are described in Figure 13 and explained in details below:

Step 1.1 – Data Collection

All the information available from past projects is clustered based on the commonality of product families. This is the very first activity of the "parametric cost estimation" process and it is needed to have a data storage that can be based on a set of already placed orders (PO) or a set of Should Cost analysis (SC). All those data will be the basis for future developed parametric cost models, and they need to be put together and classified in terms of cost items.

It is important to underline that the final value of purchase order is the results of many contributions to cost, each one based on its cost drivers and considering that, structured storage of data from orders is carried out taking into account the most significant parameters acting as cost drivers.

At the same time, should cost value is the results of some analysis assumptions, that shall be verified with the design engineering team and manufacturing team, refer also to [64] and [65].

All those data are both collected and classified during this first step and in parallel, a preliminary identification of cost drivers is performed.

It is necessary to underline that each one of the data of this storage brings its peculiarity that can influence the data value (for example market fluctuation, raw material tends, purchases of batch volume, fees paid to lower procurement lead time). All those characteristics are registered to properly evaluate them during the next process steps. Database and cost model maintenance is necessary over time and needs to be based on new data available, considering new orders placed, new quotation shared by suppliers or new should cost analysis performed.

Step 1.2 – Main Cost Driver Identification

A preliminary proposal of the cost model is made during this stage, in parallel with a preliminary definition of main cost drivers. Cost drivers are selected starting from main geometrical features and process parameters for specific components analyzed. As soon as each cost driver is identified, relative cost contribution is calculated and taken away from the total cost of Step 1.1 (data collection) to continue with other conceptual costing models contributions definition.

Generally speaking, the very first contribution identified is the cost associated with the material, and only after that, all the cost associated with machining processes, coating, and treatments if present, destructive and non-destructive tests, and controls.

With all those contributions to the total cost, it will be possible to both define the cost drivers for the specific item analyzed and the conceptual costing model.

The goal of this process step is to identify correlation among cost drivers and to understand which cost drivers are general and applicable for more than one component and in more than one situation.

Sometimes a cluster refinement may be necessary: for example, in case of low data points presents it can be preferable to add them to other similar items that have the same material type or a similar production process. Or in other scenarios, a bigger cluster can be split into two clusters, each one that can be owner separately to identify its specific set of cost drivers.

Step 1.3 – Conceptual Cost Model Definition

Considering identified cost drivers, parametric costing models are defined based on dataset defined, that can be composed by data from purchased orders or by data from the should-cost analysis, clustered as described in the previous Step 1.2.

The cost model aims to show how cost varies based on the cost driver's fluctuation. To find a simple but also effective correlation, the costing models presented will consider one cost driver variation at a time.

For each set of data and all the cost drivers, different parametric models will be proposed, and then compared in the next Step 1.4.

In particular, conceptual cost models can be categorized as follows:

- Models based on regression
    - Linear regression
    - Quadratic regression
    - Cubic regression
    - n-order polynomial regression
- Model based on Neural Network
    - Backpropagation algorithm
    - Cross-validation algorithm
    - Deep Learning algorithm
- Model based on Random Forest algorithm

For sure, models based on regressions are the simplest in terms of mathematical definition and ease of calculation, but sometimes other more complex models are needed to increase the accuracy of prediction.

On the other hand, neural networks are an innovative machine learning technique that allows to development of complex computational problems quickly. Inspired by the human brain's functionality and its structure, a neural network is made up of many "units" called neurons. Many neural network algorithms can be defined, in particular in this thesis work back-propagation, cross-validation and deep learning algorithm have been taken into account.

Random Forest is a machine learning algorithm very popular for its simplicity, ease of use, and interpretability. It is a supervised learning methodology for ensembles that belongs to the decision tree family of algorithms.

Machine learning techniques are a valid alternative to the traditional regressions technique and in some cases are expected to have better performances.

Step 1.4 – MAPE Calculation

For each Cost Model type, it is possible to calculate MAPE (Mean Absolute Percentage Error) value as described in paragraph 4.4.6 to predict which is the best cost model for each specific situation.

The mean absolute percentage error (MAPE), in statistics, is measure forecast system accuracy. It measures cost model accuracy with a percentage value, and it works best if there are no extremes to the data (and no zeros).

Consequently, MAPE has mainly two advantages:

- The absolute values avoid the positive and negative errors elision each other

- Thanks to the fact that relative errors are not referred to a specific scale of the dependent variable, MAPE measure allows comparing accuracy between differently scaled datasets.

Step 1.5 – Chose Best Fit Cost Model

Considering MAPE values calculated in Step 1.4, best-fit cost model, among the ones calculated, is selected (the one with a lower MAPE value).
For each cost model it is possible to calculate quartiles, median value, and outliers:

- Quartiles are those values that divide the population into four equal parts:
- Outliers are those values that are numerically distant from the rest of the collected data. In statistics they are defined as values outside a certain range and, since they can be misleading, it is better not to consider them.

Mathematical models are not always enough to find an accurate cost model. Sometimes value engineer needs to proceed with deeper data analysis to understand all the steps and cost impact of each phase.

For example, procurement plays an important role in the data values, and a deeper analysis can explain all the impact. The same component can be purchased at different costs, based on the type of supply chain: if realized completely from an external supplier or internally to company farm.

To take into account all the process steps, machining phases are split into base operations and estimated with proper parametric models.

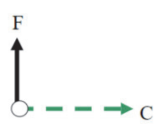
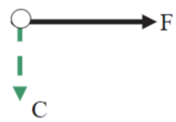
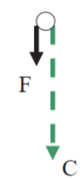
Another important factor impacting cost is batch size: purchasing together more than one component with the same design and same requirements may lower unitary cost leveraging volume opportunities.

3.3. VAVE Methodology

"VAVE methodology" is a systematic process that follows pre-defined steps to guide users to increase the product value of the component/system/process under analysis. In general, this process is followed by a multi-disciplinary team, composed of subject experts but also by not expert people that can bring new ideas leveraging their fresh and not conditioned minds.

VAVE uses functional decomposition and analysis to optimize the ratio between functions and costs, where functions identify the scope the item is designed and purchased for. There are more than one option to increase the ratio between function and cost, in particular, Table 3 summarizes five alternative ways to obtain the scope:

Table 3: Different strategies to improve product value - © 2020 Baker Hughes, LLC - All rights reserved.

					
	a)	b)	c)	d)	e)
Function	Increase	Remain Constant	Increase	Increase	Decrease with a lower rate
Cost	Remain Constant	Decrease	Decrease	Increase with a lower rate	Decrease

All five steps of this methodology are described in section 1.2, leveraging the standard process described in [12] and [9], but also adapting it to the specific case study of turbomachinery field and, in particular, of gas turbine design.

VAVE five steps are applied by the team during a five days' workshop: the team will work together in the same room and on the same pages to apply methodology, complete all the steps of VAVE process, and built one, or more than one, redesign proposals.

VAVE had been tailored on this specific case study and the modified process is shown in the next pages. The project of a new gas turbine involves a high number of complex components and to correctly analyze them standard approach was not always enough. The flowchart of Figure 13 represents all the steps of “VAVE methodology” that will be explained later on.

One of the very first activities to perform is the definition of scope and objective for the VAVE analysis. One of the main modifications introduced to the standard methodology consists of the split of the first two phases into two parallel processes. As said, for the blades case study, more than one contribution to total cost has been considered. Raw material cost, the so-called direct material (DM) is one of the major contributors to cost, and the other is direct labor cost (DL), associated with all the machining and process steps needed to obtain final geometry, starting from raw material one.

To properly consider both contributions, the team performed two information phases (information 2.1.1/2.1.2) and two function phases (function phases 2.2.1 and 2.2.2). This double Step 1 and Step 2 are crucial to understanding more in detail and better focus on the two main cost portions: DM and DL.

Another significant difference is the step total number reduction (five steps versus six steps of the standard methodology). These modifications permit the team both to address the peculiarity of gas turbine components and at the same time, to fit the standard time dedicated to a workshop, which is a working week.

Step 2.1 - Information Phase.

During this first step, the team shall improve the knowledge of the scope of work of the matter of study. VAVE engineer collects all the information starting from the weeks' antecedents the workshop and shares and explains to the team all the tools to facilitate working activities. At the end of this step, it is important to reach the following tasks:

- Share background information to all the team members;
- Define stakeholders and ask them design to cost target;
- Reach a clear understanding of current design by all the team members (VAVE engineer often leverages subject matter experts to reach this goal);
- Write all the constrain and requirements that shall be met also with alternative design proposals;
- Ask the subject expert to take part in the workshop or, in the alternative, to be available for remote support in case of need.

As anticipated, in this modified process, two information phases will occur, one focusing on blades' direct material (DM) and another on direct labor (DL).

The first one will be dedicated to raw material geometry, starting from material cost and all the rough machining associated. The other one is dedicated to DL and consequently to all the machining and control steps necessary to obtain the final geometry starting from the semi-finished one.

During this phase, a benchmark comparison is important to analyze the current design concerning previous design and, if available, to competitors' design. All these pre-work activities and phase one activities are facilitated through the usage of 2D sketches of parts under discussion.

In this phase, it is important to start defining, which is current product value, which means the parameter object of the VAVE optimization analysis.

Product Value, as stated in [7] and [8], and referring also to the European Standard EN1325:2014 [54], is defined as the "reliable performance which a product or process must achieve to make it work and sell at the least practical cost". Product value mathematically calculated with the following equation (5).

At the end of the VAVE process, the product value of the proposed redesign shall be measured with the same equation to quantify the benefit of the new solution.

$$Product\ Value = \frac{Requirements\ Satisfaction}{Product\ Cost} = \frac{Function}{Cost} \tag{5}$$

Step 2.2 - Function Analysis Phase.

This step is crucial to the success of VAVE process and it is for sure the most complicated step to be performed during the five days' workshop.

Substep of this phase are the following:

<u>Function Identification</u>: identify all verb and noun associations able to define the functions of the scope of the work item. This is the basis of the product functional analysis and shall be completed following in a precise way the correct syntax of the VAVE method as explained in [8].

<u>Function Classification</u>. Classification of functions refers to the general scope of functions. "Higher-Order Functions" are those that define the need for which the item is purchased by the customer and for which the "Basic Functions" exist. "Basic Functions" refer to the specific task an item is designed for, "Secondary Functions" represent all the functions the product absolves.

It is possible to identify also "Lower Order Functions" that are all the functions that characterize the input of the system of study, without being part of it.

A parallel type of classification is the following:

- required functions: necessary to make basic function happen
- aesthetic functions: useful to sell a product
- unwanted functions: happen in parallel to required ones but are not desired.

Resource allocation to Functions. The tool used to complete this step is the "Function-Resource" matrix. It is a matrix that helps to calculate the number of resources spent on every single function. With the use of this matrix, it is possible to create a ranking among functions based on their specific cost.

The VAVE methodology will then proceed with the functions that impact most on cost, being sure to address enough functions to explain at least 70% of the total cost.

In the modified methodology, the team is required to realize two different FAST diagrams, one for DM and the other for DL. In the end, functions with greater impact on product cost are identified to continue following steps from 3 to 5, according to standard methodology.

It is worth notice that analyze the direct labor portion of cost allowed to identify specific functions related to the production process, hardly to be recognized without the split between DM and DL performed by the team.

Step 2.3 - Creative Phase.

During this phase it is very important to generate as many innovative ideas as possible, focusing on identified functions the component shall match.

The team is stimulated to identify a huge number of ideas for each function of Step 2.2, without any judgment from team members and collecting them even if they seem to be not applicable or too disruptive.

The aim is to trigger as many new ideas as possible, and even a not applicable one can help other team members to suggest a terrific new proposal.

All the ideas collected shall be written on a flip chart by one of the team members, and this facilitates a brainstorming process.

Ideas generation cannot stop until the team has identified at least eighty different ideas for each function of the previous step.

Step 2.4 - Evaluation Phase.

During this step, the team applies the engineering judgment to all the ideas proposed in Step 2.3. At the end of this step, from the huge number of ideas proposed, the team select the most interesting ones to create from them some design proposals, during the next development step.

To optimize the evaluation phase, it is possible to proceed as follows:

- Categorize Ideas: the team may group ideas in case they are related to the same topic, for example, "material selection" or "innovative technology" or "geometry change", etc.
- "Cost-Rank" Ideas: ideas may be ranked based on their eventual impact on item cost.
- Combine Ideas: during this step, the team is allowed to combine ideas, creating new ideas, or more complex proposals.

The output of this step is the list of ranked ideas, that will be deeply analyzed during the following steps to create solid proposals.

Step 2.5 - Development Phase.

This step is dealing with a preliminary engineering development of some set of most promising ideas, building one or more redesign proposals.

For sure, at this stage it is not a complete and detailed analysis of ideas, that will be completed only for approved proposals.

At this stage, it is very important to measure the impact of proposals on product value. Referring to equation (5) *"Requirements Satisfaction" is related to* customer needs satisfaction while *"Product Cost"* deals with all the contribution to the total cost, as described in [23].

To reach a numerical calculation for product value it is possible to write the previous equation as follows:

$$Product\ Value = \frac{Requirements\ Satisfaction}{Product\ Cost} = \frac{\prod_i 1 - \frac{|Requirement\ Value_i - Design\ Value_i|}{Requirement\ Value_i}}{\frac{Estimated\ Cost}{Target}} \tag{6}$$

According to Equation (6) "*Requirement Value$_i$*" describes the value assumed according to Product Requirement Document (PRD) and the "*Design Value$_i$*" is the associated value for the variable "i" according to the suggested design. "*Estimated cost*" refers to evaluated cost and "*Target*" is cost target as per PRD.

Estimated cost can be evaluated through available parametric cost models if present, otherwise also through should cost analysis or preliminary quotations from suppliers.

Equation (6) is applicable in case "*Design Value$_i$*" assumes values between zero and double value of "*Requirement Value$_i$*". In case of value outside this range, the equation is set equal to zero and as a consequence of that also product value will be zero, considering the fact that the "i-th" requirement is not matched. Product value is a sort of numerical parameter that can be evaluated during each stage of the project to understand if the design modification proposed introduces improvements in value, as shown also in [8] and [23].

Product value is a crucial parameter during the analysis of new product introduction and for target costs definition.

After completing double Step 2.1.1/2.1.2 and 2.2.1/2.2.2, VAVE process can follow a standard methodology for what concerned the creative, evaluation, development, and presentation phases.

The team shall formalize solid proposals for re-design options to increase product value. After the workshop completion, all the proposals need detailed planning to guarantee correct implementation. In particular, the plan shall assure that all the drawings and documents impacted by the modification will be revised in accordance.

Each proposal needs to be formalized as follows:

- Proposal Sheet: contain a detailed description of the idea, showing differences between actual and new design. For each sheet, the team needs to estimate the cost of change, in terms of effort and return period.
- T-chart: table with all the advantages, drawbacks, and possible ways to overcome them listed.
- Action Plan: needed to control proper timing and implementation cost. This task will be specifically followed by the project manager.

Proposals will be formalized in a format like the one shown in Figure 14.

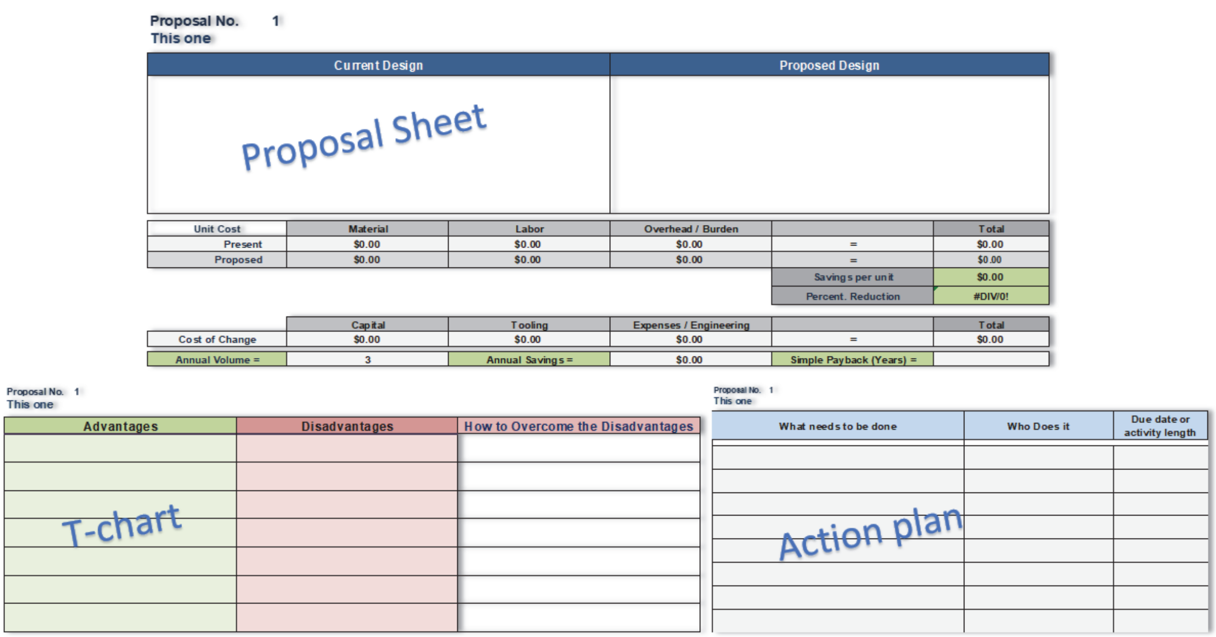


Figure 14. Proposal sheet, T-chart and Action plan templates -

Chapter 4.

Methodology Implementation: Tools

In this chapter, a proposed software platform to follow all the steps of the methodology presented in Figure 13 is described, showing system architecture, modules, use scenarios, and requirements.
The software platform can manage both conceptual costing and VAVE analysis.
Design tools are applied in industries starting from conceptual design and also in the following steps to simulate and model all the proposed product designs.
Many tools become a standard practice in design, helping design teams to save money and avoid trivial mistakes.
That is why, nowadays, the bottleneck and more time-consuming activity is the data exchange among different departments and different tools, used to finalize product design.
The proposed software platform aims to help the integration among design and cost evaluation as well as product value optimization, to reduce the time of data exchange between design and value teams and even to allow design engineers directly to perform cost and value assessment on its own, while investigating a different possible solution for its product design.
Platform architecture, requirements, and modules will be discussed, to show how it is possible to integrate all the methodology and workflow processes presented in Figure 13.
It is worth notice that some modules need one, or more than one, tool to be proficiently used and integrated into the entire platform framework. Those modules will be deeply discussed together with several tools, useful for methodology implementation, in particular:

- Tools to build Parametric Cost Model
    - Leancost
    - Rapid Miner
    - SPSS
- Tools to optimize Product Value through VAVE analysis
    - FAST Diagrams
    - Function-Resource Matrix

4.1. Proposed Software Platform

The holistic goal of this thesis is to define guidelines for the creation of a software platform able to manage both tools for cost estimation, parametric cost model definition, and VAVE analysis, interacting together as per the flow chart described in Chapter 3 and in particular in Figure 13.
First of all, Baker Hughes company defined general requirements and requirements related to the graphic interface. These requirements are listed below and have been recalled in the following paragraphs.

R1. Cost estimation provided by the tool must be parametric, based on the drivers (geometric and non-geometric);

R2. The instrument must be able to extract parametric equations from a series of data points stored in its database;

R3. The instrument must be able to adjust equations and parametric coefficients based on updated data;

R4. The conceptual costing tool must be open to extending coverage to other families of parts. The tool provides a user interface to add new families of parts, define their drivers, send the set of data point references, etc;

R5. The conceptual costing tool must allow the extension of the drivers. The tool provides a user interface to add or change drivers for an existing part family by sending an updated (expanded) reference set of data points;

R6. The conceptual costing tool must allow calculations based on a multiple series of data points (Actual Cost, Should Cost);

R7. The conceptual costing tool must perform a validation of the calculated equations against the data points;

R8. The conceptual costing tool provides the calculated cost with the applicable accuracy concerning the reference data set;

R9. The user can select the family of parts to work on;

R10. The user can select the data set (Actual Cost, Should Cost);

R11. The user enters the specific drivers of the selected family of parts;

R12. The conceptual costing tool generates costs of raw materials, mechanical processing.

R13. The VAVE tool must be able to define the Function List of a generic gas turbine component, leveraging a database of functions stored in the database.

R14. The VAVE tool must be able to build FAST diagram based on available FAST diagrams

R15. The VAVE tool must provide a preliminary idea generation based on the functions identified and data stored in the database

R16. The VAVE tool must be open to extending coverage to other families of parts. The tool provides a user interface to add new families of parts, define their functions and relative relationships, etc.

For each requirement, one or more than one dedicated module of the software platform is needed, in particular:

- **R1** DB parametric cost models
- **R2** Administrator GUI, Parametric cost Modeller
- **R3** Administrator GUI, Parametric cost Modeller
- **R4** Administrator GUI, Parametric cost Modeller
- **R5** Administrator GUI, Parametric cost Modeller
- **R6** Parametric cost Estimator
- **R7** Parametric cost Estimator
- **R8** Parametric cost Estimator, BH software systems Interface
- **R9** User GUI
- **R10** User GUI
- **R11** User GUI
- **R12** User GUI
- **R13** DB VAVE, Administrator GUI, VAVE Analysis
- **R14** DB VAVE, Administrator GUI, VAVE Analysis
- **R15** DB VAVE, Administrator GUI, VAVE Analysis
- **R16** DB VAVE, Administrator GUI, VAVE Analysis

The proposed System Architecture is shown in Figure 15 below:

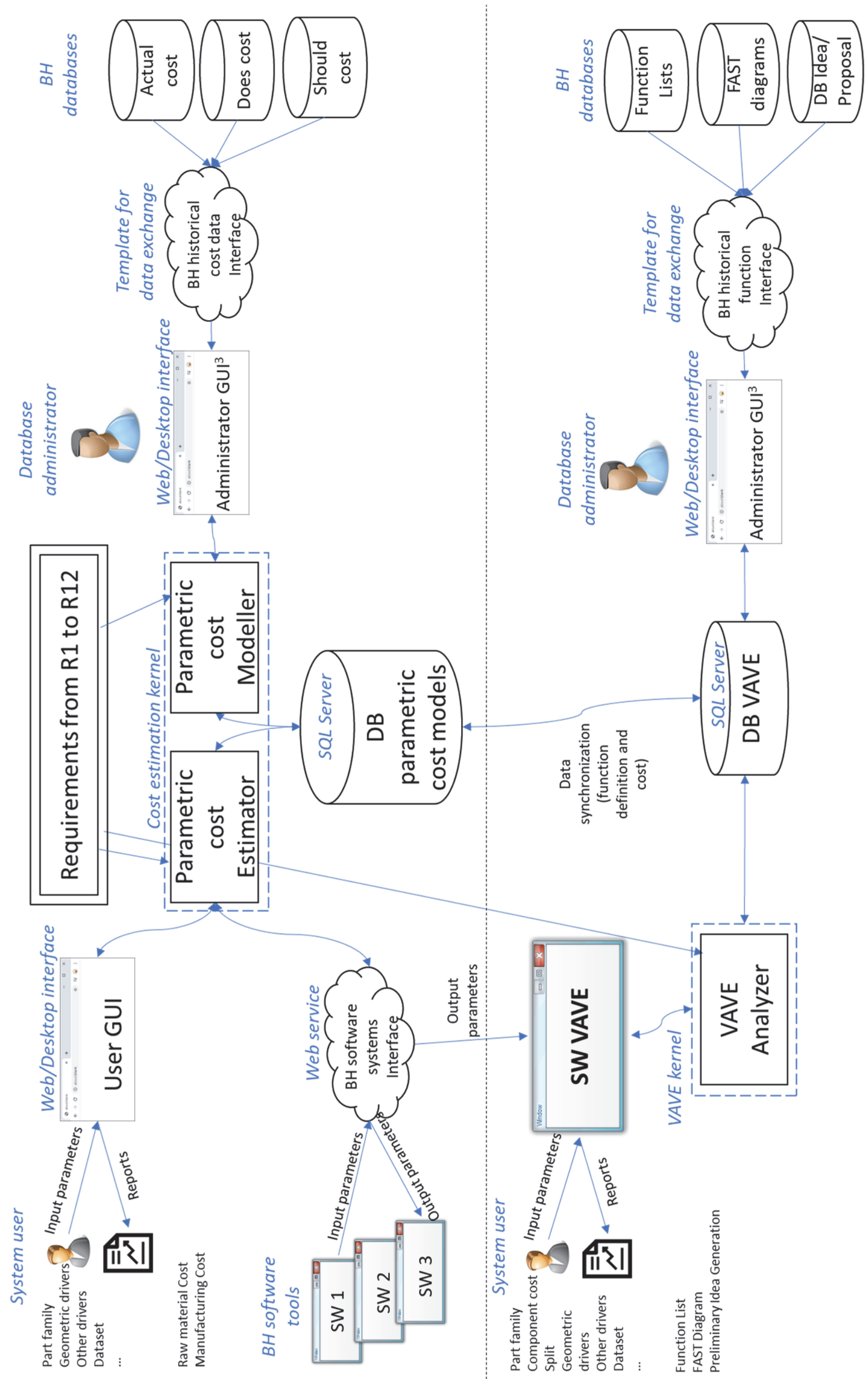


Figure 15. Software Platform Architecture - © 2020 Baker Hughes, LLC - All rights reserved.

There are three possible use scenarios for this software platform, that are described here below:

- System User: The system user will be able to interact with a proper GUI through that he can provide input information and get manufacturing and raw material cost estimates according to the cost models available from a central database as well as function lists, a standard FAST diagram and a preliminary idea generation based on VAVE database.
- System Administrator: The system administrator will be able to manage Baker Hughes historical cost data for developing or modifying customized and parametric cost models, to be stored into a central database. Moreover, will be able to manage Baker Hughes database of functions, FAST diagrams, and standard idea and proposals for each function, to be stored in VAVE database.
- Baker Hughes software tools: Baker Hughes software tools can exchange information with the Parametric cost Estimator and with VAVE Analyzer, by providing input information and getting manufacturing and raw material cost estimates, as well as a preliminary functional analysis, according to the cost models and functions available from a central database.

Referring to Figure 15, it is possible to show how the following modules interact together:

- User GUI: The graphical user interface allows the system user (for example design engineer) to evaluate the cost of a component by proving information such as the part family, geometric and non-geometric cost drivers, dataset, etc. This interface provides to the user manufacturing and raw material cost within a specific report. *Referring to Figure 13, it allows performing parametric cost estimation leveraging output of Step 1.5*

- Administrator GUI: The graphical user interface allows the system administrator to define new parametric cost models (or even update historical ones) according to proprietary data of BH. The interface aids the administrator during the workflow for parametric cost modelling. The interface will manage only a specific cost modelling approach, among regression, neural network, or random forest. For what concerns VAVE module, the interface allows the system administrator to define standard functions absolved by main BH components and their relative FAST diagrams. For each function, also a standard database of possible alternative solutions can be stored, even if Step 2.3 of VAVE process cannot be completely substituted by the tool. *Referring to Figure 13, it allows performing Platform update and maintenance*.
- BH historical cost data Interface: This interface allows data exchange between BH database containing historical cost data with the Parametric cost Modeler. It allows standardizing data elaborated by the administrator GUI and the Parametric cost Modeler. *Referring to Figure 13, it is the database needed to support Step1.1 and Step 1.2.*
- BH historical functions Interface: This interface allows data exchange between BH database containing historical function data with the VAVE Analysis module. It allows standardizing data elaborated by the administrator GUI and the VAVE Analysis module. *Referring to Figure 13, it is the database needed to support Step2.1 and Step 2.2.*
- BH software systems Interface: This interface allows data exchange between BH software tools (for example cross-section configurator software) with the Parametric Cost Estimator. *It allows standardizing data elaborated by the Parametric Cost Estimator.*
- Parametric cost Estimator: This module is in charge to compute the cost of a component according to input data defined by the system user or external software tools. Cost estimation is based on the database of

parametric cost models. *Referring to Figure 13, it leverages Step from 1.1 to 1.5 to compute cost and check if the gap between estimated cost and target cost is lower than 10%*

- Parametric cost Modeler: This module is in charge to develop parametric cost models according to historical data of BH, through a precise approach (e.g., data regression, neural network, random forest). Cost models will be stored in the parametric cost models database. *Referring to Figure 13, it performs Step 1.3 and Step 1.4.*
- DB parametric cost models: Database containing the parametric cost models defined by the Parametric cost Modeler, which can be used by the Parametric Cost Estimator. *Referring to Figure 13, it performs Step 1.5.*
- VAVE Analyzer: This module is in charge to develop VAVE Analysis (referring to steps 2.1 and 2.2 only) according to historical data of BH, through a standardized functional analysis approach. Database of standard functions for each component will be stored in the VAVE database. *Referring to Figure 13, it performs Step 2.1, Step 2.2, partially Step 2.3 for the component under analysis.*
- DB VAVE: Database containing the standard functional analysis for main BH components (all the functions associated with each component and relative FAST diagram), that can be used by VAVE analysis module. *Referring to Figure 13, it performs Step 2.2 for standard components.*

4.2. Parametric Cost Estimator

In Figure 16 it is possible to see an example of a parametric cost estimator interface. In particular, inputs are received by a geometry configurator (here below, only diameter "d" and thickness "s" are shown as parameters, but there are many others both geometrical and technological, here not disclosed for proprietary reasons). Parametric Cost Estimator interface read inputs received by configurator and, leveraging cost models available, return component cost, with the cost split shown here below.

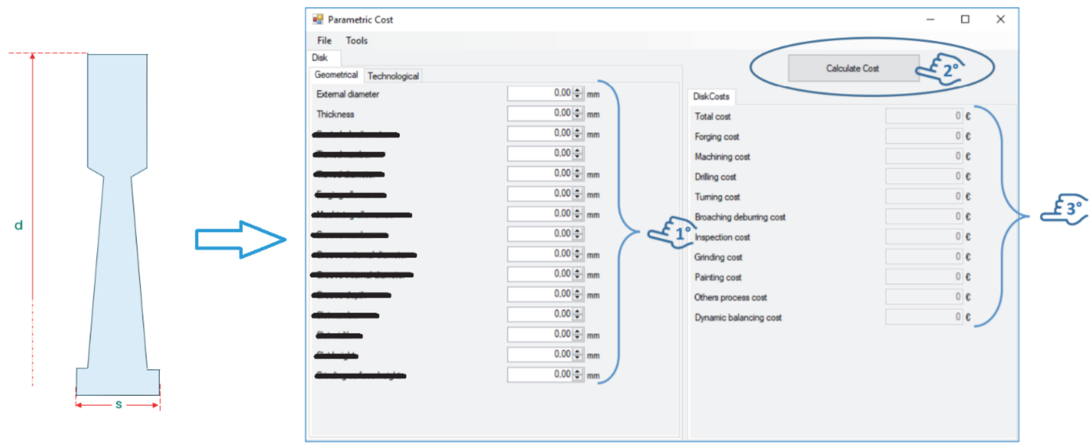


Figure 16. Parametric Cost Estimator: Input / Output Example -

4.3. BH historical cost data Interface

To build a parametric cost model, first of all, it is necessary to define a database of should cost to be analyzed. To define this set of should costs, **LeanCOST** tool has been used, thanks to the fact that it is automatic, based on 3d models and analytic.

The purpose of this software is to simplify all those operations to be performed to determine the cost of manufacturing a mechanical component or an assembly. Therefore, once a 3D CAD prototype has been developed, the designer can define a cost estimate that the technologist will then work on. The latter can make changes to the technological parameters relating to processing by varying the cost items (raw material cost, set-up, ancillary costs, and processing costs). The output of the latter part is an actual cost that will then be modified or not by the external relations office.

The structure of the program can be summarized in four modules:

1. CAD interface module analyzes the CAD model and related non-geometric information to identify construction characteristics. The software performs a topological analysis of geometric entities (faces, loops, and edges), dimensions, finishes, tolerances, and physical properties (mass and density);
2. Process allocation module converts the set of identified construction characteristics into a series of operations to determine the manufacturing process;
3. Calculation engine automatically calculates production times using the calculation functions relating to the identified processes and translates them into costs;
4. Report generation module manages the calculated data and allows the user to use them according to their needs.

A strong advantage of this software is the ability to integrate with the most well-known commercial 3D CAD systems and to independently derive the geometric characteristics of the product and define the necessary machining processes.

LeanCOST is to all effects a company platform for the management of production costs and the exchange of information between the various figures involved through dedicated user interfaces. Each figure analyzes the cost with a different level of detail and has the necessary tools to support their work efficiently and completely.

Despite the numerous potentialities, this software does not replace the work performed by the technologist or the designer: it is a useful tool to simplify and speed up the work and therefore to reduce costs without losing in terms of quality and precision. The knowledge and experience of the technologist and the designer always remain essential characteristics for the success of the costing operation.

The operating logic is as follows: starting from the 3D CAD model and/or design specifications, the software analyzes the geometry and extracts the significant geometric parameters to obtain a complete definition of the product automatically. The calculation engine then associates the geometric description with one or more production technologies thanks to rules that contain company processes and procedures (knowledge), extracts the technological parameters, and defines the technology thanks to the information that characterizes the specific context (machines and raw materials). At the end of the analysis, the system determines the production time and cost for each component, divided into five cost items which are: cost of raw material, cost of investment, set-up, accessories, and operations.

Although the procedure is automatic, the software leaves ample room for the user to customize any production detail. This is made possible by the high level of detail for each cost item; it is possible during the industrialization phase to investigate in detail all the cost contributions and it is possible to easily check and, if desired, modify any geometric and technological parameter of any operation, such as the speed of advancement of the tool of a machine or any other process variable.

Leancost tool has been used in this thesis work specifically to generate a database of should-cost analysis that constitutes the input database for Rapid Miner tool application.

4.4. Parametric Cost Modeller

RapidMiner is a data mining and predictive analytics platform that permits to perform advanced analysis on data in a simple and fast way, thanks to methods of extraction, transformation, and visualization of data that do not require special programming knowledge. It has been used to create cost models based on regression, neural networks, and random forest methods.

RapidMiner is a software platform developed by the company of the same name, which offers an integrated environment for machine learning, data mining, text mining, predictive analysis, and business analysis. It is used for business and commercial applications, as well as for research, education, training, rapid prototyping, and application development and supports all phases of the data mining process including data preparation, result visualization, validation, and optimization.
The graphic interface for the design of processes consists of:

1. The central panel for the design of the process where operators for modelling are inserted and connected;
2. The parameters box of the process and of each operator;
3. The list box of all possible operators that can be used, from process control ones to data analysis ones, up to modelling and validation operators; Each operator performs a single activity within the process and the exit door of each operator constitutes the entry door of the next;
4. The repository pane where the data used in the various work sessions are saved.

RapidMiner uses the standardized XML language for processes, which can be modified from a special window.
RapidMiner functionality can be extended with additional plug-ins made available through RapidMiner Marketplace. The RapidMiner Marketplace provides a platform for developers, where data analysis algorithms can be created and shared with the community through platform publishing. With the software, you can make connections to the main data sources including Excel.

SPSS (Statistical Package for Social Science) is an advanced statistical analysis tool that offers the ability to use a variety of functions, from planning and data collection to analysis, reporting, and distribution.

In particular, it is possible:

- Analyze a set of data to solve complex business and research problems through an intuitive interface, similar to that of Microsoft Excel;
- Understand large and complex datasets faster with advanced statistical procedures, which help ensure high accuracy and quality decision making;
- Use the extensions, the programming language code Python, R, and also RapidMiner for integration with open-source software.

The SPSS software offers in particular the possibility of carrying out an analysis of variance (ANOVA) of different types (based on the number of independent variables):

- One-way ANOVA: models that include a single independent variable;
- Factorial ANOVA: models involving two or more independent variables;
- univariate ANOVA: models that foresee a single dependent variable;
- Multivariate ANOVA (also known as MANOVA): models that include two or more dependent variables.

Specifically, the software is used to carry out the MANOVA factor analysis.

4.4.1. Parametric Costing Modelling with Rapid Miner

In this paragraph, Rapid Miner functionality is compared and linked to requirements defined to develop a parametric cost model and the software platform, both described in section 4 in detail.

Figure 17 shows the scheme of the designated methodology.

The first step of the diagram in Figure 17 is the choice of the database: starting from a set of historical data from purchase order but also from a set of should cost values from LeanCOST, the database is prepared to make it usable by the various methods. Any anomalous values are excluded and only the cost drivers with a strong influence on the output are considered. In the latter case, the

MANOVA factor analysis is used to determine the significance and therefore the weight of each input with respect to the output.

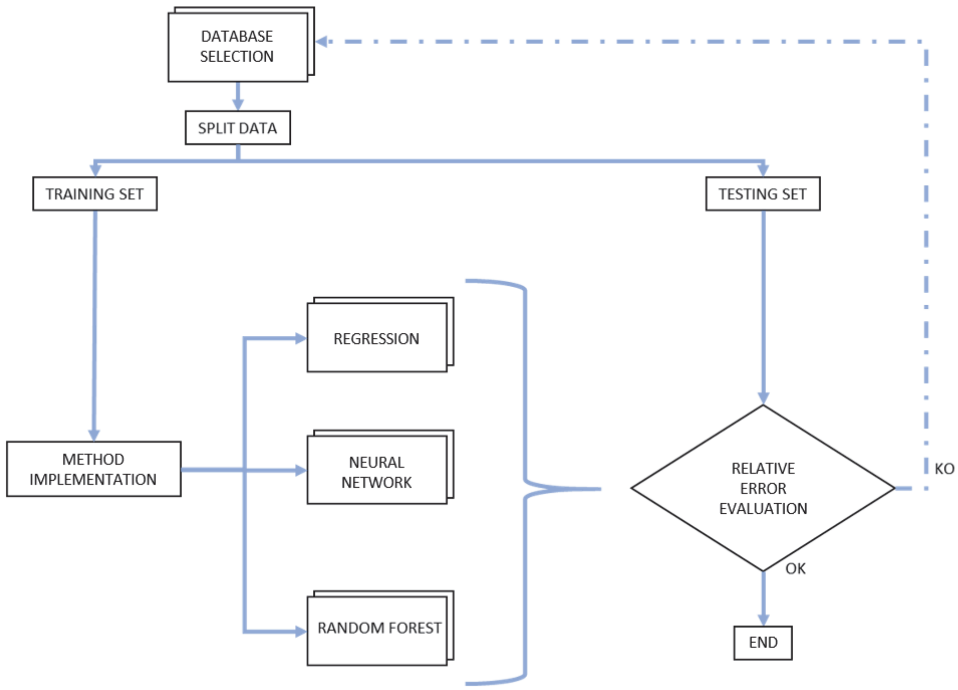


Figure 17. Methodology -

Subsequently, through the spit data, the database is divided into two groups: a training set and testing set. The former is used to implement the traditional linear regression method and non-linear machine learning methods, while the latter is used to validate the models obtained.

In RapidMiner for each method, a process is generated consisting of four sub-processes relating to: pre-processing data, training model, testing model, create prediction, and output (Figure 18).

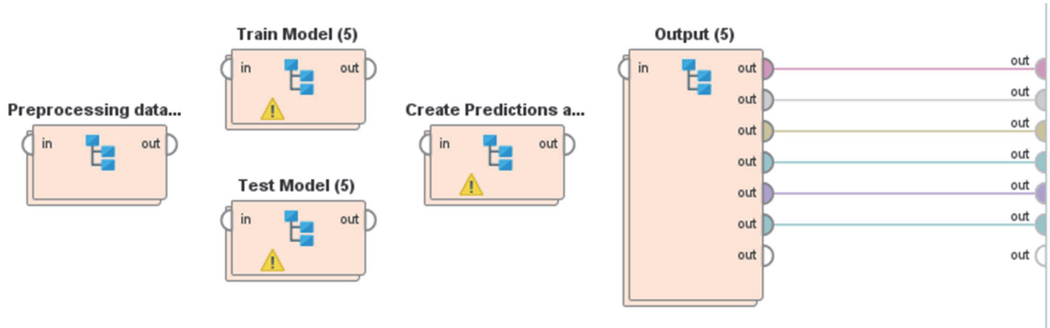


Figure 18. Sub-processes

The “pre-processing data” corresponds to the choice of the database of the schema of Figure 17: as input, the initial database is inserted and as output, the training and testing data are obtained. To link all the sub-processes together, two operators are used:

• "Remember": stores the specified object in the process object archive;

• "Recall": recalls the object from the archive.

The training data, stored in the first sub-process, are recalled in the trained model where a specific model is generated (linear regression, neural network, deep learning, or random forest) [requirement R2].

The test data are recalled, together with the model, in the test model where the algorithm is valid [requirement R7].

In the "create prediction" training data, test data and the model are called to generate the model simulator.

Finally, in the output sub-process, all the objects of the process archive are recalled and transcribed to an Excel file (as well as displaying them in the RapidMiner results).

4.4.2.Database Definition

The accuracy of a cost estimate analysis is highly dependent on the shape of the incoming historical database. It consists of input data (geometric and non-geometric drivers) [requirement R1] and related output data (costs of raw materials and production). The first step of this part is to generate the database. In some cases, the data may be available in a tabulated form (Actual Cost), in others it may be necessary to carry out a Should Cost Analysis with the help of software such as LeanCOST® or DFMA [requirements R6 - R10]. Choosing the right input drivers is not an easy operation; it is necessary to be able to identify those parameters that the designer knows in the design phase and that have a greater impact on the calculation of costs. Furthermore, identifying any outliers and eliminating them allows reducing the error of the output prediction. So, before using the various algorithms, you need to do some pre-processing operations, including:

- Multivariate factorial analysis of variance (MANOVA);
- Determine any outliers.

The multivariate factorial analysis of variance is a statistical data analysis technique that determines whether and how strongly two variables are dependent on each other. The degree of dependence is expressed by the significance value which is a number between 0 and 1; in the case of a very small value (less than 0.1, there is a strong dependence. If, on the other hand, it assumes the value 1, the two variables have no dependency relationship between them.

This technique can help in choosing the drivers with greater "weight”, but experience also plays a key role. Knowledge of production processes allows to determine any "hidden" process drivers (which derive for example from relationships between parameters) and it is essential to understand which parameters are known in the initial stages of design when cost estimates are requested.

As for the anomalous values, in this initial phase of database choice, it may happen in writing the data of Should Cost or Actual Cost in a spreadsheet, making typing errors, and consequently inserting one or more incorrect records at the internal database.

To obtain models able to estimate costs in a performing way, it is important to eliminate all those outliers due to incorrect or inconsistent records.

Outliers do not only mean incorrect records but also any duplicate or incomplete data.

Database for parametric cost model contains a set of actual cost (from latest purchase orders) or in alternative, a set of should cost values, obtained as said using software such as LeanCOST® or DFMA. Those software calculate should cost values based on analytic system of cost prediction, that leverages production know-how and manufacturing expertise, useful to create set of data in a reduced amount of time.

Once the database has been defined, the split data operation follows. As shown in Figure 17, not all data is used to generate the model but there will be a remainder dedicated to a test phase. It is precisely the test-sets that allow us to understand how a model responds and therefore validate the results.

In particular, a good percentage of the split is 60% of training data and 40% of testing data or 70 - 30% (a lot depends on the type of database, 80 - 20% could be used).

It is important to underline that machine learning techniques require a huge amount of data to be trained (especially neural networks). The data for the training should be structured and stored in a proper way to enable efficient and effective training. An important aspect is the fact that all the data used for training are proprietary information and owned by the company. In case of a different situation, it would have been much more complicated to properly training models.

RapidMiner

The “pre-processing data” sub-process contains various functions (Figure 19).

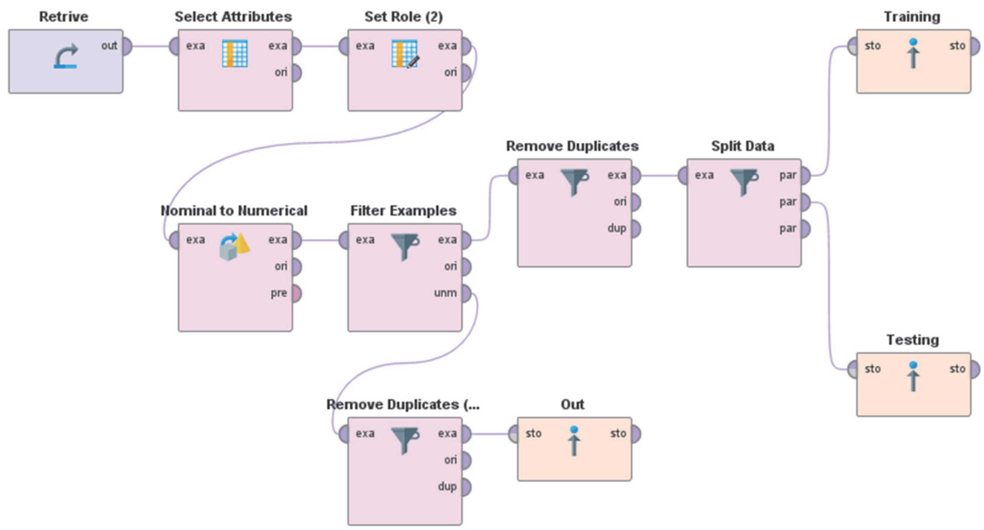


Figure 19. Pre-processing dataset

First of all, by clicking on “import data”, you can insert the initial data set in the RapidMiner repository: you can insert both numeric and qualitative variables (numerical or polynomial). Through the "retrieve" function, the database is recalled from the repository and inserted sub-process.

From Figure 19 it can be seen that each work has one or more input and output doors.

To link two operators together, you connect an output port of the first operator with an input port of the second operator.

The links are not random but depend on the type of information you want to share. RapidMiner, to simplify the connections, uses different colors according to the type of information shared to quickly locate which are the input and output ports.

If not already specified in the database, the “set role” function allows you to change the role of a variable from regular to label (i.e. the target or response variable).

A limitation of RapidMiner is that it cannot carry out the MANOVA factor analysis (there is the "ANOVA" function, but it refers to the case of classification problems and not regression). Therefore, you can use another software called SPSS to understand which are and independent variables with greater "weight" on the dependent variables.

Returning to RapidMiner, through the “select attributes” function, the inputs to be considered are selected and irrelevant ones are excluded. After this step, it is good to use "remove duplicates" as it could create duplicate records which would falsify the analysis results.

In case you want to consider a subset of the database for a second test or simply to exclude it from the next analysis, you can insert the "filter example" function.

On RapidMiner some prediction models use mixed databases (quantitative and qualitative data) and others that work only with numerical or qualitative data. In particular, for the “linear regression” and “neural network” algorithms, the “nominal to numerical” function can be used to convert a mixed database into a numerical one using the dummy or binary variable technique.

To understand the meaning of the dummy variables, refer to example in Table 4. In the case of a "material" categorical variable (first column), with two possible values M1 and M2, it is possible to generate a database made up of numerical data only by creating two new binary variables (third and fourth column) which take on the value 1 if a condition is satisfied, vice versa null value.

So, for example, row one corresponds to material M1 as the first dummy variable takes the value 1, while the second value zero.

Finally, through "split data" the database is divided into two sets: a training set and a testing set. The "remember" operator allows you to store the two datasets and reuse them in subsequent sub-processes ("recall" is used to recall them).

Table 4:: Dummy Variable -

Material	Conversion	Material = M1	Material = M2
M1	=	1	0
M2	=	0	1

4.4.3.Method implementation: Regression

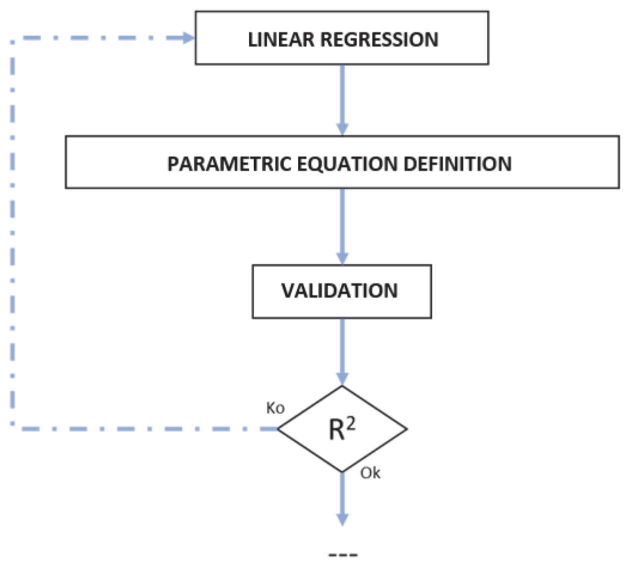


Figure 20. Linear Regression Method -

Linear regression can only be developed for numerical variables.
Figure 20 shows the various steps of this first method.
The procedure is iterative: starting from the training data (defined in the previous paragraph), simple or multiple linear regression is performed by determining the

parametric equation. This is followed by the validation in which the R^2 coefficient is calculated; in the event of an acceptable value, the analysis ends, otherwise, the procedure is repeated (making appropriate changes to the database).

R^2 can assume the following values:

- $R^2 \geq 0.85$: the analysis is satisfied;
- $0.6 < R^2 < 0.85$: it is necessary to carry out a more in-depth study of the data and identify more cost factors;
- $R^2 \leq 0.6$: the analysis is not acceptable and cannot be used for further studies.

R^2 value is used to understand if the proposed regression is acceptable or not. On the other hand, to understand, among different cost models, which one is the best solution for the specific case study, MAPE value will be taken into account, as described in Step 1.4. Some applications of regressions are described in [66][67][68].

RapidMiner

Figure 21 shows the "train model" sub-process; the training data is recalled with the "recall" function.

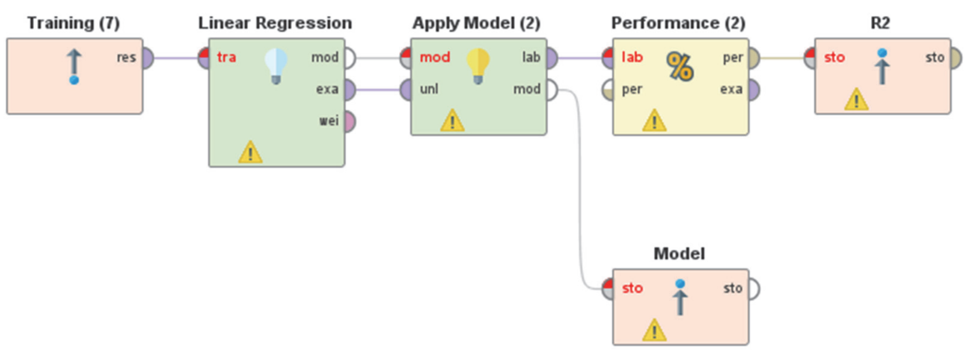


Figure 21. Linear Regression Train Model

The "linear regression" operator develops linear regression (both simple and multiple) and returns the parametric equation. Through the "apply model" the parametric equation is applied to the training data set while the "performance" operator reports the value of the R^2 coefficient (in the "parameters" window of the "performance" operator, the "squared error" must be selected correlation ").

In addition to this first validation, the model is tested with a new set of data to calculate the relative error and then to compare it with the other machine learning models developed in parallel (see paragraph 4.4.6 on Validation). A clarification must be made: the "linear regression" operator (and regression in general) has the limit of being able to be used only for numerical variables. This involves two different ways of proceeding:

1. Create a model using "linear regression" applied to the entire training set. Any qualitative variables (polynominal) must be converted into numeric using the dummy variable technique. For example, if I have material A and material B, I use the "nominal to numerical" function to generate two dummy variables, each referring to a material;
2. Create n linear regression models related to n subsets of the training set. To determine the subsets, the "filter example" operator is used n times, which filters the dataset considering only a set of data. It is the qualitative variables that guide the breakdown of the data. For example, if I have material A and material B, I make two sets each referring to a material.

The final model (or the models if subsets are created) can be stored via "remember" and then recalled in the output sub-process.

4.4.4. Method Implementation: Simple neural network and deep learning

Neural networks, as well as other machine learning techniques, can be more complicated to implement than linear parametric methods but are often more accurate. Unlike regression, they can handle numeric and categorical variables (except for the simple neural network), as described in [69][70][71].

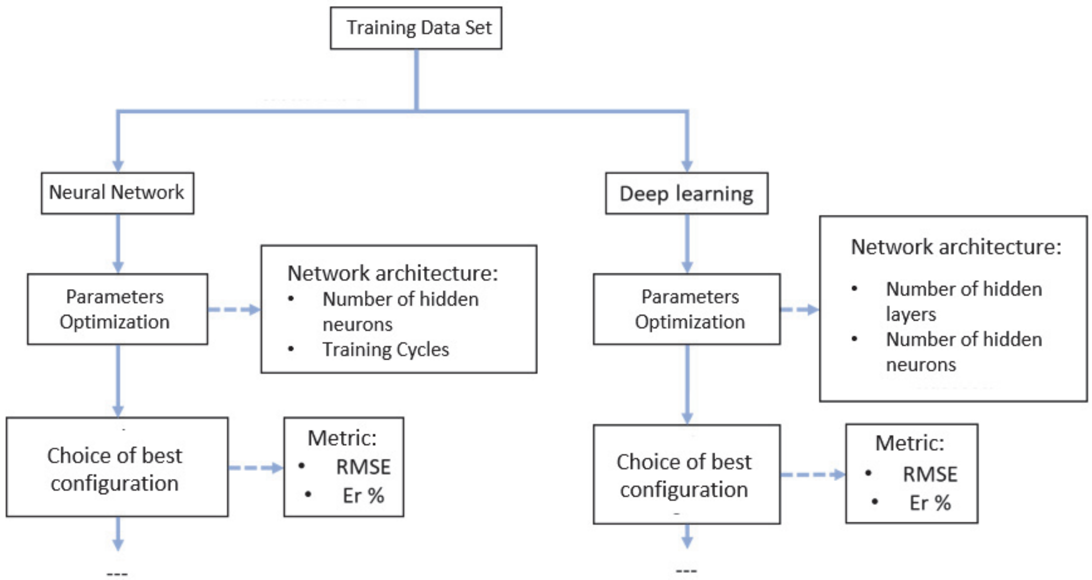


Figure 22. Simple Neural Network and Deep Learning Method -

Figure 22 shows the method schematically.

As in previous cases, also in this case the procedure is iterative and repeats until an acceptable model is obtained. Starting from the training data, the type of algorithm to be used is chosen: the simple neural network is typically characterized by a single hidden layer and is suitable for a limited number of input data; deep learning is always based on the concept of neural networks but has two or more hidden layers and therefore it is preferable to use it in the case of numerous input data. Choosing the number of neurons in a hidden layer as well as choosing the number of hidden layers is anything but trivial. When using the simple neural network, a rule of thumb allows you to determine the number of hidden neurons of the single hidden layer (3):

$$n° \, hidden \, neurons = \frac{n° \, input + n° \, output}{2} + 1 \quad (7)$$

However, equation (7) does not necessarily return the best network architecture and also, there are other parameters to specify such as the number of training cycles.

According to [72], the amount of data required for machine learning depends on many factors, such as:

- **Problem complexity:** the unknown underlying function that best relates your input variables to the output variable.
- **Learning algorithm complexity:** the algorithm used to inductively learn the unknown underlying mapping function from specific examples.

Number of hidden layers

- One or two hidden layers are most favorable to convergence, on the contrary, too more or too less may lead to bad convergence results. According to [72], empirically speaking, one layer may be chosen for the general problems and two layers may be used for the more complex problems.
- Referring to [72], authors suggest that using more than one hidden layer is rarely beneficial. The problem is related to the fact that training often slows dramatically when more hidden layers are used.

Number of Hidden nodes

- Choosing an appropriate number of hidden neurons is very important. Using too few will starve the network of the resources it needs to solve the problem. Using too many will increase the training time so much that it becomes impossible to train it adequately in an acceptable period. Also, an excessive number of hidden neurons may cause a problem called overfitting [72].

It is therefore clear that identifying the right simple or deep learning neural network model is not an immediate operation and, as there is no general rule, a sort of "exploration" must be carried out in which a computing system automatically starts multiple simultaneous executions with different parameter values and find the configuration with the best performance. Therefore, once those fundamental parameters have been identified, values are attributed to each of them and more models are generated that derive from the different combinations, as described also in [73][74][75][69].
This process is called parametric optimization and it is helpful to evaluate which configuration of the simple neural network or deep learning response best; the relative error is used as a metric.

RapidMiner

The first step, as shown in Figure 23, is to recall the training data.

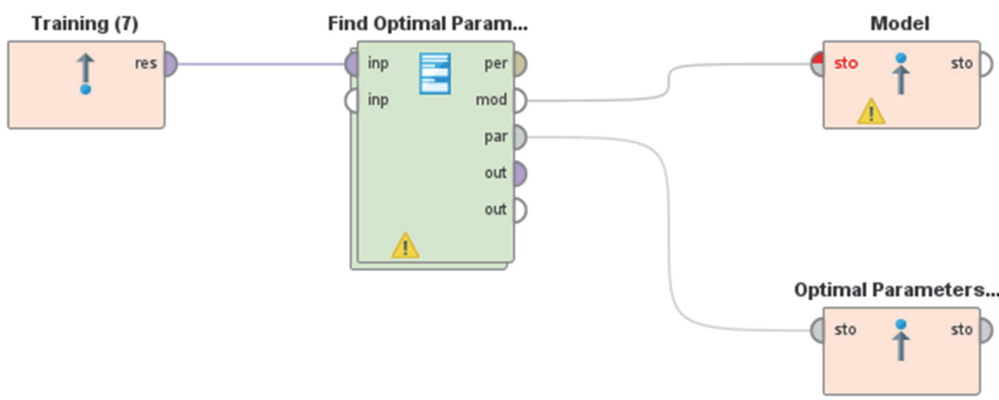


Figure 23. Neural Network train model

Parametric optimization is done with "optimize parameters (grid)": by clicking on "edit parameters setting" you select the parameters to be optimized and specify the values they can assume.
The "optimize parameters" operator is a nested operator, this means that there are additional operators within it (Figure 24).

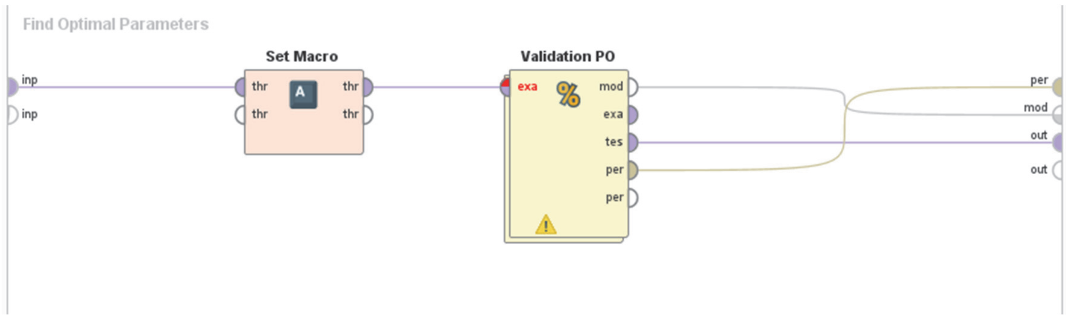


Figure 24. Parametric Optimization

"Cross-validation" is a nested operator as well and allows for cross-validation once the k-fold number has been specified.
Therefore, the training set is automatically divided into k subsets of which k-1 is used for training while the last subset is used for validation. As shown in Figure 24, the operator to be trained is entered in the "training" area while, in the "testing" area, the model found is validated from time to time and the error is calculated. Note that cross-validation is used in the train model sub-process and is different from the final validation done with the test dataset.
This technique is often used in machine learning algorithms and allows to obtain models with a lower probability of overfitting problems. Alternatively, for the calculation of the model performance (essential for the "optimal parameters" operator), "apply model" could be used combined with "performance" on the

same training data but with the risk of less performing prediction models in the "test model" phase.

A limitation of RapidMiner is that it cannot directly select the hidden layer sizes in the optimization, i.e. the number of neurons in a hidden layer. To overcome this problem, the "set macro" operator is used: one must be inserted for each hidden layer present, therefore, in the case of the classic neural network, one is enough while for deep learning two or more (Figure 25).

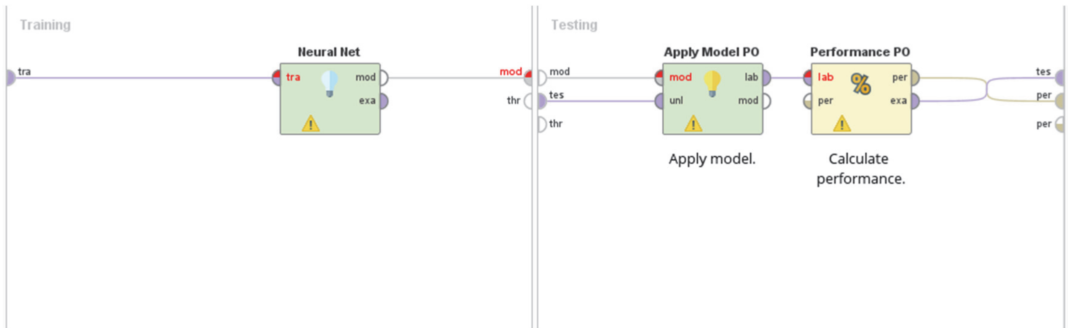


Figure 25. Cross validation

In addition to the "neural network" of Figure 24, it is also possible to use "deep learning". The latter has some advantages:

- Allows you to automatically determine the optimal number of training periods;
- Is be able to create a mixed model that considers both qualitative and quantitative variables;

And some disadvantages:

- It takes a long time to train because it is made up of multiple hidden layers and many neurons within them;
- The parametric optimization process is more complex.

In the case of the “neural network,” the database must be characterized only by numerical variables. So, just like in regression, there are two different ways of doing it:

1. Create a model via "neural network" applied to the entire training set by converting the qualitative variables (polynominal) into numerical ones (dummy variables)
2. Create n "neural netwrok" models relating to n subsets of the training set using the "filter example" operator n times

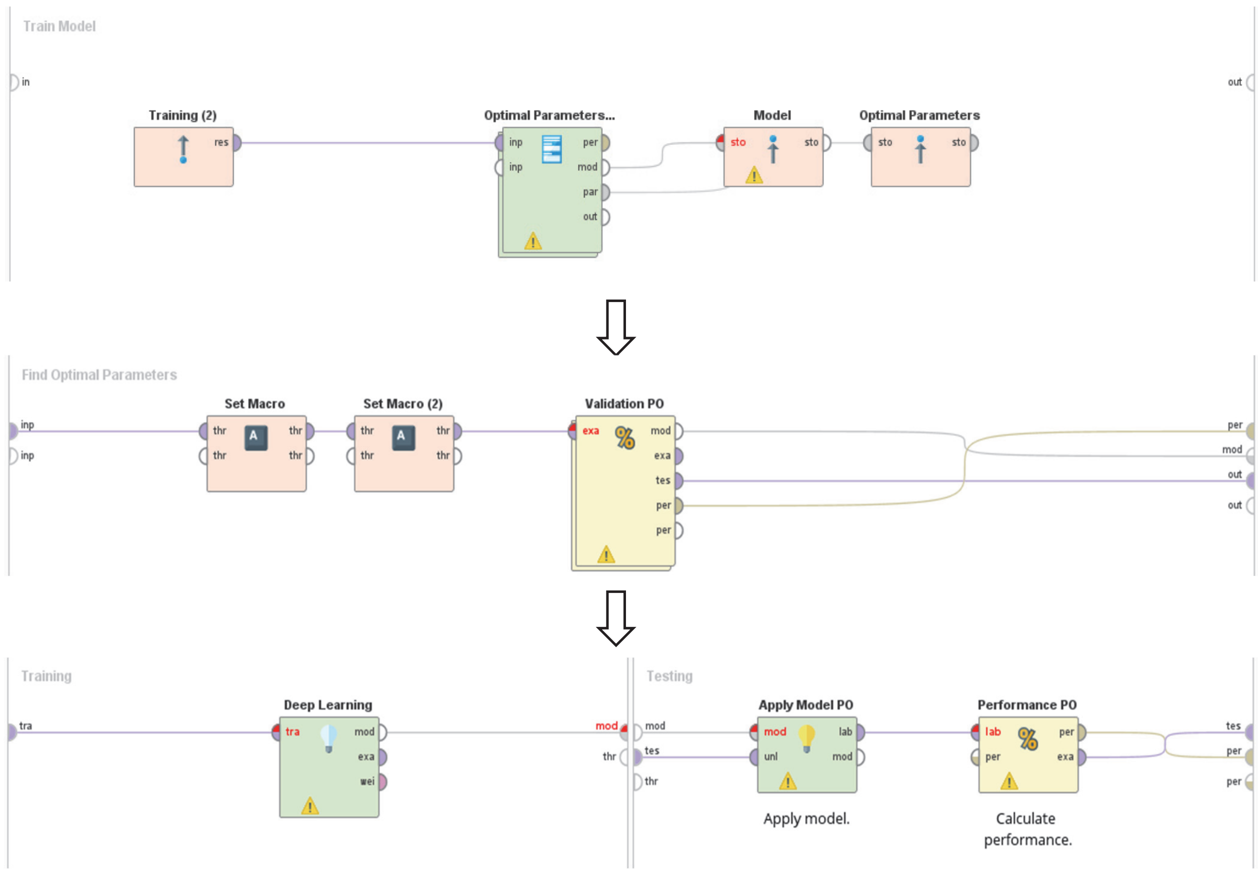


Figure 26. Deep Learning train model

4.4.5. Method implementation: Random Forest

Random Forest is part of the machine learning techniques and the modelling procedure is very similar to that described in the previous paragraph in the case of neural networks.

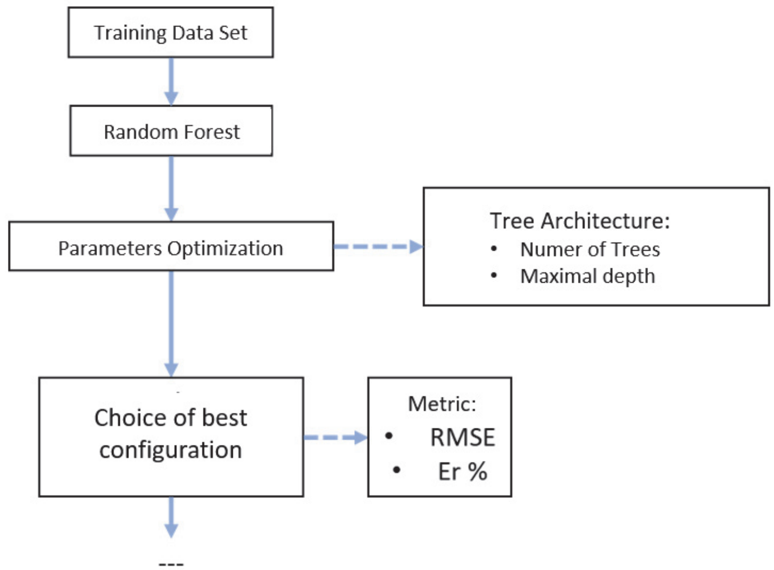


Figure 27. Random Forest Method - © 2020 Baker Hughes, LLC - All rights reserved.

As shown in Figure 26, the first step is to recall the training data and proceed with the implementation of the Random Forest method. Note that, although less complex than neural networks, also in this case it may be useful to use parametric optimization to determine the values of the ideal parameters relating to maximal depth (the depth of the tree) and the number of trees. This is followed by a selection of the model with the optimal configuration.

RapidMiner

From RapidMiner point of view the process is very similar to that of neural networks (Figure 28).

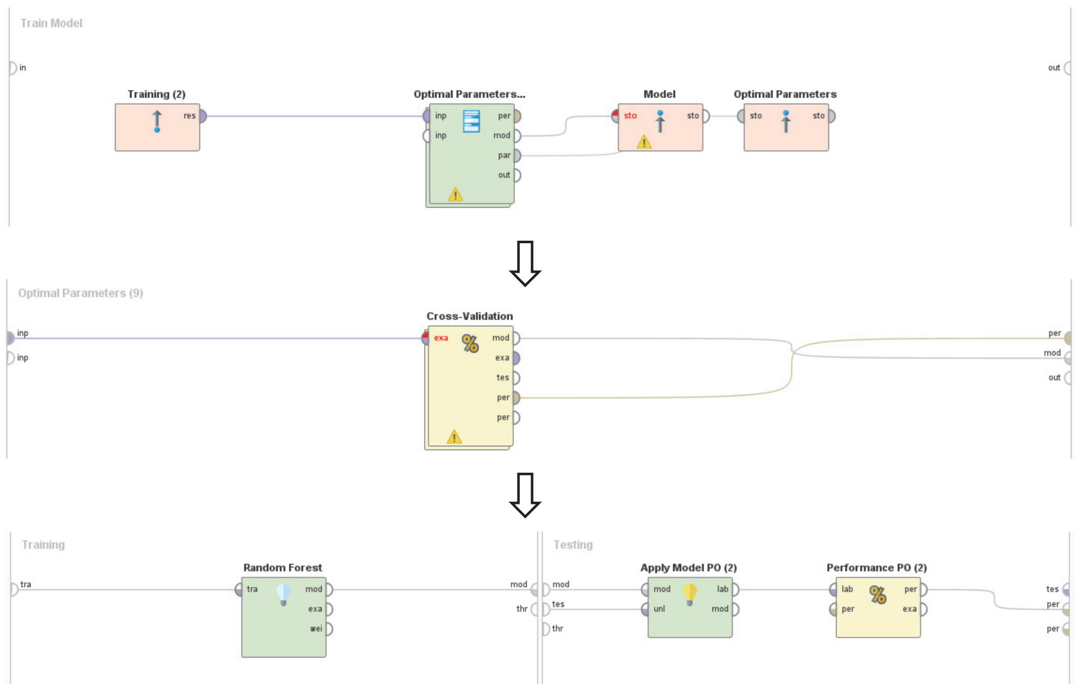


Figure 28. Random Forest train model

The procedure is like one of the neural networks but has the following differences:

- The "neural network" operator is replaced by the "random forest";
- There is no need to use the "macro set" because in this case of machine learning RapidMiner has no limits in the choice of parameters to be optimized;
- In the "optimal parameters" operator, the parameters to be optimized change (Figure 27).

4.4.6. Model Validation

Once all the possible models have been determined, the testing phase follows. As mentioned earlier, the initial database is divided into two large groups: training data and testing data. The former is used to generate the models while the latter to validate them. This step is essential in estimating costs because it allows you to understand which of the models is best suited to the type of study. Relative errors are used as metrics to compare regression, simple neural networks, deep learning, and random forest. Three among main relative error indicators are described below, the most used is the MAPE, which will be used also for the comparison among different cost models in the following chapters:

- MAPE: Mean Absolute Percentage Error.

The mean absolute percentage error (MAPE) is the mean or average of the absolute percentage errors of forecasts. Error is defined as actual or observed value minus the forecasted value. Percentage errors are summed without regard to sign to compute MAPE. This measure is easy to understand because it provides the error in terms of percentages. Considering that absolute percentage errors are used, the problem of positive and negative errors canceling each other out is avoided. Consequently, MAPE has managerial appeal and is a measure commonly used in forecasting. The smaller the MAPE the better the forecast.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_i - \hat{y}_i|}{y_i} \times 100 \right) \quad (8)$$

- Relative Error

Relative error (RE) is the ratio of the absolute error of a measurement to the measurement being taken. Consequently, this type of error is relative to the size of the item being measured. RE is expressed as a percentage as well.

$$RE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_i - \hat{y}_i|}{\hat{y}_i} \times 100 \right) \quad (9)$$

- Relative Error Lenient

The average Lenient relative error is the average of the absolute deviation of the prediction from the actual value divided by the maximum of the actual value and the prediction.

$$REl = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_i - \hat{y}_i|}{\max\{y_i, \hat{y}_i\}} \times 100 \right) \quad (10)$$

where:

- y: Estimated actual cost or should cost
- $\hat{y}$: Known actual cost or should cost
- n: total number of pairs

RapidMiner

On RapidMiner, using Figure 17, the validation is carried out in the "Test Model" sub-process (Figure 29). The idea is to apply the model in question to a new data set, therefore, as shown in the figure, the testing data is recalled and applied to the model using the "apply model". The "performance" operator allows you to calculate the errors described in the previous paragraph except for the MAPE (which can be calculated on the Excel software).

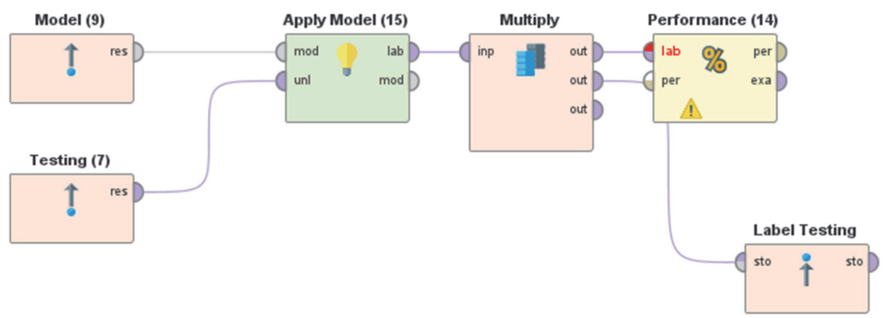


Figure 29. Test Model

The "multiply" operator in Figure 30 allows you to create a copy of the "label testing" data to store them using "remember". They report all test data with an additional final column containing the output values predicted by a model. In Figure 18, after the "Test Model" phase there is the "Create prediction" (Figure 31) where, recalling the model, the training data, and the test data, a model simulator is created.

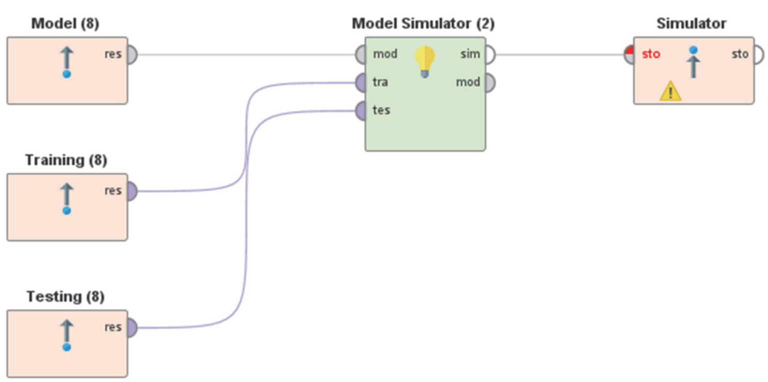


Figure 30. Create Prediction

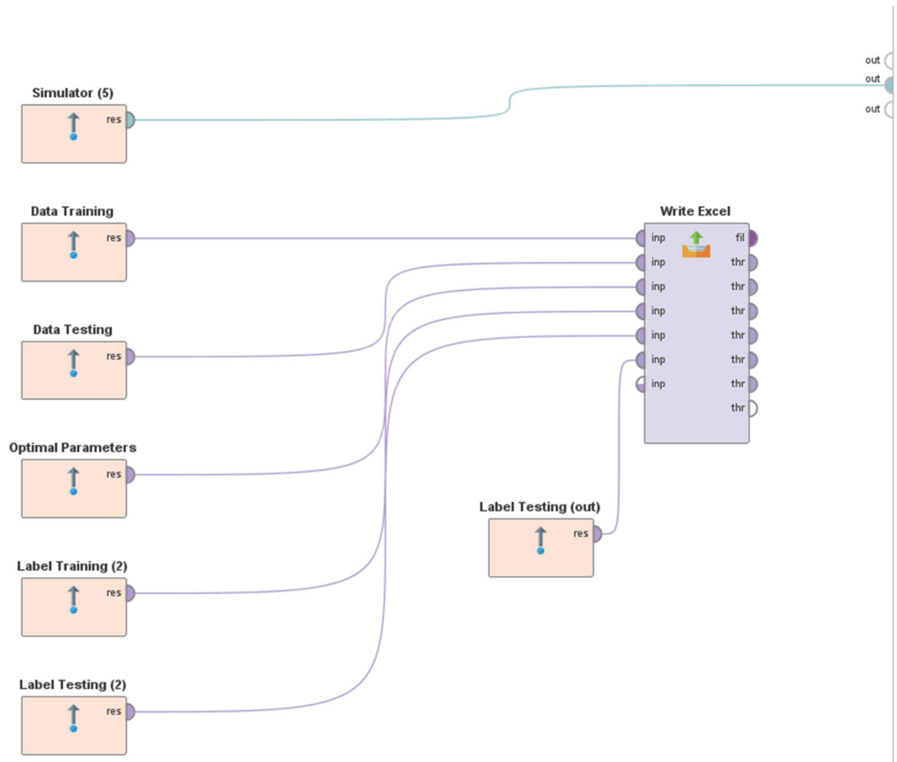


Figure 31. Output

The last "output" sub-process contains all the results obtained in the previous sub-processes (Figure 31) recalled with the "recall" function and transcribed in an excel file sheet with "write excel" (for the exception of the “model simulator” which can be displayed in the RapidMiner results).

4.5. VAVE Analyzer

Also, VAVE analyzer needs support tools to be performed and in particular, the best ones are the creation of FAST diagram, which is a diagram based on functions definition and analysis, very helpful to complete the abstraction from the specific design to the functions the component need to absolve. The second tool used, and presented in the following chapter, is Function-Resource Matrix, very helpful to associate to each function, the correct percentage of total component cost and to understand which are the most impacting functions on cost.

Functions identified in Step 2 of VAVE methodology can be connected in **FAST Diagram** (Functional Analysis System Technique).

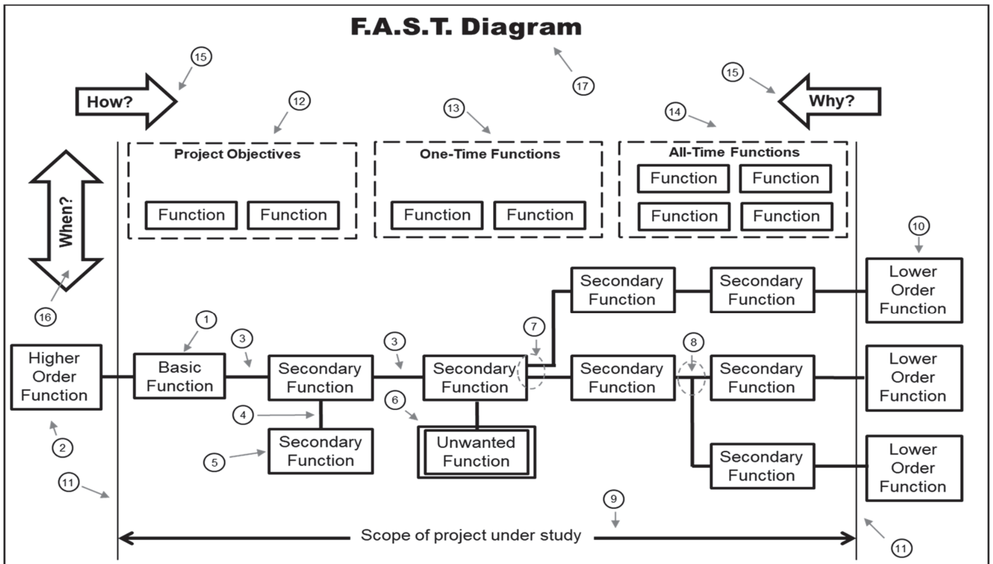
This tool permits to complete analysis of the dependency among functions within a project, product, or process, and it is very useful to identify and analyze functions to trigger creativity and innovative thinking.

This tool is a diagram that the team starts building in Step 2 of VAVE methodology. During the VAVE workshop, the team works on FAST diagram using a flipchart at the very beginning and, only once it is fixed and shared by all the team members, writing it down on an Excel or PowerPoint file.

The following steps need to be executed to correctly build a FAST Diagram, using functions identified during the functional analysis step:

- Connect Function in the "How" and "Why" directions:
- Connect along the "How" path by answering “how is the function achieved”? Put the function that answers this question at the right side (always using verb + noun format)
- Double-check the logic in the opposite direction: it shall be possible to answer the question "Why" going right to left.

- If the logic test is not passed, identify if some functions are missing or some others are redundant and correct.
- Identify also concurrent function, answering the question "during this function occurrence, is there any other function in place?"



1.Basic Function.
2.High Order Function.
3.The Critical Path (All connections on the "How?" and "Why?" Directions, linked to the Basic Function).
4."When" Direction.
5.A Secondary function that "caused by" or "happens at the same time" as the one above.
6.An "Unwanted" secondary function.
7."OR" Logic Gate.
8."AND" Logic Gate.
9.Description of the scope under study..
10."Lower Order Functions", formerly known as "Input Functions" or "Assumed functions".
11.Scope Lines.
12.Secondary Required Functions that are requested by the client, design or standard (and is not connected to the critical path).
13.Secondary Required Functions that happen only once in the life of the studied object.
14.Secondary Required Functions that are permanent, even if the basic function is not happening.
15.Visual aides, remainders of how to read the "How-Why" direction.
16.Visual aides, remainders of how to read the "When" direction.
17.A title describing what the FAST Diagram is about (The project).

Figure 32. FAST Diagram - © 2020 Baker Hughes, LLC - All rights reserved.

As said, FAST diagram is a graphical representation of the dependent relationships of functions within the product or process under analysis. The components of a FAST diagram consist of the scope lines; the labeled "How," "Why," "When," and "Scope of the project under study" arrows; and all the classified functions, including basic functions, secondary functions, unwanted functions, higher-order functions, lower-order functions, project objectives, one-time functions, and all-the-time functions.

The critical function logic path lines connect higher-order, basic, secondary, and lower order functions by their "how" and "why" relationships.

The vertical lines connect secondary functions by their "When" relationships.

To better understand all the components of FAST diagram and how to place them inside the diagram itself, it is necessary to go through some definitions:

Function

That which a project, product, or process must do to satisfy customer's needs. It shall be defined with a two-word composed of an active verb and a measurable noun structure.

The overall categories of functions that are included on FAST diagrams include Basic, Secondary, Unwanted, Higher Order, Lower Order, Project Objective, One-Time, and All-the-Time Functions.

Basic Function

The specific purpose(s) for which a project, product, or process exists. It answers the following question: "What must it do?"

Secondary Function

A function that supports the basic function and that results from the specific design solution adopted to achieve the basic function.

Unwanted Function

A negative secondary function is caused by the solution adopted to achieve the basic function.

All-time Function

A secondary function that happens continuously, anywhere in the performance of the project, product, or process.

One-time Function

A secondary function occurs only once in the performance of the project, product, or process.

Higher-order Function

The specific goal or need for which the basic function exists. Note that the higher-order function is placed outside the scope of the subject under study.

Lower order Function

The function that is selected to start the project, product, or process (it is an input). Note that also lower order function is outside the scope of the subject under study.

Project Objective

Those functions express specific compulsory requirements of the project, product, or process that shall be met.

The second tool used by VAVE Analyzer is the **Cost Function Matrix**, which helps to allocate the specific portion of the total cost to every single function realized by the product. It is constituted by an excel file, with proper sheets linked together by equations that help the team to properly allocate a portion of the total cost on each defined function.

The final scope of the Cost Function Matrix is to move attention from the cost of components to the cost contribution of each function.

Here below an example of Function-Resource Matrix is presented, the process to complete it is the following:

- List all functions
- List major parts with associated cost
- Check off which functions are impacted by each part
- Determine how much of the cost of the part belongs to each function
- Add all columns vertically to determine how much cost is allocated to each function

		DIRECT	**FUNCTION - Active Verb / Measurable Noun**								
PART or OPERATION	**QTY.**	**COST**	**CONTROL CLEARANCE**	**RECEIVE HEAT**	**PROTECT COMPONENT**	**GUIDE FLOW**	**ISOLATE COMPONENT**	**PREVENT DAMAGE**	**DELIVER PERFORMANCE**	**RESIST PRESSURE**	**RESIST TEMPERATURE**
Part 1	#	€	%		%				%		
Part 2	**#**	**€**			%						
Part 3	**#**	**€**				%			%		%
Part 4	**#**	**€**					%		%	%	
Part 5	**#**	**€**				%		%		%	%
Part 6	**#**	**€**		%	%		%			%	
TOTALS											
Direct Costs EXCLUDE all Fixed burden & Overhead, S,G,&A, Margin											
FUNCTION - PERCENTAGE:			%	%	%	%	%	%	%	%	%

Figure 33. Function-Resource Matrix- © 2020 Baker Hughes, LLC - All rights reserved.

Chapter 5.

Case Study

The methodology presented in the previous section is applied respectively to the parametric cost model and VAVE technique, in the general framework of product cost management.

In particular, the parametric Cost Model deals with cost modelling for gas turbine components, and more in general components for Oil & Gas application, both in preliminary and detailed design stages

VAVE Technique is applied to components of a gas turbine to increase product value and can be used both in preliminary and detailed design stages.

All those aspects and analysis contribute to Product Cost Management, which focuses on the cost aspects of gas turbine configurations in the early design stages.

In particular, in this chapter, case studies presented are related to gas turbine components (blades, discs, and spacers) and have been analyzed through the methodology presented in this thesis work.

Every new product developed needs to reach the target cost, to be competitive on the market. Parametric cost evaluation helps to check that product cost is aligned to market expectation since the very beginning of a new project. If cost, based on preliminary evaluation, is not aligned with item target, design engineers will leverage VAVE technique to optimize product value. In this chapter a case study of VAVE Methodology is described, focusing in particular on gas turbine blades.

The case studies presented in the following paragraphs are related to gas turbine components and have been developed during conceptual or detailed design phases. The final scope is to determine gas turbine total cost, to do that, it is necessary to estimate the cost of every single sub-component. Moreover, each component will have its target cost and, in the case estimated cost and target cost are not aligned, it is possible to proceed with VAVE Methodology techniques to optimize product value, leveraging functional analysis.

Because the software platform is still a prototype and not fully released and tested, all the steps of methodology have been followed manually to complete the whole assessment on cost and product value.

The blades case study follows each step of the flowchart presented in Figure 13, starting from cost evaluation through parametric costing models, to the application of VAVE methodology to increase product value, lowering product cost while maintaining the same functionality. Since estimated cost is not aligned with target cost, VAVE methodology has been applied to blade design in the conceptual design stage, to optimize product value.

Axial compressors and turbine discs, spacers and shaft have been analyzed with parametric costing models, finding out specific cost drivers for each component type. In this specific case cost model has been evaluated through all the possible methodologies: regression, neural network, and random forest.

VAVE methodology application has not been necessary for discs design as will be described in next paragraphs, cost evaluated with parametric cost models for discs, did already matched target cost, with a gap lower than 10%.

5.1. Baker Hughes Turbomachinery environment

Baker Hughes Gas Turbines can satisfy every application across the oil and gas market and also in many industrial applications. LNG and pipeline transmission applications are equipped with some of our most innovative solutions. The

design of each machine has the goal of optimizing environmental footprint, operating performance, as well as total cost of ownership. Beyond core equipment, a wide range of supporting systems and specially designed services and upgrades are tailored to the unique scope and specific challenge of each specific field.
Aeroderivative Gas Turbines are evolved from aircraft engine designs, a market leader in the aerospace sector, with more than 5,000 units operating and over 450 million flight hours.
BH product portfolio of aero-derivative gas turbines from 20 to 110 MW offers solutions ideal for pipeline compression, offshore platforms, gas re-injection, and liquefied natural gas plants. BH Gas Turbines guarantee high-quality standards maintaining flexibility, efficiency reaches up to 44% simple-cycle, startup time from 5 to 10 minutes, maintenance intervals of 3 to 4 years for high unit availability, extended fuel flexibility, compact lightweight design, engine swap to minimize stops.

Baker Hughes improved its experience burning several fuel mixtures with high hydrogen content, with more than 70 projects worldwide both with Frame and aero-derivative gas turbines. Aeroderivative gas turbines may burn up to 85% hydrogen, with NOx emissions that can be lowered through water injection into the combustion section.
In this diversified and complex environment, Gas Turbine design development is a challenging process, in particular during the conceptual design stage. In fact, in this stage, many requirements shall be met, such as nominal power and efficiency, manufacturability assessment, machine availability, maintenance time and schedule, and, last but not least, total cost.

5.2. Parametric Cost Estimation

5.2.1. Blades parametric cost estimation

All the process steps described in the previous chapter need to be followed for almost any major component of a gas turbine while developing a new design. The blades case study will be presented in the next paragraph, dealing with a new project dedicated to LNG application but suitable also for simple cycle applications as well as for cogeneration and combined cycles, considering both on-shore and off-shore installations.

Turbine blades (Figure 34) are those components that realize flow path geometry. They oversee the extraction of energy from gas after the combustion chamber (where gas has high temperature and pressure) and increase of flow pressure inside the axial compressor (refer also to [76]).

Blades design is subject to many requirements, from mechanical strength to life duration, and often needs high-quality materials and complex design such as cooling channels to reduce temperature both on the surface and in the core. During new turbine development, the blades design phase starts from a previous design already available, scaled in terms of geometry, and from those preliminary sketches, design engineers start an optimization phase.

Turbine blades can be produced by many technology processes, depending on size and material needed. In particular, there are blades realized from forged parts, machined from bar, or, for most complex design, obtained from the investment casting process.

Figure 34. Example of Blades -

To evaluate parametric cost estimation for blades, the total cost for those components can be split into two main contributions: material and machining cost.

Machining cost can be correlated to the volume of material removed to obtain the final geometry, starting from raw material one. Cost models for those curves can be defined considering some geometrical and process parameters. Cost Estimating Relationship (CER) is the sum of all those contributions: material cost (DM), every machining step, coating, heat treatments, visual controls, and inspections, etc. (DL).

In the equation below, Cost Estimating Relationship (CER) is defined in terms of each cost contribution C_i (example with a generic material A):

$$CER = \sum_i C_i = C_{rmA} + C_{mach} + C_{HT} + C_{coat} + C_{insp} + C_{pack} \tag{11}$$

Blades cost can be evaluated taking into account two main factors, by the application of parametric cost models as presented in Step 1 of Figure 13. The first contribution is the raw material cost (DM), which can be investment casting, forged component, or bar, depending on the blade type, generally supplied by external partners (farm-out), while the second portion is related to all machining and process phases (DL) necessary to reach the final geometry (for example surface deposition, thermal treatment, controls, etc.).

For what concerns gas turbine blades, the regression method has been used, based on a historical dataset from purchased orders (*Step 1.1*). This decision is related to the fact that for this specific component a dedicated tool for should-cost analysis was not available yet. Consequently, to define parametric costing curves, the first activity to perform is to collect all the data from purchase orders and store them into proper clusters (*Step 1.1*). In *Step 1.2*, main cost drivers are identified and in *Step 1.3* parametric costing curve including the details of all process phases and materials involved are defined. More than one proposal can be performed for each curve to be defined, for example using linear, quadratic, or polynomial regression. *Step 1.4* refers to R^2 calculation, needed to measure fitting of parametric curves to real data, and at *Step 1.5*, the best-fit regression curve is selected (choosing the lower order one that satisfies the requirement on R^2 greater than 0.85).

Parametric curves for material cost have been defined after clustering data on a material and production technology basis. Thus, *Material* and *Technology* (forging, machining, or investment casting) are two cost drivers for material cost.

In addition to those two cost drivers, other parameters can be defined as additional drivers for cost. For example, in the case of investment casting, the *Weight of the blade* acts as a secondary driver combined with the main cost driver *Material*. Instead, in the case of blades machined from bar or forged parts, the *Height of the blade* is the most effective secondary driver also in this case combined to the *material*.

After completion of the “parametric estimation” step for the blade component, a comparison between target cost, set by the company based on market needs, and cost evaluated through parametric curves is performed. For the blade case study, the cost evaluated is not matching the target cost and in particular, the gap between values is around 20%, which is not acceptable. VAVE methodology is then used to optimize component value and cost allocation, evaluating alternative designs to perform the same function (see paragraph 5.3).

In this case study, we focused on option b of Table 3 for VAVE process described in section 5.3, that consist in reducing product cost while maintaining to match same functionality for what concern performance and life.

5.2.2. Discs parametric cost estimation: a comparative analysis of parametric cost models methods

While developing Parametric Cost estimation for discs, a comparison between linear regression, neural network, deep learning, and random forest methodologies have been performed for discs components.

The figures below show in detail an example of the forged part and the finished disc for the axial compressor of a gas turbine.

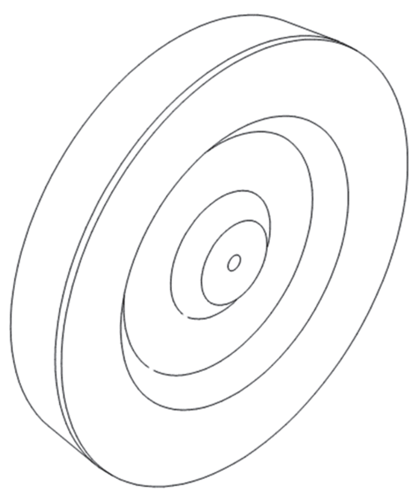


Figure 35. Forged component - © 2020 Baker Hughes, LLC - All rights reserved.

Disc in Figure 35 is the result of the forging process: the shape was obtained starting from a metal connecting rod, through subsequent upsetting actions performed through a press.
So, the cost of the forged part includes:

- The cost of the material: a function of the volume of the material, the density of the material, and the unit cost of the material;
- The cost of the forging process.

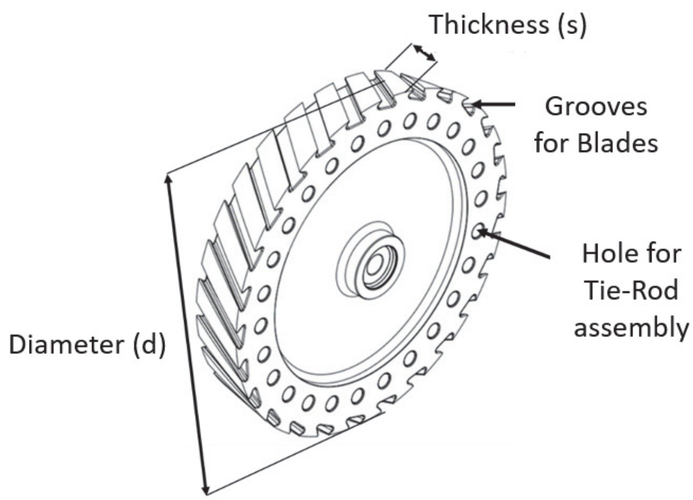


Figure 36. Disc final geometry -

Starting from the forged piece, the semi-finished disc was obtained through turning and any heat treatments.
The finished disc, on the other hand, was represented in Figure 36, compared to the semi-finished product, it has slots at the ends for positioning the axial compressor blades and intermediate holes for assembling the disc with the other components of the gas turbine.

The methodology proposed has been applied to the discs case study as well, and in particular, for this component, a detailed compared study on real data have been performed to evaluate the potential of the three costing model methods: linear regression, neural network (and deep learning) and random forest.

The components that have been studied are the discs of axial compressors and refer to two machines that we will call machine 1 and machine 2.

For what concern processing cost to obtain the finished product, the following cost items were considered:

- The cost of turning;
- The cost of drilling holes;
- The cost of broaching;
- The costs of dimensional inspection and testing.

By adding all the costs, the resulting regression equation was obtained, capable of estimating the total cost of the finished disc.
Carrying out a very detailed analysis, it is possible to identify those operations that are more onerous in terms of cost.
By testing the parametric equation on two new discs, the cost of forging was found to be the most impactful as this operation must be performed in a specific plant and requires very long setup and positioning times.
The results obtained show how the highest uncertainties were detected in the forging operation (prediction differs about 10% versus actual value). All other cost items present lower uncertainties on the final cost (prediction differs less than 2% versus actual value).
The inaccuracies and uncertainties of forging operation and related costs are closely linked to the complexity of the process, characterized by a low level of automation and a high degree of manpower, as described also in [77][78][79].

It is therefore clear that the previous study performed on the axial compressor discs is based on a very detailed empirical linear regression approach and therefore is highly dependent on the analyst's knowledge of the machining processes of the analyzed components.
In this thesis, however, a new "systematic" approach was defined in which the goal was to estimate two cost items (material cost and process cost) without going into the details of all the operations.
Furthermore, with the regression method, the same material and the same forging technique (open die) were always considered.
In this case, new materials were considered (5 in total) and the closed die forging technique was also introduced, different from the open die one, in particular:

- Closed-die forging: the piece is compressed between two shaped molds and assumes the shape of the cavity between them (saves material);
- Open-die forging: not shaped molds and consequently simple final geometries and a greater waste of material.

Consequently, the study was generalized concerning the previous work, as new qualitative (or categorical) variables were introduced and it was carried out, as well as with the traditional regression technique, with innovative non-linear artificial intelligence techniques.
More specifically, Should Cost analysis were initially conducted through the use of the LeanCOST® software.
The results of the analysis have been included in the appendix and constitute the initial database od should cost data (*Step 1.1*). A part of the table represented in the final appendix is shown in Table 5. Subsequently, following the guidelines described, the methods were implemented using the RapidMiner software. In the

following tables, sensitive data has been obscured for confidentiality reasons (in particular costs).

Table 5: Initial Database -

Machine	Cod	d	s	M	N° grooves	R	HT	Batch Size	Raw material process	Weq [Kg]	Wsf	Wr	Veq [m3]	Tot
Disc Machine1	A	AA	AAA	M1	X	I	yes	1	Open-die forging	L	M	N	O	€ -
Disc Machine1	A	AA	AAA	M1	X	I	yes	5	Open-die forging	L	M	N	O	€ -
Disc Machine1	A	AA	AAA	M1	X	I	yes	10	Open-die forging	L	M	N	O	€ -
Disc Machine1	A	AA	AAA	M1	X	I	yes	20	Open-die forging	L	M	N	O	€ -
Disc Machine1	A	AA	AAA	M1	X	I	yes	50	Open-die forging	L	M	N	O	€ -
Disc Machine1	A	AA	AAA	M1	X	J	yes	1	Open-die forging	L	M	N	O	€ -
Disc Machine1	A	AA	AAA	M1	X	J	yes	5	Open-die forging	L	M	N	O	€ -
Disc Machine1	A	AA	AAA	M1	X	J	yes	10	Open-die forging	L	M	N	O	€ -

The database consists of the following input data:

- Finished code (cod);
- Finished diameter (d);
- Finished thickness (s);
- Material (M);
- Number of grooves;
- Roughness (R);
- Presence of heat treatments (HT);
- Batch Size;
- Raw material process;
- Equivalent disc weight (Weq);
- Semi-finished weight (Wsf);
- Raw weight (Wr);

The last column, the total cost (Tot), relates to the output data and derives from the sum of the material cost and the process cost. In particular, the cost of the process is given by the cost of the forming process, processing cost, non-destructive controls and inspections (if any), and more.

Therefore, a single cost item relating to the process was considered that includes all the operations carried out to obtain the finished disc.

Since this is the same component, the production processes are the same as those at the beginning of the paragraph but, in this case, the cost estimate was developed without going into detail in the individual operations.

Each finite code corresponds to a particular type of disc characterized by a specific value of finite diameter, finite thickness, and number of slots (Figure 36); The equivalent disc volume was calculated with data relating to the diameter and thickness of the finished disc.

In particular, a dummy disc (or equivalent) was considered in which, unlike the finished disc in Figure 36, there are no slots for the blades and holes (as if it were a solid disc).

The database has about 90 records that have been generated by varying, for the same disc, the values of non-geometric data (such as the lot or the material). In table 2 nine discs have been studied, from A to I; to these is added another L which, as we will see in the following paragraphs, was used for a second test.

After Parametric cost estimation completion, described from section 5.2.2.1 to 5.2.2.6, the disc estimated cost has been compared to target cost.

Differently from the blades case study, for the discs case study it has been confirmed that the estimated cost is well aligned with the target cost defined for them. Thus in this case, according to Figure 13, it is possible to exit the flow chart proposed and the engineering design will proceed with embodiment and detailed phases as per the standard plan.

5.2.2.1. Database Definition

The initial database, as partially presented in Table 5, is not yet suitable to be used to implement the cost estimation methods but needs a pre-processing phase.

This initial operation was done to identify the weight of each independent variable and to eliminate any outliers or duplicate data.

If the database were considered in its entirety, as reported in the appendix, the linear regression would present too high errors on the final estimate while the machine learning methods would be less precise as there could be irrelevant inputs (on the cost) that would cause a white noise on the final estimate.

In general, in traditional linear regression techniques, it is important to choose only the input data necessary for the analysis; in the innovative techniques of machine learning, it is also possible to consider less "important" inputs in the estimation of costs but we must avoid considering those with zero dependence.

In this regard, a multivariate factorial analysis of variance was carried out to determine the independent variables most correlated with the dependent ones.

As can be seen from Table 6, the values significantly lower than 0.1 (highlighted in green) have been taken into consideration.

Table 6. Factorial Manova -

Input	Output	Significance
Equivalent disc Volume	Process Cost	0,000000000000007
	Material Cost	0,000000000034661
Material	Process Cost	0,000000000000000
	Material Cost	0,000000000000000
Roughness of finite geometry	Process Cost	1,000000000000000
	Material Cost	1,000000000000000
Heat Treatment	Process Cost	0,287756351108742
	Material Cost	0,880266970706251
Batch Size	Process Cost	0,000000000000023
	Material Cost	0,767459104124280
Raw Material	Process Cost	0,002319938911480
	Material Cost	0,072432694138500

Only the equivalent disc volume was considered as geometric input data as it derives from other geometric data characterizing the discs (diameter and final thickness). Note that not a single dependent variable, the total cost, was considered, but the cost of the material and the cost of the process were taken separately. Therefore, as per the requirement of BH [R12], the various models must be able to predict the individual cost items (from which it is however possible to obtain the total).

From the results of the Factorial MANOVA analysis, it is evident that the equivalent disc volume (and therefore the respective geometric data), the material, and the raw material are strongly correlated with the dependent variables. The roughness and the presence of treatments have a very low impact and are therefore negligible while the batch has a behavior dependent on the type of output.

In addition to the MANOVA factor analysis, in the initial pre-processing phase it is important to identify any outliers and duplicate data.
No anomalous data were found, i.e. initial data inconsistent with the rest of the database but, as regards the study relating to the cost of the material, all duplicate data that were created excluding the "batch" input were eliminated.
The duplicates were generated because in building the database with LeanCOST®, often the rows were created keeping all the input values constant and modifying only the batch.
From the results of the factorial MANOVA, there is no dependence between "lot" and "material cost" and consequently, eliminating all the data of the lot duplicates are generated.
A different story, however, for the “process cost” in which no situations similar to the one just described were found.
Therefore, in the case of the cost of the process, the database has the same number of initial records while, for the cost item relating to the material, the number of rows has been significantly reduced (from 90 to 33 rows).
The next step is split data operation, where the data is divided into training and testing: in the case of the cost of the material, 80% of the data was used as training and the remaining 20% as testing; in the case of the cost of the process, the percentages are 70% training and 30% testing.
The subdivision was not done randomly: in the case of the "material cost" more training data were considered as the database is considerably reduced compared to the "process cost" and the materials considered have very different unit costs. The idea was to perform training with a greater number of training data to overcome these issues.
Despite this distinction in the subdivision of data in favor of the "material cost", more precise results have been obtained with the "process cost" demonstrating that the performance of cost estimation techniques is extremely dependent on the forms of the initial database.

In summary, starting from the initial data, two databases were obtained: one relating to the cost of the material and the other relating to the cost of the processes. For both, the most important drivers (most correlated with the output) were identified and any anomalous or duplicate data was eliminated. From Table 5 it can be seen that not all independent variables are numeric; in particular, the raw material and raw materials are qualitative variables. On RapidMiner, the "random forest" and "deep learning" algorithms can work with mixed variables but this is not true for "linear regression" and "neural network" (*Step 1.2*).

Therefore, the dummy variable technique was used to generate databases consisting of numeric data only. Taking into consideration the qualitative variable "material", five dummy variables were used, each referring to a type of material. As for the starting blank, there are two binary variables, one of which is related to the open die forged case, the other to the closed die forged case. In the first line, the conditions material = M1 and raw material = Forged with closed mold are satisfied; obviously, there will never be two conditions of the same reference variable satisfied for the same row (see Table 7 for reference).

Once the subsets of data from training and testing have been defined, you can proceed with the implementation of the methods.

Table 7. Dummy Variable -

Material = M1	**Material = M2**	**Material = M3**	**Material = M4**	**Material = M5**	**Raw Material = Open-die forging**	**Raw Material = Closed-die forging**
1	0	0	0	0	0	1
0	1	0	0	0	1	0
0	0	0	1	0	1	0
1	0	0	0	0	1	0
1	0	0	0	0	0	1

5.2.2.2. Methods implementation

Starting with the training data, the various methods were implemented on the RapidMiner software. The first method analyzed is linear regression, the starting databases are the numerical ones (Table 8 and Table 9).

Table 8. Material Cost Model -

Material Cost	
Attribute	**Coefficient**
Intercept	a
Material = M1	b
Material = M2	c
Material = M3	d
Material = M4	e
Material = M5	f
Raw Material = Open-die forging	g
Raw Material = Closed-die forging	h
Diameter	i
Thickness	j
Finished weight [Kg]	k

Table 9. Process Cost Model -

Process Cost	
Attribute	**Coefficient**
Material = M1	l
Material = M2	m
Material = M3	n
Material = M4	o
Material = M5	p
Raw Material = Open-die forging	q
Raw Material = Closed-die forging	r
Batch Size	s
Disc equivalent weight [Kg]	t
Disc equivalent volume [m3]	u
Intercept	v

Where coefficients from a to v, are numerical constants, here not shared in their specific value due to intellectual properties reason.

5.2.2.3. Multiple Linear Regressions

Parametric equations of multiple linear regressions have the same form as equation 2 and are obtained from the sum of the products between the attributes of Table 8 and Table 9 and the relative coefficients (Equation 12 and Equation 13) (*Step 1.3 for regression*).

$$Cmaterial = \mathrm{a} - \mathrm{b} * M1 - \mathrm{c} * M2 + \mathrm{d} * M3 + \mathrm{e} * M4 - \mathrm{f} * M5 + \mathrm{g} * \mathrm{open\_die} - \mathrm{h} * \mathrm{closed\_die} - d * \mathrm{i} - t * \mathrm{j} + Weq * \mathrm{k} \tag{12}$$

$$Cprocess = \mathrm{v} - \mathrm{l} * M1 - \mathrm{m} * M2 + \mathrm{n} * M3 + \mathrm{o} * M4 - \mathrm{p} * M5 + \mathrm{q} * \mathrm{open\_die} - \mathrm{r} * \mathrm{closed\_die} - \mathrm{batch\_size} * \mathrm{s} - \mathrm{u} * Veq - \mathrm{t} * Weq \quad (13)$$

The goodness of the equations obtained is given by the R^2 coefficient, which in the material cost model assumes a value of 0.90 while in the process cost model it is equal to 0.84. Being acceptable values, the two models were confirmed and used in the subsequent test phases (*Step 1.4 for regression*).

As regards the innovative cost estimation techniques of machine learning, we know that determining the optimal value of the initial parameters (for example number of hidden neurons in the case of neural networks or the number of trees in the case of random forest) is not an operation. simple and requires a parametric optimization process.

Here below some cases that have been analyzed.

5.2.2.4. Simple Neural Network

On RapidMiner, the use of the "neural network" function requires the use of the numeric database with binary variables. The empirical equation 7 was used to determine the number of hidden neurons.

Figure 37 shows the architectures of the simple neural networks of both cases studied.

The segments that connect the neurons of two successive layers have different shades depending on the value of the weight of the connection.

In Figure 37, on the left, the architecture of the neural network relating to the material cost has been represented, while on the right that relating to the cost of the process (Step 1.3 for neural network).

The number of neurons contained in the various layers is identical in the two cases, while the value of the characteristic parameters is different.

As regards the “number of training cycles” parameter, the parametric optimization function was used to determine the optimal value: possible values were entered in the input and the corresponding “relative error leninet” cost item was calculated. RapidMiner automatically calculates the different iterations to determine the optimal value of training cycles.
The results of the parametric optimization were reported on two graphs in Figure 38 and Figure 39 (*Step 1.4 for neural network*).

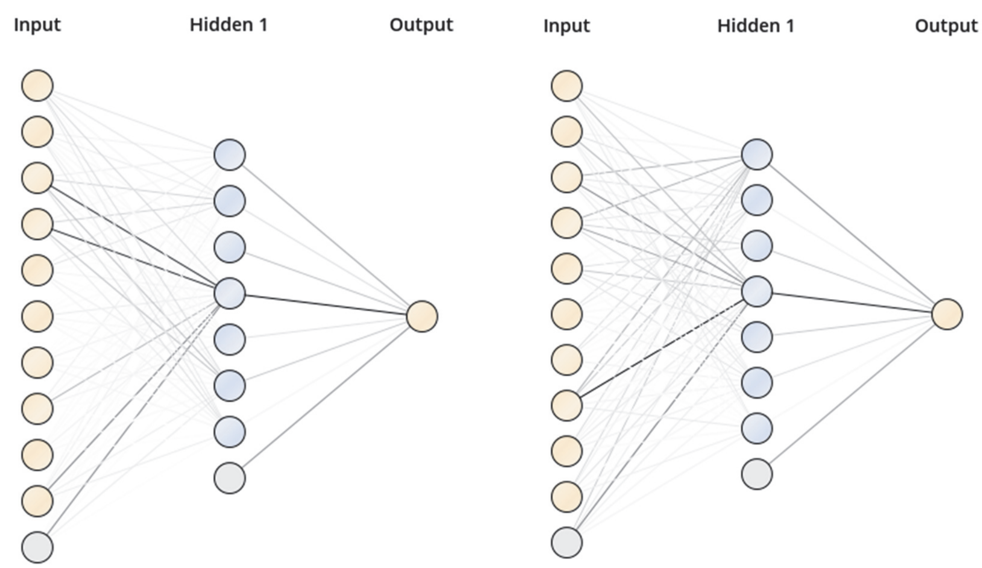


Figure 37. Neural Network architectures - © 2020 Baker Hughes, LLC - All rights reserved.

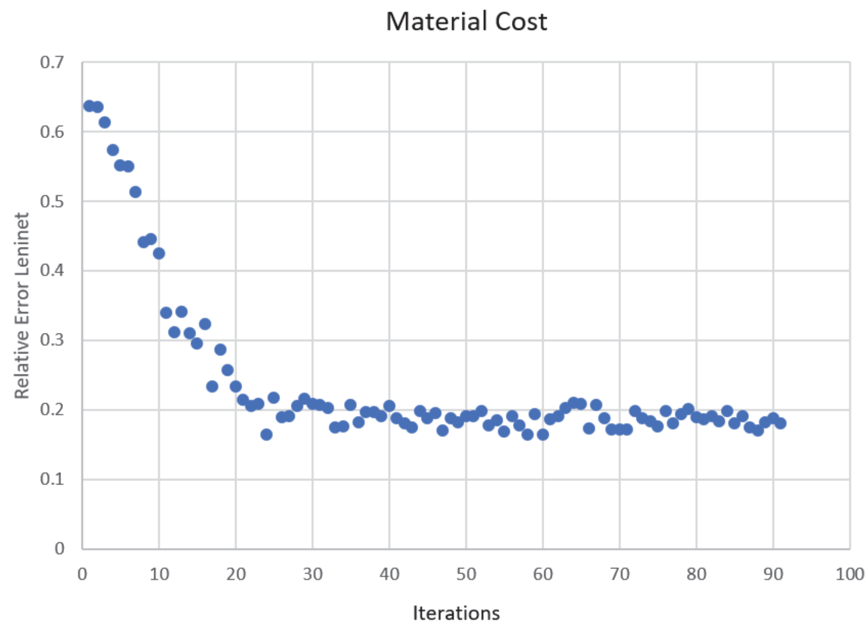


Figure 38. Neural network parametric optimization (material cost) - © 2020 Baker Hughes, LLC - All rights reserved.

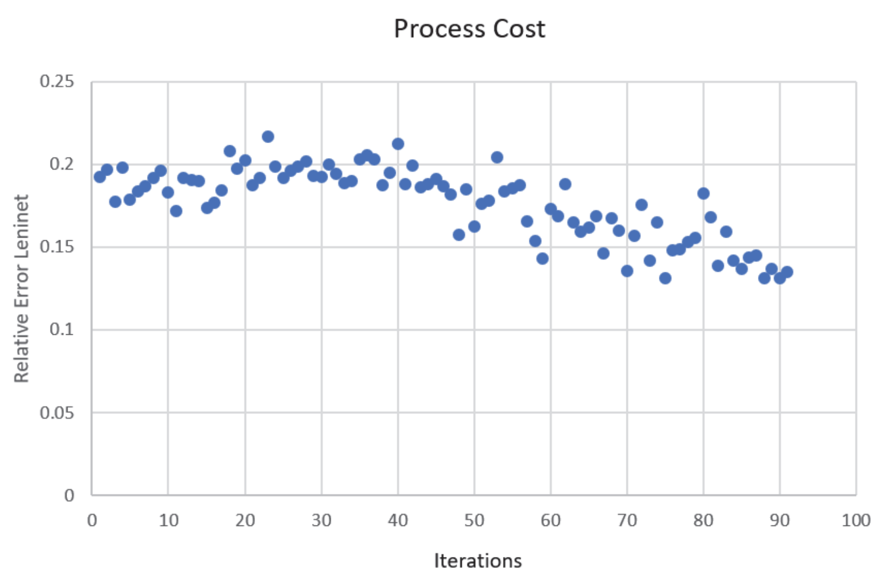


Figure 39. Neural network parametric optimization (process cost) - © 2020 Baker Hughes, LLC - All rights reserved.

These last two graphs show that an increase in the number of training cycles does not always lead to a reduction in error. The goal of this parametric optimization study was to identify the optimal number of cycles to train the neural network both in the case of the cost of the material and in the case of the cost of the process. From Figure 38 it can be seen how iteration 58, characterized by 670 training cycles, has the lowest error equal to 0.164 or 16.4%.
Instead, from Figure 39 the best case is given by iteration number 90 which has 990 training cycles and 13.1% of relative leninet errors.
These results were obtained from training data; in particular, the cross-validation technique was used to generalize the machine learning model avoiding the problem of overfitting.

5.2.2.5. Deep learning

The "deep learning" algorithm can work with a mixed database, consisting of both qualitative and quantitative data.
In this case, the definition of the number of hidden neurons requires a process of parametric optimization, also according to [72]. Although deep learning may have more than two hidden layers, only two with a minimum number of neurons of 5 and a maximum of 50 (per layer) were considered in this study.
The use of the "deep learning" algorithm on RapidMiner has some advantages over the simple "neural network":

- Can work with a mixed database;
- It’s able to autonomously determine the number of optimal training periods (through the stochastic descent of the gradient). So there is no need to do parametric optimization to find the optimal number of training cycles;
- It’s able to autonomously determine some secondary parameters characterizing neural networks. In particular, while in the simple neural

network these parameters have been left with their default value, in the case of deep learning they have been calculated case by case by enabling "adaptive rate".

In this case, the parametric optimization was done by taking as reference parameters the number of neurons hidden in the first and second hidden layer (*Step 1.3 for deep learning*). Several iterations with corresponding relative error leninet values were calculated and also in this case the results were plotted (Figure 40 and Figure 41) (*Step 1.4 for deep learning*).

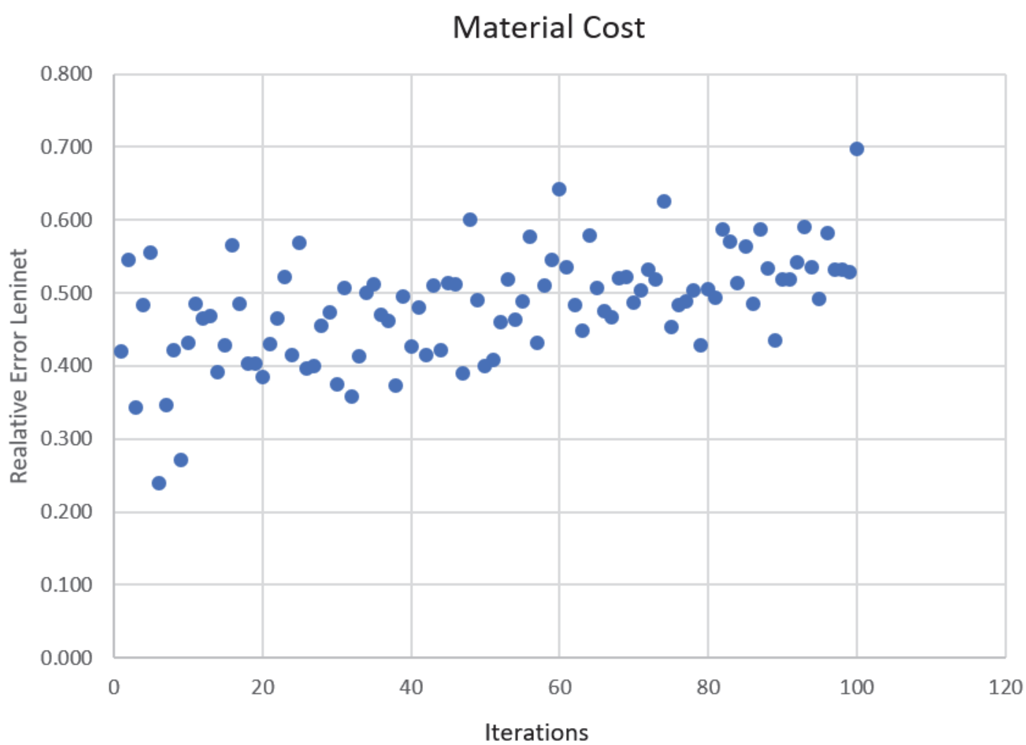


Figure 40. Deep Learning parametric optimization (material cost) -

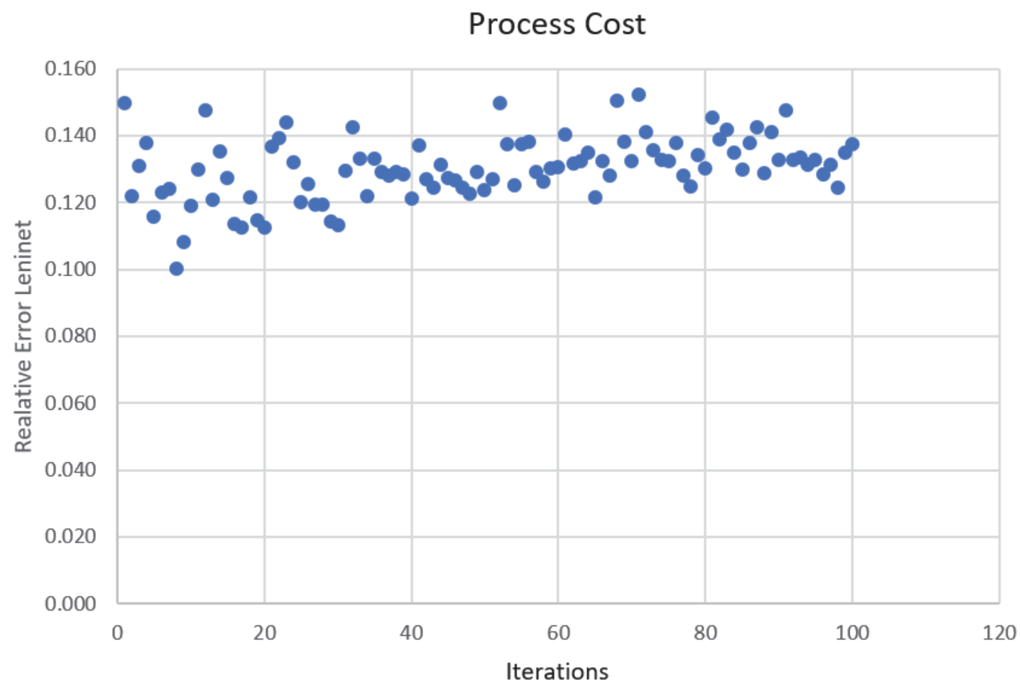


Figure 41. Deep Learning parametric optimization (process cost) -

In this case, it is difficult to interpret the results obtained: as the number of hidden neurons changes, different results are obtained, and it is not possible to predict the best configuration a priori.

For this, parametric optimization is of great help in identifying the optimal values of hidden neurons which in the case of the cost of the material are 5 for layer 1 and 30 for layer 2 (iteration 6). The error is 0.239 or about 24%.

While, as regards the cost of the process, the optimal case is with iteration number 8: 40 neurons in layer 1 and 5 neurons in layer 2. The error is approximately 10%.

For deep learning too, these results were found by applying cross-validation to training data.

5.2.2.6. Random Forest

The latest model that has been implemented is the random forest.

The initial database is mixed, so it is capable of analyzing both qualitative and quantitative variables.

The characteristic parameters that have been studied for the parametric optimization change with respect to the neural network and deep learning and are the number of trees (weak) and the tree depth or maximal depth (Step 1.2 and 1.3 for random forest).

The first indicates the number of weak trees generated by the random forest to predict the output value: the final result is the average of the forecast of each weak tree.

This parameter is directly proportional to the number of subsets generated in the training set as each of them corresponds to a weak decision tree.

As for the maximal depth, RapidMiner allows you to use a “special” value: by entering “-1”, all trees are built until other stopping criteria are met.

The graphs relating to the two cases are shown below (Figure 42 and Figure 43) (Step 1.4 for random forest).

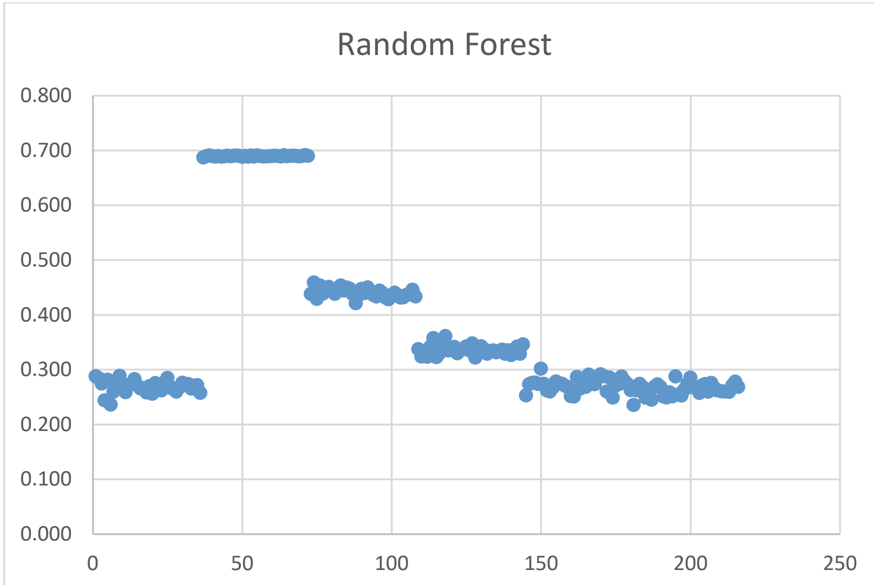


Figure 42. Random Forest parametric optimization (material cost) -

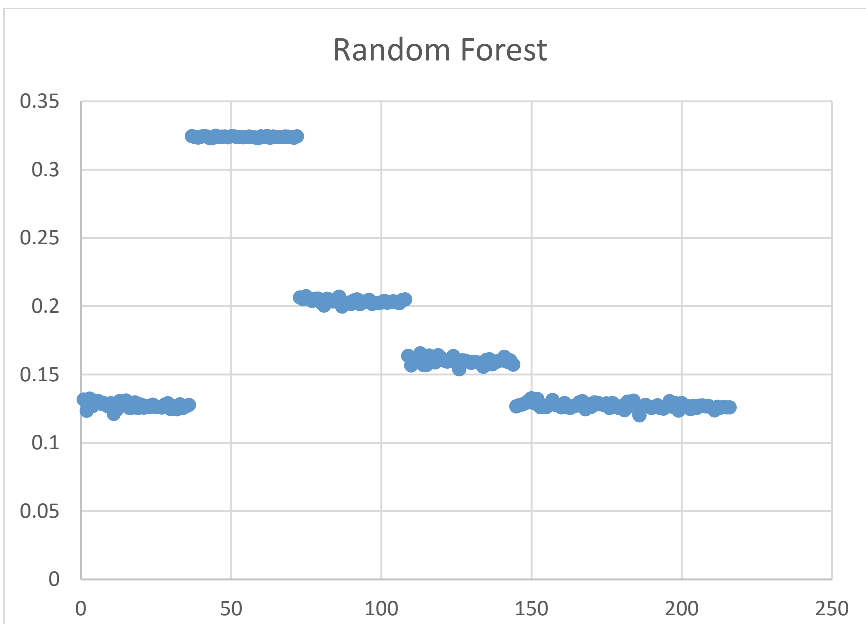


Figure 43. Random Forest parametric optimization (process cost) -

The trend of the points in the two cases of "material cost" and "process cost" is very similar.
In both graphs, point clouds can be identified: each of them corresponds to a value of the maximal depth, therefore, by increasing the value of this parameter, greater precision in the estimate is obtained (the relative error decreases).
The first point cloud is relative to the value of the tree depth of "-1" and returns good results.
As regards the number of trees parameter, a range of values between 50 and 400 was used and it is noted how the results are little dependent on this parameter (the value of the relative error leninet is approximately the same).
It could be concluded in the case of the random forest that parametric optimization is not extremely necessary as in the case of the simple neural network and deep learning since taking a high maximal depth value (for example 10 is a good value) we obtain acceptable results. regardless of the number of trees.
The performance results were obtained with the cross-validation technique.
As for the cost of the material, the lowest error is obtained with iteration number 181 characterized by 50 trees with a depth of 10 (the relative error is about 23.5%).
The most accurate estimate in the case of the cost of the process is obtained with iteration number 186 with 100 trees and with a depth of 10 (about 12% of relative Leninet error).
For what concern Step 1.5, “Chose best-fit cost model among the ones proposed”, it will be discussed more in detail in the next Chapter, comparing results of MAPE calculation for each cost model defined to understand which is the one with the best accuracy.

5.3. VAVE Methodology applied to Blades component

VAVE engineer scope of work for what concern information phase, in this case focusing for gas turbine blades (Step 2.1 of Figure 13), is the collection of all the available expertise inside the company, in particular, dealing with both engineering and manufacturing.

During this stage, it is also important to perform an internal and external benchmark, comparing the analyzed design with other gas turbine design both of company's gas turbine and, whenever possible, with other competitor's design, if relative information is public and accessible.

Moreover, a sharing session focusing on the workout scope of work, to look in detail to 3D models and drawings has been very useful, to start committing the team and generate first ideas.

VAVE methodology, as described by SAVE International, may be proficiently applied to both products and processes.

In this specific case study, the team applied VAVE methodology to both product and process in the very same workshop on gas turbine blades.

In fact, in this case, study, raw material impact on total cost (also called DM, direct material, cost of semi-finished blades) and process impact on total cost (DL, direct labor, all the machining costs) were similar.

Due to this fact, a standard method applied only on DM or only on DL wouldn't have been able to address total cost and optimize effectively the product value.

Consequently, two different FAST diagrams have been realized, one dedicated to semi-finished material cost (DM portion of cost) and the other to process cost (DL portion of the cost).

As already explained, to create a correct FAST diagram, all the functions need to be placed in the right position to be able to answer the question "how" from left to right and "why" from right to left. This function chain moves from the higher-order function to the lower one, from left to right.

DL FAST diagrams can be seen as a high-level machining cycle and manufacturing expertise played a crucial role in its definition.

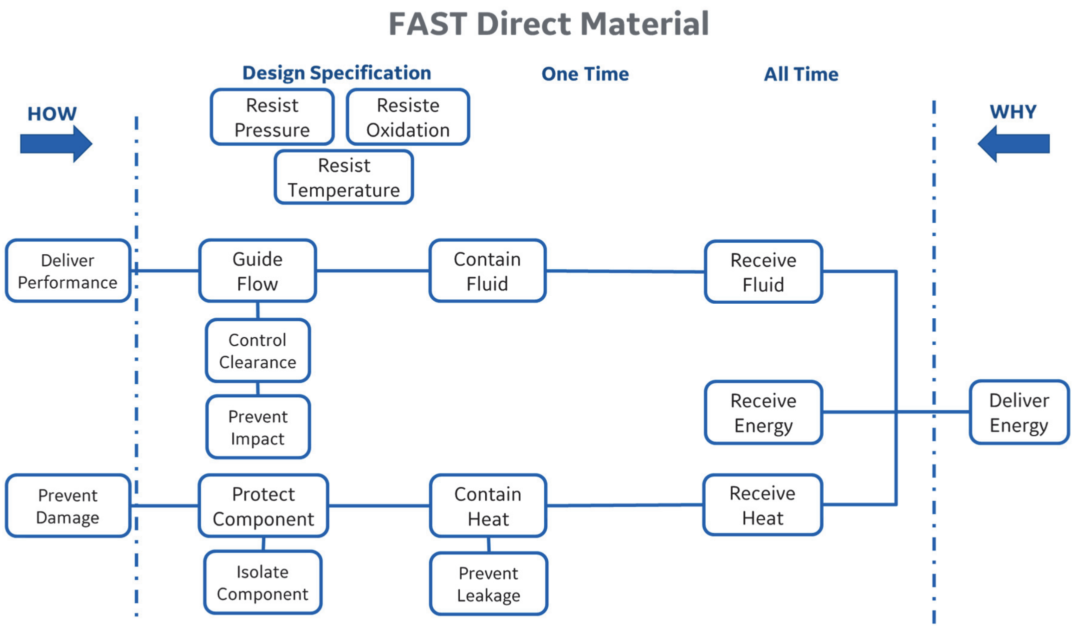
After completing Step 2.2 for both contributions (DM and DL), all the functions were correctly identified, and FAST diagrams were built. Refer to Figure 44 and Figure 45 for both DM and DL FAST Diagrams.

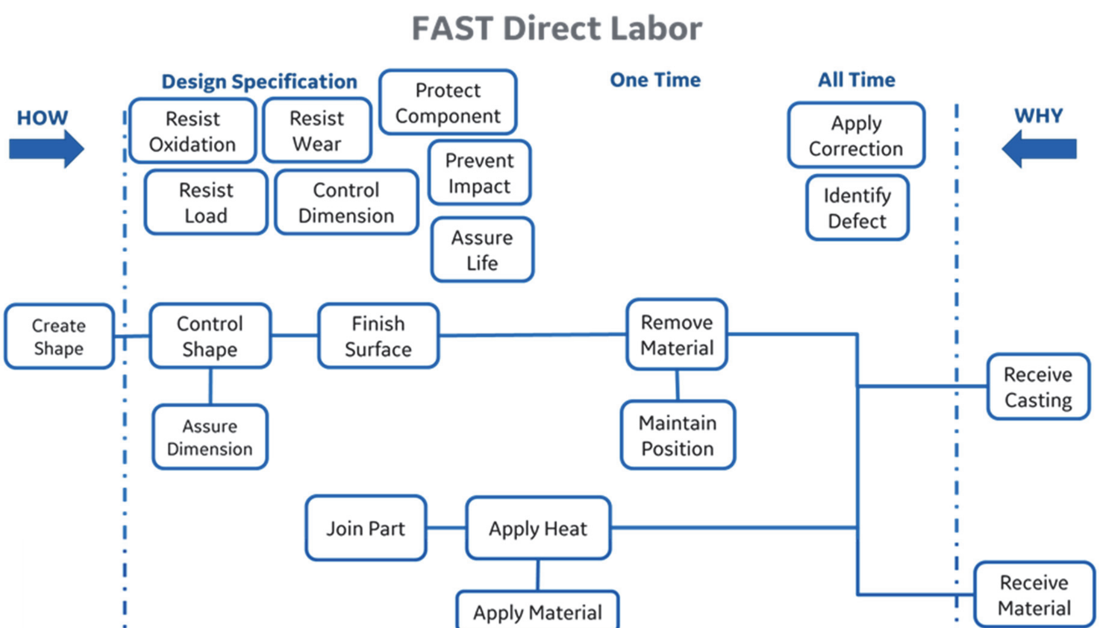
As shown in FAST Diagrams, it is possible to underline that *Higher Order function "Deliver Performance"* for DM (referring to the whole gas turbine) but also *"Prevent Damage"* (referring to casings), while is to *"Create Shape"* for DL, machining the semi-finished geometry to reach final required surfaces.

The need that guide blade design is to assure gas turbine performances, while the need for DL portion of the cost is to achieve the final geometry through process steps.

DM Basic functions are then identified as *"Guide Flow"* and *"Protect Component"* while DL Basic Function is *"Control Shape"*.

Lower Order functions are respectively for *DM "Deliver Energy"* and *"Receive Casting"* and for DL *"Receive Material".*


FAST Direct Material
Design Specification
One Time
All Time
HOW
WHY
Resist Pressure
Resiste Oxidation
Resist Temperature
Deliver Performance
Guide Flow
Contain Fluid
Receive Fluid
Control Clearance
Prevent Impact
Receive Energy
Deliver Energy
Prevent Damage
Protect Component
Contain Heat
Receive Heat
Isolate Component
Prevent Leakage


FAST Direct Labor
Design Specification
One Time
All Time
HOW
WHY
Resist Oxidation
Resist Wear
Protect Component
Prevent Impact
Resist Load
Control Dimension
Assure Life
Apply Correction
Identify Defect
Create Shape
Control Shape
Finish Surface
Remove Material
Receive Casting
Assure Dimension
Maintain Position
Join Part
Apply Heat
Apply Material
Receive Material

After completing the two FAST diagrams, one for DM and the other for DL, as per the modified methodology applied to the blades case study, the team follow steps from 3 to 5, according to standard methodology. It is important to underline that some identified functions, especially those related to the production process, wouldn't have been identified working only on the entire part, without splitting the process between DM and DL.

During the creative phase, as many innovative ideas as possible are generated, focusing on identified functions of the component.

All the ideas collected shall have been written before on a flip chart and then recorder into an excel file. As per the general rule, for each function, at least eighty different ideas have been generated.

After the creative phase, the team proceeded with the evaluation phase and analyzed critically all the ideas proposed in Step 2.3. At the end of this step, from the huge number of ideas proposed, the team select the most interesting ones to create from them some design proposals, during the next development step.

The development step is the very preliminary engineering development of the most promising ideas. Some redesign proposals are discussed and developed by the team, even if, at this stage, it is not a complete and detailed development of change proposed.

In order to proficiently discuss proposals with higher management and to obtain approval to proceed, it has been very important to evaluate the impact of a suggested design change on product value. The team defined, starting from equation (5), all the "Requirements Satisfaction" that are related to customer needs satisfaction and "Product Cost" as a contribution to the total cost, both in current and proposed new design (for sure in this case as a preliminary guess) as described in [23]. For each proposal, Proposal Sheet, Action plan and T-chart have been completed. All the proposals can be categorized into the following macro groups:

- Material change
- Production technology change
- Geometry change
- Coating change
- Insourcing Vs. Outsourcing process

For each group, more than one idea has been analyzed and, after VAVE stream, the two most promising ideas were chosen, based on the relative business case.

After VAVE methodology completion, Team checked and confirmed the applicability of the parametric costing models also for the new proposed design. The proposed design is a geometrical optimization of the current solution (with revised material selection) for which has been confirmed the chance to skip or chance few steps of the production process (regarding for example coating process that can be modified or in some cases also avoided). Referring to those design changes, the correct raw material costing model has been selected for new raw material while, for machining, the same model reported in Table 5 was applied and some items of the equation were not without the need to define new cost curves.

The starting cost estimation value versus the target cost was 30%. After VAVE methodology implementation, the gap between target cost and the new evaluated cost is lowered to the value of 5%, achieving an overall cost reduction of -25% if compared to the initial design.

Chapter 6.

Results and Discussion

This chapter is organized into three paragraphs:
In the first paragraph the full methodology, presented in this thesis work, is followed, discussing the blade case study. In particular, after parametric cost estimation for the component, considering the gap between target and estimated cost, VAVE process has been applied in order to optimize Product Value.
In the second paragraph, the discs case study is presented, which represents a deep dive on parametric cost models, and compares results obtained with different methodologies, such as regressions, artificial neural network, random forest.
The last paragraph includes a discussion that, starting from blades and discs case studies, underlines general outcomes for the gas turbine products.

6.1. Results related to Blades component

6.1.1. Blades Parametric Cost Estimation

Cost assessment is significant since the conceptual design stage, to assure avoiding re-work in later stages when the cost will be fixed by design decisions. Cost accuracy is linked to lack of information at this stage when assumptions are defined and need to be confirmed later on during the project. Design and

VAVE engineers together analyze cost and complete an estimation with its relative variability range based on those assumptions.

Parametric cost estimation for both blades and discs of the gas turbine has been presented and discussed as case studies.

Figure 46 shows parametric costing curves related to the *blade's height* cost driver, considering machining and forging technologies. In particular, the graph shows four options based on different materials (A, B, C and D).

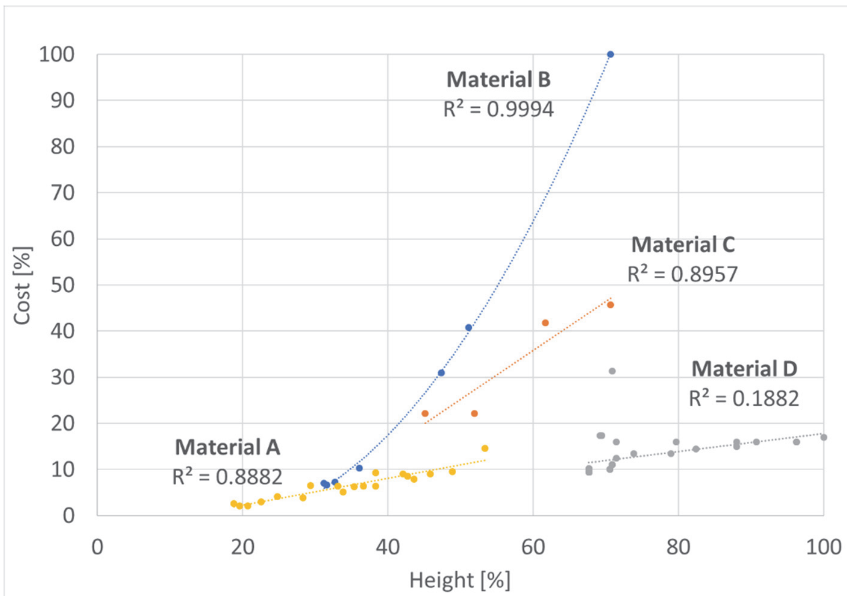


Figure 46. Blades Raw Material Cost Vs. Height (original curves including all data) - © 2020 Baker Hughes, LLC - All rights reserved.

Material A, C, and D equations are obtained by linear regression, on the contrary, Material B is described by a quadratic regression. R squared values show a good correlation has been already achieved for materials A, B, and C, while material D is currently not so well described by the regression curve defined (R^2 lower than 0.85). A poor correlation needs a deeper investigation. In the case of Material D, three data from place orders are impacted by acceleration fees, paid to the supplier to reduce the time of procurement, while one other order was placed for an important batch size quantity and this aspect helped during the commercial phase to decrease unitary cost. Having collected all this information and understood why those data are outliers, relative place order values have been removed from the analysis. Once this data refinement has been completed, Steps 1.4 and 1.5 have been repeated and the parametric costing curve calculated again. Material D is now described from the equation in Figure 47 and updated R^2 rises from 0.19 to 0.88, which is now an accurate value according to defined standards.

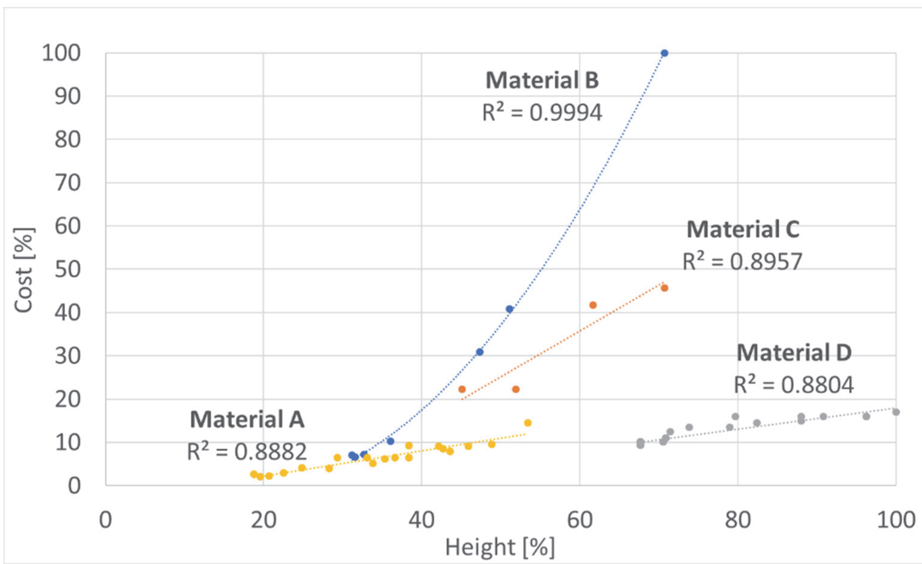


Figure 47. Blades Raw Material Cost Vs. Height (modified curves excluding outlier data for Material D) -

All the data used for the assessment of the parametric cost curves are reported in Table 10 below. In particular, details on all the cost models that contribute to CER are presented, having considered for raw material option 9 different materials, depending on technology selected (machining, forging, or investment casting). Moreover, details of cost model curves are listed for machining, heat treatment, coating, inspection, and packaging, similarly to what is described in [48].

Table 10: Cost Contributions to CER - © 2020 Baker Hughes, LLC - All rights reserved.

C_i	Technology	Input parameters	Regression Equation Type	# of Data Points	R^2
C_{rmA}: Raw Material Cost applicable only for Material A	Machining	Blades radial size (Height)	Linear	18	0.88
C_{rmB}: Raw Material Cost applicable only for Material B	Machining	Blades radial size (Height)	Quadratic	6	0.99
C_{rmC}: Raw Material Cost applicable only for Material C	Machining	Blades radial size (Height)	Linear	4	0.89
C_{rmD}: Raw Material Cost applicable only for Material D	Machining	Blades radial size (Height)	Linear	12	0.19. After Data Refinement 0.88
C_{rmE}: Raw Material Cost applicable only for Material E	Forging	Blades radial size (Height)	Linear	6	0.92
C_{rmF}: Raw Material Cost applicable only for Material F	Forging	Blades radial size (Height)	Linear	11	0.93
C_{rmG}: Raw Material Cost applicable only for Material G	Forging	Blades radial size (Height)	Linear	6	0.99
C_{rmH}: Raw Material Cost applicable only for Material H	Investment Casting	Blade Weight	Linear	7	0.95
C_{rmI}: Raw Material Cost applicable only for Material I	Investment Casting	Blade Weight	Linear	9	0.92
C_{mach}: Machining Cost	All	Cutting Speed and Delta Weight between raw and machined geometries	Quadratic	40	0.86
C_{HT}: Heat Treatment Cost	All	Heat Treatment Hours	Linear	40	0.88
C_{coat}: Coating Cost applicable only if present	All	Coating thickness	Linear	32	0.85
C_{insp}: Inspection Cost	All	Number of CTQ dimensions	Linear	40	0.87
C_{pack}: Pack and Shipping Cost	All	Weight of Blades	Linear	40	0.89

Each operation executed inside the company can be estimated in terms of cost effort leveraging internal knowledge, such us with the support of the manufacturing team, to achieve an increased accuracy.

Activities performed by external suppliers will need a deeper investigation (i.e. for investment casting processes or coatings application). Engage suppliers since the early project phase is very useful to both increase cost predictability of component purchased externally, but also to receive feedback on project decisions, leveraging their knowledge about the production process.

Thinking of products completely manufactured outside the company, cooperation and collaboration with suppliers have an important role in achieving cost targets and completing cost estimation in preliminary stages with enough accuracy.

Equations and all the data used for modelling are proprietary information and consequently, they cannot be showed, even though the methodology described can be considered universal and followed even to analyze different types of components. Case studies described demonstrating reliably that applying all the steps of the method is effective. Considering R^2 values, accuracy on blades cost evaluation between parametric and data points methods are never higher than 3%.

6.1.2. VAVE Methodology applied to Blades component

Considering that the cost estimation is higher than the target cost and the gap is more than 10%, VAVE experts supported and gave their contribution facilitating the VAVE Methodology approach usage. In this specific case, VAVE has been applied to gas turbine blades with good results.

The team allocated resources to functions on each one of the two FAST diagrams and, before proceeding with the following Step 2.3, collected all the functions together, ranking in terms of percentage of function cost on total component cost. The percentage of the cost of function on the total part cost (composed of DM and DL contributions) is shown in red in Figure 48 and Figure 49 for the main contributions.

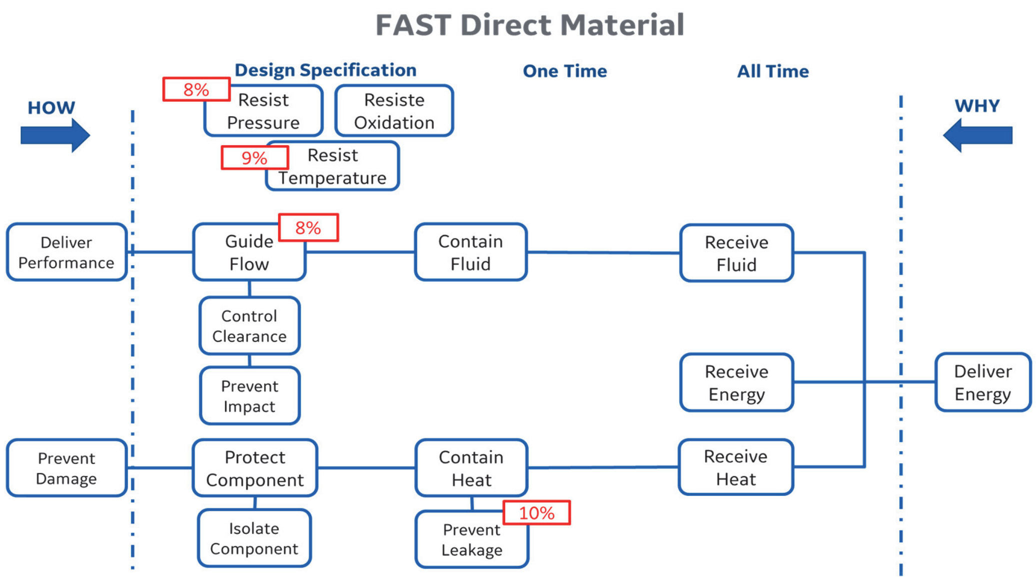


Figure 48. FAST DM for blades from investment casting technology -

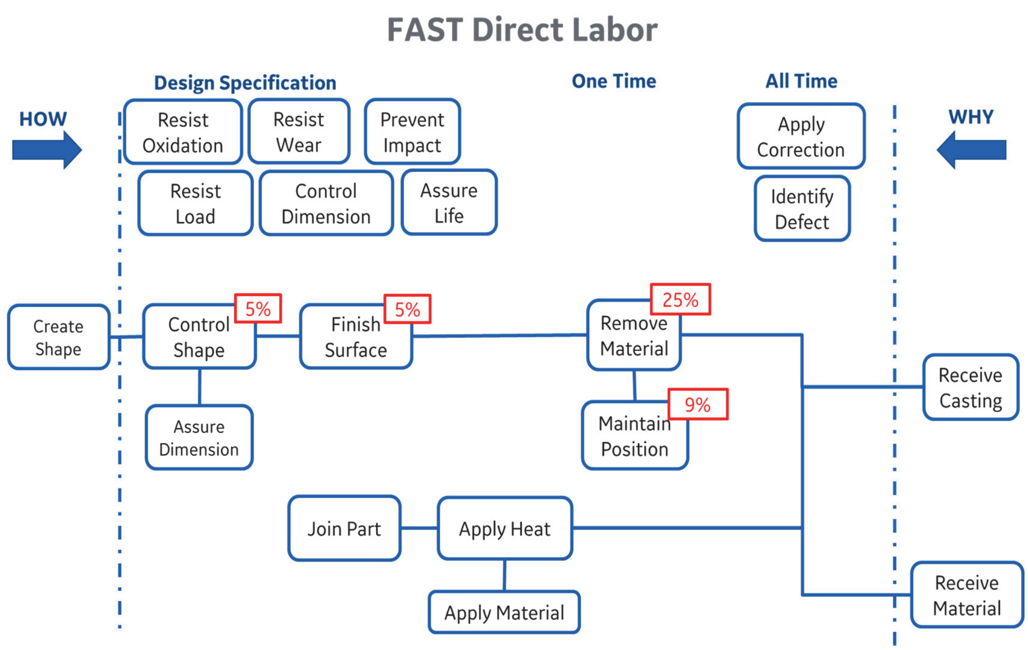


Figure 49. FAST DL for blades from investment casting technology -

Following this procedure, the team discovered which functions have a major impact on cost, independently if generated for DM or DL FAST diagram.

Four functions come from DM FAST and four from DL one, this shows once again the contribution of splitting analysis between DM and DL FAST diagrams.

Completed Step 2.2 following the enhanced methodology, VAVE process continues with Creative Phase (Step 2.3), acting on functions identified, and to the other Steps 2.4, 2.5, and 2.6 considering all the standard rules. In the blades case study, once VAVE has been completely followed, team members controlled and confirmed that the parametric curves for cost estimation were still usable for the new design. This has been possible thanks to the fact that the proposed redesign is a geometrical optimization of the current design (material change) including the removal of a few steps of the manufacturing process (surface deposition). As a consequence of that, some equation factors were removed and material cost changed, causing a minor impact on cost curves.

The team verified that gap between target cost and new estimated cost has now been lowered to *-5%*, having obtained an overall cost reduction of *-25%* through VAVE methodology application compared with the original design.

Figure 50 shows the effect of this methodology applied to gas turbine blades, after collecting DM and DL analysis, and summarizes the impact on the total cost of functions identified for this component. Figure 50 shows cost breakdown represented by a function, so important to identify which function it is worth analyzing first on the idea generation phase: it is a Pareto representation of the cost of each function the component is willing to absolve. As introduced before, in this case study, the authors focused on cost reduction as the major tool to improve product value. Pareto of cost breakdown by function is a key indicator during the following phase of idea generation: functions that impact the most on cost, are the ones to focus on to find an alternative solution to absolve the same function with a lower cost. For sure, if the function not only has a big impact on cost but also is not a primary function, it shall be prioritized to focus on during the idea generation phase.

In the blades case study, four over eight major contributions to total cost come from DM fast diagram (Prevent Leakage, Resist Pressure, Resist Temperature, Guide Flow). The other four contributions come from DL FAST, the most important is the Remove Material function, linked to 25% of the total cost. Analyzing those eight functions, and focusing on idea generation for them, some proposals came out and 77% of the total cost is impacted by design to cost ideas. Prioritization based on cost breakdown by functions is so important to guide idea generation (Step 2.3) and design to cost proposal (Step 2.4 and 2.5).

Otherwise, instead of impacting on major contributors to cost, we would have time-consuming activities addressing lower portions of the total cost.

Completed creativity phase (Step 2.3), all the ideas generated have been categorized into the following macro groups:

- Material change
- Production technology change
- Geometry change
- Coating change
- Insourcing Vs. Outsourcing process

VAVE expert led the team to remove not applicable ideas and duplicates, and then to classify the other ones into groups (Step 2.4). After this analysis, the team started building solid proposals, estimating savings, benefits, drawbacks, and the cost of implementation and defining a complete business case (Step 2.5). At the very end, twenty proposals have been presented associated with a preliminary business case, saving, the cost for investment, and planning (Step 2.6).

All proposals all described by their own Technical Impact, Risk Evaluation, and Saving Confidence parameters. All design changes shall be approved by the correct level of authority, depending on the level of impact on gas turbine design (low, medium, or high). Once all the approved ideas have been implemented, the updated cost of components decreased by 25%, and this allowed to increase total product value.

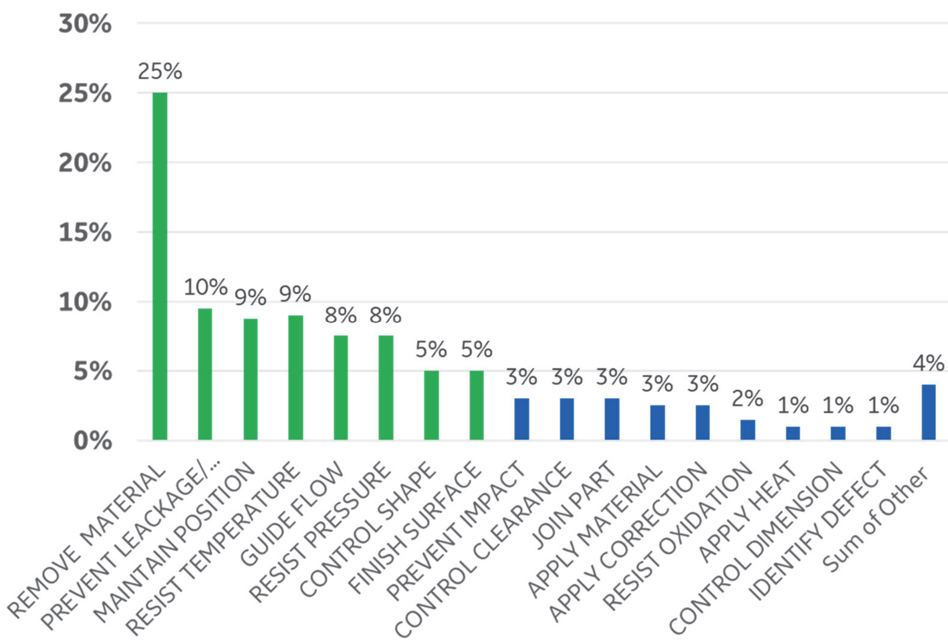


Figure 50. Current Cost Breakdown by Functions - © 2020 Baker Hughes, LLC - All rights reserved.

Considering Equation of Product Value definition and having in mind that the team lowered product cost following option b Table 3, it is possible to write (assuming to have same design response to each requirement):

$$Requirements\ Satisfaction_{afterVAVE} = Requirements\ Satisfaction_{beforeVAVE} \quad (14)$$

And also:

$$\frac{ProductValue_{after\ VAVE}}{Product\ Value_{before\ VAVE}} = \frac{\frac{Requirements\ Satisfaction_{afterVAVE}}{Product\ Cost_{afterVAVE}}}{\frac{Requirements\ Satisfaction_{beforeVAVE}}{Product\ Cost_{beforeVAVE}}} = \frac{\frac{Estimated\ Cost_{before\ VAVE}}{Target}}{\frac{Estimated\ Cost_{after\ VAVE}}{Target}} \tag{15}$$

Simplifying:

$$ProductValue_{after\ VAVE} = \frac{Estimated\ Cost_{before\ VAVE}}{Estimated\ Cost_{after\ VAVE}} \cdot Product\ Value_{before\ VAVE} \tag{16}$$

Achieving a cost reduction of 25% by, thanks to VAVE methodology application, product value of redesigned component is 33% higher.

$$ProductValue_{after\ VAVE} = \frac{Estimated\ Cost_{before\ VAVE}}{0.75 \cdot Estimated\ Cost_{before\ VAVE}} \cdot Product\ Value_{before\ VAVE} = 1.33 \cdot Product\ Value_{before\ VAVE} \tag{17}$$

6.2. Results related to Discs component

Referring to Discs cost modelling analysis, after the training phase, the various models were tested with new data, different from the training ones.
In particular, two types of tests were made:

1. Test with data within the training range;
2. Test with data outside the training range.

The data of the first test derives from the split data made initially to divide the historical database into training and test data.
In the second test, however, a new disc (finite code "L") was considered that presents data out of range compared to those used to train the models.
Starting from the first case, the "prediction" column (which represents the output value of the different methods) has not been reported in the following tables for reasons of corporate confidentiality.
The column relating to MAPE, which represents the metric used to evaluate performance, has not been obscured (Paragraph 4.4.6, equation 8).
Both in the case of the cost of the material and the cost of the process, two databases have also been built for the test data (one numeric and one mixed).
Furthermore, by excluding the "lot" input data in the case of the cost of the material, as happened in the training, also the test data are considerably reduced compared to the cost of the process.
The results relating to the material cost are as follows (refers to Table 11, Table 12, Table 13, Table 14):

Table 11: Linear Regression Results (Material Cost) -

Linear Regression										
M = M1	**M = M2**	**M = M3**	**M = M4**	**M = M5**	**Raw Material = Open-die forging**	**Raw Material = Closed-die forging**	**d**	**s**	**Weq [Kg]**	**MAPE**
0	1	0	0	0	0	1	A	AA	LL	45,0%
0	0	1	0	0	1	0	A	AA	MM	22,4%
0	0	1	0	0	0	1	A	AA	MM	42,0%
0	0	1	0	0	0	1	F	GG	LLL	11,0%
0	0	0	1	0	0	1	F	GG	LLL	21,6%
0	0	0	0	1	0	1	H	II	Q	10,0%
0	0	0	1	0	0	1	H	II	M	16,4%
0	0	0	1	0	0	1	H	EE	PP	38,1%
MAPE										**25,8%**
Standard Deviation										**13,0%**

Table 12: Simple Neural Network Results (Material Cost) -

Simple neural network										
M = M1	**M = M2**	**M = M3**	**M = M4**	**M = M5**	**Raw Material = Open-die forging**	**Raw Material = Closed-die forging**	**d**	**s**	**Weq [Kg]**	**MAPE**
0	1	0	0	0	0	1	A	AA	LL	5,77%
0	0	1	0	0	1	0	A	AA	MM	2,18%
0	0	1	0	0	0	1	A	AA	MM	8,45%
0	0	1	0	0	0	1	F	GG	LLL	16,96%
0	0	0	1	0	0	1	F	GG	LLL	12,75%
0	0	0	0	1	0	1	H	II	Q	40,83%
0	0	0	1	0	0	1	H	II	M	13,69%
0	0	0	1	0	0	1	H	EE	PP	5,88%
MAPE										**13,31%**
Standard Deviation										**11,35%**

Table 13: Deep Learning Results (Material Cost) -

Deep Learning					
d	**s**	**M**	**Raw Material**	**Weq [Kg]**	**MAPE**
A	AA	M2	Closed-die forging	LL	18,33%
A	AA	M3	Open-die forging	MM	4,85%
A	AA	M3	Closed-die forging	MM	13,23%
F	GG	M3	Closed-die forging	LLL	15,70%
F	GG	M4	Closed-die forging	LLL	3,69%
H	II	M5	Closed-die forging	Q	4,36%
H	II	M4	Closed-die forging	M	8,36%
H	EE	M4	Closed-die forging	PP	29,37%
MAPE					**12,24%**
Standard Deviation					**8,28%**

Table 14: Random Forest Results (Material Cost) -

Random Forest					
d	**s**	**M**	**Raw Material**	**Weq [Kg]**	**MAPE**
A	AA	M2	Closed-die forging	LL	21,63%
A	AA	M3	Open-die forging	MM	14,87%
A	AA	M3	Closed-die forging	MM	10,16%
F	GG	M3	Closed-die forging	LLL	15,95%
F	GG	M4	Closed-die forging	LLL	2,70%
H	II	M5	Closed-die forging	Q	20,49%
H	II	M4	Closed-die forging	M	38,65%
H	EE	M4	Closed-die forging	PP	24,88%
MAPE					**18,67%**
Standard Deviation					**10,00%**

For what concerns the results of the cost of the process, the test data are more numerous due to the absence of duplicates.

This difference between the two databases implies greater reliability of the results obtained with the "process cost" compared to the "material cost" as the number of records that have been used to carry out the test is much higher.

The tables relating to the “process cost” are shown below (Table 15, Table 16, Table 17, Table 18) and, as in the previous case, the column relating to costs has been obscured while the one relating to MAPE is visible and can be used for final considerations.

Table 15: Linear Regression Results (Process Cost) -

Linear Regression										
M = M1	**M = M2**	**M = M3**	**M = M4**	**M = M5**	**Raw Material = Open-die forging**	**Raw Material = Closed-die forging**	**Batch Size**	**Weq [Kg]**	**Veq [m3]**	**MAPE**
1	0	0	0	0	1	0	1	L	O	2,55%
1	0	0	0	0	1	0	10	L	O	33,26%
1	0	0	0	0	0	1	5	L	O	21,85%
1	0	0	0	0	0	1	20	L	O	11,25%
0	1	0	0	0	1	0	1	L	O	11,44%
0	1	0	0	0	0	1	50	L	O	94,93%
0	0	1	0	0	1	0	1	MM	O	2,11%
1	0	0	0	0	1	0	1	P	SS	9,20%
0	0	0	0	1	1	0	1	T	Z	24,33%
0	0	0	0	1	0	1	5	T	Z	2,00%
0	0	1	0	0	1	0	5	LL	OO	29,02%
0	1	0	0	0	0	1	1	PP	SS	46,76%
0	0	1	0	0	1	0	5	LLL	ZZ	0,51%
0	0	1	0	0	0	1	10	LLL	ZZ	12,70%
0	0	0	0	1	1	0	5	Q	ZZ	26,33%
0	1	0	0	0	0	1	1	Q	ZZ	6,06%
0	0	0	1	0	1	0	1	LLL	ZZZ	33,48%
1	0	0	0	0	1	0	5	MM	ZZZ	32,96%
0	0	1	0	0	1	0	5	LLL	ZZZ	6,72%
0	1	0	0	0	0	1	10	Q	ZZZZ	18,96%
1	0	0	0	0	0	1	10	Q	ZZZZ	29,37%
0	0	0	0	1	0	1	5	Q	ZZZZ	19,85%
0	0	0	1	0	0	1	5	RR	ZZZZ	15,05%
0	0	0	0	1	1	0	10	T	SS	19,08%
MAPE										**21,24%**
Standard Deviation										**19,47%**

Table 16: Simple Neural Network (Process Cost) -

Simple neural network										
M = M1	M = M2	M = M3	M = M4	M = M5	Raw Material = Open-die forging	Raw Material = Closed-die forging	Batch Size	Weq [Kg]	Veq [m3]	MAPE
1	0	0	0	0	1	0	1	L	O	12,05%
1	0	0	0	0	1	0	10	L	O	4,61%
1	0	0	0	0	0	1	5	L	O	17,68%
1	0	0	0	0	0	1	20	L	O	11,02%
0	1	0	0	0	1	0	1	L	O	14,22%
0	1	0	0	0	0	1	50	L	O	7,47%
0	0	1	0	0	1	0	1	MM	O	2,54%
1	0	0	0	0	1	0	1	P	SS	24,19%
0	0	0	0	1	1	0	1	T	Z	13,82%
0	0	0	0	1	0	1	5	T	Z	22,69%
0	0	1	0	0	1	0	5	LL	OO	16,61%
0	1	0	0	0	0	1	1	PP	SS	5,61%
0	0	1	0	0	1	0	5	LLL	ZZ	9,57%
0	0	1	0	0	0	1	10	LLL	ZZ	7,90%
0	0	0	0	1	1	0	5	Q	ZZ	17,20%
0	1	0	0	0	0	1	1	Q	ZZ	10,80%
0	0	0	1	0	1	0	1	LLL	ZZZ	6,44%
1	0	0	0	0	1	0	5	MM	ZZZ	18,90%
0	0	1	0	0	1	0	5	LLL	ZZZ	0,31%
0	1	0	0	0	0	1	10	Q	ZZZZ	3,07%
1	0	0	0	0	0	1	10	Q	ZZZZ	9,35%
0	0	0	0	1	0	1	5	Q	ZZZZ	11,96%
0	0	0	1	0	0	1	5	RR	ZZZZ	7,48%
0	0	0	0	1	1	0	10	T	SS	12,74%
MAPE										**11,18%**
Standard Deviation										**6,09%**

Table 17: Deep Learning Results (Process Cost) -

Deep learning					
Material	**Batch Size**	**Raw Material**	**Peq [Kg]**	**Veq [m3]**	**MAPE**
M1	1	Open-die forging	L	O	15,66%
M1	10	Open-die forging	L	O	2,49%
M1	5	Closed-die forging	L	O	14,47%
M1	20	Closed-die forging	L	O	3,53%
M2	1	Open-die forging	L	O	8,62%
M2	50	Closed-die forging	L	O	8,32%
M3	1	Open-die forging	MM	O	1,66%
M1	1	Open-die forging	P	SS	27,04%
M5	1	Open-die forging	T	Z	7,81%
M5	5	Closed-die forging	T	Z	14,46%
M3	5	Open-die forging	LL	OO	10,60%
M2	1	Closed-die forging	PP	SS	8,85%
M3	5	Open-die forging	LLL	ZZ	11,42%
M3	10	Closed-die forging	LLL	ZZ	4,21%
M5	5	Open-die forging	Q	ZZ	12,17%
M2	1	Closed-die forging	Q	ZZ	4,51%
M4	1	Open-die forging	LLL	ZZZ	4,68%
M1	5	Open-die forging	MM	ZZZ	23,21%
M3	5	Open-die forging	LLL	ZZZ	6,84%
M2	10	Closed-die forging	Q	ZZZZ	7,79%
M1	10	Closed-die forging	Q	ZZZZ	5,29%
M5	5	Closed-die forging	Q	ZZZZ	12,50%
M4	5	Closed-die forging	RR	ZZZZ	12,12%
M5	10	Open-die forging	T	SS	15,54%
MAPE					**10,16%**
Standard Deviation					**6,10%**

Table 18: Random forest Results (Process Cost) -

Random Forest					
Material	**Batch Size**	**Raw Material**	**Peq [Kg]**	**Veq [m3]**	**MAPE**
M1	1	Open-die forging	L	O	2,69%
M1	10	Open-die forging	L	O	2,08%
M1	5	Closed-die forging	L	O	13,17%
M1	20	Closed-die forging	L	O	3,03%
M2	1	Open-die forging	L	O	2,80%
M2	50	Closed-die forging	L	O	6,81%
M3	1	Open-die forging	MM	O	13,66%
M1	1	Open-die forging	P	SS	1,11%
M5	1	Open-die forging	T	Z	0,40%
M5	5	Closed-die forging	T	Z	20,26%
M3	5	Open-die forging	LL	OO	26,19%
M2	1	Closed-die forging	PP	SS	5,09%
M3	5	Open-die forging	LLL	ZZ	3,49%
M3	10	Closed-die forging	LLL	ZZ	8,89%
M5	5	Open-die forging	Q	ZZ	11,80%
M2	1	Closed-die forging	Q	ZZ	8,42%
M4	1	Open-die forging	LLL	ZZZ	20,16%
M1	5	Open-die forging	MM	ZZZ	5,97%
M3	5	Open-die forging	LLL	ZZZ	3,99%
M2	10	Closed-die forging	Q	ZZZZ	8,04%
M1	10	Closed-die forging	Q	ZZZZ	1,07%
M5	5	Closed-die forging	Q	ZZZZ	9,26%
M4	5	Closed-die forging	RR	ZZZZ	14,62%
M5	10	Open-die forging	T	SS	18,03%
MAPE					**8,79%**
Standard Deviation					**6,93%**

For what concern MAPE, in the last two rows of the tables just presented, two values were reported:

- The value at the top indicates the resulting error given by the average of the contributions of the individual test data;
- The bottom value indicates the standard deviation.

The results obtained in the case of the material cost are less accurate than the process cost. The reasons for this difference are attributable to:

- Different number of records of the two initial databases;
- Nature of the data: very different materials (in terms of unit cost) were considered.

The results obtained from the different methods were compared to each other to identify which of these has the best performance (Figure 51 and Figure 52).

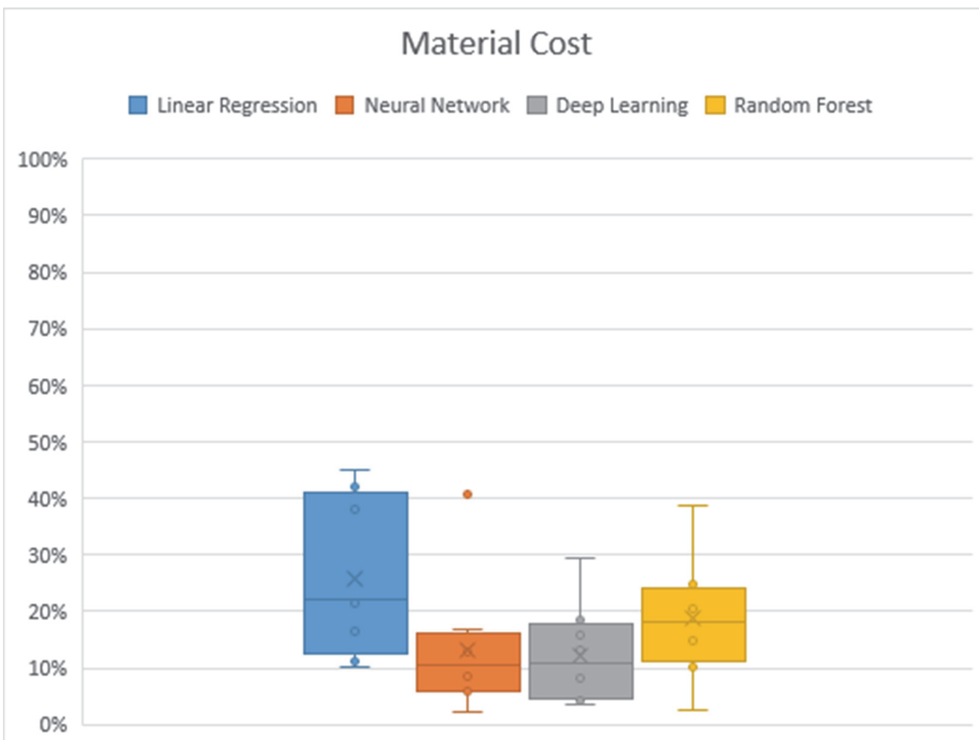


Figure 51. Material Cost Comparison -

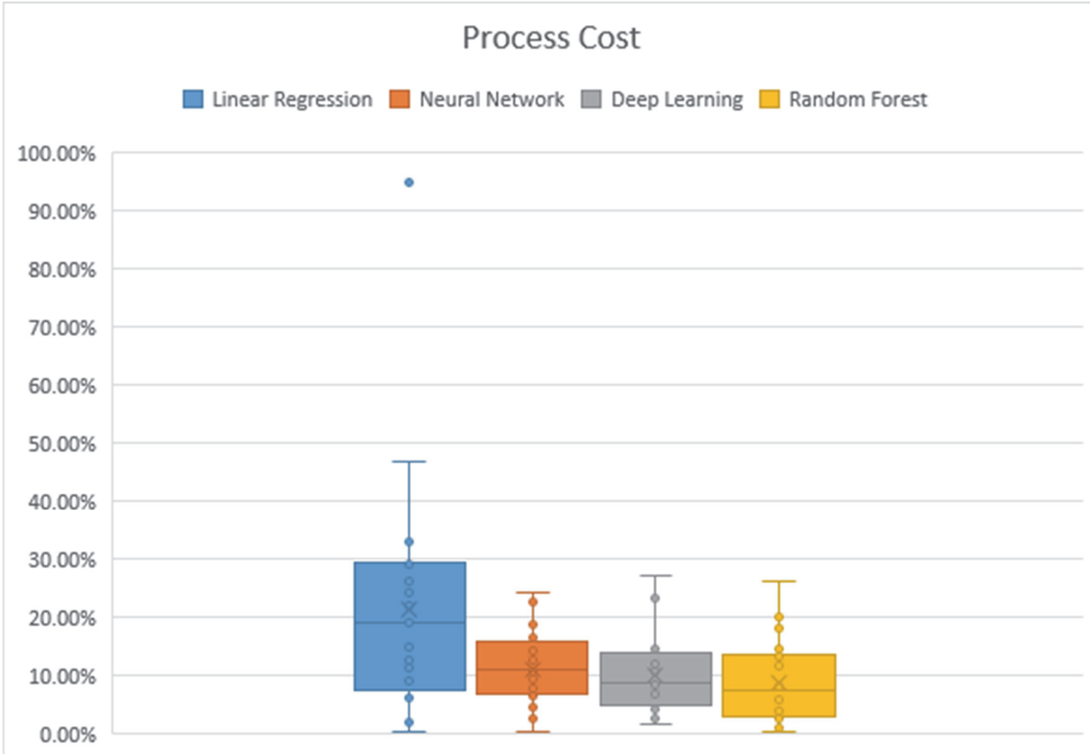


Figure 52. Process Cost Comparison -

Figure 51 and Figure 52 represent two box-and-whisker plots and show the distribution of data relating to the components of the MAPE of each model in quartiles, highlighting the median and outliers:

- Quartiles are those values that divide the population into four equal parts:
- Outliers are those values that are numerically distant from the rest of the collected data. In statistics they are defined as values outside a certain range and, since they can be misleading, it is better not to consider them.

In the graphs of Figure 51 and Figure 52, the rectangles are called "boxes" while the lines that extend vertically are called "whiskers".

The boxes represent the interquartile range, which is the difference between the third quartile (upper border) and the first quartile (lower border).
Within the interquartile range, 50% of the observations fall (therefore the most frequent values are contained in this range).
The line inside the boxes corresponds to the median or second quartile.
An outlier value is a value with a positive deviation from the third quartile greater than 1.5 times the interquartile range or, symmetrically, a value with a negative deviation from the first quartile greater than (in absolute value) than 1.5 times the interquartile range.
Instead, whiskers correspond to the minimum value (lower whisker) and the maximum value (upper whisker) observed after excluding the outliers.
Both in the case of the cost of the material and the cost of the process, the innovative estimation techniques of machine learning were found to be more precise than the traditional linear regression technique.
In the case of the cost of the material, neural networks are more precise than random forest (in particular deep learning). As for the cost of the process, however, random forest is slightly more precise than deep learning (the two have very similar performances).
Overall, a tool could be developed that uses deep learning in the case of the cost of the material and random forest in the cost of the process.
Or, considering that deep learning still has excellent performance even in the second case, a tool could be developed that uses only this algorithm.
The graphs just described allowing you to test the algorithms on a set of test data. In addition to this first check, a second test was made on a new set of data relating to a new disc (finished code "L").
The new disc is part of machine 2 and the database was shown in the next Table 19.

Table 19: Disc “L” Database -

Cod	d	s	M	N° grooves	Batch size	Raw Material	Weq [Kg]	Veq [m3]	Tot
L	I	FF	M2	Y	1	Open-die forging	1P	OOO	€
L	I	FF	M2	Y	5	Open-die forging	1P	OOO	€
L	I	FF	M2	Y	10	Open-die forging	1P	OOO	€
L	I	FF	M3	Y	1	Open-die forging	1Q	OOO	€
L	I	FF	M3	Y	5	Open-die forging	1Q	OOO	€
L	I	FF	M3	Y	10	Open-die forging	1Q	OOO	€
L	I	FF	M1	Y	1	Open-die forging	1R	OOO	€
L	I	FF	M1	Y	5	Open-die forging	1R	OOO	€
L	I	FF	M1	Y	10	Open-die forging	1R	OOO	€
L	I	FF	M2	Y	1	Closed-die forging	1P	OOO	€
L	I	FF	M2	Y	5	Closed-die forging	1P	OOO	€
L	I	FF	M2	Y	10	Closed-die forging	1P	OOO	€
L	I	FF	M3	Y	1	Closed-die forging	1Q	OOO	€
L	I	FF	M3	Y	5	Closed-die forging	1Q	OOO	€
L	I	FF	M3	Y	10	Closed-die forging	1Q	OOO	€
L	I	FF	M1	Y	1	Closed-die forging	1R	OOO	€
L	I	FF	M1	Y	5	Closed-die forging	1R	OOO	€
L	I	FF	M1	Y	10	Closed-die forging	1R	OOO	€

As can be seen in the appendix, 9 discs have been inserted in the initial database that has a thickness between a maximum value of AA and a minimum value EE. In reality, there may be a need to estimate the cost of a component characterized by one or more input data outside the training range. For example, in this case, the disc "L" has a thickness equal to FF, a value greater than the range of the initial database of Table 5 [EE; AA].

For this reason, a second test was made, different from the previous one, in which a disc was used that has new values and in particular a very high thickness and outside the range of the training data.

Also in this case the sensitive data has been obscured. To evaluate the results, Table 20, Table 21, Table 22, Table 23 below present the last column with the values of the deviations, or the difference between the cost generated by the algorithms and the real cost.

Table 20: Test 2 Linear Regression Results (Material Cost) -

Linear Regression										
M = M1	**M = M2**	**M = M3**	**M = M4**	**M = M5**	**Raw Material = Open-die forging**	**Raw Material = closed-die forging**	**d**	**s**	**Weq [Kg]**	**Delta**
0	1	0	0	0	1	0	I	FF	1P	232,32%
0	0	1	0	0	1	0	I	FF	1R	-0,32%
1	0	0	0	0	1	0	I	FF	1Q	512,78%
0	1	0	0	0	0	1	I	FF	1P	340,99%
0	0	1	0	0	0	1	I	FF	1R	33,23%
1	0	0	0	0	0	1	I	FF	1Q	723,90%

Table 21: Test 2 Simple Neural Network Results (Material Cost) -

Simple neural network										
M = M1	M = M2	M = M3	M = M4	M = M5	Raw Material = Open-die forging	Raw Material = closed-die forging	d	s	Weq [Kg]	Delta
0	1	0	0	0	1	0	I	FF	1P	109,76%
0	0	1	0	0	1	0	I	FF	1R	-6,73%
1	0	0	0	0	1	0	I	FF	1Q	212,97%
0	1	0	0	0	0	1	I	FF	1P	113,64%
0	0	1	0	0	0	1	I	FF	1R	19,75%
1	0	0	0	0	0	1	I	FF	1Q	207,48%

Table 22: Test 2 Deep Learning Results (Material Cost) -

Deep Learning					
d	s	M	Raw Material	Weq [Kg]	Delta
I	FF	M2	Open-die Forging	1P	77,32%
I	FF	M3	Open-die Forging	1R	-26,45%
I	FF	M1	Open-die Forging	1Q	248,53%
I	FF	M2	Closed-die Forging	1P	79,81%
I	FF	M3	Closed-die Forging	1R	-4,81%
I	FF	M1	Closed-die Forging	1Q	313,61%

Table 23: Test 2 Random Forest Results (Material Cost) -

Random forest					
d	s	M	Raw Material	Weq [Kg]	Delta
I	FF	M2	Open-die Forging	1P	192,00%
I	FF	M3	Open-die Forging	1R	-30,13%
I	FF	M1	Open-die Forging	1Q	431,04%
I	FF	M2	Closed-die Forging	1P	280,77%
I	FF	M3	Closed-die Forging	1R	-9,43%
I	FF	M1	Closed-die Forging	1Q	603,89%

In this case as well, as regards the cost of the process, the test data have more records and the results are more precise (see Table 24, Table 25, Table 26, Table 27):

Table 24: Test 2 Linear Regression Results (Process Cost) -

Linear Regression										
M = M1	**M = M2**	**M = M3**	**M = M4**	**M = M5**	**Raw Material = Open-die forging**	**Raw Material = Closed-die forging**	**Batch Size**	**Weq [Kg]**	**Veq [m3]**	**Delta**
0	1	0	0	0	1	0	1	1P	OOO	-2,13%
0	1	0	0	0	1	0	5	1P	OOO	-27,31%
0	1	0	0	0	1	0	10	1P	OOO	-27,72%
0	0	1	0	0	1	0	1	1Q	OOO	19,77%
0	0	1	0	0	1	0	5	1Q	OOO	5,23%
0	0	1	0	0	1	0	10	1Q	OOO	4,77%
1	0	0	0	0	1	0	1	1R	OOO	-25,32%
1	0	0	0	0	1	0	5	1R	OOO	-72,70%
1	0	0	0	0	1	0	10	1R	OOO	-75,85%
0	1	0	0	0	0	1	1	1P	OOO	6,18%
0	1	0	0	0	0	1	5	1P	OOO	-22,89%
0	1	0	0	0	0	1	10	1P	OOO	-23,94%
0	0	1	0	0	0	1	1	1Q	OOO	13,19%
0	0	1	0	0	0	1	5	1Q	OOO	-7,55%
0	0	1	0	0	0	1	10	1Q	OOO	-8,76%
1	0	0	0	0	0	1	1	1R	OOO	-9,30%
1	0	0	0	0	0	1	5	1R	OOO	-59,01%
1	0	0	0	0	0	1	10	1R	OOO	-63,00%

Table 25: Test 2 Simple Neural Network Results (Process Cost) -

Simple neural network										
M = M1	M = M2	M = M3	M = M4	M = M5	Raw Material = Open-die forging	Raw Material = Closed-die forging	Batch Size	Weq [Kg]	Veq [m3]	Delta
0	1	0	0	0	1	0	1	1P	OOO	-16,43%
0	1	0	0	0	1	0	5	1P	OOO	-26,34%
0	1	0	0	0	1	0	10	1P	OOO	-6,57%
0	0	1	0	0	1	0	1	1Q	OOO	0,17%
0	0	1	0	0	1	0	5	1Q	OOO	-11,67%
0	0	1	0	0	1	0	10	1Q	OOO	-1,97%
1	0	0	0	0	1	0	1	1R	OOO	-55,36%
1	0	0	0	0	1	0	5	1R	OOO	-85,84%
1	0	0	0	0	1	0	10	1R	OOO	-56,84%
0	1	0	0	0	0	1	1	1P	OOO	4,55%
0	1	0	0	0	0	1	5	1P	OOO	-12,65%
0	1	0	0	0	0	1	10	1P	OOO	-1,74%
0	0	1	0	0	0	1	1	1Q	OOO	-2,17%
0	0	1	0	0	0	1	5	1Q	OOO	-16,43%
0	0	1	0	0	0	1	10	1Q	OOO	-3,29%
1	0	0	0	0	0	1	1	1R	OOO	-21,53%
1	0	0	0	0	0	1	5	1R	OOO	-57,96%
1	0	0	0	0	0	1	10	1R	OOO	-43,15%

Table 26: Test 2 Deep Learning Results (Process Cost) -

Deep learning					
Material	**Batch size**	**Raw Material**	**Weq [Kg]**	**Veq [m3]**	**Delta**
M2	1	Open-die Forging	1P	OOO	-6,60%
M2	5	Open-die Forging	1P	OOO	-14,41%
M2	10	Open-die Forging	1P	OOO	-2,02%
M3	1	Open-die Forging	1Q	OOO	6,35%
M3	5	Open-die Forging	1Q	OOO	-3,55%
M3	10	Open-die Forging	1Q	OOO	7,62%
M1	1	Open-die Forging	1R	OOO	-38,26%
M1	5	Open-die Forging	1R	OOO	-69,96%
M1	10	Open-die Forging	1R	OOO	-60,21%
M2	1	Closed-die Forging	1P	OOO	-3,61%
M2	5	Closed-die Forging	1P	OOO	-14,23%
M2	10	Closed-die Forging	1P	OOO	2,03%
M3	1	Closed-die Forging	1Q	OOO	-1,79%
M3	5	Closed-die Forging	1Q	OOO	-18,53%
M3	10	Closed-die Forging	1Q	OOO	-6,85%
M1	1	Closed-die Forging	1R	OOO	-26,58%
M1	5	Closed-die Forging	1R	OOO	-58,95%
M1	10	Closed-die Forging	1R	OOO	-43,75%

Table 27: Test 2 Random Forest Results (Process Cost) -

Random forest					
Material	**Batch size**	**Raw Material**	**Weq [Kg]**	**Veq [m3]**	**Delta**
M2	1	Open-die Forging	1P	OOO	-12,30%
M2	5	Open-die Forging	1P	OOO	-24,85%
M2	10	Open-die Forging	1P	OOO	-18,78%
M3	1	Open-die Forging	1Q	OOO	15,21%
M3	5	Open-die Forging	1Q	OOO	9,80%
M3	10	Open-die Forging	1Q	OOO	14,48%
M1	1	Open-die Forging	1R	OOO	-37,37%
M1	5	Open-die Forging	1R	OOO	-64,10%
M1	10	Open-die Forging	1R	OOO	-60,76%
M2	1	Closed-die Forging	1P	OOO	-5,81%
M2	5	Closed-die Forging	1P	OOO	-22,82%
M2	10	Closed-die Forging	1P	OOO	-18,77%
M3	1	Closed-die Forging	1Q	OOO	10,73%
M3	5	Closed-die Forging	1Q	OOO	0,91%
M3	10	Closed-die Forging	1Q	OOO	4,09%
M1	1	Closed-die Forging	1R	OOO	-28,02%
M1	5	Closed-die Forging	1R	OOO	-62,65%
M1	10	Closed-die Forging	1R	OOO	-60,23%

In the second test, the MAPE was not calculated but a graph was drawn (for the cost of the material and the cost of the process) in which the real values predicted by the algorithms are compared (Figure 53 and Figure 54).

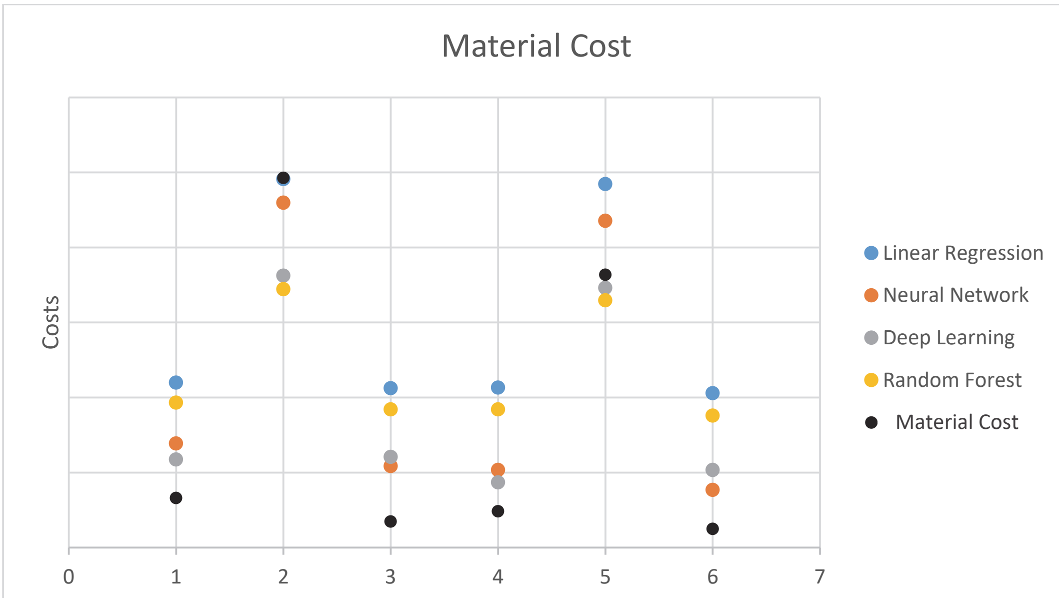


Figure 53. Test 2 material cost - © 2020 Baker Hughes, LLC - All rights reserved.

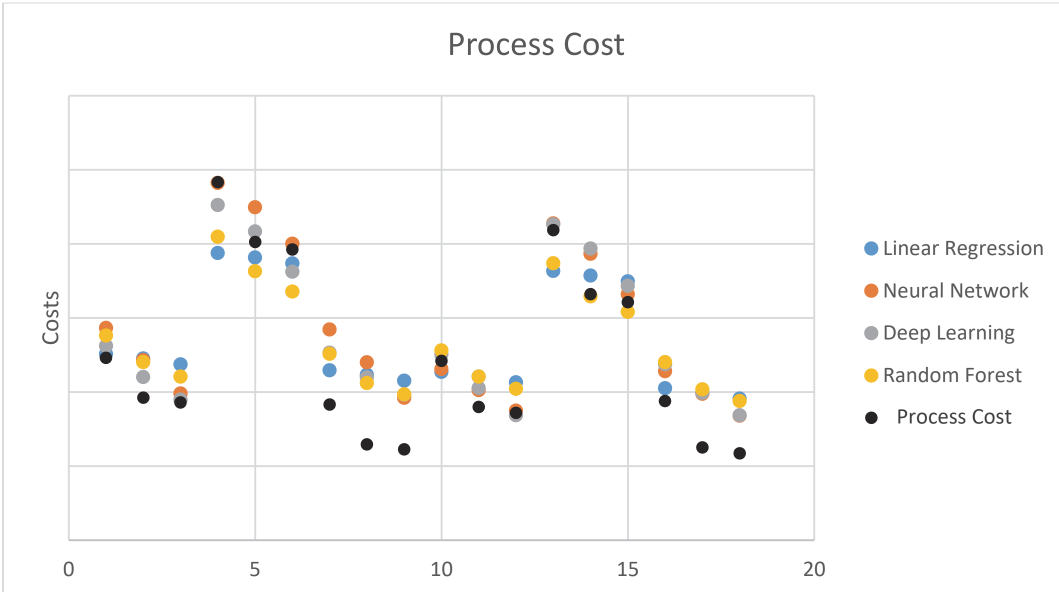


Figure 54. Test 2 process cost -

In Figure 53 and Figure 54, colored dots were used to represent the cost values predicted by the algorithms and the real cost (in black). It is noted how the neural networks (orange and gray color) provide good performance in case of data outside the starting range.

In the case of the cost of the material, the result is consistent with that of the first test. Neural networks can better estimate costs. Increased error in cost estimation, concerning test 1, is mainly driven by higher variability of cost depending on material al limited numbers of records (only 25 records) with respect to independent parameters number (5 independent parameters).

The values of the deviations can be both positive and negative: in the first case, there is an overestimation as the cost generated by the algorithm is greater than the real cost. Conversely, in the second case, there is an underestimation as the cost generated by the algorithm is less than the real cost.

The cost estimation techniques, applied to the material cost case, show very high deviations.

In the case of the cost of the process, however, even if the random forest in the first test was slightly more precise, in this variant, which considers test data outside the starting range, the neural networks can be more precise than the random forest (not always but in most cases).

Therefore, this particular case can be explored in subsequent works to verify whether the simple neural network and deep learning methods are more precise than the other machine learning algorithms in the case of test data outside the starting range.

For process cost, the results obtained from the different methods were compared to each other based on MAPE value, to identify which of these has the best performance (Figure 55). It is possible to see how different methods lower their accuracy and performances, except regression that almost confirms its behavior.

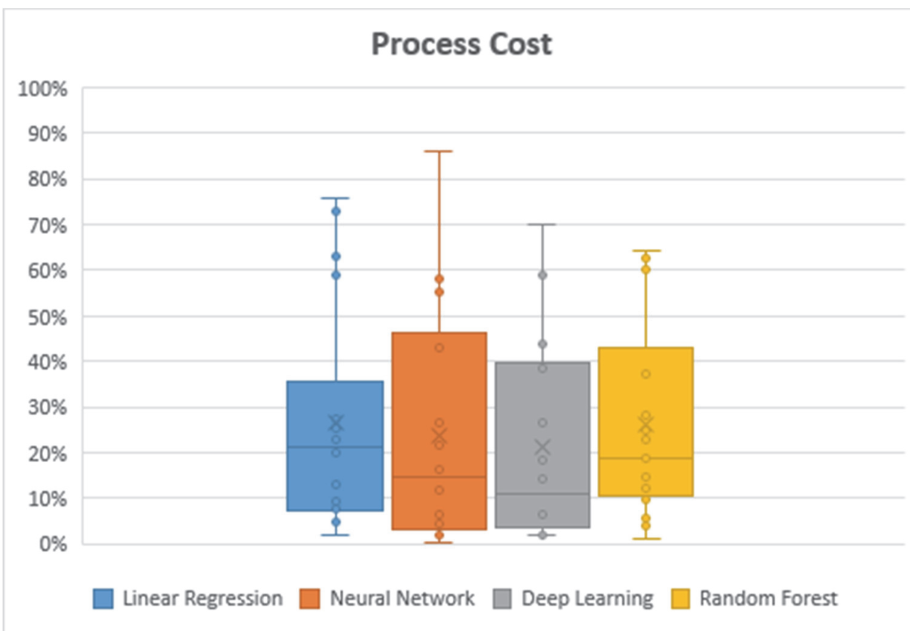


Figure 55. Test 2 MAPE process cost - © 2020 Baker Hughes, LLC - All rights reserved.

6.3. Discussion

Design to cost and VAVE Methodology shall be included in a new product project plan since the very beginning of the program and shall be applied to all the major components of a gas turbine.

VAVE tools facilitate design trade-offs based on cost impact as well as performance indicators and take into account all the Total Cost of Ownership drivers. VAVE experts maintain the Product Cost model up to date, following each design stage. VAVE engineer is also responsible to understand and show the gap between current cost estimation and target cost, acting both on total product cost and on each sub-item cost.

However, the methodology presented shows some limitations, such as it needs a quite huge effort and a long time to be followed. Only for VAVE workshop, five working days are needed for the whole team, without considering also pre-work time to collect and prepare all the material and post-workshop time to follow proposals implementation.

Definition of costing models is not a quick task as well: each step of the stream proposed shall be followed meticulously and it can take weeks or even months to be completed and reviewed.

To overcome those limitations, it is possible to prepare standard functions definition and FAST diagrams for major Gas Turbine components. Having those databases will for sure help in further exercises, on similar components but presenting different characteristics based on specific gas turbines analyzed.

For what concern the second topic addressed in this thesis work, machine learning cost estimation techniques applied to axial compressor discs have proved to be a valid alternative to the traditional linear regression technique, and those techniques can be proficiently extended to all the components of a gas turbine.

In particular, in the case of the cost of the material, the results were less performing due to the reduced amount of training data and the nature of the materials considered (very different unit costs).
In the case of the cost of the processes, the results were much more satisfactory as very low relative errors were detected in the case of test data within the starting range (about 10%), and also in the second test, the deviations were acceptable.
In this thesis work, the R6 requirement was partially answered, since the analysis was carried out only on Should Cost data.

Anyhow, considering the analysis performed, it is evident that the innovative techniques of machine learning constitute a very good alternative to the traditional linear regression technique and demonstrated to have better performances.

Conceptual and parametric cost models can be applied to estimate alternative architectures considering cost as a design parameter, leveraging automated tools, since the very beginning of the project plan. Cases study discussed in this thesis, focusing on gas turbine blades and discs, show how this methodology represents an important achievement for preliminary cost evaluation

For sure, this methodology can give important results if applied to all the components of a gas turbine, covering the entire product architecture configuration.

On the other hand, a limitation is represented by the need for a periodic maintenance effort during the time.

Chapter 7.

Concluding Remarks

The major contribution of this research work has been to establish a novel cost estimation and value optimization approach with the purpose to consider cost as a design parameter in the early phases of the product development process. The approach here described leads to new applications for the product value optimization in the conceptual design stage. A software platform has been defined and proposed to support design activities and to standardize parametric cost models and VAVE analysis practice in industrial design departments.

The advance beyond the state of art has already been discussed; it is interesting to mention that the proposed methodology presents big advantages also with regards to previously implemented solutions. In particular, parametric cost estimation is rigorous and mathematical and allows to overcome some limitations of previous methods that were based on similarities with other components and required and important executor expertise.

VAVE permits to increase product value by following a systematic step by step process, on the contrary, previously applied methods were standard brainstorming sessions that are not repeatable and often do not allow to reach product value improvement targets.

Two of the main results presented in this thesis work will be discussed following, and in particular on one hand cost reduction and value optimization achieved through the full methodology implementation (referring to the blades case study) and, on the other hand, the evaluation and comparison between different machine learning techniques to lower cost estimation error of cost models performed (referring to the discs case study).

For what concern the blades case study, a list of the main outcome reached with the full developed methodology developed, are listed as follows:

- Address cost constrains and value requirements early in the design process when changes matter most and cost least
- Design more sustainable and cost-effective products
- Identify product main functions, with special focus on ones the customer is paying for
- Increase product value and customer satisfaction
- Customize products to reduce the total cost.
- Achieve competitive advantage through design to cost approach and product value optimization
- Standardize process steps to provide engineers with a systematic procedure to be followed to obtain the above goals

The main limitations, referring to the blades case study, are the following:

- The methodology proposed requires a huge effort and a long time to be followed

- Sometimes it is not easy to find all the expertise and skilled people to complete all the steps of the assessment proficiently (i.e. when the production process know-how is on suppliers' hand)

To overcome the main limitations of VAVE methodology that are the huge effort, the long time to be implemented and the lack of expertise, it is possible to define standard templates for common components of gas turbines, as well as, to train reference experts inside the company to supervise the methodology correct implementation.

Since phase 1 and phase 2 have been performed two times, even if crucial for methodology success, the time of implementation of VAVE methodology increased. But on the other hand, splitting into two different steps both Step 2.1 and Step 2.2 helped the team to concentrate deeply on both DM and DL portions of cost, which for sure is a novelty and important progress of the described method.

To overcome this drawback and help to complete Steps 2.1 and 2.2 more quickly, it is possible to prepare standard function definition and FAST diagrams for Gas Turbine components. Having those databases will for sure help in further exercises, on similar components but presenting different characteristics based on specific gas turbines addressed.

VAVE methodology demonstrated to be a very valid support to engineering design, especially in case the target cost is not matching to the estimated cost of components. In fact, leveraging the potentiality of this methodology, it is possible to help the design team to find alternative solutions maintaining the same functionality at a lower cost.

Referring to the discs case study, a list of the main outcome during the comparison of different machine learnings techniques are listed as follows:

- They have better performances with respect to traditional methods such as regressions
- Given their non-linearity, they can manage very complex problems
- Random forest can be carried out in less time (more immediate parametric optimization process) and more easily
- Neural networks have proved to be more accurate in most cases of estimating data outside the training range

The main limitations are the following:

- All the machine learning techniques can be considered "black boxes"
- Sometimes time to train is not negligible
- Require a huge amount of data to be trained (especially neural networks).

Regarding the last point, the data for the training should be structured and stored in a proper way to enable accurate and profitable training. This point is connected to how the company databases/repository are structured and how data can be easily retrieved: this is why it is so important the database definition as described in 4.4.2.

In other business environments, this limitation can be a real roadblock, especially in the case, but this is not the situation, if the training data are not owned by the company itself. This could be the main barrier to the application of this method in some other working environment.

On the contrary, in this case, databases contain data owned by the company: or a set of actual cost (from latest purchase orders) or in alternative, a set of should cost values, obtained using software such as LeanCOST® or DFMA. Leveraging those tools, it is possible to calculate should cost values using analytic system for cost prediction, based on production know-how and manufacturing expertise. This approach permits to create set of data in a reduced amount of time, in particular for discs case study, starting from 38 codes of different discs, it has been possible to generate a database of over 700 records in an automatic manner, using Leancost tool and varying some parameters such as material and batch size and reducing time for database creation and, consequently, for parametric cost modelling completion.

From the analysis carried out, it is evident that the innovative techniques of machine learning, in addition, to be a valid alternative to the traditional linear regression technique, have better performances.

Despite this advantage, it is not always certain to use them because they have the limit of being considered "black boxes": it is not possible to give a theoretical interpretation of the results, especially in the case of unexpected or unjustified values.

Linear regression, on the other hand, is characterized by being deducible from technical considerations and, consequently, easily interpretable (it is a relationship between the dependent and independent parameters).

Machine learning techniques, given their non-linearity, can manage very complex problems thanks to the fact that:

- It is not necessary to carefully check each potential cost driver to verify its correlation with the output
- There is no need to make initial assumptions on the shape of the approximation functions.

For regression, it is possible to obtain greater precision in the results if the problem is simplified (for example by considering only one material instead of five).

As regards the differences between random forest and neural networks (simple and deep learning), it emerged that the former is carried out in less time (more immediate parametric optimization process) and more easily.
However, in the cases of estimating data outside the training range, neural networks have proved to be more accurate in most cases than the random forest.
It is therefore evident that the algorithm to be used to estimate costs in the conceptual design phase is highly dependent on the type of problem or more specifically on the shape of the input database.

As future developments, linear and non-linear methods could be implemented for other turbomachinery components (such as vanes) and a study on Actual Cost data could be carried out.
The latter, being more random and unpredictable than Should Cost data, would go to test the algorithms for even more complex problems.
Furthermore, the use of the Monte Carlo method could be of support to generate a more robust database of actual cost data.

Further progress in cost estimation and optimization is to improve supplier involvement and collaboration: suppliers hold all the know-how and expertise necessary to optimize cost without any impact on quality. Cooperation with them could bring benefits and can be boosted by the development of an automated tool for cost estimation, based on a set of information and a database that includes all the best practices and lessons learned from previous projects. For sure, it is not easy to implement that kind of collaboration with all the suppliers. Each company has the need to protect their "know-how" and interoperability of information systems for exchanging such information are not easily available on the market. However, selecting some strategic suppliers and defining with them a long-term collaboration also in terms of volume commitment, it is possible to define a consolidated commercial partnership, within which it is easier and accepted to share information on the production process of components.

In conclusion, this work has allowed us to confirm the potential of machine learning non-linear regression techniques. Therefore, the study can be deepened to develop in a short time a tool able to estimate the costs of a new product in the conceptual design phase and consequently discard or modify unsuitable projects before significant economic resources have been invested for their realization. The tool that will be developed, and the relative methodologies behind it, shall be compliant to ASTM E2516 standard [80] and also to European Standards [54][56] and Government Auditing Standards [57].

Chapter 8.

References

[1] Pahl, G., Beitz, W., Feldhusen, J., Grote, K.-H.: Engineering Design. A Systematic Approach. Springer-Verlag, London (2007)

[2] Svengren L., (1997) Industrial design as a strategic resource— a study of industrial design methods and approaches for companies' strategic development in The Design Journal, Vol. 0, No. 1, pp. 3–11.

[3] Ulrich K., and Eppinger S., (1995) Product Design and Development, McGrawhill, New York.

[4] Ulrich K., (1995) The Role of Product Architecture In The Manufacturing Firm in Research Policy, Vol. 24.

[5] Ulrich, K., Eppinger, S.A.: Product Design and Development. McGraw-Hill Education, New York (2011).

[6] Boothroyd, G., Dewhurst, P., Knight, W.A.: Product Design For Manufacture and Assembly 3rd Edition. CRC Press (2011).

[7] INCOSE Systems Engineering Handbook, A Guide for System Lifecycle Processes and Activities, INCOSE-TP-2003-002-03, version 3, International Council on Systems Engineering (INCOSE), June 2006 – version 4 was just issued in July 2015

[8] Value Methodology Standard, SAVE International, October 1998. The Society of American Value Engineers (SAVE) International is devoted to the advancement and promotion of the value methodology. Information can be found at http://www.value-eng.org/

[9] Cicconi, P., Germani, M., Mandolini, M.: How to Support Mechanical Product Cost Estimation in the Embodiment Design Phase. Advanced Concurrent Engineering. (2010) 10.1007/978-0-85729-024-3_45

[10] Panchenko, Y., Moustapha, H., Mah, S., Patel, K., Dowhan, M.J., Hall, D.: Preliminary Multi-Disciplinary Optimization in Turbomachinery Design. RTO-MP-089 (2002)

[11] Irene Martinelli, Federico Campi, Emanuele Checcacci, Giulio Marcello Lo Presti, Francesco Pescatori, Antonio Pumo, Michele Germani (2019). «Cost Estimation Method for Gas Turbine in Conceptual Design Phase», 29th CIRP Design Conference 2019.

[12] Zhongsheng, H. Jie, Y. Solomani, C., Bin, Z.: Integration TRIZ with problem-solving tools: A literature review from 1995 to 2006. International Journal of Business Innovation and Research - Int J Bus Innovat Res (2006). 10.1504/IJBIR.2006.011091

[13] Chan, L.K., Wu, M.L.: Quality function deployment: A literature review. European Journal of Operational Research 143 463–497 (2002).

[14] Favi, C., Germani, M., Mandolini, M.: A Multi-objective Design Approach to Include Material, Manufacturing and Assembly Costs in the Early Design Phase. Procedia CIRP, Vol. 52, pp. 251-256 (2016). https://doi.org/10.1016/j.procir.2016.07.043.

[15] Favi, C., Germani, M., Mandolini, M.: Analytical Cost Estimation Model in High Pressure Die Casting. Procedia Manufacturing, 11: 526-535 (2017).

[16] Favi, C., Germani, M., Mandolini, M.: Design for Manufacturing and Assembly vs. Design to Cost: Toward a Multi-objective Approach for Decision-making Strategies during Conceptual Design of Complex Products. Procedia CIRP, 50: 275-280 (2016).

[17] Favi, C., Germani, M., Mandolini, M.: Development of complex products and production strategies using a multi-objective conceptual design approach.

The International Journal of Advanced Manufacturing Technology, Vol. 95, pp.1281–1291 (2018). https://doi.org/10.1007/s00170-017-1321-y

[18] Otto, K.N. and Wood, K.L.: Product Evolution: A Reverse Engineering and Redesign Methodology. Research in Engineering Design, 10(4): 226–243 (1998).

[19] Stone, R.B., Wood, K.L. and Crawford, R.H.: A heuristic method for identifying modules for product architectures. Design Studies, 21(1):5–31 (2000).

[20] Mikkola, J. H.: Management of Product Architecture Modularity for Mass Customization: Modelling and Theoretical Considerations. IEEE Transactions on Engineering Management, 54(1): 57–69 (2007).

[21] de Jesus Pacheco, D.A., Schwengber ten Caten, C., Jung, C.F., Guitiss Navas, H.V., Cruz-Machado, V.A., Tonetto, L.M.: State of the art on the role of the Theory of Inventive Problem Solving in Sustainable Product-Service Systems: Past, Present, and Future, Journal of Cleaner Production, Volume 212, 2019, Pages 489-504, ISSN 0959-6526 (2019) https://doi.org/10.1016/j.jclepro.2018.11.289.

[22] Ocampo, L.A., Labrador, J.J.T., Jumao-as, A.M.B., Rama, A.M.O.: Integrated multiphase sustainable product design with a hybrid quality function deployment – multi-attribute decision-making (QFD-MADM) framework, Sustainable Production and Consumption, Volume 24, Pages 62-78 (2020) ISSN 2352-5509, https://doi.org/10.1016/j.spc.2020.06.013.

[23] Hastak, M., Association for the Advancement of Cost Engineering International, Skills and Knowledge of Cost Engineering 6th Edition, 2015, https://books.google.nl/books?id=jPErjwEACAAJ

[24] Malmiry, R.B., Dantan, J.Y., Pailhes, J., Antoine, J.F.: A product functional modelling approach based on the energy flow by using characteristics-properties modelling. Journal of Engineering Design, Taylor & Francis, 2016, 27 (12), pp.817-843. ff10.1080/09544828.2016.1235261ff. ffhal-01394715f

[25] Malmiry, R.B., Pailhès, J., Qureshi, A.J., Antoine, J.F., Dantan, J.Y.: Management of product design complexity due to epistemic uncertainty via

energy flow modelling based on CPM, CIRP Annals, Volume 65, Issue 1, Pages 169-172, ISSN 0007-8506 (2016), https://doi.org/10.1016/j.cirp.2016.04.048.

[26] Magni, G.: Application of Value Analysis/Value Engineering (VAVE) Methodology to Gas Turbine Auxiliary Systems in General Electric (GE) Oil & Gas. SAVE 2017 Value Summit: Celebrating 70 Years of Value (2017)

[27] Peng, F., Jiang, X.: A Value-Event Path Model Based Value Engineering Analysis Methodology. Third International Symposium on Intelligent Information Technology Application (2009). 10.1109/IITA.2009.216

[28] Kumar, A., Zope, N.R., Lokku, D.S.: An Approach for Services Design by Understanding Value Requirements, Identifying Value Carriers, Developing Value Proposition, and Subsequently Realizing Value. 2014 Annual SRII Global Conference (2014). 10.1109/SRII.2014.5.

[29] Bing, X.: Application Research on the Investment Control of Construction Project with Value Engineering at Design Stage (2019). 10.1109/ICIII.2009.408

[30] Rehman, S., Guenov, M. D.: A Methodology for modelling manufacturing costs at conceptual design. Computers and Industrial Engineering, 35(3–4): 623–626 (1998).

[31] Mileham, A. R., Currie, G. C., Miles, A. W., Bradford, D. T.: A Parametric Approach to Cost Estimating at the Conceptual Stage of Design. Journal of Engineering Design, 4(2): 117-125 (1993).

[32] Mauchand, M., Siadat, A., Bernard, A., Perry, N.: Proposal for tool-based method of product cost estimation during conceptual design. Journal of Engineering Design, 19(2): 159-172 (2008).

[33] Hari, A., Shoval, S., Kasser, J.: Conceptual design to cost: A new systems engineering tool. 18th Annual International Symposium of INCOSE, 6th Biennial European Systems Engineering Conference, 18(1): 543-557 (2008).

[34] Saravi, M., Newnes, L., Roy, A., Yee, M., Goh, M.: Estimating Cost at the Conceptual Design Stage to Optimize Design in terms of Performance and

Cost. Collaborative Product and Service Life Cycle Management for a Sustainable World: 123-130 (2008).

[35] Davide Durastanti, «Metodi e strumenti per il Design to Cost di gruppi funzionali nell'ambito di macchine per l'Oil & Gas», Corso di Laurea Magistrale in Ingegneria Meccanica, Università Politecnica delle Marche, 2018.

[36] Hoque, A.S.M., Halder, P.K., Parvez, M.S., Szecsi, T.: Integrated manufacturing features and Design-for-manufacture guidelines for reducing product cost under CAD/CAM environment. Computers & Industrial Engineering: 66: 988-1003 (2013).

[37] Niazi, A., Dai, J.S., Balabani, S., Seneviratne, L.: Product Cost Estimation: Technique Classification and Methodology Review. Journal of Manufacturing Science and Engineering, 128 (2): 563-575 (2005).

[38] Koini, G., Sotirios, S., Ioannis, N.: A software tool for parametric design of turbomachinery blades. Advances in Engineering Software. 40. 41-51 (2009). 10.1016/j.advengsoft.2008.03.008.

[39] Lanzotti, A., Carbone, F., Grazioso, S., Renno, F., Staiano, M.: A new interactive design approach for concept selection based on expert opinion. International Journal on Interactive Design and Manufacturing. Vol. 12, pages 1189–1199 (2018).

[40] Budiono, H.D.S., Kiswanto, G., Soemardi, T.P.: Method and Model Development for Manufacturing Cost Estimation during the early design phase related to the complexity of the machining processes. International Journal of Technology 2014; 2: 183-192 (2014).

[41] Hooshmand, Y., Köhler, P., Korff-Krumm, A.: Cost Estimation in Engineer-to-Order Manufacturing. Open Engineering formerly Central European Journal of Engineering 2016; 6:22-34 (2016).

[42] Lukic, D., Milosevic, M., Borojevic, S., Durdev, M., Vukman, J., Antic, A.: Manufacturing Cost Estimation in the Conceptual Process Planning. Machine Design 2016; 8(3):83-60 (2016)

[43] Xua, Y., Elghb, F., Erkoyuncua, J.A., Bankolea, O., Gohc, Y., Cheungd, W.M., Baguleya, P., Wange, Q., Arundachawata, P., Shehaba, E., Newnesf, L., Roya, R.: Cost Engineering for manufacturing: Current and future research. International Journal of Computer Integrated Manufacturing. 1–15 (2011)

[44] Lin, H., Li, Y., Li, W.: An Empirical Analysis of Activity Based Costing in Chinese Enterprises. Journal of Finance and Accounting. Vol. 4, No. 5, pp. 301-309 (2016). 10.11648/j.jfa.20160405.17

[45] Barone, G., Frangopol, D.M.: Life-cycle maintenance of deteriorating structures by multi-objective optimization involving reliability, risk, availability, hazard and cost. Structural Safety, 48, pp.40-50 (2014).

[46] Roy, R., et al.: Detailed cost estimating in the automotive industry: Data and information requirements. International Journal of Production Economics (2011) 10.1016/j.ijpe.2011.05.018

[47] Filippazzo, G.: Parametric Cost Modelling of an Activity Based Proposal for a System of Systems. IEEEAC paper #1123, Version 2, December 15, 2004.

[48] Ardiansyah, R., Sutopo, W., Nizam, M.: A Parametric Cost Estimation Model to Develop Prototype of Electric Vehicle based on Activity-based Costing. (2013) IEEE. 978-1-4799-0986-5/13.

[49] Duverlie, P., Castelain, J.M: Cost Estimation During Design Step: Parametric Method versus Case Based Reasoning Method, Journal of Advanced Manufacturing Technology 15,895-906 (1999)

[50] Kavitha S, Varuna S, Ramya R. A Comparative Analysis on Linear Regression and Support Vector Regression. 2016 Online International Conference on Green Engineering and Technologies (IC-GET). 1-5. 10.1109/GET.2016.7

[51] S. Deng, Tsung-Han Yeh, Using least squares support vector machines for the airframe structures manufacturing cost estimation, International Journal of Production Economics, Volume 131, Issue 2, 2011, Pages 701-708, ISSN 0925-5273, https://doi.org/10.1016/j.ijpe.2011.02.019.

[52] Hirtz J., Stone R.B., McAdams D.A., Szykman S., and Wood K.L., (2002) A Functional Basis for Engineering Design: Reconciling and Evolving Previous Efforts NIST Technical Note 1447.

[53] UNI EN 16271 : 2013, Value management - Functional expression of the need and functional performance specification - Requirements for expressing and validating the need to be satisfied within the process of purchasing or obtaining a product, 2013.

[54] EN1325:2014, Value Management - Vocabulary - Terms and definitions, 2014.

[55] Haik, Yousef; Shahin, Tamer; Engineering Design Process, Book, 2010, 10:0-495-66816-8

[56] Pahl, G., Beitz, W., Feldhusen, J., Grote, K.-H.; Engineering Design, A Systematic Approach; Book, 2007.

[57] GAO-20-195G, COST ESTIMATING AND ASSESSMENT GUIDE: Best Practices for Developing and Managing Program Costs, 2020.

[58] Evans, D.K., Lanham1, J.D., Marsh, R.: Cost estimation method selection: matching user requirements and knowledge availability to methods, Systems Engineering and Estimation for Decision Support (SEEDS) Group, University of West of England, Bristol, UK.

[59] Claudio Favi, Federico Campi, Marco Mandolini, Irene Martinelli, Michele Germani. "Conceptual cost estimation of multistage Axial Compressor module". ASME 2019 International Mechanical Engineering Congress & Exposition IMECE 2019

[60] Mandolini Marco, Cicconi Paolo and Germani Michele How to support mechanical product cost estimation in the embodiment design phase [Conference] // New world situation: new directions in concurrent engineering. - 6-10 September 2010, Cracow, Poland : Jerzy Pokojski, Shuichi Fukuda, Jòzef Salwìnski, 2010. - pp. 419-431. - ISBN 978-0-85729-023-6.

[61] Mandolini Marco, Cicconi Paolo and Germani Michele Knowledge-based tool for cost estimation in agile product design [Conference] // 3rd Internation Conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV 2009). - 5-7 October 2009, Munich, Germany : Markus Wiedemann, Frèdèric-Felix Lacour, 2009. - pp. 510-521. - ISBN 978-3-8316-0933-8.

[62] Germani M., Mengoni M. and Peruzzini M. A method to define a co-design platform to support cooperative work in SMEs [Conference] // Proceedings of the 6th international conference on product lifecycle management PLM09, 6-8 july 2009. - Bath, UK : [s.n.], 2009.

[63] Favi, C., Germani, M.: A method to optimize assemblability of industrial product in early design phase: From product architecture to assembly sequence. International Journal for Interactive Design and Manufacturing (2012).

[64] Mandolini, M., Favi, C., Peruzzini, M., Germani, M.: A should costing approach for manufacturing companies Int. J. Agile Systems and Management, Vol. 12, No. 4 (2019). 10.1504/IJASM.2019.10026303.

[65] Russo, A.C., Landi, D., Germani, M.: A Design Methodology and Tool to Support Ecodesign Implementation in Induction Hobs (Version 10006967) (2017). http://doi.org/10.5281/zenodo.1130179

[66] Birkler, John, Jeffrey B. Garfinkle, and Kenneth E. Marks, Development and Production Cost Estimating Relationships for Aircraft Turbine Engines, Santa Monica, Calif.: RAND Corporation, N-1882-AF, 1982. As of November 10, 2020: https://www.rand.org/pubs/notes/N1882.html

[67] Li Qian, David Ben-Arieh, Parametric cost estimation based on activity-based costing: A case study for design and development of rotational parts, International Journal of Production Economics, Volume 113, Issue 2, 2008, Pages 805-818, ISSN 0925-5273, https://doi.org/10.1016/j.ijpe.2007.08.010.

[68] Dale T. Masel, Robert P. Judd; Using Bottoms-Up Cost Estimating Relationships in a Parametric Cost Estimation System; Presented at the 2007 ISPA/SCEA Joint Annual International Conference and Workshop - www.iceaaonline.com

[69] D.J Kim, B.M Kim, J.C Choi, Determination of the initial billet geometry for a forged product using neural networks, Journal of Materials Processing Technology, Volume 72, Issue 1, 1997, Pages 86-93, ISSN 0924-0136, https://doi.org/10.1016/S0924-0136(97)00134-9.

[70] Fangwei Ning, Yan Shi, Maolin Cai, Weiqing Xu, Xianzhi Zhang, Manufacturing cost estimation based on a deep-learning method, Journal of Manufacturing Systems, Volume 54, 2020, Pages 186-195, ISSN 0278-6125, https://doi.org/10.1016/j.jmsy.2019.12.005.

[71] Fangwei Ning, Yan Shi, Maolin Cai, Weiqing Xu, Xianzhi Zhang, Manufacturing cost estimation based on the machining process and deep-learning method, Journal of Manufacturing Systems, Volume 56, 2020, Pages 11-22, ISSN 0278-6125, https://doi.org/10.1016/j.jmsy.2020.04.011.

[72] Timothy Masters, Practical neural network recipes in C++, Academic Press Professional, Inc., ISBN:978-0-12-479040-7, July 1993

[73] Jean-Loup Loyer, Elsa Henriques, Mihail Fontul, Steve Wiseall, Comparison of Machine Learning methods applied to the estimation of manufacturing cost of jet engine components, International Journal of Production Economics, Volume 178, 2016, Pages 109-119, ISSN 0925-5273, https://doi.org/10.1016/j.ijpe.2016.05.006.

[74] Xiaonan Chen, Jun Huang, Mingxu Yi, Cost estimation for general aviation aircrafts using regression models and variable importance in projection analysis, Journal of Cleaner Production, Volume 256, 2020, 120648, ISSN 0959-6526, https://doi.org/10.1016/j.jclepro.2020.120648.

[75] H.S. Wang, Y.N. Wang, Y.C. Wang, Cost estimation of plastic injection molding parts through integration of PSO and BP neural network, Expert

Systems with Applications, Volume 40, Issue 2, 2013, Pages 418-428, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2012.01.166.

[76] Turner, M. G., Merchant, A., and Bruna, D.: A Turbomachinery Design Tool for Teaching Design Concepts for Axial-Flow Fans, Compressors, and Turbines. ASME. J. Turbomach. July 2011; 133(3): 031017 (November 18, 2010). https://doi.org/10.1115/1.4001240

[77] S. V. Tammineni, A. R. Rao, J. P. Scanlan, P. A.S. Reed, A. J. Kean; A knowledge-based system for cost modelling of aircraft gas turbines; Journal of Engineering Design 20(3):289-305; 2009

[78] Masel, D., Dowler, J.D., & Judd, R. (2017). Adapting Bottoms-up Cost Estimating Relationships to New Systems.

[79] Stephan Langmaak, Stephen Wiseall, Christophe Bru, Russell Adkins, James Scanlan, András Sóbester, An activity-based-parametric hybrid cost model to estimate the unit cost of a novel gas turbine component, International Journal of Production Economics, Volume 142, Issue 1, 2013, Pages 74-88, ISSN 0925-5273, https://doi.org/10.1016/j.ijpe.2012.09.020.

[80] ASTM E2516-11(2019), Standard Classification for Cost Estimate Classification System, ASTM International, West Conshohocken, PA, 2019, www.astm.org

Appendix A.

cod = machined code

d = machined diameter

s = machined thickness

M = material

R = roughness

Tratt = heat treatment presence

Peq = equivalent weight

Psl = semi machined weight

Pg = raw weight

Veq = equivalent disc volume

Machine	Cod	d	s	M	N° grooves	R	Tratt	batch	Raw material	Peq [Kg]	Psl	Pg	V eq [m3]	Tot
Disc mach 1	A	AA	AAA	M1	X	I	si	1	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	si	5	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	si	10	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	si	20	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	si	50	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	J	si	1	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	J	si	5	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	J	si	10	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	J	si	20	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	0,8	si	50	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	no	1	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	no	5	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	no	10	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	no	20	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	no	50	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	no	1	Closed die	0L	0M	0N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	no	5	Closed die	0L	0M	0N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	no	10	Closed die	0L	0M	0N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	no	20	Closed die	0L	0M	0N	O	€ -
Disc mach 1	A	AA	AAA	M1	X	I	no	50	Closed die	0L	0M	0N	O	€ -
Disc mach 1	A	AA	AAA	M2	X	I	no	1	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M2	X	I	no	20	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M2	X	I	no	50	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M2	X	I	no	1	Closed die	0L	0M	0N	O	€ -
Disc mach 1	A	AA	AAA	M2	X	I	no	20	Closed die	0L	0M	0N	O	€ -
Disc mach 1	A	AA	AAA	M2	X	I	no	50	Closed die	0L	0M	0N	O	€ -
Disc mach 1	A	AA	AAA	M3	X	I	no	1	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M3	X	I	no	10	Open die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M3	X	I	no	1	Closed die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M3	X	I	no	10	Closed die	L	M	N	O	€ -
Disc mach 1	A	AA	AAA	M4	X	I	no	10	Open die	L	M	N	O	€ -
Disc mach 1	B	BB	BBB	M1	Y	I	no	1	Open die	P	Q	R	S	€ -
Disc mach 1	B	BB	BBB	M1	Y	I	no	20	Open die	P	Q	R	S	€ -
Disc mach 1	B	BB	BBB	M1	Y	I	no	1	Closed die	P	Q	R	S	€ -
Disc mach 1	B	BB	BBB	M1	Y	I	no	20	Closed die	P	Q	R	S	€ -
Disc mach 1	B	BB	BBB	M4	Y	I	no	1	Closed die	P	Q	R	S	€ -

Disc mach 1	B	BB	BBB	M4	Y	I	no	20	Closed die	P	Q	R	S	€ -
Disc mach 1	C	CC	CCC	M5	Z	I	no	1	Open die	T	U	V	Z	€ -
Disc mach 1	C	CC	CCC	M5	Z	I	no	5	Open die	T	U	V	Z	€ -
Disc mach 1	C	CC	CCC	M5	Z	I	no	1	Closed die	T	U	V	Z	€ -
Disc mach 1	C	CC	CCC	M5	Z	I	no	5	Closed die	T	U	V	Z	€ -
Disc mach 1	C	CC	CCC	M4	Z	I	no	1	Open die	T	U	V	Z	€ -
Disc mach 1	C	CC	CCC	M4	Z	I	no	5	Open die	T	U	V	Z	€ -
Disc mach 1	D	DD	DDD	M3	Z	I	no	1	Open die	LL	MM	NN	OO	€ -
Disc mach 1	D	DD	DDD	M3	Z	I	no	5	Open die	LL	MM	NN	OO	€ -
Disc mach 1	E	CC	EEE	M2	W	I	no	1	Closed die	PP	QQ	RR	SS	€ -
Disc mach 1	E	CC	EEE	M2	W	I	no	10	Closed die	PP	QQ	RR	SS	€ -
Disc mach 1	F	EE	FFF	M5	J	I	no	1	Open die	TT	UU	VV	Z	€ -
Disc mach 1	F	EE	FFF	M5	J	I	no	10	Open die	TT	UU	VV	Z	€ -
Disc mach 1	F	EE	FFF	M1	J	I	no	1	Closed die	TT	UU	VV	Z	€ -
Disc mach 1	F	EE	FFF	M1	J	I	no	10	Closed die	TT	UU	VV	Z	€ -
Disc mach 2	G	FF	GGG	M3	Z	I	no	1	Open die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M3	Z	I	no	5	Open die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M3	Z	I	no	10	Open die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M3	Z	I	no	1	Closed die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M3	Z	I	no	5	Closed die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M3	Z	I	no	10	Closed die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M5	Z	I	no	1	Open die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M5	Z	I	no	5	Open die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M5	Z	I	no	10	Open die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M2	Z	I	no	1	Open die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M2	Z	I	no	10	Open die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M2	Z	I	no	1	Closed die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M2	Z	I	no	10	Closed die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M2	Z	I	no	50	Closed die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M4	Z	I	no	1	Closed die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M4	Z	I	no	10	Closed die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	G	FF	GGG	M4	Z	I	no	50	Closed die	LLL	MMM	NNN	ZZ	€ -
Disc mach 2	H	GG	HHH	M4	Z	I	no	1	Open die	LLL	QQQ	RRR	ZZZ	€ -
Disc mach 2	H	GG	HHH	M4	Z	I	no	5	Open die	LLL	QQQ	RRR	ZZZ	€ -
Disc mach 2	H	GG	HHH	M1	Z	I	no	1	Open die	TTT	UUU	VVV	ZZZ	€ -
Disc mach 2	H	GG	HHH	M1	Z	I	no	5	Open die	TTT	UUU	VVV	ZZZ	€ -
Disc mach 2	H	GG	HHH	M5	Z	I	no	1	Open die	TTT	UUU	VVV	ZZZ	€ -
Disc mach 2	H	GG	HHH	M5	Z	I	no	5	Open die	TTT	UUU	VVV	ZZZ	€ -
Disc mach 2	H	GG	HHH	M3	Z	I	no	1	Open die	LLLL	MMMM	NNNN	ZZZ	€ -
Disc mach 2	H	GG	HHH	M3	Z	I	no	5	Open die	LLLL	MMMM	NNNN	ZZZ	€ -
Disc mach 2	I	HH	III	M2	W	I	no	1	Closed die	PPPP	QQQQ	RRRR	ZZZZ	€ -
Disc mach 2	I	HH	III	M2	W	I	no	5	Closed die	PPPP	QQQQ	RRRR	ZZZZ	€ -
Disc mach 2	I	HH	III	M2	W	I	no	10	Closed die	PPPP	QQQQ	RRRR	ZZZZ	€ -
Disc mach 2	I	HH	III	M1	W	I	no	1	Closed die	TTTT	UUUU	VVVV	ZZZZ	€ -
Disc mach 2	I	HH	III	M1	W	I	no	5	Closed die	TTTT	UUUU	VVVV	ZZZZ	€ -
Disc mach 2	I	HH	III	M1	W	I	no	10	Closed die	TTTT	UUUU	VVVV	ZZZZ	€ -
Disc mach 2	I	HH	III	M5	W	I	no	1	Closed die	TTTT	UUUU	VVVV	ZZZZ	€ -
Disc mach 2	I	HH	III	M5	W	I	no	5	Closed die	TTTT	UUUU	VVVV	ZZZZ	€ -
Disc mach 2	I	HH	III	M5	W	I	no	10	Closed die	TTTT	UUUU	VVVV	ZZZZ	€ -
Disc mach 2	I	HH	III	M4	W	I	no	1	Closed die	1L	1M	1N	ZZZZ	€ -
Disc mach 2	I	HH	III	M4	W	I	no	5	Closed die	1L	1M	1N	ZZZZ	€ -
Disc mach 2	I	HH	III	M4	W	I	no	10	Closed die	1L	1M	1N	ZZZZ	€ -