PARAMETRIC AND NEURAL METHODS FOR COST ESTIMATION OF PROCESS VESSELS

Antonio C. Caputo
Department of Mechanical, Energy and Management Engineering, University of L'Aquila, 67040 Monteluco, L'Aquila, Italy
caputo@ing.univaq.it

Pacifico M. Pelagagge
Department of Mechanical, Energy and Management Engineering, University of L'Aquila, 67040 Monteluco, L'Aquila, Italy
pelmar@ing.univaq.it

Abstract. *In the paper a comparison is made between artificial neural networks and parametric functions for estimating manufacturing cost of large sized pressure vessels in engineering to order manufacturing systems for bidding purposes. The developed methods have been tested with reference a wordl leading manufacturer and both greatly outperformed the manual methods currently adopted. Average estimation error is about 10% which is quite a satisfactory result for this kind of production context.*

Keywords: Cost estimation, process vessels, artificial neural networks, parametric cost modeling

1. Introduction

Production cost assessment is a relevant issue in any industrial management activity. Typically, cost calculations may be classified as pre-calculations, intermediate calculations and post calculations (Layer et al., 2002). Pre-calculation involves the estimate of future costs before the actual production, in order to allow cost-based decision making; intermediate calculations are carried out for cost controlling purposes during the product development cycle, while post-calculations include cost accounting methods utilized to determine the actual costs incurred to serve as base data for future precalculations (Dewhurst and Boothroyd, 1989; Geiger and Dilts, 1996). Since large part of the product life cycle costs are defined during the design stage there is a growing concern to integrate cost estimation in the early design stage (French, 1990), thus implementing the so-called concurrent engineering approach (Creese and Moore, 1990; Noble and Tanchoco, 1990; Oh and Park, 1993). This enables an early identification of the sensitive factors affecting the manufacturing costs (materials, fixtures and toolings, processes) in order to timely change design decisions and parts configuration. Production cost estimation is therefore widely adopted in new product development processes, in the development of budget estimates, in profitability analyses, as a criterion to define selling prices, and for bidding purposes in project-oriented companies as happens in make-to-order and engineering-to-order sectors. For such applications quantitative cost estimation techniques are increasing their utilization and supersede qualitative expert judgement. However, in such cases only incomplete or uncertain product describing data are usually available, and suitable cost calculation methods are required if future costs are to be predicted fairly accurately from such unreliable information. In particular, the capability of rapidly and correctly estimating manufacturing costs for bidding purposes is critical for engineer to order manufacturers of non standard equipment with customer defined designs and specifications. In this case any error may lead to overestimating the cost with the risk of being uncompetitive and loosing a customer, or underestimating the cost with the result of winning a contract but incurring in a financial loss. This also implies a difficulty in estimating delivery times which are instead critical for project management purposes and as a competitive advantage respect competitors. Therefore, correct cost estimating leads to both correct resource utilization and effective planning of the manufacturing phase, but also to a greater contractual power with the customer and competing suppliers. Beside estimation accuracy issues, cost estimation methods are also required to be rapid in order to ensure a quick response to customers, but also cheap to utilize so that a fairly accurate preliminary cost estimate can be obtained without the need for a detailed design of the product and in a short time. Traditional cost estimating techniques in this kind of application decompose the workload in tasks assigned to workcentres and try to estimate the required each resource utilization time by basing on standard times data or previous experience. Then the materials cost is added. However, this process is quite time consuming, suffers from the lack of previous experience in case of non standard components and asks for a rather detailed design of the equipment to be preliminary carried out if the estimates are to be accurate. Since a detailed design is usually not available at the time of bidding, such kind of estimates are rather unsatisfactory. This can have adverse effects on the competitiveness of the company and on the budget management of the procurement order.

Therefore, in order to contribute to a solution of this problem the utilization of historical data to obtain accurate and rapid estimates through the development of both parametric models and artificial neural networks (ANN) has been attempted in this paper. In order to consider a significant application context and to obtain the required field data, reference has been made to the production of large sized pressure vessels for the chemical process industry which is typical of the engineering to order sector. The methods were developed referring to one of the leading world companies

in this sector, gathering data on 68 process units manufactured in the past two years, to determine actual consutive costs and work hours in comparison to estimated values, besides technical details of the equipments and the manufacturing process, thus creating an historical data base. In the paper, following a literature review and a description of the traditional estimation methodology adopted, the data gathering and classification activity is detailed. Parametric and ANN models are described and the procedures adopted for models tuning based on historical data are discussed. Verification of the performances of the models is then carried out and a critical appraisal of obtained results is finally provided to justify the residual errors and identify further improvement strategies for future refinement of the models.

2. A review of cost modeling techniques

From a methodological point of view cost estimation may be based on qualitative or quantitative approaches (Layer et al, 2002). Qualitative approaches rely on expert judgement or heuristic rules and will not be dealt with in this work as only state if an alternative is better or worse that the other without specifying absolute values. Quantitative methods instead may be further classified in statistical models, analogous models or generative-analytical models (Layer, et al, 2002; Asiedu et al., 2000). Statistical or analogous models are also called "lump-sum" methods as do not consider the characteristics of the production process or do not show the details of the cost structure. In fact such methods attempt to establish an overall correlation (called CER, Cost Estimating Relationship) between the total manufacturing cost and some cost-effective product characteristics (i.e. variables related to the product configuration or physical characteristics such as weight, size etc., which may act as cost drivers). Parametric cost models belong to the family of statistical methods in that statistical criteria are utilized to identify the causal links and correlate costs and product characteristics in order to obtain a parametric function with one or more variables. Statistical methods can rely on formulas or alternative approaches to link product characteristics to costs. For example regression analysis (Dean, 1995; DoD, 1999; Schreve et al., 1999) or optimisation methods (Pickel, 1989) have been widely utilized, but artificial neural networks have been also employed to extend the field of statistical methods thanks to their ability in classifying, summarizing and extrapolating collections of data (Bode, 1998 and 2000). Seo et al. (2002) also utilize ANN and statistical correlation methods in Life Cycle Costing for use in conceptual design stages, while the same approach was adopted by Cavalieri et al. (2004) for the estimation of manufacturing cost of mechanical components (disk brakes). Zhang et al. (1996) and Zhang and Fuh (1998) utilized ANN to estimate packaging costs based on product dimensions. A number of papers compared the performances of ANN and parametric regression models, in generic context (Mason and Smith, 1997), in assembly industries (Shtub and Zimmermann, 1993) or for mechanical components (Cavalieri et al., 2004). These works confirm that ANN may show better performances than regression models as already pointed out by Hill et al. (1994). ANN models accept in input shape-describing and semantic product characteristics and give in output the product cost. The advantage of ANN is that they can effectively extrapolate or generalize and can model cases where the functional relationship between case data is hidden or can not be expressed in polynomial form, because an input-output mapping is allowed without understanding the functional relationship between variables. However, ANN require a large set of training cases, while the cost driving parameters or product attributes should be known when developing parametric models. Nevertheless, PM or ANN have the advantage of not requiring a detailed definition of the single process phases and ANN are also dynamically adaptive in that the ANN training set may be extended as new data are available to reflect changes or performance improvement in the manufacturing operations and the related resources through a continuous acquisition of knowledge. On the contrary, the regression models correlation must be re-estimated whenever new data arrive. However, statistical methods do not transparently show how costs are derived, are not able to identify the cost-driving product characteristics (Wierda, 1988) and can not be utilized for generative design when new manufacturing technologies are introduced. Furthermore, owing to the low level of detail in characterizing the product they usually do not allow a cost-based comparison between alternative products. Finally, they require historical data which prevents including innovative technologies or new resources.

Analogous methods instead identify a similar product, and reuse the cost information to estimate the future cost in analogy by adjusting the cost for the differences between the products. Analogous models thus infer a similarity in the cost structure from a functional or geometrical similarity among products features with the strength of the similarity being proportional to the correspondence of the relevant characteristics (Layer et al., 2002; Shields and Young, 1991), for example measured as the distance between the points of a multi-dimensional features space. However, the method is only as reliable as one is capable to correctly identify the differences between the studied product and the reference one. Generative-analytical methods are the most accurate in that try to depict the actual product creation process. A detailed analysis of the production process and decomposition in the single manufacturing operations is carried out, and specific models analytically estimate the cost of each processing phase attributing a monetary value to the resources consumption on the basis of the technical parameters characterizing the operation. A bottom-up approach is then utilized to properly aggregate the costs incurred during the process of fabrication through summation of each cost item. A detailed model uses estimates of labour time and rates, material quantities and prices to estimate the direct costs of a product or activity and an allocation rate is used to allow for indirect/overhead costs (Shields and Young, 1991). Therefore, a detailed costing estimate results from a generative process plan which also allows to expose specific cost drivers, while alternatives to adjust products cost can be derived. However, such methods require a very large amount of information, are extremely time consuming as require a detailed design of the product and knowledge of the processes, and may be therefore difficult to carry out. Process oriented methods often include direct integration with CAD models to extract cost-impacting geometrical product features (Ou-Yang and Lin, 1997; Wierda, 1991; Zondervan, 1998) or

rely on data bases of standard times, cost rates and best-practice manufacturing methods, which may be integrated with computer aided process planning software and knowledge-based methods (Shehab and Abdalla, 2002 a,b). Analytical methods in practice form the basis of the widely employed Design-to-Cost or Design-for-Manufacturing methods (Boothroyd et al., 2001; Eversheim, 1998, Kiritsis et al., 1999 and 2000; Poli, 2001) used in concurrent engineering approaches, which provide detailed models for single technological processes such as casting, machining, sheet metal stamping, injection moulding etc., even if such detailed models often rely on parametric cost functions derived using regression analysis. Some application examples are cast and forged products (Venkatachalam et al., 1993), injection moulded products (Chen and Liu, 1999; Chin and Wang, 1996; Shing, 1999), metal stamping (Tang et al., 2004). Analytical process oriented methods are usually applied to small sized components and assembled mechanical products or even printed circuit board assembly (Ong, 1995). In case of large non standard products in engineering to order or make to order fabrication environments it is also frequent to refrain from analytically modelling the actual manufacturing processes but to rather assign an hourly cost to each work center involved by the process plan and to analytically estimate the utilization time of the resources of each cost center. In this case this general cost model follows.

$$C_p = C_{MP} + \sum_{i=1}^{n} \lambda_i T_i + \lambda_G \sum_{i=1}^{n} T_i \qquad (1)$$

where C_p is the final product cost, C_{MP} is the cost of raw materials, n the number of work centers, λ_i the hourly cost of the *i-th* work center, and T_i the number of fabrication hours at each cost center. λ_G is the average overhead hourly cost of cost centers, and accounts for general expenses and indirect cost allocated to each product with reference to the total duration of time that its fabrication process utilizs the resources of the firm. The computation of λ_G value can be carried out resorting to the following equation

$$\lambda_G = \frac{1}{n} \sum_{i=1}^{N} \frac{C_{Ai}}{T_{Ai}} \qquad (2)$$

where n is the total number of cost centers, C_{Ai} is the annual cost of the *i-th* cost center , and T_{Ai} is the annual working time of the *i-th* work center. While C_{MP} is easily and quite accurately estimated as $C_{MP} = PT\ C_{MU}$ where PT (kg) is the total weight of the vessel and C_{MU} is the unit cost of the material (€/kg), after a preliminary sizing of the vessel has been carried out in order to define the thickness of the vessel walls, the estimation of the fabrication hours at each cost center remains a hard task left to the experience of the costing department and to the reliance on historical data. Estimation errors may be quite significant where a wide variability of vessels sizes and configuration occurs and when few historical data are available for any kind of vessel configuration as often happens in engineering to order manufacturing where any product is different from the others because is made to customers specification and design. Therefore, the criticality of this traditional method lies in the fabrication hours estimation.

However, despite the large number of methods developed in the literature and the availability of commercial cost estimating software tools, such techniques have not yet found widespread application within industry owing to a number of practical limitations as highlighted by Layer et al. (2002) in their review of cost estimating methods and available software. As a general remark the methods are not as accurate as they should. It is difficult to include company specific issues such as manufacturing technologies, while treatment of complex parts (for which generation of analytical models is difficult or the unambiguous assignment of features to operations is not possible) remains precluded. Moreover, the models have difficulty in considering the level of maturity or the degree of experience in product development while and the acquisition and maintenance of the knowledge is still an unresolved issue. Finally, the time it takes to produce a cost estimate is also a critical factor.

3. The cases tudy

This paper is aimed at comparing traditional cost estimation techniques adopted in engineering to order industries with parametric and ANN techniques. Reference is made to the production of large sized and complex process vessels. This is a particularly demanding application because non standard single items are assembled based on customer defined design and specifications. In order to obtain the required historical data regarding manufacturing cost and product configuration, reference has been made to an Italian company which is a world leader in the sector, producing complex process units or storage tanks up to 40 m long and with a diameter of up to 5-6 m. In this kind of context the development of reliable cost estimation methods is critical to allow successfull bidding and profitable operations, and becomes a strategic objective of the company.

The production process of pressure vessels includes the construction of vessels heads, the realization walls plates, the preparation of nozzles and fittings, the assembly and plates welding, and a final non destructive testing phase. In greater detail the following work centres (WC) are identified which act as cost centres for accounting purposes and production cost and/or time estimation. WC1 carries out sheet metal plates cutting; plane sheets are cut to shape, their edges contoured for subsequent welding and finally sandblasted to remove traces of oxy-fuel cutting. WC 2 carries out

calendering of the metal plates to produce the elements of vessels walls. In WC3 the vessel heads are formed through plastic deformation of a contoured metal plate. Elliptic or emispherical ends are usually adopted. WC4 is devoted to the manual preparation of ancillaries, fittings, supports, skirts etc. WC5 carries out the main carpentry operation of vessel body assembly. Manual welding (MIG/MAG) is carried out in WC6, while submerged arc automated welding (Electroslag/SAW) is carried out in WC7. Following vessel assembly the final non destructive testing operations are carried out in WC8. WC9 performs thermal treatments, while WC10 the hydraulic testing. Finally, painting and packaging take place in WC11 and WC12 respectively.

At first a data gathering campaign was carried out in order to build a data base of configurations of the process units produced in the past. Original quotations supplied to prospective customers from the marketing department for bidding purposes were collected at first, while actual costs incurred and allocated after production was completed to each cost center on the basis of the utilization time were recorded as supplied from the accounting department. In overall, data related to 68 different process units representative of a wide range of different vessels types were classified, from simple pressurized storage tanks, to columns, heath exchangers and reaction vessels. Many vessels were composed of multiple sections (up to four) with different diameters and a variety of auxiliary external or internal fittings, openings, and nozzles. An illustrative example of the possible configuration and the complexity level of the produced vessels is shown in Fig. 1. The entire vessels data base has been then partitioned in a training set and a test set. The former has been used to obtain the parametric model and to train the neural network while the latter to test the cost estimation accuracy of the developed models. Great care has been devoted to make the training set representative of the wide variability of vessels configuration through the inclusion of significant cases, in order to increase the generalization capability of the models to be developed. However, the fairly low number of available vessels data prevented from developing specific correlations for each category of vessels separately, which would surely improve the estimation performances. Nevertheless this problem will be solved as new data become available and the database is expanded.
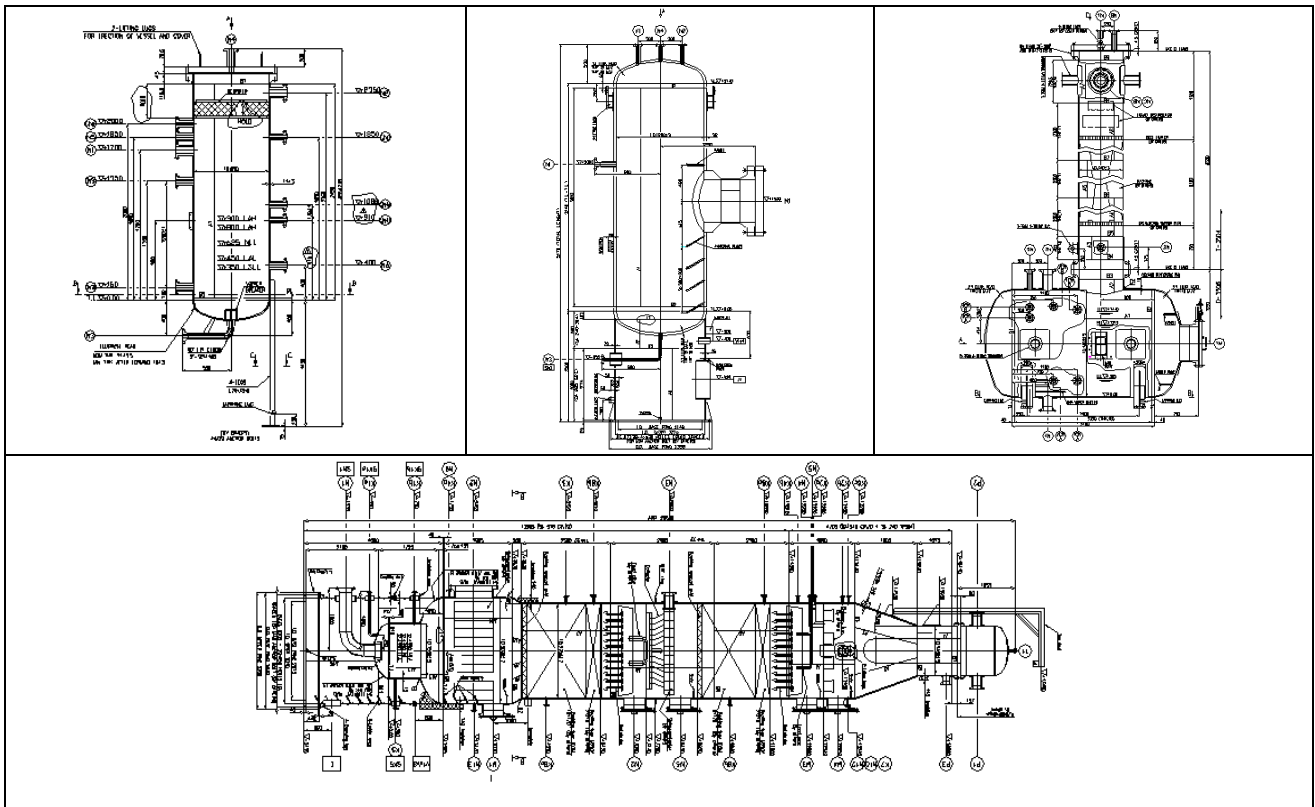

Figure 1. Examples of process vessels.

The company currently performs preliminary cost quotations for bidding purposes utilizing the previously described method (eq. 1 and 2). This involves a rough cut vessel sizing and the resort to expert judgement to estimate workhours at each work centre. In order to assess the performance of this method the available estimates have been compared with the actual production cost as recorded by the accounting department. The percentual error PE_i committed in the estimation of the fabrication hours for the *i-th* vessel according to the manual method, has been computed as

$$PE_i = \frac{NOF_{Si} - NOF_{Ci}}{NOF_{Ci}} \times 100 \tag{3}$$

where NOF_{Si} is the estimated number of hours while NOF_{Ci} is the actual recorded number of hours. Results are shown in Fig. 2 where peak PE values are about +80 and -60%, while the mean absolute percentual error (MAPE) defined as

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\left| NOF_{S,i} \quad NOF_{C,i} \right|}{NOF_{C,i}} \times 100 \right) \tag{4}$$
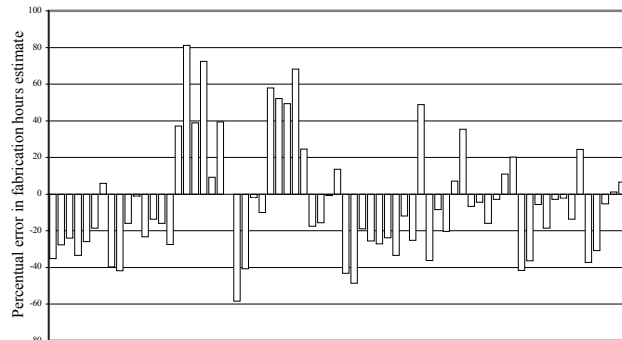
is 26%.

Figure 2. Percentual error in the fabrication hours estimate (manual method).

The above data highlight the unsatisfactory performances of the current cost estimating method. The causes of this excessive error range are the following. The products are non standard and often requiring peculiar manufacturing and assembly task, therefore the utilization of standard times methods is quite difficult and is not possible to define a reference cost for the various categories of vessels. During the bidding phase the detailed design of the unit is not yet available. Therefore, many data required for precisely estimating the work centers utilization are not available and have to be judged from past experience which, anyhow, usually refers to different kind of vessels. Moreover, the marketing department which is responsible for issuing the bid quotations may not be aware with enough details of the actual costs of some operations or resources. Finally, any new vessel configuration for which the manufacturer lacks specific experience may require in an unpredictable manner reworking or outsourcing to thid party components suppliers, which adds to the time and cost uncertainty.

This uncertainty leads to the following problem areas: risk of overestimating the quotation, with the possibility of loosing the bid. Risk of underestimating the quotation. In this case the bid may be won but an economical loss results for the manufacturer. Risk of not correctly quoting the delivery lead time. An excessive lead time promise may discourage the prospective buyer who may switch supplier, while an underestimation of the lead time may expose to penalty due to late delivery.

In order to contribute to the solution of this problem in this peculiar industrial sector, this work is aimed at developing alternative cost estimating methods, both more accurate and quicker to apply respect the ones currently utilized, which can support the manual quotation process for bidding purposes. This is to be achieved through the switch from detailed design approaches to statistical methods based on historical data of the manufacturer, which seem more suited to this application. In particular parametric regression models and artificial neural networks based methods will be compared in order to assess their applicability in this context and evaluate their performances.

4. Parametric cost estimation

Given the great variability of configuration and fittings of the process vessels, a large number of variables are required to thoroughly classify and characterize any given design. In this application 62 different variables were found to be necessary. In this way vessels with up to four sections with different diameters could be considered, with the adopted variables describing either physical characteristics of the vessel (nozzles, heads, plate thickness and so on) and process related information (i.e. thermal treatment: Yes/No). In particular the main required information are vessel volume and surface, plating surface and thickness, vessel body diameter, height and thickness of single plates, number of circumferential welds, diameter and thickness of heads, number of sectors of assembled heads, characteristics of the conical sections reducing the diameter from one cylindrical section to another (type, height and number of required metal plates), characteristics of the supporting skirt (diameter, height, number of metal plates and thickness), number of nozzles and diameter. Qualitative information are also required such as type of vessel, choice of construction material, type of heads (elliptic or hemispherical) and supports (brackets, legs, skirt, saddles), type of reducers (conical or contoured conical) and the required thermal treatment. However, after a close scrutiny only 42 variables were directly put in relationship with the WC utilization and the manufacturing process.

In order to choose the significant variables to be included in the parametric cost function, all of the 42 selected variables were checked singularly to verify if a linear correlation existed with the available cost data. However, it was found that only 6 variables, as depicted for example in Fig. 3, showed a fairly good level of linear correlation with the actual fabrication hours recorded in the historical database, namely the vessel volume V, vessel external surface S, the weight of the support PS, the weight of wall plates and heads PB, the total weight PT and the weight of the auxiliary fittings PA (nozzles, external ladders etc.).
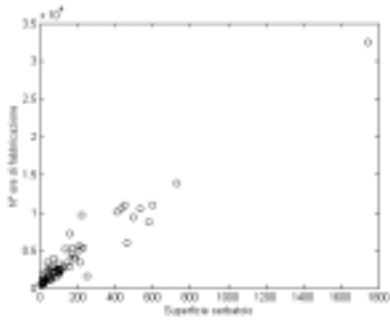
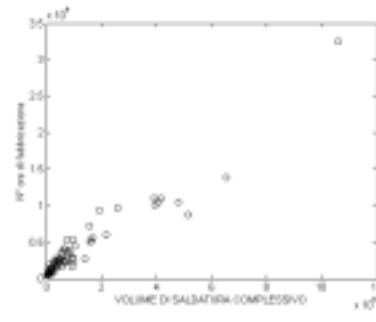Figure 3. Manufacturing hours vs vessel surface   Figure 4. Manufacturing hours vs overall welding lenght

In order to account for the remaining variables and in particular for the total welding lenght, and the type of vessels heads and openings, which have a direct impact on the manufacturing costs, the following three new variables grouping have been defined. The total welding length VSC, computed as

$$VSC = \sum_{i=1}^{4} \left(T_i^{0.7} \cdot \left(\pi \cdot D_i \cdot C_i + F_i \cdot H_i \right) \right) \cdot EG + TG^{0.9} \cdot \left(\pi \cdot DG \cdot (VG + 1) + FG \cdot HG \right) \cdot \left(\frac{SP \cdot 10^6}{40} \cdot TP \right)^{0.75} \tag{5}$$

The nozzles welding length is

$$SB = \frac{\sum_{i=1}^{4} \left(T_i \cdot H_i \right)}{\sum_{i=1}^{4} T_i} \cdot \sum_{j=1}^{n} \left(DB_j \cdot NB_j \right) \tag{6}$$

The characteristics of vessel heads are expressed as

$$CF = \sum_{i=1}^{3} DF_i \cdot SF_i \tag{7}$$

In the above equations, with the subscript i referring to the *i-th* section of a multisection vessel, T is the wall plate thickness, D the vessel diameter, C the number of circumferential welds, F the type of heads, H the body length, EG the efficiency of welded joints, TG the tickness of the skirt plates, DG the diameter of the skirt, VG the number of plates utilized for the skirt assembly, FG the number of single foils for each skirt plate, HG the skirt lenght, SP the plating surface and TP the product type, NB the number of nozzles, DB the nozzle diameter, SF the number of sectors of hemispherical ends, and DF the diameter of heads. These aggregate parameters showed quite a good correlation level with the manufacturing hours as shown for example in Fig. 4.

In overall the 9 chosen variables showed the correlation index values of Tab. 1. All values are greater than 0.85 which is considered a lower limit for the linear correlation hypothesis, thus confirming the validity of the choice.

Table 1. Linear correlation coefficient values for the identified variables.

Variable	Description	ρ_{xy}	Variable	Description	ρ_{xy}
PB	Weight of wall plates and heads	0,94	S	External surface	0,97
PT	Total weight	0,97	VSC	Overall welding volume	0,96
PA	Total fittings and nozzles weight	0,95	SB	Nozzles welding lenght	0,94
PS	Weight of supports	0,91	CF	Heads type	0.87
V	Volume	0,91			

However, rather than adopting a simple linear correlation with the selected variables, it was found preferable to adopt a power law dependence, resulting in a parametric cost model for the computation of the number of fabrication hours (NOF) having the general epression

$$NOF = K + \sum_{i=1}^{9} A_i X_i^{a_i} \tag{8}$$

where NOF is the estimated number of fabrication hours, K is a constant, X_i are the parameters to be accounted for, while A_i are the regression coefficients and a_i are the power laws constants.

The parametric cost model has been developed on a subset of the historical database comprising 23 different vessels. The remaining vessels data were left to be utilized as test data in order to check the accuracy of the model. The

coefficients a_i have been determined for each *i-th* parameter by numerical methods in order to minimize the estimation error on the vessels testing subset.

The resulting parametric cost model is

$$NOF = C + A_1 \cdot PB + A_2 \cdot PT + A_3 \cdot PA^{0.15} + A_4 \cdot V^{0.52} + A_5 \cdot S^{0.95} + A_6 \cdot VSC + A_7 \cdot PS^{0.8} + A_8 \cdot SB + A_9 \cdot CF^{0.55} \qquad (9)$$

In this equations the A_i coefficients were instead obtained through the solution of the system of the 23 equations (8) which could be written for the reference vessels. In matrix form this can be written as $y^* = X \cdot \alpha$ where α is the coefficient column vector, so that $\alpha = (X'X)^{-1} X' y$. The final coefficient values are witheld for privacy reasons because they represent actual production data of a single manufacturer.

Following the CER coefficients determination process, the resulting percentual error PE_i is shown in Fig. 5 referring to the training subset of vessels, while the MAPE has been computed to be 9.4%. In overall the correlation function showed a quite high value of the coefficient of determination, with a value $R^2 = 0.975$ computed over the entire set of 68 vessels, showing the capability of the CER in interpreting the effects of design variable variations.
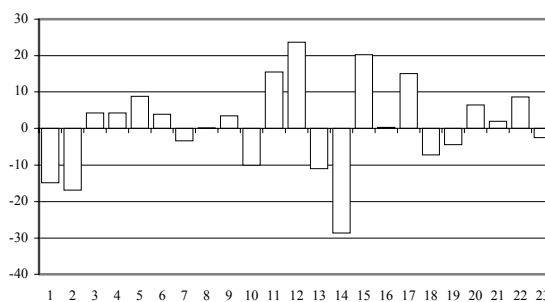

Figure 5. Percentual error in the fabrication hours estimate (parametric model building set).

5. ANN cost modeling

Artificial neural networks are information processing models constituted by a large number of elementary computational units (neurones) linked by weighed connections (Chester, 1993). ANN are able to learn from a set of training data and perform classification, clustering, function approximation and control tasks (Hassoun, 1995; Zahedi, 1991). In particular, ANN may be considered as "universal regression tools" (Hornik et al., 1989), which justifies their utilization in cost estimating applications.

A multilayer perceptron network has been utilized as this configuration gives the best results as a function approximator (Hornik et al., 1989), while the detailed ANN structure has been defined in a trial and error process. The resulting network has four layers, with 66 nodes in the input layer, two and three nodes respectively in the hidden layers and a single node in the output layer. Fig. 6 shows the structure of the ANN network utilized in this study. The adopted learning algorithm was the error back-propagation. The ANN inputs include the values of all the product characterizing variables (normalized in the 0 to 1 range) while the single output is the number of required fabrication hours. Qualitative variables (type of material, type of reducers, heads and supports) have been assigned predefined numerical values. The ANN model was implemented in the Matlab numerical computation environment resorting to the dedicated ANN Toolbox (The Mathworks, 2000). The network has been trained on a subset of the historical database including 28 different vessels through 500 training cycles.
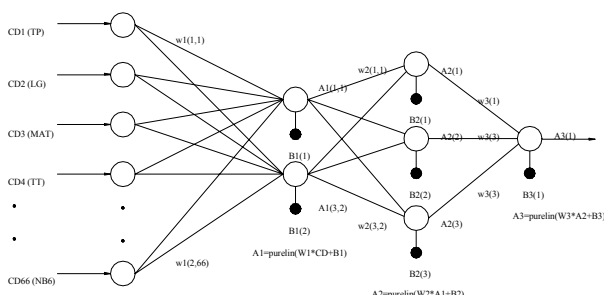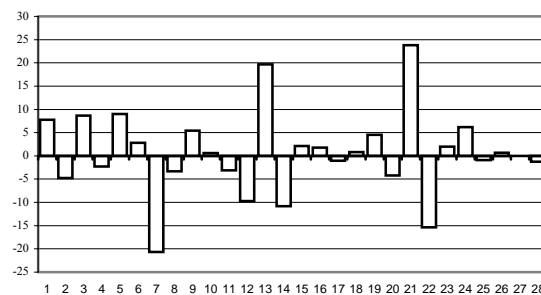

Figure 6. Scheme of the adopted ANN

Figure 7. Percent error in the fabrication hours estimate (ANN training set).

This fairly low number has been chosen in order to avoid overtraining and preserve the network generalization capability. Following the training process the resulting percent error in the estimation of the fabrication hours referring to the training set is shown in Fig. 7, while a MAPE of 6.2% has been also computed. This level of error has been considered acceptable because it was chosen to avoid the overtraining of the network, capable of further reducing the MAPE, in order to preserve the generalization capabilities of the network and possibly make it as robust as possible.

6. Results discussion

In order to assess the performances of the two developed cost estimation methods both have been utilized to estimate the manufacturing hours for the test sets of 43 and 40 vessels respectively which compose the remaining available historical database and compared with the actual recorded worktime hours and the estimates obtained by the manual costing method currently employed (which showed a MAPE of 26% and maximum error of 80% as already stated). Application of the parametric method yields the estimate error values shown in Fig. 8. The resulting average error is 14.1%, which is quite satisfactory for a budgetary estimate and is much better than the average error obtained from the previously adopted manual costing method, with all values within the ±33% range. Error values for the ANN method are instead shown in Fig. 9. In this case the average estimation error has been computed as 10.6% with a +33% - 22% variability range. In overall the parametric and ANN model MAPE computed over the entire database of vessels is 12.5% and 8.7% respectively. Therefore the ANN seems to outperform the parametric model at least as far as the MAPE is concerned.
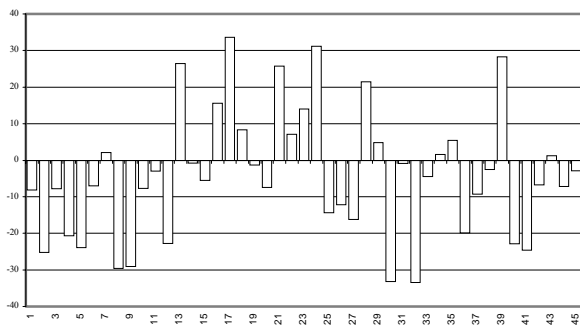

Figure 8. Percentual error in the fabrication hours estimate (parametric model).
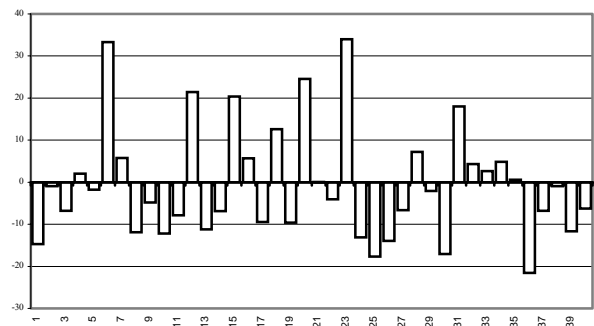
Figure 9. Percentual error in the fabrication hours estimate (ANN model).

In order to further test the predictive capability of the models two new vessels have been considered, respectively a process column and areaction vessel, which were not part of the utilized historical data base. The results of work hours estimation in comparison to the actual manufacturing time is shown in Tab. 2. The cost estimation error was about 3.5% and 5.5% thus confirming the methods capabilities.

Therefore, both methods are capable to reduce the cost estimation error to less than one third of the value obtained when using the manual method, which is a significant result, with the ANN method showing fairly better estimation capabilities (in terms of reduced MAPE and lower variability of absolute percentual errors) because it can utilize the entire set of 66 variables to estimate the production cost.

Table 2. Comparison of work hours estimation performances for new vessels.

	Vessel #1	Vessel #2
Estimated manufacturing time (hours)	4185	6705
Actual manufacturing time (hours)	3970	6480
Error (%)	5.41	3.47

Even if the absolute error is still significant, it can be justified by the difficulty in correlating with a single function or ANN mapping the data referred to such diverse configuration of products. Anyhow the average absolute error value is quite acceptable in this industrial sector for bidding purposes, given the scarcity of data and the unreliability of information available at the moment of submitting a bid. The ANN in particular shows to be quite robust. However, an experiment was carried out by feeding the ANN with only the values of the 9 variables utilized by the CER. In this case the same numerical results of the CER were obtained, thus confirming the correctness of the choice of the CER formulation.

While the obtained values of estimates accuracy are quite satisfactory, still better results could be obtained by expanding the historical data base utilized to train the ANN or to define the parametric model coefficients. In fact the utilized set of vessels could be considered rather limited given the high variability of vessels configuration it included. However, the estimate accuracy would anyway suffer from a number of factors which is difficult to account for during any preliminary cost assessment. For example the model would underestimate in case of unexpected problems during manufacturing leading to delays or necessity of remanufacturing, while could overestimate in case some components

are bought from third parties instead of being made in house. This is the major cause that was found to justify the largest observed errors in cost estimation. In this respect, the quality of data supplied by the accounting department also played a major role in the estimation uncertainty, because the utilization of third parties supplier or changes in the design during construction or even extensive remanufacturing were included in the overall value of time or cost accounted but were not declared nor justified in detail. As a consequence widely different accounted data may be supplied for roughly similar vessels without the possibility of determining the causes. Apart from the extension of the historical database, an improvement of the results may be thus obtained through a greater accuracy of the accounting process utilized to allocate and record actual costs, in order to obtain a reliable data base. Furthernore, some new variables could be introduced to better characterize a vessel design. In particular it is advisable to explicitly account for the degree of experience the workers have with any kind of vessels, as the manufacturing of a vessel having an unfamiliar configuration adds to the uncertainty of the estimate. Genetic algorithms or other non statistical techniques could be instead adopted to set the values of the coefficients in the parametric model, while neuro fuzzy techniques could improve the estimation capability of the ANN. In both cases the integration with expert systems would prove useful. Finally, an analysis of the actual impact of the single cost centres would provide as a means to more accurately modeling the work centres where the estimation error are the greatest. Anyway, the obvious advantages of the ANN are that the preliminary data analysis activity is much simpler as it is not required to define the functional form of the CER and that it is not required to closely screen each potential cost driver to check for its correlation with the output variable. This infact allowed to feed the ANN with all the input variables while the parametric model had to be built only on the 9 most significant parameters. Finally, as more data became available the ANN can be rapidly retrained on an extended data set, while the parametric model should be completely recalculated. Nevertheless, while the CER may structure be deduced from technical considerations and engineering reasoning, which also helps to interpretate results and identify the main cost drivers, the ANN gives no information obout the kind of correlation among input and output values.

7. Conclusions

In this paper parametric or artificial neural network based cost estimating approaches have been developed for application in engineering to order companies with particular reference to the heavy carpentry sector and the manufacturing of large sized pressure vessels. This application is quite significant owing to the relevant cost involved and the wide variability of vessels configurations and size. Both the developed methods have been tested over a six months period in the facilities of one of the world leading manufacturers with very encouraging results if compared with the manual estimation method previously utilized. While with the manual method an average estimation error of 26% had occurred with maximum error values of +81% and -60%. The parametric function method enabled to reduce the average estimation error to 14%, with extremal values within the ±30% range and the neural network approach allowed to further reduce the average error 10% with a +30% - 20% variability range.

The models have been also tested with two completely new vessels design showing cost estimation errors within 5% of actual consuntive costs, thus showing a much better performance than the methods currently practiced. This performance level could be further improved by extending the historical data base as new data become available and by improving the quality of the data in the database through a more precise accounting of actual cost data. As a consequence the capability of both methods to suitably estimate manufacturing costs for large carpentry equipment in engineering to order companies for bidding purposes has been demonstrated.

8. Acknowledgements

9. References

Asiedu, Y., Besant, R.W., Gu, P., 2000, "Simulation-based Cost Estimation Under Economic Uncertainty Using Kernel Estimators", Int. J. of Production Research, Vol. 38, No. 9, pp. 2023-2035.
Bode, J., 1998, "Decision Support with Neural Networks in the Management of Research and Development. Concepts and Applications to Cost Estimation", Information & Management, Vol. 34, No. 1, pp. 33-40.
Bode, J., 2000, "Neural Networks for Cost Estimation: Simulation and Pilot Applications", Int. J. of Production Research, Vol. 38, No. 6, pp. 1231-1254.
Boothroyd, G., Dewhurst, P., Knight, W., 2001, Product Design for Manufacture & Assembly, Marcel Dekker, USA.
Cavalieri, S., Maccarrone, P., Pinto, R., 2004, "Parametric vs. Neural Network Models for the Estimation of Production Costs: A Case Study in the Automotive Industry", Int. J. of Production Economics, Vol. 91, pp. 165-177.
Chen, Y.N., Liu, J.J., 1999, "Cost Effective Design for Injection Molding", Robotics and Computer-Integrated Manufacturing, Vol. 15, pp. 1-21.
Chester, J., 1993, Neural Networks. A tutorial, Prentice Hall, New York.
Chin, K.S., Wang, T.N., 1996, "Developing a Knowledge-based Injection Mold Cost Estimation by Decision Table", Int. J. of Advanced Manufacturing Technology, Vol. 11, n. 6, pp. 353-364.
Creese, R.C., Moore, L.T., 1990, "Cost modeling for Concurrent Engineering", Cost Engineering, Vol. 32, No. 6, pp. 23-27.
Dean, E.P., 1995, "Parametric cost deployment", Proc. of the 7th Symposium on Quality Function Deployment, June, Usa, pp. 27-34.

Department of Defense (DoD), 199, Parametric Estimating Handbook, available from the International Society of Parametric Analysts (ISPA), http://www.ispa-cost.org.

Dewhurst, P., Boothroyd, G., 1989, "Early Cost Estimating in Product Design", J. of Manufacturing Systems, pp. 183-191.

Eversheim, W., Neuhausen, J., Sesterhenn, M., 1998, "Design-to-Cost for Production Systems", CIRP Annals, Vol. 47, No. 1, pp. 357-360.

French, M.J., 1990, "Function-costing: A Potential Aid to Designers", J. of Engineering Design, Vol. 1, No. 1, pp. 47-53.

Geiger, T.S., Dilts, D.M., 1996, "Automated Design-to-Cost: Integrating Costing into the Design Decisions, Computer Aided Design, Vol. 28, pp. 423-438.

Hassoun, M.H., 1995, Fundamentals of Artificial Neural Networks, MIT Press, Cambridge, Massachussets, USA.

Hill, T., Marquez, L., Oconnor, M., Remus, W., 1994, "Artificial Neural Networks Models for Forecasting and Decision Making", Int. J. of Forecasting, V. 10, pp. 5-15, 1994.

Hornik, K., Stinchcombe, M., White, H., 1989, "Multilayer Feedforward Networks are Universal Approximators, Neural Networks, Vol. 2, pp. 359-366.

Kiritsis, D., Neuendorf, K.P., Xirouchakis, P., 1999, "Petri Net Techniques for Process Planning Cost Estimation" Advances in Engineering Software, Vol. 30, No. 6, pp. 375-387.

Kiritsis, D., Xirouchakis, P., 2000, "DEVIPLAN: a Bid Preparation System for Mechanical SME", Proc. of Design Technical Conference, Baltimore, Maryland 10-13 September.

Layer, A., Brine, E.T., Van Houten, F., Kals, H., Haasis, S., 2002, "Recent and Future Trends in Cost Estimation", Int. J. Computer Integrated Manufacturing, Vol. 15, No. 6, pp. 499-510.

Mason, A.K., Smith, A.E., 1997, "Cost Estimation Predictive Modeling: Regresssion vs Neural Network", The Engineering Economist, Vol. 42, No. 2, pp. 137-162.

Noble, J.S., Tanchoco, J.M.A.,1990, "Concurrent Design and Economic Justification in Developing a Product", Int. J. of Production Research, Vol. 28, No. 1, pp. 1225-1238.

Oh, C.J., Park, C.S., 1993, "An Economic Evaluation Model for Product Design Decisions Under Concurrent Engineering", The Engineering Economist, Vol. 38, No. 4, pp. 275-296.

Ong, N.S., 1995, "Manufacturing Cost Estimation for PCB Assembly: an Activity-based Approach", Int. J. of Production Research, Vol. 38, pp. 159-172.

Ou-Yang, C., Lin, T.S., 1997, "Developing an Integrated Framework for Feature-based Early Manufacturing Cost Estimation", Advanced Manufacturing Technology, Vol. 13, No. 9, pp. 618-627.

Pickel, H., 1989, "Kostenmodelle als Hilfsmittel zum Kostengünstigen Konstruieren", Hanser, München, Germany.

Poli, C., Design for Manufacturing, Butterworth-Heinemann, 2001.

Schreve, K., Schuster, H.R., and Bason ,A.H., 1999, "Manufacturing Cost Estimation During Design of Fabricated Parts", Proc. of the Institution of Mechanical Engineers, Vol. 217, No. 3, pp. 731-735.

Seo, K.K., Park, J.H., Jang, D.S., Wallace, D., 2002, "Prediction of the Life Cycle Cost Using Statistical and Artificial Neural Network Methods in Conceptual Product Design", Int. J. Computer Integrated Manufacturing, Vol. 15, No.6, pp. 541-554.

Shehab, E., Abdall,a H., 2002a, "A Design to Cost System for Innovative Product Development", Proc. Instn. Mech. Engrs, Part B, Journal of Engineering Manufacture, Vol. 216, pp. 999-1019.

Shehab, E., Abdalla, H., 2002b, "An Intelligent Knowledge-based System for Product Cost Modelling", Int. J. of Advanced Manufacturing Technology, Vol. 19, pp. 49-65.

Shields, M.D., Young, S.M., 1991, "Managing Product Life Cycle Cost of an Organizational Model", Cost Management, Autumn, pp. 39-52.

Shing, O.N., 1999, "Design for Manufacture of a Cost-based System for Moulded Parts", Advances in Polymer Technology, Vol. 18, No. 1, pp. 33-42.

Shtub, A., Zimmermann, Y., 1993, "A Neural Network Based approach for Estimating the cost of Assembly Systems", Int. J. of Production Economics, Vol. 32, n. 2, pp. 189-207.

Tang, D., Eversheim, W., Schuh, G., 2004, "Qualitative and Quantitative Cost Analysis for Sheet Metal Stamping", Int. J. Computer Integrated Manufacturing, Vol. 17, No. 5, pp. 394-412.

The MathWorks Inc., 2000, Neural Network Toolbox for Use with MATLAB.

Venkatachalam, A.R., Mellichamp, J.M., Miller, D.M., 1993, "A Knowledge-based Approach to Design for Manufacturability", Journal of Intelligent Manufacturing, Vol. 4, pp. 355-366.

Wierda, L.S., 1988, "Product-cost Estimation by the Designer", Engineering Costs and Production Economics, Vol. 14, pp. 189-198.

Wierda, L.S., 1991, "Linking Design, Process Planning and Cost Information by Feature Based Alignment", Journal of Engineering Design, Vol. 2, No. 1, pp. 3-18.

Zahedi, F., 1991, "An Introduction to Neural Networks and a Comparison with Artificial Intelligence and Expert Systems", Interfaces, Vol. 21, pp. 25-38.

Zhang, Y.F., Fuh, J.Y., Chan, W.T., 1996, "Feature-based Cost Estimation for Packaging Products Using Neural Networks", Computers in Industry, Vol. 32, pp. 95-113.

Zhang, Y. F., Fuh, J. Y. H., 1998, "A Neural Network Approach for Early Cost Estimation of Packaging Products", Computers Ind. Engng, Vol. 34, No. 2, pp. 433-450.

Zondervan, P.R., 1998, "Cost Estimating Integrated with CAD/CAM for Sheet Metal Products", Proc. of the 6th International Conference on Sheet Metal: SheMet'98, Enschede, Netherland, Vol. 2, pp. 243-252.

10. Responsibility notice

The authors are the only responsible for the printed material included in this paper.